Current Research in the Semantics/ Pragmatics Interface

MEANING AND THE DYNAMICS OF INTERPRETATION

SELECTED WRITINGS OF HANS KAMP

Edited by Klaus von Heusinger
and Alice ter Meulen

BRILL

Meaning and the Dynamics of Interpretation

Current Research in the Semantics/Pragmatics Interface

Series Editors

Klaus von Heusinger
Ken Turner

Editorial Board

Nicholas Asher, Université Paul Sabatier, France
Johan van der Auwera, University of Antwerp, Belgium
Betty Birner, Northern Illinois University, USA
Claudia Casadio, Università degli studi G. d'Annunzio Chieti Pescara, Italy
Ariel Cohen, Ben Gurion University, Israel
Marcelo Dascal, Tel Aviv University, Israel
Paul Dekker, University of Amsterdam, the Netherlands
Regine Eckardt, University of Göttingen, Germany
Markus Egg, Humbolt University Berlin, Germany
Donka Farkas, University of California, Santa Cruz, USA
Bruce Fraser, Boston University, USA
Thorstein Fretheim, Norwegian University of Science and Technology, Norway
Brendan Gillon, McGill University, Canada
Jeroen Groenendijk, University of Amsterdam, the Netherlands
Yueguo Gu, Chinese Academy of Social Sciences, PRC
Larry Horn, Yale University, USA
Yan Huang, University of Auckland, New Zealand
Asa Kasher, Tel Aviv University, Israel
Manfred Krifka, Humboldt University, Germany
Susumu Kubo, Matsuyama University, Japan
Chungmin Lee, Seoul National University, South Korea
Stephen Levinson, Max Planck Institute for Psycholinguistics, the Netherlands
Claudia Maienborn, University of Tübingen, Germany
Tony McEnery, Lancaster University, UK
Alice ter Meulen, University of Geneva, Switzerland
François Nemo, University of Orléans, France
Peter Pelyvas, University of Debrecen, Hungary
Jaroslav Peregrin, Czech Academy of Sciences and University of Hradec Králové, Czech Republic
Allan Ramsay, University of Manchester, UK
Rob van der Sandt, Radboud University Nijmegen, the Netherlands
Kjell Johan Sæbø, University of Oslo, Norway
Robert Stalnaker, Massachusetts Institute of Technology, USA
Martin Stokhof, University of Amsterdam, the Netherlands
Gregory Ward, Northwestern University, USA
Henk Zeevat, University of Amsterdam, the Netherlands
Thomas Ede Zimmermann, University of Frankfurt, Germany

VOLUME 29

The titles published in this series are listed at brill.com/crispi

Meaning and the Dynamics of Interpretation

Selected Papers of Hans Kamp

Edited by

Klaus von Heusinger
Alice ter Meulen

BRILL

LEIDEN · BOSTON
2013

This is an open access title distributed under the terms of the CC BY-NC-ND 4.0 license, which permits any non-commercial use, distribution, and reproduction in any medium, provided no alterations are made and the original author(s) and source are credited. Further information and the complete license text can be found at https://creativecommons.org/licenses/by-nc-nd/4.0/

The terms of the CC license apply only to the original material. The use of material from other sources (indicated by a reference) such as diagrams, illustrations, photos and text samples may require further permission from the respective copyright holder.

Library of Congress Cataloging-in-Publication Data

Meaning and the dynamics of interpretation : selected papers of Hans Kamp / Edited by Klaus von Heusinger, Alice ter Meulen.
pages cm. – (Current Research in Semantics / Pragmatics Interface ; 29)
Includes bibliographical references.
ISBN 978-90-04-25191-5 (hardback : alk. paper) – ISBN 978-90-04-25288-2 (e-book) 1. Semantics. 2. Meaning (Philosophy) 3. Context (Linguistics) 4. Discourse analysis. I. Heusinger, Klaus von, editor of compilation. II. Meulen, Alice G. B. ter, editor of compilation. III. Kamp, Hans.

P325.M385 2013
401'.43–dc23

2013026460

This publication has been typeset in the multilingual "Brill" typeface. With over 5,100 characters covering Latin, IPA, Greek, and Cyrillic, this typeface is especially suitable for use in the humanities. For more information, please see www.brill.com/brill-typeface.

ISSN 1472-7870
ISBN 978-90-04-25191-5 (hardback)
ISBN 978-90-04-25288-2 (e-book)

Copyright 2013 by Klaus von Heusinger and Alice ter Meulen. Published by Koninklijke Brill NV, Leiden, The Netherlands.
Koninklijke Brill NV incorporates the imprints Brill, Global Oriental, Hotei Publishing, IDC Publishers and Martinus Nijhoff Publishers.
Koninklijke Brill NV reserves the right to protect this publication against unauthorized use.

This book is printed on acid-free paper.

CONTENTS

Introduction to Meaning and the Dynamics of Interpretation 1
Alice ter Meulen and Klaus von Heusinger

PART ONE

EVENTS, TEMPORAL REFERENCE, AND DISCOURSE

Introduction to Events, Temporal Reference, and Discourse 7
Formal Properties of 'Now' 11
Events, Instants and Temporal Reference 53
Deixis in Discourse. Reichenbach on Temporal Reference 105

PART TWO

SEMANTICS AND PRAGMATICS

Introduction to Semantics and Pragmatics 163
Free Choice Permission 169
Semantics versus Pragmatics 185

PART THREE

VAGUENESS

Introduction to Vagueness 221
Two Theories about Adjectives 225
The Paradox of the Heap 263

PART FOUR

DISCOURSE REPRESENTATION

Introduction to Discourse Representation 323
A Theory of Truth and Semantic Representation 329
Context, Thought and Communication 371
Disambiguation in Discourse 393

PART FIVE

PRESUPPOSITION

Introduction to Presupposition 421
Presupposition Computation and Presupposition Justification:
One Aspect of the Interpretation of Multi-Sentence Discourse 425
The Importance of Presupposition 455

PART SIX

PROPOSITIONAL ATTITUDES

Introduction to Propositional Attitudes 505
Prolegomena to a Structural Account of Belief and Other Attitudes 513
Temporal Reference Inside and Outside Propositional Attitudes 585

Interview with Hans Kamp 629
Alice ter Meulen and Klaus von Heusinger

INTRODUCTION TO MEANING AND THE DYNAMICS OF INTERPRETATION

Alice ter Meulen and Klaus von Heusinger

This selection of original papers written by Hans Kamp—philosopher-linguist-logician-cognitive-scientist and computer-scientist of great repute and lasting influence—presents to you the core of his scientific research on natural language semantics and its relation to logic, philosophy and linguistics. In selecting his best papers we often faced some difficult decisions, but we maintained as editorial guiding principles that (1) the paper should have Hans Kamp as sole author, (2) the paper should have had a significant impact in its field upon first publication, (3) the paper should still have lasting importance for various readerships, and last but not least, (4) easy access to the paper must be guaranteed for future generations. Although the selected papers have been printed here in their original version, some minor corrections, adjustments in formal details or removal of typos have been allowed. Each paper contains a first footnote acknowledging its original place of publication, thanking its publisher for its kind permission to include it in this selection.

The tremendous breadth of Hans Kamp's research is reflected in the distribution of the selected papers over six parts, ranging from philosophical reflection on the foundational issues in the ancient Sorites Paradox with a formal account of what its solution is, to a detailed review of current analyses of presuppositions in dynamic semantic theories. Each part is introduced by Hans Kamp with a new text in which the papers are positioned. Part One ***Events, Temporal Reference, and Discourse*** contains three influential papers on the logic of time and temporal reference, including Kamp's early and justly very famous paper *Formal Properties of 'Now'*, which established Kamp's reputation at a relatively young age in 1971 as a model-theoretic logician, shortly after completing his Ph.D. at UCLA with Richard Montague. *Events, Instants and Temporal Reference* addresses, among other topics, how instants may be derived from events, following in the footsteps of a historical logical analysis advocated by Bertrand Russell, Norbert Wiener and Alfred Whitehead in the beginning of the 20th century. *Deixis in Discourse. Reichenbach on Temporal Reference* contains a detailed account of how the historic theory of temporal reference from Hans Reichenbach—a logician in the Vienna Circle (±1924–1936), who showed a surprising

© ALICE TER MEULEN AND KLAUS VON HEUSINGER, 2013 | DOI:10.1163/9789004252882_002
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

sensitivity to natural language at a time this was definitely out of fashion—currently may help linguists to understand the deictic aspects of temporal reference times and their dynamics. It reviews also the shortcomings of the Priorean tense logics, that Montague had adopted in his work on natural language semantics.

Part Two ***Semantics and Pragmatics*** contains two classical papers on the interface of semantics, as the logical theory of truth-functional meaning, and the pragmatic aspects of interpretation in context. *Free Choice Permission* addresses the logical properties of disjunctive permission, arguing that giving permission cannot be reduced to asserting that something is permitted, as had previously been equated in von Wright's deontic logic. Kamp investigates in more depth the intensional logical relations between performative acts and assertions in terms of options and continuations of given situations in *Semantics versus Pragmatics*, showing that even assertions cover a diverse set of acts whose meaning cannot be adequately captured by concepts of truth and satisfaction only, while maintaining a universal semantics for logical vocabulary. The paper also addresses Kaplan's notion of *character* in discussing the essential indexical aspects of making assertions in context, ending on a more cognitive science note of whether ways in which people may differ in computing meaning in context should bear on theories of linguistic competence. With our current hindsight both papers may be viewed as early precursors of the more comprehensive, dynamic theory of meaning Kamp was to develop a decade later.

Part Three ***Vagueness*** contains two early papers addressing the semantics of adjectives. In *Two Theories about Adjectives* Kamp first reviews as possible solutions multi-valued logics, supervaluations, and vague models with degrees of satisfaction and a probability distribution over possible completions, before specifying the notion of a graded context-dependent model as a first theory of adjectives. His second theory of adjectives proposes to take coherence of contexts as core notion in the model theory and exploits the dynamic modification of context as a new semantic tool, leaving the metalogical question just what kind of logic it yields wide open. The *Paradox of the Heap* investigates the consequences of contextual fixing of the extensions of vague predicates. Its semantics for a formal language containing vague predicates is highly non-standard, worth pursuing for even someone who questions the assumptions the semantics is based on. Some technical aspects of this semantics are made explicit in an appendix, added in this reprint. If a reasonable solution could be found using the fixed point logics the paper defines with some reasonable inferential properties, such a solution might still qualify as illuminating, since it has been obtained under what must be about the most unaccommodating preconditions. The two

papers can both be regarded as precursors of a growing trend of seeing context-dependence as a central aspect of vagueness in natural language predicates and to identify and disentangle the different contextual factors that can influence the use and interpretation of such predicates.

Part Four ***Discourse Representation*** starts with the famous 1981 paper *A Theory of Truth and Semantic Representation* which first outlined Discourse Representation Theory (DRT) as the dynamic theory of interpretation, relying essentially on a level of representation to characterize not only inference and entailment, but also cross-sentential pronominal dependencies (anaphora) and other forms of constraints on binding. The paper explicitly sets out to unite truth-functional aspects of meaning with the more pragmatic or epistemic aspects of what language users do when they come to understand what a clause means in the context in which it is used, presenting the account as a rule based fragment of a syntax-semantics mapping reminiscent of Montague Grammar. The paper *Context, Thought and Communication*, reverts to some classical puzzles in the philosophy of language concerning referential identity, suggesting that anchoring a discourse referent in DRT may provide us with the right concept to identify referents across belief contexts and other attitudes. *Disambiguation in Discourse* was written some ten years after the '81 paper on DRT, and argues for the importance of lexical meaning in determining content and the fact that cognition and interpretation are necessarily and intrinsically intertwined.

Part Five ***Presupposition*** contains two influential papers on how contexts may be adjusted to fit the requirements of a phrase to be interpreted by not only verifying whether its presuppositions are satisfied by the context, but also repairing it whenever it does not satisfy these by accommodating information that is otherwise thought of as entailed by the discourse. In *Presupposition Computation and Presupposition Justification: One Aspect of the Interpretation of Multi-Sentence Discourse* this is accomplished by constructing preliminary DRSs and interpreting the novel phrase against these as context. In *The Importance of Presupposition* the interaction between various presuppositions triggered by one phrase are investigated and computation of presupposition is distinguished fundamentally from their justification in context. Part Six ***Propositional Attitudes*** presents two papers on the logic of epistemic attitudes, where *Prolegomena to a Structural Account of Belief and Other Attitudes* goes back to older arguments that any theory that identifies epistemic attitudes with sets of possible worlds will be unable to explain behavioral differences that may come about, if logically equivalent propositions remain indistinguishable. To account for how beliefs guide our actions, an extension of DRT is proposed to deal with the more structural aspects of attitudes, as well as with indexicals, interactive forms of conversational

exchange and shared information. The final paper *Temporal Reference Inside and Outside Propositional Attitudes* combines the DRT account of presupposition with its analysis of epistemic attitudes and attributions thereof to others, exploiting external anchors to formally effectuate referential binding across such contexts. Notions of doxastic strength and persistence of beliefs provide novel ways to formally account for dynamic belief revision in DRT and how people actually may come to understand each other. A long, in-depth interview with Hans Kamp is included at the end of the book. It reviews Kamp's entire intellectual development in its historical and often very personal context, while discussing core examples that drove the development of linguistic semantics and assessing their impact on the way DRT became seen as the linguist's tool of preference to analyze meaning, interpretation and inference as the triad of human language understanding.

This editorial project has demanded a huge amount of support from many of our good colleagues. For their expedient help in proofreading we express our profound gratitude to Daniel Altshuler, Atle Grønn, Fritz Hamm, Katja Jasinskaja, Ewan Klein, Emar Maier, David Nicolas, Tillmann Pross, Galit Sassoon, Dominik Schlechtweg, Ken Turner, Frank Veltman and Tyler Wilbers. The production process, shared between Köln, Stuttgart and Konstanz, has been very efficiently supported by Susanne Trissler, who also researched and perfected all the bibliographic details. To Katrin Schneider (U. Stuttgart) we owe much more as editors, for she sustained us superbly over two years with her expertise in producing beautiful Latex copy, managing a myriad of successive versions of various parts of the book. We thank Stephanie Paalvast, our CRiSPI editor at the publisher Brill in Leiden (NL), for her steadfast professional support, patient understanding during production and also for some material support of it, and the team from TAT Zetwerk for the very professional typesetting of the book. We also acknowledge the generous financial support of the Institute für Deutsche Sprache und Literatur I at the University of Cologne and the collaborative research center 732 *Incremental Specification in Context* at the University of Stuttgart, which hosts the comprehensive archive for Hans Kamp's work and publications, http://www.ims.uni-stuttgart.de/archiv/kamp, with much more content supplementing the present selection. With his characteristic vigor Hans Kamp played the recursive role of author, editor, and critical commentator on his own original texts. The eternal braiding of looping historical narrative could have been extended much longer, had we not ended it by editorial fiat, for which only we deserve any blame.

Geneva and Stuttgart/Köln, March 2013

PART ONE

EVENTS, TEMPORAL REFERENCE, AND DISCOURSE

INTRODUCTION TO EVENTS, TEMPORAL REFERENCE, AND DISCOURSE

Hans Kamp

These three articles cover between them considerably more time that any other of the six parts of this collection.

Formal Properties of 'Now' is the oldest and perhaps also the most demanding paper in the collection, at least for someone who has the patience and stamina to have a closer look at the theorems and their proofs. From a formal point of view its results are still intriguing in that they show that the operator (*it is*) *now* (*the case that*) is eliminable from systems of propositional logic in an extremely general sense, but when quantifiers are added this is no longer so. Once you have seen why quantifiers block the elimination procedures for the propositional case, you may feel like you have understood why elimination of the *now* operator from formulas of tense predicate logic isn't possible in general. But it turns out to be quite hard to prove that this is so. There may be simpler ways to establish this result than the proofs given in the paper, but so far I haven't seen one.

In a way this paper is strangely at odds with developments in the domain to which it was meant to be relevant—the properties of and interactions between temporal and aspectual expressions in natural languages such as English. The central idea of using double indexing to treat *now* as a sentential operator, rather than as an ('indexical') individual constant, dates back to 1967, when tense logic had only recently been discovered and looked like a very powerful and conceptually plausible tool for investigating the role of time and tense in language and thought. But it was realized not long after, and even less time after the appearance of the paper in print, that from a conceptual and linguistic point of view treating *now* as an operator doesn't have much to recommend it. Using double indexing may have been a bit of a technical innovation at the time and computer scientists have picked it up for other reasons and purposes. But if we want to do justice to what *now* does in language, and arguably also in thought, the older and simpler assumption that it is an indexical constant will do much better.

By the time *Events, Instants and Temporal Reference* was written, this battle had already been irrevocably lost by the aficionados of operators. Adverbials that look like they refer to dates, such as *August 1914*, *tomorrow*

© HANS KAMP, 2013 | DOI:10.1163/9789004252882_003
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

night, *the first of April*, *half an hour ago*, do not only look like referring terms; refer to instants and periods of time is what they actually do. And *now* is in this regard no different from all the others. It is part and parcel of this view that what such expressions refer to are first class citizens of the ontology, along with *you, me, Mount Everest* and the rest. What was still much in dispute at this time was what tenses, as well as expressions like temporal adverbials and temporal conjunctions, exactly have to do with states and events. The literature on aspect seemed to take it for granted that verbs and their projections describe such entities as states, events and processes. The question here was just what kinds of entities are described by different verbs and by the phrases built from them. But the scientific community that had only recently struggled free from the nearly compulsive idea that everything in the temporal domain should be understood in terms of overt or concealed operators hadn't yet quite graduated to such a conception. Times had been, if reluctantly, accepted, but they were still thought of as times at which certain propositions are true, or at which certain objects satisfy certain predicates. That verbs, tenses, temporal adverbs were all bound together in a complex grammatical structure for describing eventualities of various kinds and for locating them temporally—in relation to each other, to us speakers and hearers and along the line of our unfolding narrative history—was a further step, which I at least wasn't able to take without a further struggle. *Events, Instants and Temporal Reference* was part of that struggle. It developed out of the intuition that tense forms often have to tell us as much about aspect as they tell us about temporal location. Crucially for my own further development, some of the clearest manifestations of how they can do those two things at the same time is their capacity to connect what is described by their verbs to what was described in preceding sentences of the text or discourse. Making discourse the focus of linguistic and conceptual analysis emerged as the key to getting a theoretical grip on this aspect of the contributions that tenses and other linguistic devices of temporal reference can make to discourse meaning and interpretation. This concern with the discourse-related dimensions of tense and aspect naturally led to the conviction that discourse must play a central part in the theory of linguistic meaning generally. It would be another two years before I found what seemed convincing evidence for this suspicion in anaphora. That discovery also enabled me to give a formal discourse-based semantics the more definitive shape of DRT in *A Theory of Truth and Semantic Representation*.

The main thrust of *Deixis in Discourse. Reichenbach on Temporal Reference* is backward-looking. The paper is an elaboration of the Reichenbach Lecture of 1999, delivered at UCLA in the spring of that year. It tries to trace and

explain the importance and huge influence of the comparatively short chapter on tense in Hans Reichenbach's ***Elements of Symbolic Logic***. It always struck me as curiously at odds with what seem to be the general aims of the book, and precisely because of that, a testimony to the author's genius. What Reichenbach discovered when he looked into the semantics of tense more closely may not have been at all what he was expecting, when he tackled this topic. But that was the way it was and whether that fitted the general conception of the book, or the general conception with which the project of the book was undertaken, was secondary. What is conceptual unity as compared to truth?

Reichenbach's contribution to the semantics of temporal reference in natural languages is his two-dimensional theory of tense, according to which each tense form of English makes a semantic contribution consisting of two temporal relations, one between speech time and reference time and one between reference time and event time. *Deixis in Discourse. Reichenbach on Temporal Reference* acknowledges this signal contribution, but is at the same time a critical appraisal of it and in particular of Reichenbach's notion of reference time. Sometimes, my colleagues and I came to realize in the early eighties, the interpretation of a tense involves more than one reference time, with different reference times playing different roles. We saw the need to differentiate between *Temporal Perspective Times*—these play the part of Reichenbach's reference times in his two-dimensional schemata for the semantics of tense forms—and *Reference Times* which, for instance, mark temporal progress in simple forms of past tense narrative prose. Recently—I am speaking of the past couple of years—I have come to have my doubts whether this complication is really desirable. A thorough review of what we had been thinking and saying for many years about the Perfect now makes me think that no distinction between temporal perspective time and reference time is needed after all. This new perspective also entails that the tenses aren't really two-dimensional in the way Reichenbach thought. There remains the central discovery he made, that reference times distinct from the time of speech are indispensable to a correct analysis of the range of tenses we find in languages like English, French and many others. I am firmly persuaded that it will be with us for good.

FORMAL PROPERTIES OF 'NOW'[a]

Hans Kamp

To the memory of Arthur Prior

This paper is about the word 'now'. It is closely related to the article *"Now"* by professor A.N. Prior (1968). In that article Prior gives an extensive, and undoubtedly correct analysis of the semantical function of the word 'now' in ordinary discourse. He then develops a number of logical calculi which contain formal counterparts of the word 'now', as well as of certain other temporal notions and the truth-functional connectives. In some of these calculi 'now' is formalized as a 1-place propositional connective, while in others it is represented by a propositional constant. As I believe that 'now'—like so many adverbs and adverbial clauses—should be regarded as a propositional modifier, I prefer the calculi of the first sort; and thus I will restrict my attention to them.

I will for these, and similar, calculi formulate the semantics, and then prove a number of metamathematical results about them. The most important of these results have the following form: Let $\mathcal{L}(N)$ be such a calculus, and let $\mathcal{L}$ be the calculus obtained by omitting the 'now'-operator from $\mathcal{L}(N)$. If an axiom system $\mathcal{A}$ for $\mathcal{L}$ is semantically complete then so is a closely to $\mathcal{A}$ related axiom system $\mathcal{A}'$ for $\mathcal{L}(N)$.

These results provide us with a number of different completeness theorems for some of the 'now'-calculi considered, since for the corresponding calculi without the 'now'-operator, many complete axiom systems are already available.

---

[a] Editors' note: This article was originally published in *Theoria*, 37(3), 1971, 227–273. We would like to thank John Wiley and Sons for their kind permission to reprint the article. We carefully edited the article, corrected some typos, added full bibliographical references and brought some formulas in line with now more common usage. Hans Kamp provided some additional footnotes, indicated by roman letters, which clarify some of the arguments in the original text. We would like to thank Frank Veltman and Tyler Wilbers for careful proofreading and constructive comments.

© HANS KAMP, 2013 | DOI:10.1163/9789004252882_004
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

For propositional calculi these results are quite easily obtained. For in these calculi the 'now'-operator is, as it will turn out, always eliminable—i.e., every formula containing the operator is equivalent to a formula in which the operator does not occur. Thus, to obtain from a given axiom system $\mathcal{A}$ for the calculus without 'now' a complete axiom system for the calculus with 'now' it suffices to add axioms which make it possible toprove the equivalence between any formula and its 'now'-free equivalent.

In the predicate calculi which we will consider a formula containing 'now' is not always equivalent to a formula without 'now'. This fact is intuitively obvious, but nonetheless somewhat difficult to show. A proof is given in the last section. Because not all formulae of our predicate calculi are equivalent to 'now'-free formulae the completeness result has to be proved by means other than those used for the propositional calculi. The results obtained are somewhat less general than those for the propositional cases.

Before I proceed with the formal part of this chapter I will, on the danger of repeating some of the points of Prior's article, give a short informal discussion of the behaviour of 'now' in English. My reason for this is twofold. In the first place I hope that a few informal remarks will make it easier to understand the formal definitions of the semantics which will follow later. In the second place I want to make it clear—before I embark upon technical developments which otherwise might seem pointless—that the word 'now' is not vacuous, in the sense that whenever someone makes a true, or false, statement by uttering a certain sentence in which the word 'now' occurs, he would also have made a true, or false, statement if he had uttered instead the sentence which is obtained if the word 'now' is omitted from the first sentence.

Some people have indeed thought that 'now' is vacuous in this sense. If they had been right this paper should not have been written. But they are not. To see this let us consider an argument which is sometimes given in support of the view that 'now' is vacuous. It starts from the following observation: Suppose that I make a true statement by uttering at time t a certain sentence, e.g., the sentence 'it is raining'. Then I would also have made a true statement if I had uttered at t the words 'it is now raining'. Similarly, if the statement made by uttering the first sentence had been false, then so would have been the statement made by uttering the second sentence. This observation is certainly correct. And it remains correct if we replace the words 'it is raining' by any other English sentence in the present tense. But it is wrong to conclude from this that all occurrences of the word 'now' are vacuous. In fact, consider the sentences:

(1) 'I learned last week that there would be an earthquake.'
(2) 'I learned last week that there would now be an earthquake.'

Obviously there could be circumstances under which I would make a true statement if I uttered the first sentence, but a false one if I uttered the second.

The function of the word 'now' in (2) is clearly to make the clause to which it applies—i.e., 'there would be an earthquake'—refer to the moment of utterance of (2), and *not* to the moment, or moments (indicated by other temporal modifiers that occur in the sentence) to which the clause would refer (as it does in (1)) if the word 'now' were absent. A little reflection shows that this principle correctly describes the function of the word 'now' in all of its occurrences. It explains in particular why the occurrence of 'now' in 'it is now raining' is vacuous. For there the clause to which 'now' applies, viz., 'it is raining', is understood in any case to refer to the moment of utterance, whether 'now' be present or not.

This establishes that the word 'now' does not always occur vacuously. However, we have also seen that an occurrence of 'now' can be only non-vacuous if it occurs within the scope of another temporal modifier. Thus a formal analysis of 'now' will be of any interest only if it takes also other temporal operators into account. As a matter of fact the most interesting non-vacuous occurrences of 'now' are in sentences which contain besides such other temporal operators also propositional modifiers of a non-temporal character, e.g., modal, epistemic, or deontic operators. Sentence (2), in which the operator 'I learn that' occurs, is a case in point.

In this paper I will nonetheless consider besides 'now' only operators of a purely temporal nature. The reason is mainly one of expedience: For those operators which I will consider a comprehensive and satisfactory analysis has already been carried out, and we know a great deal about the systems to which this analysis has led. I will make use of that information to obtain the formal results which will follow. On the other, hand relatively few formal systems are so far available in which temporal as well as non-temporal operators are represented. I believe, however, that the analysis given here will make the extension of such systems with a 'now'-operator a straightforward matter once they will have been developed without 'now'.

My brief statement of the general function of the word 'now' may suggest that, even if the word is not vacuous, its semantical behaviour is too simple to justify a formal analysis. However, it turns out that a proper treatment of 'now' together with the other temporal operators which I will consider is not

totally trivial, as the semantics developed earlier for these operators cannot be extended in an entirely automatic manner so as to cover 'now'.

Nonetheless the reader may still have the feeling that the amount of attention paid to the word 'now' in this paper is excessive. I will end this introduction with a few remarks aimed at dispelling that feeling.

In the first place it should be observed that the feature which distinguishes 'now' from those temporal modifiers which had already been satisfactorily treated previously is to be found also in a number of other temporal concepts, e.g., in those expressed by the words 'today', 'yesterday', 'last week', 'next year', etc., as well as, to some extent, in the word 'then'. Thus the analysis of 'now' given here is also a paradigm for similar analyses of those other concepts.

In the second place I want to point at a phenomenon which is connected with the use in English of the ordinary past and future tenses, but which, to my knowledge, has so far been overlooked by tense logicians, and which is intimately related to the analysis of 'now' given here. I have said repeatedly in this introduction that a satisfactory analysis already exists for a number of temporal notions. Among these notions are the past and future tenses. The past tense is in this analysis represented by a 1-place propositional operator P, which should be thought of as transforming, in particular, sentences in the present tense into the corresponding sentences in the past tense. Thus, $P\varphi$ can be read as 'it was the case that φ', and, in particular, if φ is, e.g., the sentence 'it rains', as 'it rained'. Similarly the future tense is represented by the 1-place operator F, such that $F\varphi$ can be read as 'it will be the case that φ'. The semantics stipulates that a formula $P\varphi$ is true at a moment t if and only if φ is true at some moment preceding t; and that $F\varphi$ is true at t if and only if φ is true at some moment following t. Thus, in particular, $PF\varphi$ is true at t if and only if there is a moment t' before t such that φ is true at some moment later than t'. Now compare the following two sentences:

(3) 'A child was born that would become ruler of the world.'
(4) 'A child was born that will become ruler of the world.'

It is clear that while the first sentence would be true if at some past time t a child was born to become ruler of the world at some time t' later than t—whether that time t' be before, identical with, or later than the present—the second sentence would be true *only* if the child is to become ruler at a time later than the present. It follows that (3) can be correctly rendered in the following form:

$$P(\exists x)(x \text{ is born} \wedge F(x \text{ is ruler of the world})).$$

But for (4) no correct symbolization with the help of only P, F, and the apparatus of ordinary predicate logic is possible.[1] However, if we have at our disposal also the 'now'-operator N, (which will be introduced in Section 2), we can symbolize (4) properly as

$$P(\exists x)(x \text{ is born} \wedge NF(x \text{ is ruler of the world})).$$

The semantics developed in Section 2 will indeed show that this symbolization is adequate.

§ 1

I will assume throughout this paper the existence of an infinite class $\mathcal{E}$ and a well-founded concatenation function $\mathcal{C}$ on $\mathcal{E}$.[2] $\mathcal{E}$ will be referred to as the *set of expressions*. $\mathcal{S}$ will be the class $\mathcal{E} \setminus \text{Range } \mathcal{C}$; its members will be referred to as the *symbols* of $\mathcal{E}$. Whenever $e_1, e_2 \in \mathcal{E}$, we will write $e_1 e_2$ instead of $\mathcal{C}(\langle e_1, e_2 \rangle)$.

We assume that $\mathcal{S}$ contains symbols (,) (called *parentheses*); q_1, q_2, ... (called *propositional constants*); and for each $n \in \omega$ symbols $C_0^n, C_1^n, C_2^n, \ldots$ (called *n-place connectives*). We will refer to $C_0^1, C_1^1, C_2^1, C_3^1, C_0^2$ as $N, \sim, G, H, \wedge$, respectively.

By a *language for propositional tense logic* we understand a set consisting of the symbols (,), q_i ($i = 1, 2, \ldots$) and some of the symbols C_i^n. The *formulae* of a language $\mathcal{L}$ for propositional tense logic are defined by:

Definition 1 (i) q_i is a *formula* of $\mathcal{L}$; (ii) if $C_i^n \in \mathcal{L}$ and $\varphi_1, \ldots, \varphi_n$ are formulae of $\mathcal{L}$ then $C_i^n(\varphi_1, \ldots, \varphi_n)$ is a *formula* of $\mathcal{L}$.

---

[1] I of course exclude the possibility of symbolizing the sentence by means of explicit quantification over moments. Such a symbolization of (2) would certainly be possible; and it would even make the operators P and F superfluous. Such symbolizations, however, are a considerable departure from the actual form of the original sentences which they represent—which is unsatisfactory if we want to gain insight into the semantics of English. Moreover, one can object to symbolizations involving quantification over such abstract objects as moments, if these objects are not explicitly mentioned in the sentences that are to be symbolized.

[2] By a *well-founded concatenation function* $\mathcal{F}$ on a class $\mathcal{S}$ we understand a function from $\mathcal{S} \times \mathcal{S}$ into $\mathcal{S}$, such that:

(i) $\mathcal{F}(\langle \mathcal{F}(\langle s_1, s_2 \rangle), s_3 \rangle) = \mathcal{F}(\langle s_1, \mathcal{F}(\langle s_2, s_3 \rangle)\rangle)$;

(ii) if $\mathcal{F}(\langle s_1, s_2 \rangle) = \mathcal{F}(\langle s_3, s_4 \rangle)$ then either ($s_1 = s_3$ and $s_2 = s_4$) or there is an $s \in \mathcal{S}$ such that $s_1 = \mathcal{F}(\langle s_3, s \rangle)$ and $s_4 = \mathcal{F}(\langle s, s_2 \rangle)$ or there is an $s \in \mathcal{S}$ such that $s_3 = \mathcal{F}(\langle s_1, s \rangle)$ and $s_2 = \mathcal{F}(\langle s, s_4 \rangle)$;

(iii) there is no infinite sequence $s_0, s_1, s_2, \ldots$ of elements of $\mathcal{S}$ such that for each $n \in \omega$ there is a $t \in \mathcal{S}$ such that $s_n = \mathcal{F}(\langle t, s_{n+1} \rangle)$ or $s_n = \mathcal{F}(\langle s_{n+1}, t \rangle)$.

I will always write $(\varphi \wedge \psi)$ instead of $\wedge(\varphi, \psi)$. Furthermore, $(\varphi \vee \psi)$ will stand for $\sim(\sim(\varphi) \wedge \sim(\psi))$; $(\varphi \rightarrow \psi)$ will stand for $(\sim(\varphi) \vee \psi)$; $(\varphi \leftrightarrow \psi)$ will stand for $((\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi))$; $P\varphi$ for $\sim(H(\sim(\varphi)))$; and $F\varphi$ for $\sim(G(\sim\varphi))$. Parentheses will be omitted whenever no confusion is possible. In particular I will always write $\sim\varphi, N\varphi, H\varphi, G\varphi$ instead of $\sim(\varphi), N(\varphi), H(\varphi), G(\varphi)$, respectively. The formulae $N\varphi, H\varphi, G\varphi$ are read as 'it is now the case that φ', 'it has always been the case that φ', and 'it will always be the case that φ', respectively.

T will always be a non-empty set and $<$ a partial ordering (i.e., a transitive and asymmetric relation) on T. $\mathcal{T}$ will be the pair $\langle T, <\rangle$. We think of T as the set of moments and of $<$ as the earlier-later relation between them.[3] The numbers 0 and 1 will be used as truth-values, 1 for truth and 0 for falsehood.

Before turning to the 'now'-calculus itself, I will first consider a simpler system, $\mathcal{L}_1$, the connectives of which are $\sim, \wedge, H, G$. As the syntax of $\mathcal{L}_1$ has already been defined, I will proceed at once with the semantics.

I feel that this paper—which is concerned with the specific properties of 'now'—is not the place where I should give an extensive justification of the principles on which tense logic is based and which underlie in particular its semantics. Therefore I will restrict myself to a few explanatory remarks to make the following formal definitions more understandable.

Tense logic belongs to a type of logic in which is explicitly adopted a feature of natural languages that has no part in the standard systems of mathematical logic (as, e.g., the ordinary predicate calculus): In natural languages we can in many cases use the same linguistic form under different circumstances—and, in particular, at different times—to make different statements. So we may use the expression 'it rains' at one time to make a true statement and at another to make a false one. Since the truth-value of the statement made by using the expression at t varies with t, the expression can not be said to have a truth-value all by itself. However, if we assume that the truth-value of the statement is completely determined by (a) the expression used and (b) the time of utterance,[4] we may conclude that the truth-value

[3] Unlike Prior I assume, throughout this paper, that the relation 'earlier than' between moments of time is a partial ordering. My reason is this: Tense logic is concerned with the analysis of the logical properties of certain temporal notions. Such an analysis must perforce start from the naïve intuitions which we have about these notions. Now, in my opinion, our belief that the earlier-later relation between moments is a partial ordering plays such a central role in the network of intuitions which we have about temporal notions, that unless we uphold that belief we have no ground for relying, to any degree whatever, on our intuitions about the notions which we want to analyze.

[4] This assumption is of course an idealization, which precludes the proper treatment of

of the statement made can be regarded as a function of only these two factors; and thus we may consider for every (appropriate) expression the concept of 'the *truth-value of* that expression *at* an arbitrary moment *t*'. With the help of this concept of 'being true at *t*' we can give a semantical analysis of those temporal notions with which tense logic is concerned, e.g. the notion represented in $\mathcal{L}_1$ by the operator H. For we can state in a general and systematic way how the truth-value of a formula $H\varphi$ at t depends on the truth-values of φ at *t and at other moments*. For this purpose we need semantic structures which are not the models of standard model theory—which simply specify a truth-value for each atomic formula of the language that they interpret—but rather structures which specify for every formula a truth-value at each moment of time. This is just what the interpretations defined below do.

Definition 2 $\mathcal{M}$ is an *interpretation for* $\mathcal{L}_1$ *relative to* $\mathcal{T}$, iff $\mathcal{M}$ is a function such that

(1) the domain of $\mathcal{M}$ consists of $q_1, q_2, \ldots$; and
(2) for each $i > 0$ $\mathcal{M}(q_i)$ is a function with domain $\mathcal{T}$ and range included in $\{0, 1\}$.

On the basis of the preceding remarks the truth definition for complex formulae is straightforward given the intended meanings of the connectives $\sim, \wedge, H, G$

Definition 3 Let $\mathcal{M}$ be an interpretation for $\mathcal{L}_1$, relative to $\mathcal{T}$. For any formula φ of $\mathcal{L}_1$ and $t \in T$ the *truth-value of* φ *in* $\mathcal{M}$ *at t, relative to* $\mathcal{T}$, (in symbols: $[\varphi]^{\mathcal{T}}_{\mathcal{M},t}$) is defined as follows:

(1) $[q_i]^{\mathcal{T}}_{\mathcal{M},t} = 1$ iff $\mathcal{M}(q_i) = 1$;
(2) if φ, ψ are formulae of $\mathcal{L}_1$, then

(i) $[\sim\varphi]^{\mathcal{T}}_{\mathcal{M},t} = \begin{cases} 1 \text{ if } [\varphi]^{\mathcal{T}}_{\mathcal{M},t} = 0 \\ 0 \text{ otherwise} \end{cases}$

(ii) $[\varphi \wedge \psi]^{\mathcal{T}}_{\mathcal{M},t} = \begin{cases} 1 \text{ if } [\varphi]^{\mathcal{T}}_{\mathcal{M},t} = 1 \text{ and } [\psi]^{\mathcal{T}}_{\mathcal{M},t} = 1 \\ 0 \text{ otherwise} \end{cases}$

(iii) $[H\varphi]^{\mathcal{T}}_{\mathcal{M},t} = \begin{cases} 1 \text{ for if all } t' \in T \text{ such that } t' < t\ [\varphi]^{\mathcal{T}}_{\mathcal{M},t'} = 1 \\ 0 \text{ otherwise} \end{cases}$

---

large portions of natural languages, viz. of those portions where the truth-value of the statement made depends not only on the form of the expression used and the time of utterance, but also on other aspects of the situation, e.g. the identity of the speaker—compare, e.g., the sentence 'I am hungry'. However, to the analysis of the purely temporal notions with which tense logic is concerned this idealization will do no harm.

(iv) $[G\varphi]^{\mathcal{T}}_{\mathcal{M},t} = \begin{cases} 1 \text{ if for all } t' \in T \text{ such that } t < t' \ [\varphi]^{\mathcal{T}}_{\mathcal{M},t'} = 1 \\ 0 \text{ otherwise} \end{cases}$

Definition 4 A formula φ of $\mathcal{L}_1$ is *valid, relative to* $\mathcal{T}$ iff for every interpretation $\mathcal{M}$ for $\mathcal{L}_1$, relative to $\mathcal{T}$, and every $t \in T$, $[\varphi]^{\mathcal{T}}_{\mathcal{M},t} = 1$.

Clauses 2.(iii) and 2.(iv) of definition 3 suggest that the set of valid formulae of $\mathcal{L}_1$ may depend on the structure of $\mathcal{T}$. This is indeed the case: There are partial orderings $\mathcal{T}$ and $\mathcal{T}'$ such that the set of valid formulae, relative to $\mathcal{T}$, is different from the set of valid formulae, relative to $\mathcal{T}'$. Thus, intuitively, we may, in as far as we are ignorant of the structure of time, well be unable to determine which formulae are to be regarded as intuitively 'tense logically valid' (i.e., valid on the basis only of how propositional connectives and tenses occur in them). However, once we assume that time has *certain* properties (e.g., that it is dense) we can regard at least some formulae as tense-logically valid, viz., those which are valid relative to all partial orderings which have these properties. It is thus natural to introduce the following notion of validity.

Definition 5 Let $\mathcal{K}$ be a non-empty class of partial orderings. A formula φ of $\mathcal{L}_1$ is $\mathcal{K}$*-valid* iff for every $\mathcal{T} \in \mathcal{K}$, φ is valid, relative to $\mathcal{T}$.

There is a slightly different, but obviously equivalent, way in which we can develop the semantics for $\mathcal{L}_1$. This development makes use of a new kind of interpretation. In order to avoid ambiguity I will call interpretations of this new kind *interpretations*$_2$, and I will refer to the interpretations of Definition 3 as *interpretations*$_1$.

Definition 6 An *interpretation*$_2$ *for* $\mathcal{L}_1$, *relative to* $\mathcal{T}$, is an ordered pair consisting of an interpretation$_1$ for $\mathcal{L}_1$, relative to $\mathcal{T}$, and a member of T.

Definition 7 For every interpretation$_2$ $\langle \mathcal{M}, t_0 \rangle$ for $\mathcal{L}_1$, relative to $\mathcal{T}$, any formula φ of $\mathcal{L}_1$ and any $t \in T$, the *truth-value of* φ *in* $\langle \mathcal{M}, t_0 \rangle$ *at* t, *relative to* $\mathcal{T}$ (in symbols $[\varphi]^{\mathcal{T}}_{\langle \mathcal{M},t_0\rangle,t}$) is the truth-value of φ in $\mathcal{M}$ at t, relative to $\mathcal{T}$.

Definition 8 φ is *valid*$_2$, *relative to* $\mathcal{T}$ iff for every interpretation$_2$ $\langle \mathcal{M}, t_0 \rangle$ for $\mathcal{L}_1$, relative to $\mathcal{T}$, $[\varphi]^{\mathcal{T}}_{\langle \mathcal{M},t_0\rangle,t_0} = 1$.

As can be seen from Definitions 2, 3, 4, 6, 7, 8 the two developments differ only in their respective characterizations of validity in terms of truth. While according to the first characterization (Definition 4) a formula is valid only if it is true in each interpretation at each moment, the second characterization (Definition 8) demands only that the formula be true in each interpretation at one particular moment. One may think of that moment as the 'present' of the interpretation in question. Thus Definition 4 is based on the idea that

we are, so to speak, interested only in what formulae are true at the present time; the truth-values of formulae at other moments are important only in so far as they determine the truth-values of certain more complex formulae at this present time.

At this point the two approaches are of course trivially equivalent. But we will see later that when we add the operator N to $\mathcal{L}_1$, to represent 'now', the difference between the two approaches becomes important; and that while the first is probably intuitively the more natural, the second leads to a considerable technical simplification of which we will make use in subsequent proofs.

Finding axiom systems which generate the 'valid' formulae of $\mathcal{L}_1$ is a complex task. Indeed, every particular partial ordering $\mathcal{T}$ gives rise to an axiomatization problem of its own—viz., the problem of finding an axiom system that will generate all and only those formulae which are valid relative to $\mathcal{T}$—and so does every non-empty class of partial orderings. Problems of this sort, however, will not be of my concern in this paper. Rather I will show how an axiom system $\mathcal{A}$ or $\mathcal{L}_1$ can be modified into an axiom system $\mathcal{A}'$ for the 'now'-calculus defined below, so that if $\mathcal{A}$ generates the set of all formulae of $\mathcal{L}_1$ which are valid relative to $\mathcal{T}$ [or $\mathcal{K}$-valid] then $\mathcal{A}'$ will generate the set of all formulae of the 'now'-calculus which are valid relative to $\mathcal{T}$ [or $\mathcal{K}$-valid]. Since for several partial orderings and classes of partial orderings axiom systems generating the corresponding sets of valid formulas have already been given by others, this procedure will provide us with an equal number of complete axiom systems for the 'now'-calculus.

We now turn to the 'now'-calculus itself. Let $\mathcal{L}_2$ be the language $\mathcal{L}_1 \cup \{N\}$. The *interpretations for* $\mathcal{L}_2$, *relative to* $\mathcal{T}$, are simply the interpretations for $\mathcal{L}_1$ relative to $\mathcal{T}$.

The truth definition for $\mathcal{L}_2$, however, cannot be obtained by a straightforward adaptation of the corresponding definition for $\mathcal{L}_1$ (Definition 3). The difficulty stems from the peculiar behavior of the word 'now', which our truth definition should reflect. An essential feature of the word 'now' is that it always refers back to the moment of utterance of the sentence in which it occurs, even if it stands itself in that sentence within the scope of one or more tenses. It is this feature that makes the English counterparts of, e.g., $\varphi \rightarrow N\varphi \wedge HN\varphi \wedge GN\varphi$ ('if it is the case that φ, then it is now the case that φ, it always has been the case that it is now the case that φ, and it always will be the case that it is now the case that φ') logically true. Thus our truth definition should be such that it makes in particular this formula valid. Such a truth definition cannot be obtained simply by adding to Definition 3 a clause of the form:

(1) $[N\varphi]_{\mathcal{M},t} = \begin{cases} 1 \text{ iff } \Phi; \\ 0 \text{ otherwise} \end{cases}$

For if the definition correctly reflects the behavior of 'now' then it should make the formula

(2) $q_0 \leftrightarrow Nq_0$

valid, relative to any partial ordering, as any English counterpart of this formula is clearly true, irrespective of the structure of time. This implies that the condition Φ in (1) should be equivalent to the condition that $[\varphi]^{\mathcal{T}}_{\mathcal{M},t} = 1$. But if Φ is equivalent to this condition, then the formula $q_1 \rightarrow Nq_1 \wedge HNq_1 \wedge GNq_1$ will be not valid relative to any partial ordering $\mathcal{T}$ which has at least two points. For in any interpretation based upon such a $\mathcal{T}$ in which q_1 is true at only one moment, t say, Nq_1 will, according to (1), also be true only at t, and therefore $q_1 \rightarrow Nq_1 \wedge HNq_1 \wedge GNq_1$ will be false at t. Thus an adequate truth definition of this form cannot be found.

This argument may suggest where the difficulty lies: the truth-values of $HNq_1(GNq_1)$ depend on the truth values of Nq_1 in a way which is different from the manner in which the truth values of $Hq_1(Gq_1)$ depend on the truth values of q_1; and any definition obtained by adding a clause of the form (1) to Definition 3 will be incapable of doing justice to that difference. In order to find an appropriate truth definition let us recall the remark, made above, that 'the word 'now' refers back to the moment of utterance'. In view of this fact we should, if we want to analyze the truth of formulae that contain N in terms of the truth-values of their components, 'keep track' during this analysis of the moment of utterance of the entire expression. The concept we ought to analyze is not simply 'the truth-value of φ at t', but rather 'the truth-value of φ at t when part of an utterance made at t''. Of course, our real interest is in the truth-value of a sentence at the moment of *its* utterance. But the analysis of this truth-value in terms of the truth-values of the components of the sentence will automatically lead to the consideration of truth-values of formulas at moments different from the moment of their utterance. We thus come to the following definition:

Definition 9 Let $\mathcal{T}$ be a partial ordering. Let $\mathcal{M}$ be an interpretation of $\mathcal{L}_2$, relative to $\mathcal{T}$. For any formula φ of $\mathcal{L}_2$, and $t, t' \in T$, *the truth-value of φ in $\mathcal{M}$ at t when part of an expression uttered at t', relative to $\mathcal{T}$* (in symbols: $[\varphi]^{\mathcal{T}}_{\mathcal{M},t,t'}$) is defined as follows:

(1) $[q_i]^{\mathcal{T}}_{\mathcal{M},t,t'} = \mathcal{M}(q_i)(t)$;
(2) if φ, ψ are formulas of $\mathcal{L}_2$, then

(i) $[\sim\varphi]^{\mathcal{T}}_{\mathcal{M},t,t'} = \begin{cases} 1 \text{ if } [\varphi]^{\mathcal{T}}_{\mathcal{M},t,t'} = 0; \\ 0 \text{ otherwise;} \end{cases}$

(ii) $[(\varphi \wedge \psi)]^{\mathcal{T}}_{\mathcal{M},t,t'} = \begin{cases} 1 \text{ if } [\varphi]^{\mathcal{T}}_{\mathcal{M},t,t'} = 1 \text{ and } [\psi]^{\mathcal{T}}_{\mathcal{M},t,t'} = 1; \\ 0 \text{ otherwise;} \end{cases}$

(iii) $[H\varphi]^{\mathcal{T}}_{\mathcal{M},t,t'} = \begin{cases} 1 \text{ if for all } t'' \in T, \text{ such that } t'' < t, [\varphi]^{\mathcal{T}}_{\mathcal{M},t'',t'} = 1; \\ 0 \text{ otherwise;} \end{cases}$

(iv) $[G\varphi]^{\mathcal{T}}_{\mathcal{M},t,t'} = \begin{cases} 1 \text{ if for all } t'' \in T \text{ such that } t < t'', [\varphi]^{\mathcal{T}}_{\mathcal{M},t'',t'} = 1; \\ 0 \text{ otherwise;} \end{cases}$

(v) $[N\varphi]^{\mathcal{T}}_{\mathcal{M},t,t'} = \begin{cases} 1 \text{ if } [\varphi]^{\mathcal{T}}_{\mathcal{M},t',t'} = 1; \\ 0 \text{ otherwise;} \end{cases}$

A formula should be regarded as valid if in every interpretation it is *true when uttered*. Thus validity should be defined as follows:

Definition 10 A formula φ of $\mathcal{L}_2$ is *valid, relative to* $\mathcal{T}$, iff for every interpretation $\mathcal{M}$ for $\mathcal{L}_2$ relative to $\mathcal{T}$, and every $t \in T$, $[\varphi]^{\mathcal{T}}_{\mathcal{M},t,t} = 1$.

Since $\mathcal{L}_1$ and $\mathcal{L}_2$ have many formulae in common, it is conceivable that Definitions 9 and 10 clash with Definitions 3 and 4; a formula of both $\mathcal{L}_1$ and $\mathcal{L}_2$ could be valid, relative to $\mathcal{T}$, according to Definition 4, but not valid according to Definition 10; or vice versa. Such a clash would of course imply that at least one of the definitions is inappropriate. However, this is not the case. One can easily show (we omit the proof) that if φ is a formula of both $\mathcal{L}_1$ and $\mathcal{L}_2$ then φ is valid, relative to $\mathcal{T}$, according to Definition 4 if and only if φ is valid, relative to $\mathcal{T}$, according to Definition 10.

We will now consider the analogues for $\mathcal{L}_2$ of the interpretations for $\mathcal{L}_1$.

Definition 11 An *interpretation*$_2$ for $\mathcal{L}_2$, *relative to* $\mathcal{T}$, is a pair consisting of an interpretation$_1$ for $\mathcal{L}_2$, relative to $\mathcal{T}$, and a member of T.

As I said before when considering interpretations, we are primarily interested in the truth-values of formulae at only one particular moment. Therefore we can avoid the complications that arose in connection with the truth definition for $\mathcal{L}_2$ applicable to interpretations$_1$. Indeed, the truth definition can be given as a simple extension of Definition 7.

Definition 12 Let $\mathcal{M} = \langle \mathcal{M}', t_0 \rangle$ be an interpretation$_2$ for $\mathcal{L}_2$, relative to $\mathcal{T}$. For any formula φ of $\mathcal{L}_2$ and $t \in T$, the *truth-value of* φ *in* $\mathcal{M}$ *at* t, *relative to* $\mathcal{T}$, (in symbols: $[\varphi]^{\mathcal{T}}_{\mathcal{M},t}$) is defined as follows:

(1) (2.i–iv): as in Definition 3,

(2) (v) $[N\varphi]^{\mathcal{T}}_{\mathcal{M},t} = \begin{cases} 1 \text{ if } [\varphi]^{\mathcal{T}}_{\mathcal{M},t_0} = 1; \\ 0 \text{ otherwise.} \end{cases}$

As before validity$_2$ is defined by:

Definition 13 A formula φ of $\mathcal{L}_2$ is *valid*$_2$, *relative to* $\mathcal{T}$, or $\mathcal{T}$*-valid*$_2$, if for every interpretation$_2 \langle \mathcal{M}, t \rangle$ relative to $\mathcal{T}$, $[\varphi]_{\langle \mathcal{M},t \rangle,t} = 1$.

It is obvious that this definition does not conflict with Definition 8: every formula of $\mathcal{L}_2$ which is also a formula of $\mathcal{L}_1$ is valid$_2$ in the sense of Definition 8 iff it is valid$_2$ in the sense of Definition 13. This follows from the fact that for every interpretation$_1$ $\mathcal{M}$ for $\mathcal{L}_2$, relative to $\mathcal{T}$ and $t_0, \in T$, and every formula φ of $\mathcal{L}_1$,

$$[\varphi]^{\mathcal{T}}_{\mathcal{M},t} = [\varphi]^{\mathcal{T}}_{\langle \mathcal{M}, t_0\rangle, t}$$

(which can be shown by an easy induction argument, omitted here).

In the proofs of the theorems below it will be somewhat more convenient to work with interpretations$_2$ than with interpretations$_1$. Therefore we now drop interpretations$_1$ altogether and will refer to interpretations$_2$ simply as *interpretations*. Also we will speak of *validity* instead of validity$_2$.

§ 2

As I have said in the introduction, the results presented in this paper apply to arbitrary axiom systems. Also, I will later introduce unusual forms of proof from axiom systems. In view of these two facts it will be necessary to give a precise account of the general notion of an *axiom system* and of a *proof from* an axiom system which will be used in the sequel. Definition 15 provides this account. Definition 14, which precedes it, is concerned with the notion of substitution of formulae for propositional constants which is essential for that account.

Definition 14 Let $\varphi, \psi_1, \ldots, \psi_k$ be formulae of some language $\mathcal{L}$ for propositional tense logic, and let $i_1, \ldots, i_k$ be positive integers. By $[\varphi]\psi_1/q_{i_1}, \ldots, \psi_k/q_{i_k}$ we understand the result of replacing in φ q_{i_1} everywhere by $\psi_1, \ldots, q_{i_k}$ by ψ_k. $[\varphi]\psi_1/q_{i_1}, \ldots, \psi_k/q_{i_k}$ is called an *instance of* φ *in* $\mathcal{L}$.[b]

Definition 15

(1) An *inference rule in* $\mathcal{L}$ is a pair consisting of a finite set of formulae of $\mathcal{L}$ and a formula of $\mathcal{L}$. If $R = \langle \Sigma, \varphi\rangle$ is an inference rule in $\mathcal{L}$, we call the members of Σ the *premisses* of R and φ the *conclusion* of R. In case Σ is empty, R is called an *axiom* (*in* $\mathcal{L}$).

(2) An *axiom system for* $\mathcal{L}$ is a set of inference rules in $\mathcal{L}$.

(3) Let $R = \langle \Sigma, \varphi\rangle$ be an inference rule in $\mathcal{L}$. Let Γ be a set of formulae of $\mathcal{L}$, ψ a formula of $\mathcal{L}$. We say that ψ *follows from* Γ *by* R iff there are propositional constants $q_1, \ldots, q_k$, and formulae $\psi_1, \ldots, \psi_k$ of $\mathcal{L}$, such that $\psi = [.. [[\varphi]\psi_1/q_1 \ldots] \psi_k/q_k]$, and for each φ' in Σ $[.. [[\varphi']\psi_1/q_1 \ldots]\psi_k/q_k]$ belongs to Γ.

---

[b] By $[\varphi]\beta/\alpha$ I understand the result of proper substitution of β for α in φ. '$\phi \vee \psi$', '$\phi \rightarrow \psi$', '$\phi \leftrightarrow \psi$' are now abbreviations for '$\sim(\sim\phi \wedge \sim\psi)$', '$\sim(\phi \wedge \sim\psi)$' and '$(\phi \leftrightarrow \psi) \wedge (\psi \leftrightarrow \phi)$'.

(4) Let $\mathcal{A}$ be an axiom system for $\mathcal{L}$. A *proof from $\mathcal{A}$ in $\mathcal{L}$* is a finite sequence of formulae of $\mathcal{L}$ such that each member of the sequence follows from the preceding members in the sequence by one of the inference rules in $\mathcal{A}$. A formula of $\mathcal{L}$ is said to be *provable from $\mathcal{A}$ in $\mathcal{L}$* iff it occurs in some proof from $\mathcal{A}$ in $\mathcal{L}$.

If $\mathcal{A}$ is an axiom system for $\mathcal{L}$ and φ is a formula of $\mathcal{L}$ which is provable from $\mathcal{A}$ in some other language $\mathcal{L}'$, then φ is provable from $\mathcal{A}$ in $\mathcal{L}$. I will therefore omit reference to $\mathcal{L}$ when speaking of provability.

For the remainder of this section we will limit our attention to axiom systems for $\mathcal{L}_1$ and $\mathcal{L}_2$. The characterization of the notions of *$\mathcal{K}$-consistency* and *$\mathcal{K}$-completeness* of an axiom system or $\mathcal{L}_1$ is almost straightforward. However, it is worthwhile to note that the distinction between what is usually referred to as *strong completeness* and *weak completeness* is here more important than in connection with ordinary propositional (or even predicate) calculus, since for $\mathcal{L}_1$ and $\mathcal{L}_2$ these two notions diverge. An example to show that the notions do not coincide in the case of $\mathcal{L}_1$ will follow the next definition.

Definition 16 Let $\mathcal{K}$ be a non-empty class of partial orderings, $\mathcal{A}$ an axiom system for $\mathcal{L}_1$.

(1) $\mathcal{A}$ is *$\mathcal{K}$-consistent in $\mathcal{L}_1$* iff every formula of $\mathcal{L}_1$ which is provable from $\mathcal{A}$ is $\mathcal{K}$-valid.[c]

(2) $\mathcal{A}$ is *weakly $\mathcal{K}$-complete in $\mathcal{L}_1$* iff every formula of $\mathcal{L}_1$ which is $\mathcal{K}$-valid is provable from $\mathcal{A}$.

(3) A set Δ of formulae of $\mathcal{L}_1$ is *consistent relative to $\mathcal{A}$* iff there are no number $n \geq 1$ and formulae $\varphi_1, \ldots, \varphi_n \in \Delta$ such that $\sim(\varphi_1 \wedge \ldots \wedge \varphi_n)$ is provable.

(4) $\mathcal{A}$ is *strongly $\mathcal{K}$-complete in $\mathcal{L}_1$* iff
   (i) for every set Δ of formulae of $\mathcal{L}_1$ which is consistent relative to $\mathcal{A}$ there is a $\mathcal{T} \in \mathcal{K}$ and interpretation $\langle \mathcal{M}, t_0 \rangle$ relative to $\mathcal{T}$ such that for all $\varphi \in \Delta$ $[\varphi]^{\mathcal{T}}_{\langle \mathcal{M}, t_0 \rangle, t_0} = 1$; and
   (ii) if any formula $\sim\sim\varphi$ of $\mathcal{L}_1$ is provable from $\mathcal{A}$ then so is φ.

One easily verifies that if $\mathcal{A}$ is strongly $\mathcal{K}$-complete then it is weakly $\mathcal{K}$-complete. The converse is not true, for we know of an axiom system for $\mathcal{L}_1$ which is $\{\mathcal{J}\}$-consistent and weakly $\{\mathcal{J}\}$-complete, where $\mathcal{J} = \langle J, < \rangle, J$ is the set of integers and $<$ their natural ordering (see Bull 1968). That such an axiom system cannot be strongly $\{\mathcal{J}\}$-complete follows from the fact that the notion of truth at the point 0 in an interpretation relative to

[c] Missing here is an explicit definition of $\mathcal{K}$-validity. A formula is *$\mathcal{K}$-valid* iff it is true at every time of every interpretation relative to every partial ordering in $\mathcal{K}$. When $\mathcal{K}$ consists of a single partial ordering $\mathcal{T}$ (i.e. $\mathcal{K} = \{\mathcal{T}\}$), then we refer to $\mathcal{K}$-validity as '$\mathcal{T}$-validity'.

$\mathcal{J}$ is not compact.[5] That it is not follows from the fact that for the set $\Delta = \{Fq_1, FFq_1, FFFq_1, \ldots, FG{\sim}q_1\}$ there is no interpretation $\langle \mathcal{M}, 0\rangle$ relative to $\mathcal{J}$ such that for all $\varphi \in \Delta$, $[\varphi]^{\mathcal{J}}_{\langle \mathcal{M},0\rangle,0} = 1$, while for every finite subset Δ' of Δ there does exist an interpretation $\langle \mathcal{M}, 0\rangle$, relative to $\mathcal{J}$, such that $[\varphi]^{\mathcal{J}}_{\langle \mathcal{M},0\rangle,0} = 1$ for all $\varphi \in \Delta'$.

The next definition tells us how to extend an axiom system $\mathcal{A}$ for $\mathcal{L}_1$ to an axiom system $\mathcal{A}'$ for $\mathcal{L}_2$ so that, roughly speaking, $\mathcal{A}'$ will be $\mathcal{K}$-consistent and $\mathcal{K}$-complete for $\mathcal{L}_2$ if $\mathcal{A}$ is $\mathcal{K}$-consistent and $\mathcal{K}$-complete for $\mathcal{L}_1$.

Definition 17 Let $\mathcal{A}$ be an axiom system for $\mathcal{L}_1$. Then $\mathcal{A}'$ is the axiom system $\mathcal{A} \cup \{\langle\{q_1, q_1 \to q_2\}, q_2\rangle, \langle\{q_1\}, Lq_1\rangle, \langle\emptyset, L(Nq_1 \to LNq_1)\rangle, \langle\emptyset, Lq_1 \to Nq_1\rangle, \langle\emptyset, {\sim}Nq_1 \leftrightarrow N{\sim}q_1\rangle, \langle\emptyset, N(q_1 \to q_2) \to (Nq_1 \to Nq_2)\rangle, \langle\emptyset, q_1 \leftrightarrow Nq_1\rangle\}$.[6]

N.B. we will refer to the axioms $Lq_1 \to Nq_1, {\sim}Nq_1 \leftrightarrow N{\sim}q_1, N(q_1 \to q_2) \to (Nq_1 \to Nq_2), q_1 \leftrightarrow Nq_1, L(Nq_1 \to LNq_1)$ as $N_L, N_{\sim}, N_{\to}, N_1$ and N_2, respectively. The rule $\langle\{q_1, q_1 \to q_2\}, q_2\rangle$ is called M(odus) P(onens).

One easily verifies that if $\mathcal{K}$ is a non-empty class of partial orderings and $\mathcal{A}$ is an axiom system for $\mathcal{L}_1$ which is $\mathcal{K}$-consistent in $\mathcal{L}_1$, then $\mathcal{A}'$ will not necessarily be $\mathcal{K}$-consistent in $\mathcal{L}_2$. In fact, suppose that $\mathcal{K}$ contains linear orderings with more than one point, that $\mathcal{A}$ is $\mathcal{K}$-consistent and $\mathcal{K}$-complete, and that $\langle\{q_1\}, Gq_1\rangle$ belongs to $\mathcal{A}$. ($\mathcal{K}$-complete axiom systems that contain this rule have actually been given.) Then the formula $q_1 \to Gq_1$ will be provable from $\mathcal{A}'$, though it is not $\mathcal{K}$-valid. (The provability of $q_1 \to Gq_1$ from $\mathcal{A}'$ can be easily verified).[d] We will, however, select among the proofs from $\mathcal{A}'$ a certain group, which we will call (by lack of a better term) *sound*

[5] A notion of truth for a language $\mathcal{L}$ is *compact* if the following is the case: Let $\mathcal{K}$ be any set of models for $\mathcal{L}$ such that whenever Δ is a set of formulae of $\mathcal{L}$ and there is a model for $\mathcal{L}$ in which all members of Δ are true, then there is a model in $\mathcal{K}$ for which this is also the case. Let $\mathcal{T}$ be the topology on $\mathcal{K}$ defined by the condition that $\mathcal{K}' \subseteq \mathcal{K}$ *is open* iff there is a set Δ of formulae of $\mathcal{L}$ such that $\mathcal{K}'$ is the set of models in $\mathcal{K}$ in which all members of Δ are true. Then $\langle \mathcal{K}, \mathcal{T}\rangle$ is compact.

[6] For any formula φ, $L\varphi$ stands for $\varphi \wedge H\varphi \wedge G\varphi$.

[d] This claim isn't quite accurate as stated, even if it might be said to be correct in spirit. First, the assumption that the rule $\langle\{q_1\}, Gq_1\rangle$ is part of $\mathcal{A}'$ isn't really important. Independently of its presence in or absence from the system $\mathcal{A}'$ as a primitive rule, it is possible to derive the formula $q_1 \to Lq_1$, which is no less invalid than $q_1 \to Gq_1$. A proof of $q_1 \to Lq_1$ could take the following form:

proofs, in such a way that if $\mathcal{A}$ is $\mathcal{K}$-consistent then all formulae of $\mathcal{L}_2$ occurring in sound proofs from $\mathcal{A}'$ will be $\mathcal{K}$-valid, whereas on the other hand, if $\mathcal{A}$ is $\mathcal{K}$-complete then the formulae of $\mathcal{L}_2$ occurring in sound proofs from $\mathcal{A}'$ will include the set of all $\mathcal{K}$-valid formulae of $\mathcal{L}_2$.

Definition 18 Let $\mathcal{A}$ be an axiom system for $\mathcal{L}_1$. A proof from $\mathcal{A}'$ is *sound* if each of its members that is preceded in the proof by an instance of N_1 or N_2 is itself an instance of N_1 or N_2, or else follows from its predecessors by MP. A formula is said to be *sound, provable from* $\mathcal{A}'$, if it occurs in a sound proof from $\mathcal{A}'$.

The following facts concerning provability from $\mathcal{A}$, sound provability from $\mathcal{A}$, sound provability from $\mathcal{A}'$ are easily verified, and will be most useful in further developments. We state them without proof.

(1) If $\mathcal{A}$ is an axiom system for $\mathcal{L}_1$, φ is provable from $\mathcal{A}$ [soundly provable from $\mathcal{A}$] and ψ is an instance of φ, then ψ is provable from $\mathcal{A}$ [soundly provable from $\mathcal{A}$].
(2) If $\mathcal{A}$ is an axiom system for $\mathcal{L}_1$ and φ is provable from $\mathcal{A}$ then φ is soundly provable from $\mathcal{A}'$.
(3) If $\mathcal{A}$ is an axiom system for $\mathcal{L}_2$, $\varphi_1, \ldots, \varphi_n$ are soundly provable from $\mathcal{A}$ and $(\varphi_1 \wedge \ldots \wedge \varphi_n) \rightarrow \varphi$ is soundly provable from $\mathcal{A}$, then φ is soundly provable from $\mathcal{A}$.

The notions of *$\mathcal{K}$-consistency in $\mathcal{L}_2$ relative to $\mathcal{A}$*, of *$\mathcal{K}$-consistency in $\mathcal{L}_2$* (of $\mathcal{A}$) and of *weak* and *strong $\mathcal{K}$-completeness in $\mathcal{L}_2$* are defined as before, except that in the definitions 'provable' should everywhere be replaced by 'soundly provable'.

Intuitively one would demand of an inference rule R which is used in an axiomatization of (all or some of) the $\mathcal{K}$-valid formulae of $\mathcal{L}_1$ that whenever ψ comes by an application of R from $\psi_1, \ldots, \psi_k$, and $\psi_1, \ldots, \psi_k$ are $\mathcal{K}$-valid,

---

1.	$q_1 \leftrightarrow Nq_1$	N_1
2.	$q_1 \rightarrow Nq_1$	PC
3.	$Nq_1 \rightarrow q_1$	PC
4.	$L(Nq_1 \rightarrow q_1)$	N_2
5.	$Nq_1 \rightarrow LNq_1)$	N_L
6.	$L(Nq_1 \rightarrow q_1)$	R_L
7.	$L(Nq_1 \rightarrow q_1) \rightarrow (LNq_1 \rightarrow Lq_1)$	Theorem of $\mathcal{A}$
8.	$LNq_1 \rightarrow Lq_1$	MP, lines 6,7
9.	$q_1 \rightarrow Lq_1$	from lines 2,5,8 using PC

Here we rely on the completeness of $\mathcal{A}$ (which entails in particular that $\mathcal{A}$ allows derivations of all tautologies of the classical Propositional Calculus.

It should be evident that this isn't a sound proof in the sense of Definition 18.

then so is ψ. That an axiom system containing R is $\mathcal{K}$-consistent in $\mathcal{L}_2$ is in itself no guarantee that R has this property. For example, the axiom system consisting only of the rule $\langle\{q_1 \rightarrow q_1\}, \sim(q_1 \rightarrow q_1)\rangle$ is $\mathcal{K}$-consistent since one can prove from it no formula whatsoever. But the rule itself clearly fails to have the mentioned property. On the other hand, $\mathcal{A}$ will indeed be $\mathcal{K}$-consistent in $\mathcal{L}_1$ if all its members have this property. We will call a rule with this property *$\mathcal{K}$-valid*, and an axiom system all members of which are $\mathcal{K}$-valid, *strongly $\mathcal{K}$-consistent*.

With respect to the language $\mathcal{L}_2$ matters are a little more complicated. For we just saw that this language contains valid formulae, e.g. $q_1 \leftrightarrow Nq_1$, from which we can obtain invalid formulae, as, e.g., $G(q_1 \leftrightarrow Nq_1)$ by application of a rule (in this case $\langle\{q_1\}, Gq_1\rangle$) which is perfectly acceptable in the sense discussed above when considered as a rule for $\mathcal{L}_1$. However, every inference rule for $\mathcal{L}_1$ which is $\mathcal{K}$-valid in the sense discussed above will have the property that it leads from formulae of $\mathcal{L}_2$ which are true in every model relative to every member of $\mathcal{K}$ *at every moment*, only to formulae which satisfy this same condition (such formulae will be called *strongly $\mathcal{K}$-valid*). It is this last property which we will need in connection with $\mathcal{L}_2$. Thus we come to the following.

Definition 19

(1) A formula of $\mathcal{L}_2$ is *strongly $\mathcal{K}$-valid* iff $[\varphi]^{\mathcal{T}}_{\langle\mathcal{M},t_0\rangle,t} = 1$ for every $\mathcal{T} \in \mathcal{K}$, interpretation $\langle\mathcal{M}, t_0\rangle$ relative to $\mathcal{T}$, and $t \in T$.

(2) An inference rule $\langle S, \varphi\rangle$ for $\mathcal{L}_1$ [for $\mathcal{L}_2$] is *$\mathcal{K}$-valid for* $\mathcal{L}_1$ [for $\mathcal{L}_2$] iff whenever $\psi, \psi_1, \ldots, \psi_k$ are formulae of $\mathcal{L}_1$ [of $\mathcal{L}_2$], $\psi_1, \ldots, \psi_k$ are $\mathcal{K}$-valid [strongly $\mathcal{K}$-valid], and ψ comes from $\psi_1, \ldots, \psi_k$ by an application of $\langle S, \varphi\rangle$ then ψ is $\mathcal{K}$-valid [strongly $\mathcal{K}$-valid].

(3) An axiom system $\mathcal{A}$ for $\mathcal{L}_2$ is *strongly* $\mathcal{K}$-consistent in $\mathcal{L}_1$ [in $\mathcal{L}_2$] iff all its members are $\mathcal{K}$-valid for $\mathcal{L}_1$ [for $\mathcal{L}_2$]. Note that every formula of $\mathcal{L}_1$ is strongly $\mathcal{K}$-valid if it is $\mathcal{K}$-valid; and that a rule $\langle\emptyset, \varphi\rangle$ is $\mathcal{K}$-valid iff φ is strongly $\mathcal{K}$-valid.

It is now possible to state the result which partly confirms the mentioned conjecture on completeness which occurs in Prior's article (Prior 1968).

Theorem 1 *Let $\mathcal{K}$ be a non-empty class of linear orderings; let $\mathcal{A}$ be an axiom system for $\mathcal{L}_1$ which is strongly $\mathcal{K}$-consistent in $\mathcal{L}_2$ and weakly* [*strongly*] *$\mathcal{K}$-complete in $\mathcal{L}_1$. Then $\mathcal{A}'$ is $\mathcal{K}$-consistent in $\mathcal{L}_2$ and weakly* [*strongly*] *$\mathcal{K}$-complete in $\mathcal{L}_2$.*

Theorem 1 will be proved in the next section. In fact, it will be a corollary to a number of much more general results, from which I will also derive a complete confirmation of the conjecture, made in Prior's article, that a

certain axiom system presented there is complete (cf. Prior 1968: p. 113). I have stated Theorem 1 already here, as it is a paradigm of all the theorems to follow in the next two sections.

It is worth noting that if the axiom system $\mathcal{A}$ is $\mathcal{K}$-consistent and $\mathcal{K}$-complete and the only members of $\mathcal{A}$ which are not axioms are the rules MP, $\langle\{q_1\}, Gq_1\rangle$ and $\langle\{q_1\}, Hq_1\rangle$ then

(1) $\mathcal{A}'$ is $\mathcal{K}$-consistent and $\mathcal{K}$-complete.
(2) The rule $\langle\{q_1\}, Lq_1\rangle$, and the axioms N_L, $N_\sim$ and $N_\rightarrow$ are redundant.[e]

---

[e] The following argument is flawed. But the result is right, provided we correct for the obvious misprint in the definition of $\mathcal{A}''$ in the next paragraph. What is wanted here is: $\mathcal{A}'' = \mathcal{A} \cup \{N_1, N_2\}$.

The problem with the argument as stated is the argumentation under (d) on p. 28. For one thing it contains an entirely incomprehensible sentence—'But we observed under (a) that for each i $\rho_i \rightarrow L\rho_i$ is provable'. (a) says no such thing. What follows this sentence is therefore inconclusive. What can be inferred is that when in the proof P we pass from ϕ to $G\phi$ and in our replacement proof P' we have reached the corresponding line $\rho_1 \rightarrow (\rho_2 \rightarrow \ldots(\rho_k \rightarrow \phi)\ldots)$, then we can add a derivation of the formula $G\rho_1 \rightarrow (G\rho_2 \rightarrow \ldots(G\rho_k \rightarrow G\phi)\ldots)$ which uses only rules from $\mathcal{A}$. Since there may be any number of applications of the rules $\langle\{q_1\}, Gq_1\rangle$ and $\langle\{q_1\}, Hq_1\rangle$ before the reconstruction of P has reached the first occurrence of an instance of N_1 or N_2 in P, the formulas ρ_i that occur as antecedents in the conditionals in P' that correspond to the lines of P and that occur in P' as it has been constructed up to this point will be instances of N_L, $N_\sim$ or $N_\rightarrow$ preceded by strings consisting of (0 or more) G's and H's. To complete the construction of a sound proof from $\mathcal{A}''$ from this point onwards it is necessary to assume that a number of lines have been added at this point to P' so that when we then go on with the reconstruction, which may involve 'old' applications of N_1 or N_2 (i.e. applications that occur in P) as well as new ones, the remaining steps of the proof other than these applications of N_1 or N_2 will all be applications of MP.

How this can be accomplished is best illustrated by an example. But first an observation that is crucial to the remainder of this footnote and that the presentation on pp. 26 and 28 of the text fails to make. It is this: because of the definition of L, because the rules $\langle\{q_1\}, Gq_1\rangle$ and $\langle\{q_1\}, Hq_1\rangle$ are part of $\mathcal{A}$ and because $\mathcal{A}$, since it is by assumption complete, must allow for the derivation of all (instances of) tautologies of classical propositional logic, the rule $\langle\{q_1\}, Lq_1\rangle$ is redundant in any case, and each application of it can be replaced by a derivation using the rules of $\mathcal{A}$ (and thus *without* using N_1 or N_2) of the conclusion obtained through that application. In virtue of this fact we can add at the point reached in the construction of P' formulas that will enable us later on to get rid of all antecedents of conditionals in P' that correspond to lines on P.

To see how this can be done consider for an example of a possible such antecedent the formula, in which an instance of $N_\sim$ is preceded by the prefix GH.

$$GH(\sim N\phi \leftrightarrow N\sim\phi)$$

Consider the following formula:

(1) $(\phi \leftrightarrow N\phi) \rightarrow (L(N\phi \rightarrow LN\phi) \rightarrow (L(N\sim\phi \rightarrow LN\sim\phi) \rightarrow GH(\sim N\phi \leftrightarrow N\sim\phi)))$

(1) follows from the fact that $\mathcal{A}$ is strongly $\mathcal{K}$-consistent (all its rules are $\mathcal{K}$-valid). (2) can be argued as follows. Let $\mathcal{A}'' = \mathcal{A} \cup \{N_2\}$. Let P be a sound proof of the formula φ_0 of $\mathcal{L}_2$ from $\mathcal{A}'$. In the first place we may eliminate all applications of the rule $\langle\{q_1\}, Lq_1\rangle$ from P in view of the presence of the rules $\langle\{q_1\}, Gq_1\rangle$ and $\langle\{q_1\}, Hq_1\rangle$. Let us assume therefore that P contains no applications of $\langle\{q_1\}, Lq_1\rangle$.

Let P_1 be that part of P which precedes the first instance of N_1 or N_2 in P (or all of P if no such instance occurs). First we produce a proof P' from $\mathcal{A}''$ in which occurs, for every line φ of P_1, a formula of the form $\sigma_1 \to (\sigma_2 \to \ldots(\sigma_k \to \varphi)\ldots)$, where each of the σ_i is an instance of $N_L, N_\sim$ or $N_\to$. The construction of P' proceeds inductively down the lines of P_1 and is possible in view of the following facts:

(a) If φ is an instance of an axiom of $\mathcal{A}'$ other than $N_1, N_L, N_\sim$, or $N_\to$ then φ is provable from $\mathcal{A}''$.

(b) If φ is an instance of $N_L, N_\sim$ or $N_\to$, then $\varphi \to \varphi$ is provable from $\mathcal{A}''$.

(c) If φ comes by MP from ψ and $\psi \to \psi$, and $\varrho_1 \to (\varrho_2 \to \ldots(\varrho_k \to \psi)\ldots)$ and $\sigma_1 \to (\sigma_2 \to \ldots(\sigma_n \to (\psi \to \varphi)\ldots))$ are provable from $\mathcal{A}''$, then $\varrho_1 \to (\varrho_2 \to \ldots(\varrho_k \to (\sigma_1 \to (\sigma_2 \to \ldots(\sigma_n \to \varphi)\ldots))\ldots))$ is provable from $\mathcal{A}''$.

(d) If φ is of the form $G\psi$ and comes from ψ by an application of $\langle\{q_1\}, Gq_1\rangle$, and $\varrho_1 \to (\varrho_2 \to \ldots(\varrho_k \to \psi)\ldots)$ is provable from $\mathcal{A}''$, then $G(\varrho_1 \to (\varrho_2 \to \ldots(\varrho_k \to \psi)\ldots))$ is provable from $\mathcal{A}''$, and since every formula $G(\chi_1 \to \chi_2) \to (G\chi_1 \to G\chi_2)$ is provable from $\mathcal{A}$ (which is the case because $G(q_1 \to q_2) \to (Gq_1 \to Gq_2)$ is $\mathcal{K}$-valid), $G\varrho_1 \to (G\varrho_2 \to \ldots(G\varrho_k \to G\psi)\ldots)$ is provable from

---

This formula can be proved without use of N_1 or N_2. The crucial steps of the derivation are these:

1. From $(\phi \leftrightarrow N\phi)$ we can derive $(\sim N\phi \leftrightarrow N\sim\phi)$ using Propositional Logic. That is, using just PC we can derive $(\phi \leftrightarrow N\phi) \to (\sim N\phi \leftrightarrow N\sim\phi)$ as a theorem (and thus as a line in P').
2. With the help of $L(N\phi \to LN\phi)$ and $L(N\neg\phi \to LN\neg\phi)$ we can derive from $N\neg\phi \leftrightarrow \neg N\phi$ the formula $L(N\neg\phi \leftrightarrow \neg N\phi)$ and from that the formula $LL(N\neg\phi \leftrightarrow \neg N\phi)$. From this last formula $GH(N\phi \leftrightarrow \neg N\phi)$ can be derived using PC (in view of the definition of L). The derivation of $L(N\neg\phi \leftrightarrow \neg N\phi)$ from $N\neg\phi \leftrightarrow \neg N\phi$, $L(N\phi \to LN\phi)$ and $L(N\neg\phi \to LN\neg\phi)$ involves valid inference principles concerning L which (also because of the definition of L in terms of G and H) can be derivd using the G- and H-rules.

The upshot of all this is that (1) can be derived using just these rules and principles and thus without the use of N_1 or N_2.

Theorems similar to (1) can be proved (without N_1 or N_2) for all the formulas ρ that occur as antecedents of the (usually nested) conditionals that occur as lines in the replacement proof P': for each such ρ we can prove a nested conditional all antecedents of which are instances of N_1 or N_2. After proofs of all these conditionals have been added to P' at the point we had reached, we can then insert into P' all their antecedents (as instances of N_1 or N_2). And after all this has ben accomplished we can now obtain all the antecedents ρ by applications of MP and then use the lines with the ρ's to derive the final line of P, again using only MP.

$\mathcal{A}''$. But we observed under (a) that for each i, $\varrho_i \to L\varrho_i$ is provable from $\mathcal{A}''$. Since $Lq_1 \to Gq_1$ is a $\mathcal{K}$-valid formula of $\mathcal{L}_1$, $L\varrho_i \to G\varrho_i$ is provable, and so $\varrho_i \to G\varrho_i$ is provable. It follows that $\varrho_1 \to (\varrho_2 \to \ldots(\varrho_k \to G\psi)\ldots)$ is provable from $\mathcal{A}''$.

(e) If φ is of the form $H\psi$ and comes by an application of the rule $\langle\{q_1\}, Hq_1\rangle$ and $\varrho_1 \to (\varrho_2 \to \ldots(\varrho_k \to \psi)\ldots)$ is provable from $\mathcal{A}''$ one argues in the same way that $\varrho_1 \to (\varrho_2 \to \ldots(\varrho_k \to H\varphi)\ldots)$ is provable from $\mathcal{A}''$.

For each instance χ of N_L, $N_\sim$, or $N_\to$ there are instances $\chi_1, \ldots, \chi_k$ of N_1 such that $\chi_1 \to (\chi_2 \to \ldots(\chi_k \to \chi)\ldots)$ is provable from $\mathcal{A}''$, for the formulae $(q_1 \leftrightarrow Nq_1) \to (Lq_1 \to Nq_1)$, $(q_1 \leftrightarrow Nq_1) \wedge (\sim q_1 \leftrightarrow N\sim q_1) \to (\sim Nq_1 \leftrightarrow N\sim q_1)$, and $(q_1 \leftrightarrow Nq_1) \wedge (q_2 \leftrightarrow Nq_2) \wedge ((q_1 \to q_2) \leftrightarrow N(q_1 \to q_2)) \to (N(q_1 \to q_2) \to (Nq_1 \to Nq_2))$ are provable from $\mathcal{A}''$. Thus we can extend P' to a proof P'' in which occurs for each line ψ of P_1 a formula of the form $\sigma_1 \to (\sigma_2 \to \ldots(\sigma \to \psi)\ldots)$, where the σ_i are instances of N_1. We can then extend P'' by adding all the instances of N_1 which occur as antecedents of lines in P'', all lines which result from detaching these antecedents by applications of MP, and further all remaining lines of P. The resulting proof will be a sound proof of ψ_0 from $\mathcal{A}'' \cup \{N_1\}$.

The case just described is of some importance, since all axiom systems for $\mathcal{L}_1$ which are known (to me) have indeed no other inference rules than MP, $\langle\{q_1\}, Gq_1\rangle$ and $\langle\{q_1\}, Hq_1\rangle$.

§ 3

It is natural to ask for a justification of the specific choice of the operators G and H (with their given semantic interpretations) as primitives for a system of tense logic. A similar question can be asked within the context of ordinary propositional calculus about the primitive truth-functional connectives of any given language. But there the answer is in all 'standard' cases rather straightforward. Many combinations of well-known truth-functional connectives (e.g., $\{\sim, \wedge\}$, $\{\sim, \vee\}$, $\{\sim, \to\}$, as well as all sets that include any of these) are *functionally complete*, in the sense that every truth-function can be expressed by a formula of a propositional language containing the connectives of that combination.

A similar justification cannot be given for the choice of the operators G and H. In fact, rather simple and natural tense operators, as, e.g., the operator 'it has been the case uninterruptedly for some time' cannot be expressed by a formula of $\mathcal{L}_1$. It is therefore desirable to consider besides $\mathcal{L}_1$ other languages which have different tense operators as primitives.

In order to give a uniform characterization of these languages, it is necessary to characterize the general notion of a 'tense operator'. Unlike for the general concept of a truth-function there is for this notion no entirely straightforward definition. What the definition ought to be is a question which I will not discuss here exhaustively. But I will treat the problem in so far as is necessary for the analysis of the operator 'now' with which this paper is concerned.

First this: Truth-functional connectives in natural languages (e.g., the expressions 'and', 'not', 'neither ... nor' of English) are expressions with the property that the truth-value (at a given moment t) of a sentence formed by means of such an expression out of other sentences is determined completely by the truth-values (at that same moment t) of the component sentences. In so far as truth is concerned the behavior of a truth-functional—say, 2-place—connective, is completely characterized by some 2-place *truth-function*, i.e., a function which assigns to each pair of truth-values a truth-value; and thus the study of truth-functional connectives can properly be reduced to the study of truth-functions.

With tense operators of natural languages (e.g., the tenses of English, or expressions like 'it has sometimes been the case that', 'it has been the case that ... ever since ...'; etc.) the situation is more complicated. Here the truth-value, at a given time t, of a sentence formed by applying a tense, or such an expression, to (an) other sentence(s) does not depend only on the truth-value(s) of that (those) sentence(s) at t, but also (in some cases even exclusively) on their truth-values at moments other than t. On the other hand one easily verifies that for the expressions which were just listed as examples *nothing more* is required to determine the truth-value of the compound sentence at t than the truth-values of the component sentence(s) at t and other moments. Thus with respect to (momentary) truth-value the behavior of such an expression is completely characterized by a function which takes both as arguments and as values 'courses of truth-values through time', i.e., functions from moments to truth-values. (If $\mathcal{T}$ is the structure of time we will refer to functions with such arguments and values as $\mathcal{T}$-tenses.) I maintain that this principle is an essential aspect of what should be regarded as a 'tense operator'. Therefore the study of tense operators reduces to the study of tenses, just as the study of truth-functional connectives comes down to the study of truth-functions.

The above considerations naturally lead to the following definition.

Definition 20 Let $\mathcal{T}$ be a partial ordering. An *n-place $\mathcal{T}$-tense* is an *n*-place function from $\{0,1\}^T$ into $\{0,1\}^T$.[7]

The study of tenses is complicated by the fact that we do not know what the structure of time is. I said already that for this reason we want to consider classes of partial orderings, as well as single partial orderings. For given tense operators of English it is often fairly clear what the characterizing $\mathcal{T}$-tense is, independently of any exact determination of the structural properties of $\mathcal{T}$. So, for example, will the $\mathcal{T}$-tense corresponding to the expression 'it has always been the case that' be the function $F_{\mathcal{T}}$ which assigns to a function f from moments to truth-values that function g such that for any moment $t \in T\, g(t) = 1$ iff for all t' preceding $t, f(t') = 1$. Thus there corresponds to this tense operator a $\mathcal{T}$-tense $F_{\mathcal{T}}$ for every possible time structure $\mathcal{T}$—i.e., a function which assigns to each possible time structure $\mathcal{T}$ a $\mathcal{T}$-tense. I take such functions to embody the most general idea of a *tense*.

Definition 21 An *n-place tense* is a function which assigns to each partial ordering $\mathcal{T}$ an *n*-place $\mathcal{T}$-tense.

For any class of partial orderings $\mathcal{K}$, an *n*-place $\mathcal{K}$-tense is the restriction to $\mathcal{K}$ of an *n*-place tense.

It is doubtful whether every function which is a $\mathcal{T}$-tense according to Definition 20 is intuitively acceptable as the semantic characterization of a possible tense operator. Even less plausible is it that every tense can be so regarded. For example, a tense which assigns to the structure *Re* of the real numbers the *Re*-tense which corresponds to the expression 'it has always been the case that' and to the structure *Rat* of the rational numbers the *Rat*-tense corresponding to the expression 'it will some time be the case that' clearly is a monstrosity. It is therefore desirable to further limit the notion of a tense. I maintain that this can indeed be done in a satisfactory way. (Cf. Kamp 1968.) However, as the results on 'now' which are proved in this article hold true even for the excessively general notion of a tense developed here, we will leave this problem aside.

To each *n*-place truth-function f corresponds a tense $\mathcal{F}$ such that for every partial ordering $\mathcal{T}$, $t \in T$ and members $P_1, \ldots, P_n$ of $\{0,1\}^T$, $\mathcal{F}_{\mathcal{T}}\ (P_1(t), \ldots, P_n(t)) = (\mathcal{F}_{\mathcal{T}}(P_1, \ldots, P_n))(t)$. This fact enables us to treat truth-functional connectives—and in particular the standard connectives which occur, e.g., in $\mathcal{L}_1$—as special tense operators.

---

[7] For any sets U, V, understand by U^V the set of all functions from V into U.

Definition 22

(1) For any countable indexed family $\mathcal{F}$ of tenses let $\mathcal{L}_{\mathcal{F}}$ be the language whose connectives are those symbols C_i^n such that $i \in \mathrm{Dom}\ \mathcal{F}$ and n is the number of places of $\mathcal{F}(i)$.[8]

(2) The notion of an *interpretation for* $\mathcal{L}_{\mathcal{F}}$ *relative to* $\mathcal{T}$ is the same as that of an interpretation for $\mathcal{L}_1$ relative to $\mathcal{T}$ as defined in Section 1 (Definition 6).

(3) The *truth-value of* a formula φ of $\mathcal{L}_{\mathcal{F}}$ *in* an interpretation $\langle \mathcal{M}, t_0 \rangle$ *relative to* $\mathcal{F}$ and $\mathcal{T}$ *at* a moment $t \in T$, $[\varphi]^{\mathcal{F},\mathcal{T}}_{\langle \mathcal{M},t_0\rangle,t}$, is defined by:

(i) $[q_1]^{\mathcal{F},\mathcal{T}}_{\langle \mathcal{M},t_0\rangle,t} = \mathcal{M}(q_1)(t)$

(ii) $[C_i^n(\varphi_1, \ldots, \varphi_n)]^{\mathcal{F},\mathcal{T}}_{\langle \mathcal{M},t_0\rangle,t} = \mathcal{F}(i)_{\mathcal{T}}\ (\bar{\varphi}_1, \ldots, \bar{\varphi}_n)(t)$, where for $i = 1, \ldots, n$, $\bar{\varphi}_i = \{t' \in T : [\varphi_i]^{\mathcal{F},\mathcal{T}}_{\langle \mathcal{M},t_0\rangle,t'} = 1\}$.

(4) A formula φ of $\mathcal{L}_{\mathcal{F}}$ is $\mathcal{T}$*-valid, relative to* $\mathcal{F}$, if for every interpretation $\langle \mathcal{M}, t_0 \rangle$ for $\mathcal{L}_{\mathcal{F}}$ relative to $\mathcal{T}$, $[\varphi]^{\mathcal{F},\mathcal{T}}_{\langle \mathcal{M},t_0\rangle,t_0} = 1$. φ is $\mathcal{K}$*-valid relative to* $\mathcal{F}$ iff φ is $\mathcal{T}$-valid, relative to $\mathcal{F}$, for all $\mathcal{T} \in \mathcal{K}$.

(5) For any language $\mathcal{L}_{\mathcal{F}}$ let $\mathcal{L}_{\mathcal{F}}(N) = \mathcal{L}_{\mathcal{F}} \cup \{N\}$. The *interpretations for* $\mathcal{L}_{\mathcal{F}}(N)$ *relative to* $\mathcal{T}$ are the interpretations for $\mathcal{L}_{\mathcal{F}}$ relative to $\mathcal{T}$, and $[\varphi]^{\mathcal{F},\mathcal{T}}_{\langle \mathcal{M},t_0\rangle,t}$ is defined for formulae of $\mathcal{L}_{\mathcal{F}}(N)$ by the clauses (i) and (ii) above together with the clause

(iii) $[N\varphi]^{\mathcal{F},\mathcal{T}}_{\langle \mathcal{M},t_0\rangle,t} = [\varphi]^{\mathcal{F},\mathcal{T}}_{\langle \mathcal{M},t_0\rangle,t_0}$.

$\mathcal{F}$*-validity* and $\mathcal{K}$*-validity* are then defined as above.

We will, for simplicity, assume that all the languages to be considered henceforth are truth-functionally complete. We will be even more specific and consider only indexed families $\mathcal{F}$ of tenses the domains of which contain the numbers 1 and 2 and where $\mathcal{F}(1), \mathcal{F}(2)$ are the tenses which correspond to the standard meaning of the connectives $\sim$, $\wedge$, respectively.

Our first result of this section states that any formula of a language $\mathcal{L}_{\mathcal{F}}(N)$ which contains N is equivalent to a formula that does not contain N.

Theorem 2 *Let $\mathcal{F}$ be an indexed family of tenses. There is a (primitive recursive) function $R_{\mathcal{F}}$ from the formulae of $\mathcal{L}_{\mathcal{F}}(N)$ to formulae of $\mathcal{L}_{\mathcal{F}}$, such that, for every partial ordering $\mathcal{T}$, $\varphi \leftrightarrow R_{\mathcal{F}}(\varphi)$ is $\mathcal{T}$-valid relative to $\mathcal{F}$.*

Proof. By a *special formula* of $\mathcal{L}_{\mathcal{F}}(N)$ understand a formula of the form $\bigvee_{i=1}^{k}(\varphi_i \wedge N\psi_i)$, where none of the formulae φ_i, ψ_i contains N. (By $\bigvee_{i=1}^{k}\sigma$ understand $\sigma_1 \vee \sigma_2 \vee \ldots \vee \sigma_k$. Similarly for $\bigvee_{\sigma\in S}\sigma$.) We first define inductively

[8] By a *countable indexed family* of objects of a certain kind I understand here a function which is defined on some set of positive natural numbers and which assigns to each member of its domain an object of the kind in question.

a function $R'_{\mathcal{F}}$ which maps the formulae of $\mathcal{L}_{\mathcal{F}}(N)$ into special formulae of $\mathcal{L}_{\mathcal{F}}(N)$:

(i) $R'_{\mathcal{F}}(q_i) = q_i$;

(ii) suppose that $i \in \mathrm{Dom}\,\mathcal{F}$, that $\mathcal{F}(i)$ is n-place, that $R'_{\mathcal{F}}(\varphi_1), \ldots, R'_{\mathcal{F}}(\varphi_m)$ have been defined, and that for $i = 1, \ldots, n$ $R'_{\mathcal{F}}(\varphi_i) = \bigvee_{j=1}^{k_i}(\varphi_{ij} \wedge N\psi_{ij})$

For any $i \leq n$ let S_i be the set of all formulae $\chi_1 \wedge \ldots \wedge \chi_{k_i}$ where for $j = 1, \ldots, k_i, \chi_j$ is ψ_{ij} or χ_j is $\sim\psi_{ij}$. Let S be the set of all formulae of the form $(\sigma_1) \wedge \ldots \wedge (\sigma_n)$, where for $i = 1, \ldots, n, \sigma_i \in S_i$. For each $\sigma = \chi_1 \wedge \ldots \wedge \chi_{k_i} \in S_i$ let

$$\varphi_i(\sigma) = \begin{cases} \text{the disjunction of all those } \varphi_{ij}, \text{ for which } \chi_j = \psi_{ij}, \\ \quad \text{provided some } \chi_j \text{ is } \psi_{ij}\,; \\ \sim(\varphi_{i1} \rightarrow \varphi_{i1}), \text{ if for all } j \leq k_i, \chi_j \text{ is } \sim\psi_{ij} \end{cases}$$

For each $\sigma \in S$, let $\varphi_i(\sigma)$ be $\varphi_i(\sigma_i)$ where σ_i is the i^{th} conjunct of σ. We put

$$R'_{\mathcal{F}}(C_i^n(\varphi_1, \ldots, \varphi_n)) = \bigvee\nolimits_{\sigma \in S}(C_i^n(\varphi_1(\sigma), \ldots, \varphi_n(\sigma)) \wedge N\sigma)$$

(iii) Suppose that $R'_{\mathcal{F}}(\varphi) = \bigvee_{j=1}^{k}(\varphi_j \wedge N\psi_j)$. Then put $R'_{\mathcal{F}}(N\varphi) = \bigvee_{j=1}^{k}((\varphi_j \rightarrow \varphi_j) \wedge N(\varphi_j \wedge \psi_j))$.

Clearly $R'_{\mathcal{F}}$, as defined by (i)-(iii), is a function which maps every formula of $\mathcal{L}_{\mathcal{F}}(N)$ onto a special formula. It is tedious but straightforward to prove by induction that for all φ of $\mathcal{L}_{\mathcal{F}}(N)$, $\varphi \leftrightarrow R'_{\mathcal{F}}(\varphi)$ is $\mathcal{T}$-valid (for arbitrary $\mathcal{T}$).

For any formula φ of $\mathcal{L}_{\mathcal{F}}$ let $R_{\mathcal{F}}(\varphi)$ be the formula which we obtain by deleting all occurrences of N from $R'_{\mathcal{F}}(\varphi)$. Clearly $R_{\mathcal{F}}(\varphi)$ is a formula of $\mathcal{L}_{\mathcal{F}}$. It is also obvious that $R'_{\mathcal{F}}(\varphi) \leftrightarrow R_{\mathcal{F}}(\varphi)$ is $\mathcal{T}$-valid for all $\mathcal{T}$. It follows that $\varphi \leftrightarrow R_{\mathcal{F}}(\varphi)$ is $\mathcal{T}$-valid. q.e.d.

The general concepts of an *inference rule*, an *axiom*, an *axiom system* for a language $\mathcal{L}_{\mathcal{F}}$, as well as the notion of *provability*, have already been given in § 2, pp. 22,23. We now define for an arbitrary axiom system $\mathcal{A}$ for $\mathcal{L}_{\mathcal{F}}$ an extension $\mathcal{A}_j$ which will stand to $\mathcal{A}$ in the same relation as, for any axiom system $\mathcal{B}$ for $\mathcal{L}_1$, the system $\mathcal{B}'$ for $\mathcal{L}_2$ stands to $\mathcal{B}$.

Definition 23 Let $\mathcal{F}$ be an indexed family of tenses.

(1) Let $\mathcal{N}(\mathcal{F})$ be the set containing

(a) all pairs $\langle \emptyset, C_i^n(q_1, \ldots, q_{j-1}, ((Nq_j \wedge q_{j+1}) \vee q_{j+2}), q_{j+3}, \ldots, q_{n+2}) \leftrightarrow [(Nq_j \wedge C_i^n(q_1, \ldots, q_{j-1}, (q_{j+1} \vee q_{j+2}), q_{j+3}, \ldots, q_{n+2})) \vee (N{\sim}q_j \wedge C_i^n(q_1, \ldots, q_{j-1}, q_{j+2}, q_{j+3}, \ldots, q_{n+2}))]\rangle$,
where $i \in \mathrm{Dom}\,\mathcal{F}$, n equals the number of places of $\mathcal{F}(i)$ and $1 \leq j \leq n+2$;

(b) all pairs $\langle \{q_1 \leftrightarrow q_2\}, \varphi \leftrightarrow \varphi' \rangle$ where φ is a formula of $\mathcal{L}_{\mathcal{F}}$ and φ' results from replacing an occurrence of q_1 in φ by q_2.

(2) For any axiom system $\mathcal{A}$ for $\mathcal{L}_{\mathcal{F}}$ let $\mathcal{A}^* = \mathcal{A} \cup \mathcal{N}(\mathcal{F}) \cup \{\mathrm{MP}, N_\sim, N_1, N_2\}$. Again a *sound proof* from $\mathcal{A}^*$ is a proof in which every line following an instance of N_1 is itself an instance of N_1 or else comes from previous lines by MP. The concepts of *$\mathcal{K}$-consistency, strong $\mathcal{K}$-consistency, weak $\mathcal{K}$-completeness* and *strong $\mathcal{K}$-completeness* in $\mathcal{L}_{\mathcal{F}}(N)$, of $\mathcal{A}$ or $\mathcal{A}^*$ (where $\mathcal{A}$ is an axiom system for $\mathcal{L}_{\mathcal{F}}$) are straightforward generalizations of the notions defined earlier for $\mathcal{L}_1$, $\mathcal{L}_2$, $\mathcal{A}$ and $\mathcal{A}'$; I will not spell them out.

The next theorem is similar to Theorem 1, but has a much wider range of application.

Theorem 3 *Let $\mathcal{K}$ be a non-empty class of partial orderings; let $\mathcal{F}$ be an indexed set of tenses; let $\mathcal{A}$ be an axiom system for $\mathcal{L}_{\mathcal{F}}$ which is strongly $\mathcal{K}$-consistent in $\mathcal{L}_{\mathcal{F}}(N)$ and weakly [strongly] $\mathcal{K}$-complete in $\mathcal{L}_{\mathcal{F}}$. Then $\mathcal{A}^*$ is $\mathcal{K}$-consistent in $\mathcal{L}_{\mathcal{F}}(N)$ and weakly [strongly] $\mathcal{K}$-complete in $\mathcal{L}_{\mathcal{F}}(N)$.*

Proof. One easily verifies that if $\mathcal{A}$ is weakly $\mathcal{K}$-complete in $\mathcal{L}_{\mathcal{F}}$ then $\varphi \leftrightarrow R_{\mathcal{F}}(\varphi)$ is soundly provable from $\mathcal{A}^*$. For $\varphi \leftrightarrow R'_{\mathcal{F}}(\varphi)$ (where $R'_{\mathcal{F}}$ defined as in the proof of Theorem 2) is provable from $\mathcal{A}^* \setminus \{N_1\}$, and $R'(\varphi) \leftrightarrow R(\varphi)$ is soundly provable from $\mathcal{A}^*$. Suppose first that $\mathcal{A}$ is weakly $\mathcal{K}$-complete in $\mathcal{L}_{\mathcal{F}}$. Let φ be a $\mathcal{K}$-valid formula of $\mathcal{L}_{\mathcal{F}}(N)$. Then, by Theorem 2, $R_{\mathcal{F}}(\varphi)$ is $\mathcal{K}$-valid. Since $R_{\mathcal{F}}(\varphi)$ is a formula of $\mathcal{L}_{\mathcal{F}}$ and $\mathcal{A}$ is $\mathcal{K}$-complete in $\mathcal{L}_{\mathcal{F}}$, $R_{\mathcal{F}}(\varphi)$ is provable from $\mathcal{A}$. Thus by the above observation φ is soundly provable from $\mathcal{A}^*$; and thus $\mathcal{A}^*$ is weakly $\mathcal{K}$-complete in $\mathcal{L}_{\mathcal{F}}(N)$.

Now suppose that $\mathcal{A}$ is strongly $\mathcal{K}$-complete in $\mathcal{L}_{\mathcal{F}}$. Let Δ be a set of formulae of $\mathcal{L}_{\mathcal{F}}(N)$ which is consistent relative to $\mathcal{A}^*$. Let $R_{\mathcal{F}}(\Delta)$ be the set of all $R_{\mathcal{F}}(\varphi)$, where $\varphi \in \Delta$. Since $\varphi \leftrightarrow R_{\mathcal{F}}(\varphi)$ is soundly provable from $\mathcal{A}^*$ for all $\varphi \in \Delta$, $R_{\mathcal{F}}(\Delta)$ is consistent relative to $\mathcal{A}$. Thus there is a $\mathcal{T} \in \mathcal{K}$ and an interpretation $\langle \mathcal{M}, t_0 \rangle$ relative to $\mathcal{T}$ such that for all $\psi \in R_{\mathcal{F}}(\Delta)$, $[\psi]^{\mathcal{F},\mathcal{T}} = 1$. So, by Theorem 2, $[\varphi]^{\mathcal{F},\mathcal{T}}_{\langle \mathcal{M},t_0\rangle,t_0} = 1$ for all $\varphi \in \Delta$.

To show that $\mathcal{A}^*$ is $\mathcal{K}$-consistent in $\mathcal{L}_{\mathcal{F}}(N)$ we first show by induction on the length of proofs that every formula which is provable from $\mathcal{A}^* \setminus \{N_1\}$ is strongly $\mathcal{K}$-valid. This is true since all inference rules in $\mathcal{A}^* \setminus \{N_1\}$ (including the axioms!) are strongly $\mathcal{K}$-valid. Thus a sound proof from $\mathcal{A}^*$ will always consist of a number of lines which are (strongly) $\mathcal{K}$-valid, followed by a number of instances of N_1 and applications of MP. It is clear that N_1 is $\mathcal{K}$-valid and that MP preserves $\mathcal{K}$-validity. Thus all soundly provable formulae are $\mathcal{K}$-valid.

Theorem 4 *Let $\mathcal{K}$, $\mathcal{F}$, $\mathcal{A}$ be as in Theorem 3. Suppose there is a formula λ of $\mathcal{L}_{\mathcal{F}}$ such that for every $\mathcal{T} \in \mathcal{K}$, interpretation $\langle \mathcal{M}, t_0 \rangle$ relative to $\mathcal{T}$ and $t \in \mathcal{T}$,*

$$[\lambda]^{\mathcal{F},\mathcal{T}}_{\langle \mathcal{M},t_0\rangle,t} = 1 \text{ iff for all } t' \in T, [q_1]^{\mathcal{F},\mathcal{T}}_{\langle \mathcal{M},t_0\rangle,t'} = 1.$$

(We will write 'Lφ' for $[\lambda]\varphi/q_1$.) Let $\mathcal{A}' = \mathcal{A} \cup \{MP, N_1, N_2, N_L, N_\sim, N_\to, \langle\{q_1\}, Lq_1\rangle\}$. If $\mathcal{A}$ is weakly [strongly]$\mathcal{K}$-complete in $\mathcal{L}_\mathcal{F}$, then $\mathcal{A}'$ is weakly [strongly]$\mathcal{K}$-complete in $\mathcal{L}_\mathcal{F}(N)$.

PROOF. Every axiom in $\mathcal{N}(\mathcal{F})$ is provable from $\mathcal{A}' \setminus \{N_1\}$. For

$$L(q_{j+1} \to Lq_{j+1}) \to [C_i^n(q_1, ..., q_{j-1}, ((q_j \wedge q_{j+1}) \vee q_{j+2}), q_{j+3}, ..., q_{n+2}) \leftrightarrow [(q_{j+1} \wedge C_i^n(q_1, ..., q_{j-1}, (q_j \vee q_{j+2}), q_{j+3}, ..., q_{n+2})) \vee (\sim q_{j+1} \wedge C_i^n(q_1, ..., q_{j-1}, q_{j+2}, q_{j+3}, ..., q_{n+2}))]]$$

is a $\mathcal{K}$-valid formula of $\mathcal{L}_\mathcal{F}$, and thus all its instances are provable from $\mathcal{A}$. Since moreover $L(Nq_{j+1} \to LNq_{j+1})$ and $\sim Nq_{j+1} \leftrightarrow N\sim q_{j+1}$ are provable from $\mathcal{A}'' \setminus \{N_1\}$, it follows that

$$C_i^n(q_1, ..., q_{j-1}, ((q_j \wedge Nq_{j+1}) \vee q_{j+2}), q_{j+3}, ..., q_{n+2}) \leftrightarrow [(Nq_{j+1} \wedge C_i^n(q_1, ..., q_{j-1}, (q_j \vee q_{j+2}), q_{j+3}, ..., q_{n+2})) \vee (N\sim q_{j+1} \wedge C_i^n(q_1, ..., q_{j-1}, q_{j+2}, q_{j+3}, ..., q_{n+2}))]$$

is provable from $\mathcal{A}' \setminus \{N_1\}$.

Further we observe that for all formulae φ, φ' of $\mathcal{L}_\mathcal{F}(N)$, where φ' comes from φ by replacement of an occurrence of q_1 by q_2, $L(q_1 \leftrightarrow q_2) \to L(\varphi \leftrightarrow \varphi')$ is provable from $\mathcal{A}' \setminus \{N_1\}$. The proof is by induction on the number of occurrences of N in φ. If φ does not contain N, then $L(q_1 \leftrightarrow q_2) \to L(\varphi \leftrightarrow \varphi')$ is a $\mathcal{K}$-valid formula of $\mathcal{L}^\mathcal{F}$ and thus provable from $\mathcal{A}$. Now assume the assertion has been demonstrated for formulae with at most k occurrences of N and suppose that φ contains $k+1$ such occurrences. Let Oc be a particular occurrence of q_1 in φ and let φ' be the formula which results if we replace Oc by q_2. Let n be the maximum of the indices of propositional constants occurring in φ or φ'. If φ has a subformula of the form $N\chi$ which does not contain Oc, let ψ be the formula which we obtain by replacing that subformula by q_{n+1}, and ψ' the result of replacing Oc in ψ by q_2. By induction hypothesis $L(q_1 \leftrightarrow q_2) \to L(\psi \leftrightarrow \psi')$ is provable from $\mathcal{A}' \setminus \{N_1\}$. Since $L(q_1 \leftrightarrow q_2) \to L(\varphi \leftrightarrow \varphi')$ is an instance of this formula, it too must be provable from $\mathcal{A}' \setminus \{N_1\}$. If on the other hand all occurrences of N in φ are in front of subformulae containing Oc, then φ must have a subformula $N\varrho$ where ϱ contains Oc but no occurrence of N. Let ϱ' be the result of replacing Oc in ϱ by q_2. Then $L(q_1 \leftrightarrow q_2) \to L(\varrho \leftrightarrow \varrho')$ is provable from $\mathcal{A}$. Further, with the help of $N_\sim, N_\to$ and $\langle\emptyset, Lq_1 \to Nq_1\rangle$ we can prove $L(\varrho \leftrightarrow \varrho') \to L(N\varrho \leftrightarrow N\varrho')$. Thus $L(q_1 \leftrightarrow q_2) \to L(N\varrho \leftrightarrow N\varrho')$ is provable from $\mathcal{A}' \setminus \{N_1\}$. Let ψ be the result of replacing $N\varrho$ in φ by q_{n+1} and let ψ be the same formula with q_{n+2} instead of q_{n+1}. Then $L(q_{n+1} \leftrightarrow q_{n+2}) \to L(\psi \leftrightarrow \psi')$ is by induction hypothesis provable from $\mathcal{A}' \setminus \{N_1\}$, and so is therefore its instance $L(N\varrho \leftrightarrow N\varrho') \to L(\varphi \leftrightarrow \varphi')$. It follows that $L(q_1 \leftrightarrow q_2) \to L(\varphi \leftrightarrow \varphi')$ is provable from $\mathcal{A}' \setminus \{N_1\}$.

It is now obvious that we can produce for any proof P from $\mathcal{A}^* \setminus \{N_1\}$ a proof P' in which, for every line φ of P, there are occurrences of both φ and $L\varphi$. Every step in P can either be automatically imitated in P' or else, in view of the preceding remarks, in the case of an application of a rule in $\mathcal{N}(\mathcal{F})$, be replaced by a proof from $\mathcal{A}' \setminus \{N_1\}$ of the appropriate theorem and, if necessary, some applications of MP. A sound proof P from A^* can then be converted into a sound proof from $\mathcal{A}'$ because we can first convert the part of P preceding the first instance of N_1 into a proof P' from $\mathcal{A}' \setminus \{N_1\}$ and then simply add to P' the remaining lines of P. q.e.d.

Corollary. Theorem 1.

Proof. If $\mathcal{K}$ consists of linear orderings then the formula $Hq_1 \wedge q_1 \wedge Gq_1$ satisfies the condition imposed on the formula λ in Theorem 4. So Theorem 4 applies to $\mathcal{L}_1$.

Prior conjectures in his article on 'now' that a certain axiom system—consisting of the axioms A1.1, A1.2, A2.1, A2.2 (p. 106); the rule MP; the rule RL (p. 111); the axioms L1–L5 (p. 111); and the axioms J1–J6 (p. 113)—is complete. The notion of validity which Prior has in mind is that of being true (at the present) in every interpretation relative to any structure $\langle T, R\rangle$ where R is a binary relation on, but not necessarily a partial ordering of, T. (Cf. Prior (1968: p. 113).)

For the class $\mathcal{K}_1$ of all *partial orderings* his axiom system is *not* complete, for the formula $Gq_1 \rightarrow GGq_1$, which is $\mathcal{K}_1$-valid, cannot be proved from that system. However, if we add this formula as well as its counterpart for the past, $Hq_1 \rightarrow HHq_1$, to his system we obtain an axiom system which is indeed $\mathcal{K}_1$-complete in the language which Prior discusses.

To be precise, let *Alw* be the 1-place tense which assigns to each partial ordering $\mathcal{T}$ the function $Alw(\mathcal{T})$ defined in the following way:

(i) if f in $\{0,1\}^T$ takes the value 0 for some $t \in T$, then $Alw(\mathcal{T})(f)$ is the function in $\{0,1\}^T$ with constant value 0;
(ii) if f is the function in $\{0,1\}^T$ with constant value 1, then $Alw(\mathcal{T})(f) = f$.

Let $\mathcal{F}_0$ be an indexed set of tenses with domain consisting of the numbers 1, ..., 5 which 'assigns' to the connectives $\sim, \wedge, G, H$ their intended meanings and such that $\mathcal{F}_0(5) = Alw$. Let us write 'L' for C^1_5. Let $\mathcal{A}_0$ be the axiom system for $\mathcal{L}_{\mathcal{F}_0}$ consisting of the rules MP and $\langle\{q_1\}, Lq_1\rangle$, and the axioms $G(q_1 \rightarrow q_2) \rightarrow (Gq_1 \rightarrow Gq_2), H(q_1 \rightarrow q_2) \rightarrow (Hq_1 \rightarrow Hq_2), PGq_1 \rightarrow q_1, FHq_1 \rightarrow q_1, Lq_1 \rightarrow q_1, Lq_1 \rightarrow Gq_1, Lq_1 \rightarrow Hq_1, L(q_1 \rightarrow q_2) \rightarrow (Lq_1 \rightarrow Lq_2), \sim Lq_1 \rightarrow L\sim Lq_1, Lq_1 \rightarrow LLq_1$, together with a complete set of axioms for ordinary propositional logic. Let $\mathcal{A}_1$ be $\mathcal{A}_0$ together with the axioms

$Gq_1 \rightarrow GGq_1$ and $Hq_1 \rightarrow HHq_1$. Let $\mathcal{A}_0^+$ $[\mathcal{A}_1^+]$ be the axiom system for $\mathcal{L}_{\mathcal{F}_0}(N)$ consisting of $\mathcal{A}_0$ $[\mathcal{A}_1]$ together with the axioms $q_1 \rightarrow Nq_1, Nq_1 \rightarrow q_1, L(Lq_1 \rightarrow Nq_1), L(Nq_1 \rightarrow Lq_1), L(N{\sim}q_1 \rightarrow {\sim}Nq_1), L({\sim}Nq_1 \rightarrow N{\sim}q_1)$ and $L(N(q_1 \rightarrow q_2) \rightarrow (Nq_1 \rightarrow Nq_2))$. One easily verifies that $\mathcal{A}_1$ is strongly $\mathcal{K}$-consistent in $\mathcal{L}_{\mathcal{F}_0}(N)$, that Lq_1 satisfies the condition on λ of Theorem 4, and that for $i = 0, 1$ $\mathcal{A}_i'$ (as defined in Definition 17) is included in $\mathcal{A}_i^+$. Moreover, it is known that $\mathcal{A}_1$ is strongly $\mathcal{K}_1$-complete in $\mathcal{L}_{\mathcal{F}_0}$. (Cf. Lemmon & Scott 1966.) Thus it follows by Theorem 4 that $\mathcal{A}_1^+$ is strongly $\mathcal{K}_1$-complete in $\mathcal{L}_{\mathcal{F}_0}(N)$.

As a matter of fact, we can just as easily obtain a complete confirmation of Prior's conjecture. It suffices to observe that the restriction to partial orderings—observed throughout this paper for reasons mentioned in Section 1—plays no role whatever in the proofs. Thus Theorem 4 is equally true if we weaken the hypothesis that $\mathcal{K}$ is a class of partial orderings to the supposition that $\mathcal{K}$ is an arbitrary class of structures $\langle T, < \rangle$ where T is a non-empty set and $<$ an (arbitrary) binary relation on T. Let $\mathcal{K}_0$ be the class of all such structures. It is known that $\mathcal{A}_0$ is strongly $\mathcal{K}$-complete in $\mathcal{L}_{\mathcal{F}_0}$. (Cf. Lemmon & Scott 1966.) Thus $\mathcal{A}_0^+$ is strongly $\mathcal{K}_0$-complete in $\mathcal{L}_{\mathcal{F}_0}(N)$.

§ 4

We now turn to languages for first order predicate tense logic. For the remainder of this paper we assume that the class $\mathcal{S}$ mentioned in Section 1 contains besides the symbols already listed there a denumerable set of symbols $v_0, v_1, v_2, \ldots$ (called 'individual variables'); for each $n \geq 1$ an uncountable number of symbols $\mathcal{Q}_\alpha^n$ (called 'n-place predicate symbols'); a proper class of symbols c_α (called 'individual constants'); the symbols $\forall$ and $\exists$ (called the 'universal quantifier' and the 'existential quantifier', respectively); and the symbol $=$ (called the 'equality sign'). For any countable (possibly empty) set Q of predicate symbols and individual constants, and any indexed family of tenses $\mathcal{F}$, $\mathcal{L}_{\mathcal{F},Q}$ is the set consisting of the symbols of $\mathcal{L}_{\mathcal{F}}$ together with those of Q, the individual variables, the quantifiers, and $=$. We put $\mathcal{L}_{\mathcal{F},Q}(N) = \mathcal{L}_{\mathcal{F},Q} \cup \{N\}$. By $\mathcal{L}_{1,Q}$ we understand the set $\mathcal{L}_{\mathcal{F}_1,Q}$ where $\mathcal{F}_1$ is the family of tenses such that $\mathcal{L}_{\mathcal{F}_1}$ is the language $\mathcal{L}_1$ (i.e., where $\mathcal{F}_1$ gives the appropriate interpretations for the connectives and tense operators of $\mathcal{L}_1$). The *formulae* of $\mathcal{L}_{\mathcal{F},Q}$ are defined by:

(i) q_i is a *formula*;
(ii) if $\mathcal{Q}_\alpha^n \in Q$ and $\beta_1, \ldots, \beta_n$ are either variables or individual constants in Q, then $\mathcal{Q}_\alpha^n(\beta_1, \ldots, \beta_n)$ is a *formula*;

(iii) if α and β are variables or individual constants in Q, then $\alpha = \beta$ is a *formula*;
(iv) if $i \in \mathrm{Dom}\ \mathcal{F}$, $\mathcal{F}(i)$ is n-place and $\varphi_1, \ldots, \varphi_n$ are *formulae*, then $C^n_1(\varphi_1, \ldots, \varphi_n)$ is a *formula*;
(v) if φ is a *formula*, then $(\forall v_i)\varphi$ and $(\exists v_i)\varphi$ are *formulae*.

The formulae of $\mathcal{L}_{\mathcal{F},Q}(N)$ are defined by clauses (i)–(v) above together with the clause

(iv) if φ is a *formula*, then $N\varphi$ is a *formula*.

An occurrence of a variable α in a formula φ is *bound* if it is within a subformula of φ which is of the form $(\forall\alpha)\psi$ or $(\exists\alpha)\psi$. Other occurrences of variables are called *free*. A formula without free occurrences is called a *sentence*. The *universal closure* of a formula φ is the sentence $(\forall\alpha_1)\ldots(\forall\alpha_n)\varphi$ where $\alpha_1, \ldots\alpha_n$ are all the variables which have free occurrences in φ, listed according to the magnitudes of their indices.

Notice that the formulae of $\mathcal{L}_{\mathcal{F}}$ and $\mathcal{L}_{\mathcal{F},\emptyset}$ coincide. More important is it to observe that if Q contains infinitely many individual constants then we can give a recursive characterization of all sentences of $\mathcal{L}_{\mathcal{F},Q}$ (as well as of those of $\mathcal{L}_{\mathcal{F},Q}(N)$) without reference to formulae which are not sentences as follows:

(i′) q_i is a *sentence of* $\mathcal{L}_{\mathcal{F},Q}$;
(ii′) if $\mathcal{Q}^n_\alpha \in Q$ and $\beta_1, \ldots, \beta_n$ are individual constants in Q then $\mathcal{Q}^n_\alpha(\beta_1, \ldots, \beta_n)$ is a *sentence of* $\mathcal{L}_{\mathcal{F},Q}$;
(iii′) if α and β are individual constants in Q then $\alpha = \beta$ is a *sentence of* $\mathcal{L}_{\mathcal{F},Q}$;
(iv′) if $i \in \mathrm{Dom}\ \mathcal{F}$, $\mathcal{F}(i)$ is n-placed and $\varphi_1, \ldots, \varphi_n$ are *sentences of* $\mathcal{L}_{\mathcal{F},Q}$ then $C^n_i(\varphi_1, \ldots, \varphi_n)$ is a *sentence of* $\mathcal{L}_{\mathcal{F},Q}$;
(v′) if φ is a *sentence of* $\mathcal{L}_{\mathcal{F},Q}$ and α an individual constant in Q, then $(\forall v_i)[\varphi]v_i/\alpha$ and $(\exists v_i)[\varphi]v_i/\alpha$ are *sentences of* $\mathcal{L}_{\mathcal{F},Q}$.[9]

An *interpretation for* a language $\mathcal{L}_{\mathcal{F},Q}(N)$ *relative to* $\mathcal{T}$ is a triple $\langle U, R, t\rangle$, where:

(i) U is a non-empty set;
(ii) R is a function with domain consisting of Q together with the symbols $q_1, q_2, \ldots$;
(iii) $R(q_i)$ is a function from T into $\{0, 1\}$;
(iv) if $\mathcal{Q}^n_\alpha$ is an n-place predicate constant in Q, then $R(\mathcal{Q}^n_\alpha)$ is a function from T into $\mathcal{P}(U^n)$;[10]

[9] By $[\varphi]\beta/\alpha$ I understand the result of proper substitution of β for α in φ.
[10] For any set U, U^n is the nth Cartesian power of U, and $\mathcal{P}(U)$ is the power set of U.

(v) if c_α is an individual constant in Q, then $R(c_\alpha)$ is a function from T into U;
(vi) $t \in T$.

A *referentially complete interpretation* for $\mathcal{L}_{\mathcal{F},Q}(N)$ *relative to* $\mathcal{T}$ is a quadruple $\langle U, R, H, t_0\rangle$ where $\langle U, R, t_0\rangle$ is an interpretation for $\mathcal{L}_{\mathcal{F},Q}(N)$ relative to $\mathcal{T}$, and H is a 1–1 function from a set of individual constants which are not in Q to constant functions from $\mathcal{T}$ into U.[11]

Let $\mathcal{M} = \langle U, R, H, t_0\rangle$ be a referentially complete interpretation for $\mathcal{L}_{\mathcal{F},Q}(N)$ relative to $\mathcal{T}$. For any sentence φ of $\mathcal{L}_{\mathcal{F},Q\cup DomH}(N)$ the *truth-value relative to* $\mathcal{F}, \mathcal{T}$ *in* $\mathcal{M}$ *at* t, $[\varphi]^{\mathcal{F},\mathcal{T}}_{\mathcal{M},t}$, is defined by the following clauses:

(i) $[q_i]^{\mathcal{F},\mathcal{T}}_{\mathcal{M},t} = R(q_i)(t)$;
(ii) $[Q^n_\alpha(c_{\alpha_1}, \ldots, c_{\alpha_n})]^{\mathcal{F},\mathcal{T}}_{\mathcal{M},t} = 1$ iff $\langle (R \cup H)(c_{\alpha_1})(t), \ldots, (R \cup H)(c_{\alpha_n})(t)\rangle \in R(Q^n_\alpha)(t)$;
(iii) $[c_\alpha = c_\beta]^{\mathcal{F},\mathcal{T}}_{\mathcal{M},t} = 1$ iff $(R \cup H)(c_\alpha)(t) = (R \cup H)(c_\beta)(t)$;
(iv) $[C^n_i(\varphi_1, \ldots, \varphi_n)]^{\mathcal{F},\mathcal{T}}_{\mathcal{M},t} = (\mathcal{F}(i)(\bar{\varphi}_1, \ldots, \bar{\varphi}_n))(t)$, where for $j = 1, \ldots, n$ $\bar{\varphi}_j = \{t' \in T : [\varphi_j]^{\mathcal{F},\mathcal{T}}_{\mathcal{M},t'} = 1\}$
(v) $[(\exists \nu_i)[\varphi]\nu_i/\alpha]^{\mathcal{F},\mathcal{T}}_{\mathcal{M},t} = 1$ iff for some $\beta \in \mathrm{Dom}\ H$, $[[\varphi]\beta/\alpha]^{\mathcal{F},\mathcal{T}}_{\mathcal{M},t} = 1$; $[(\forall \nu_i)[\varphi]\nu_i/\alpha]^{\mathcal{F},\mathcal{T}}_{\mathcal{M},t} = 1$ iff for all $\beta \in \mathrm{Dom}\ H$, $[[\varphi]\beta/\alpha]^{\mathcal{F},\mathcal{T}}_{\mathcal{M},t} = 1$;
(vi) $[N\varphi]^{\mathcal{F},\mathcal{T}}_{\mathcal{M},t} = [\varphi]^{\mathcal{F},\mathcal{T}}_{\mathcal{M},t_0}$.

Now let $\mathcal{M} = \langle U, R, t_0\rangle$ be an (arbitrary) interpretation for $\mathcal{L}_{\mathcal{F},Q}(N)$, relative to $\mathcal{T}$. Then for any sentence φ of $\mathcal{L}_{\mathcal{F},Q}(N)$ and $t \in T$ the *truth-value of* φ *at* t *in* $\mathcal{M}$ *relative to* $\mathcal{T}$, $[\varphi]^{\mathcal{T}}_{\mathcal{M},t}$ is the truth value of φ at t in any referentially complete interpretation $\langle U, R, H, t_0\rangle$, relative to $\mathcal{T}$.

I will not give here any philosophical justification for the treatment of individuals and quantification for tense logic which is implicit in the preceding formal definitions. I only want to say that in my opinion the system given here is certainly the most natural of those I have seen; and that it is an adequate frame for most philosophical investigations in this area. The decision to limit attention to predicate symbols and individual constants—instead of considering also function symbols of arbitrarily many places—was dictated by considerations of convenience only. Theorem 5 below holds also for languages which do contain function symbols.

A formula φ of $\mathcal{L}_{\mathcal{F},Q}(N)$ is *valid relative to* $\mathcal{T}$ iff for every interpretation $\mathcal{M} = \langle U, R, t_0\rangle$ for $\mathcal{L}_{\mathcal{F},Q}(N)$ relative to $\mathcal{T}$, $[\psi]^{\mathcal{F},\mathcal{T}}_{\mathcal{M},t_0} = 1$, where ψ is the

[11] A *constant* function from T into U is a function f from T into U for which there is a $u \in U$ such that $f(t) = u$ for all $t \in T$.

universal closure of φ. *$\mathcal{K}$-validity* is defined as before. For a language $\mathcal{L}_{\mathcal{F},Q}$ the definitions of an *interpretation* and the *truth-value* of a sentence *at* a moment *in* an interpretation, and *validity* of a formula are the same as those given above for $\mathcal{L}_{\mathcal{F},Q}(N)$, the only difference being that now clause (vi) of the truth definition above is superfluous.

An *inference rule* for a language $\mathcal{L}_{\mathcal{F},Q}$ [or $\mathcal{L}_{\mathcal{F},Q}(N)$] is again a pair $\langle \Sigma, \varphi \rangle$ where Σ is a finite set of formulae of $\mathcal{L}_{\mathcal{F},Q}$ [of $\mathcal{L}_{\mathcal{F},Q}(N)$] and φ a formula of $\mathcal{L}_{\mathcal{F},Q}$ [of $\mathcal{L}_{\mathcal{F},Q}(N)$]. The notions of an *axiom* and of an *axiom system* are defined as before. We say that a formula ψ of $\mathcal{L}_{\mathcal{F},Q}$ [of $\mathcal{L}_{\mathcal{F},Q}(N)$] *comes from* the set of formulae Δ of $\mathcal{L}_{\mathcal{F},Q}$ [of $\mathcal{L}_{\mathcal{F},Q}(N)$] *by an application of* the inference rule $\langle \Sigma, \varphi \rangle$ for $\mathcal{L}_{\mathcal{F},Q}$ [for $\mathcal{L}_{\mathcal{F},Q}(N)$] if there are variables $x_1, \ldots, x_k, x_1', \ldots, x_k'$, predicate letters $\mathcal{Q}_1^{i(1)}, \ldots, \mathcal{Q}_n^{i(n)}$, and formulae $\chi_1, \ldots, \chi_n$, φ of $\mathcal{L}_{\mathcal{F},Q}$ [of $\mathcal{L}_{\mathcal{F},Q}(N)$] such that ψ comes by proper substitution in φ of x_1' for x_1, ..., of x_k' for x_k, of χ_1 for $\mathcal{Q}_1^{i(1)}$, ..., and of χ_n for $\mathcal{Q}_n^{i(n)}$, and for each $\varphi' \in \Sigma$ there is a ψ' in Δ which comes by proper substitution in φ' of x_1' for x_1, ..., of x_k' for x_k, of χ_1 for $\mathcal{Q}_1^{i(1)}$, ..., and of χ_n for $\mathcal{Q}_n^{i(n)}$.[12] The notion of *strong $\mathcal{K}$-validity* of a rule for $\mathcal{L}_{\mathcal{F},Q}$ [$\mathcal{L}_{\mathcal{F},Q}(N)$] and of *$\mathcal{K}$-consistency in* $\mathcal{L}_{\mathcal{F},Q}$ [$\mathcal{L}_{\mathcal{F},Q}(N)$], *strong $\mathcal{K}$-consistency in* $\mathcal{L}_{\mathcal{F},Q}$ [$\mathcal{L}_{\mathcal{F},Q}(N)$], and *weak* and *strong $\mathcal{K}$-completeness* of an axiom system $\mathcal{A}$ for $\mathcal{L}_{\mathcal{F},Q}$ [$\mathcal{L}_{\mathcal{F},Q}(N)$] are defined as before. For any axiom system $\mathcal{A}$ for $\mathcal{L}_{\mathcal{F},Q}$ let

$$\mathcal{A}' = \mathcal{A} \cup \{\, \text{MP}, \langle \{\varphi\}, (\forall \nu_0)\varphi \rangle, N_1, N_2 \}.$$

A *sound proof* in $\mathcal{L}_{\mathcal{F},Q}(N)$ from $\mathcal{A}'$ is a proof in $\mathcal{L}_{\mathcal{F},Q}$ from $\mathcal{A}'$ in which every line following an instance of N_1 is itself such an instance, or else comes from previous lines by an application of MP or $\langle \{\varphi\}, (\forall \nu_0)\varphi \rangle$.

Our result for languages for first order predicate logic will be somewhat less general than those for languages of propositional tense logic. On the other hand, they are in my opinion more interesting, since the fact expressed by Theorem 2, that every formula containing *N* is equivalent to one which does not contain *N*, fails for languages of predicate tense logic.

Theorem 5 *Let $\mathcal{K}$ be a non-empty class of partial orderings, $\mathcal{F}$ an indexed set of tenses, Q a countable set of individual constants and predicate symbols. Let*

[12] The notions of proper substitution of a variable for a variable, of proper substitution of a formula for predicate letter, and of an alphabetic variant, referred to here, are the usual ones. I do not give their rather involved definitions which can be found for example in Kalish & Montague (1964: p. 148, 157–159, 155–156).

λ be a formula of $\mathcal{L}_{\mathcal{F},Q}$ which satisfies the condition of Theorem 4. (*Again we write 'Lφ' for $[\lambda]\varphi/q_1$.*) *Let $\mathcal{A}$ be an axiom system for $\mathcal{L}_{\mathcal{F},Q}$ which is strongly $\mathcal{K}$-consistent in $\mathcal{L}_{\mathcal{F},Q}(N)$ and strongly $\mathcal{K}$-complete in $\mathcal{L}_{\mathcal{F},Q}$. Then $\mathcal{A}'$ is strongly $\mathcal{K}$-complete in $\mathcal{L}_{\mathcal{F},Q}(N)$.*

The proof of Theorem 5 has the irritating—if common—property that when you try to write it up either you don't write enough, or else you write far too much. I have chosen the second alternative. However, since the idea behind the proof is very simple, I will first give a rough description of it. He who is satisfied with this informal explanation may save himself the trouble of reading the proof itself.

To show that a consistent set Δ of sentences of $\mathcal{L}_{\mathcal{F},Q}(N)$ has an interpretation in which all the sentences of Δ are true simultaneously, we proceed as if the subformulae of those sentences that begin with N were atomic formulae. In this way we can regard Δ as a set of sentences of a language $\mathcal{L}_{\mathcal{F},Q'}$ (where $Q \subseteq Q'$) for which, since it is a consistent set relative to $\mathcal{A}$, there is an interpretation in which all sentences are simultaneously true. This interpretation gives us, in a completely straightforward way, an interpretation in which, at one moment, all the members of Δ are true according to the truth definition for $\mathcal{L}_{\mathcal{F},Q}(N)$.

PROOF. Let us call a formula φ of $\mathcal{L}_{\mathcal{F},Q}(N)$ an *n-place irreducible predicate* if

(1) for each i such that $1 \leq i \leq n$ there is exactly one free occurrence in φ of the variable v_i;
(2) for $i = 1, \ldots, n-1$ the free occurrence of v_{i+1} in φ is the first free occurrence in φ to the right of the free occurrence of v_i in φ;
(3) no other variables have free occurrences in φ; and
(4) φ contains no individual constants.

For any n-place irreducible predicate φ and individual symbols (constants or variables) $\alpha_1, \ldots, \alpha_n$ we write $\varphi(\alpha_1 \ldots \alpha_n)$ for $[\varphi]\alpha_1/v_1, \ldots, \alpha_n/v_n$. Let G be a function which assigns to each n-place irreducible predicate of $\mathcal{L}_{\mathcal{F},Q}(N)$ a different n-place predicate symbol not in Q. Let $Q' = Q \cup \text{Ran } G$. If ψ is any formula of $\mathcal{L}_{\mathcal{F},Q}$ and n is the total number of occurrences which are either occurrences of individual constants or free occurrences of variables, then there is a unique n-place irreducible predicate φ of $\mathcal{L}_{\mathcal{F},Q}$ such that $\psi = \varphi(\alpha_1 \ldots \alpha_n)$, where each α_i is a variable or individual constant. The function G therefore induces a function G' from the formulae of $\mathcal{L}_{\mathcal{F},Q}$ to atomic formulae of $\mathcal{L}_{\mathcal{F},Q'}$ defined by the condition

$$G'(\psi) = G(\varphi)(\alpha_1 \ldots \alpha_n),$$

where φ, n are such that φ is an n-place irreducible predicate and $\psi = \varphi(\alpha_1 \ldots \alpha_n)$. The function G' determines in turn a function G'' from the formulae of $\mathcal{L}_{\mathcal{F},Q}(N)$ to formulae of $\mathcal{L}_{\mathcal{F},Q'}$, defined recursively by:

(i) if $\psi \in \mathcal{L}_{\mathcal{F},Q}$ then $G''(\psi) = \psi$.
(ii) $G''(N\psi) = G'(\psi)$.
(iii) $G''(C_i^n(\psi_1, \ldots, \psi_n) = C_i^n(G''(\psi_1), \ldots, G''(\psi_n))$;
$G''((\forall \nu_i)\psi) = (\forall \nu_i)G''(\psi)$;
$G''((\exists \nu_i)\psi) = (\exists \nu_i)G''(\psi)$.

Now let Δ be a set of sentences of $\mathcal{L}_{\mathcal{F},Q}(N)$ which is consistent relative to $\mathcal{A}'$. Let Δ' be the union of Δ and the set of all universal closures of formulae of the forms $\varphi \leftrightarrow N\varphi$ and $L(N\varphi \rightarrow LN\varphi)$, where φ is a formula of $\mathcal{L}_{\mathcal{F},Q}(N)$. Then Δ' is consistent relative to $\mathcal{A}'$, since all the added formulae are soundly provable from $\mathcal{A}'$. Clearly $G''(\Delta')$ is a set of formulae of $\mathcal{L}_{\mathcal{F},Q'}$ which is consistent relative to $\mathcal{A}$. Since $\mathcal{A}$ is strongly $\mathcal{K}$-complete in $\mathcal{L}_{\mathcal{F},Q'}$ there is a $\mathcal{T} \in \mathcal{K}$ and an interpretation $\mathcal{M} = \langle U, R, t_0 \rangle$ relative to $\mathcal{T}$ such that for all $\varphi \in G''(\Delta')$, $[\varphi]^{\mathcal{F},\mathcal{T}}_{\mathcal{M},t_0} = 1$. Let H be a function which assigns to each constant function from $\mathcal{T}$ to an element of U a different individual constant which is not in Q. Let $\mathcal{M}' = \langle U, R, H, t_0 \rangle$; $Q'' = Q \cup \mathrm{Ran}\, H$; $Q''' = Q' \cup Q''$. $\mathcal{M}'$ is a referentially complete interpretation for $\mathcal{L}_{\mathcal{F},Q''}$, relative to $\mathcal{T}$. For $t \in T$, let Γ_t be the set of all sentences φ of $\mathcal{L}_{\mathcal{F},Q'''}$ such that $[\varphi]^{\mathcal{F},\mathcal{T}}_{\mathcal{M}',t} = 1$. The function G'' can be extended in the obvious manner to a function from the formulae of $\mathcal{L}_{\mathcal{F},Q''}(N)$ to formulae of $\mathcal{L}_{\mathcal{F},Q'''}$. We will refer to this extension also as G''. Let R' be the restriction of $R \cup H^{-1}$ to Q''. Then $\mathcal{M}'' = \langle U, R', H, t_0 \rangle$ is a referentially complete interpretation for $\mathcal{L}_{\mathcal{F},Q''}$ relative to $\mathcal{T}$. I claim that:

(1) for every sentence $\varphi \in \Delta$, $[\varphi]^{\mathcal{F},\mathcal{T}}_{\mathcal{M}'',t_0} = 1$.

Since $\Delta \subseteq (G'')^{-1}(\Gamma_{t_0})$, (1) follows from

(2) for every sentence φ of $\mathcal{L}_{\mathcal{F},Q'''}(N)$ and $t \in T$, $[\varphi]^{\mathcal{F},\mathcal{T}}_{\mathcal{M}'',t} = 1$ iff $G''(\varphi) \in \Gamma_t$.

(2) is proved by induction on sentences of $\mathcal{L}_{\mathcal{F},Q'''}(N)$. It is clear that (2) holds for atomic sentences, that if (2) holds for $\varphi_1, \ldots, \varphi_n$ then it holds for $C_i^n(\varphi_1, \ldots, \varphi_n)$, and that if (2) holds for φ then it holds for $(\forall \nu_i)[\varphi]\nu_i/\alpha$ and for $(\exists \nu_i)[\varphi]\nu_i/\alpha$. Suppose that $\varphi = N\psi$ and that (2) holds for ψ. First assume that $[\varphi]^{\mathcal{F},\mathcal{T}}_{\mathcal{M}'',t} = 1$. Then $[\psi]^{\mathcal{F},\mathcal{T}}_{\mathcal{M}'',t_0} = 1$. So by induction hypothesis $G''(\psi) \in \Gamma_{t_0}$. Let ψ' be a formula of $\mathcal{L}_{\mathcal{F},Q}(N)$ such that for some $\nu_{i_1}, \ldots, \nu_{i_k}$ and constants $\alpha_1, \ldots, \alpha, \psi = [\psi']\nu_{i_1}/\alpha_1, \ldots, \nu_{i_k}/\alpha_k$. Δ' contains the formula $(\forall \nu_{i_1}) \ldots (\forall \nu_{i_k})(\psi' \leftrightarrow N\psi')$. Therefore Γ_{t_0} contains $G''((\forall \nu_{i_1}) \ldots (\forall \nu_{i_k})(\psi' \leftrightarrow N\psi'))$, which is equal to $(\forall \nu_{i_1}) \ldots (\forall \nu_{i_k})(G''(\psi') \leftrightarrow G''(N\psi'))$. Then Γ_{t_0} also contains $[G''(\psi') \leftrightarrow G''(N\psi')]\alpha_1/\nu_{i_1}, \ldots, \alpha_k/\nu_{i_k}$, which is $[G''(\psi) \leftrightarrow$

$G''(N\psi)]$. So $G''(N\psi) \in \Gamma_{t_0}$. By an analogous argument we can infer that $LG''(N\psi) = G''(LN\psi) \in \Gamma_{t_0}$. It follows that $G''(N\psi) \in \Gamma_t$.

In a similar fashion one shows that if $G''(\varphi) \in \Gamma_t$ then $[\varphi]^{\mathcal{F},\mathcal{T}}_{\mathcal{M}'',t} = 1$. It follows that $\mathcal{A}^*$ is strongly $\mathcal{K}$-complete in $\mathcal{L}_{\mathcal{F},Q}(N)$.

It is clear that Theorem 5 applies to the language $\mathcal{L}_{1,Q}$ when $\mathcal{K}$ is a class of *linear* orderings.

§ 5

In this section I will show that in general there is not for every sentence φ of $\mathcal{L}_{\mathcal{F},Q}(N)$ a sentence ψ of $\mathcal{L}_{\mathcal{F},Q}$ such that $\varphi \leftrightarrow \psi$ is valid. To this purpose I will give two theorems, which together will show that the elimination of N (in the sense of Theorem 2) is indeed impossible for most reasonable choices of $\mathcal{F}$, Q, and $\mathcal{K}$. Undoubtedly similar theorems could be obtained for combinations of $\mathcal{F}$, Q and $\mathcal{K}$ which are not covered by the theorems presented here. But it seems to me that these two give a sufficiently clear indication of along what lines such results can be proved.

The first theorem is concerned with a large class of tenses, viz., all those which are *invariant*, according to the following definition.

Definition 24 An n-place tense $\mathcal{F}$ is $\mathcal{K}$-*invariant* iff whenever $\mathcal{T}, \mathcal{T}' \in \mathcal{K}$, g is an order-isomorphism from $\mathcal{T}$ to $\mathcal{T}'$, and $p_1, \ldots, p_n \in \{0,1\}^T$, then $\mathcal{F}(\mathcal{T}')(g(p_1), \ldots, g(p_n)) = g(\mathcal{F}(\mathcal{T})(p_1, \ldots, p_n))$, where by $g(p_i)$ we understand that function $\{0,1\}^{T'}$ such that for any $t \in T'$, $g(p_i)(t) = p_i(g^{-1}(t))$.

It is my opinion that any intuitively plausible tense operator ought to have an invariant tense for its meaning. Thus the restriction to invariant tenses is from an intuitive point of view no essential limitation at all.

The time structures that the first theorem considers are the dense linear orderings without endpoints. Let $\mathcal{K}_0$ be the class of these orderings. Let $\mathcal{F}_0$ be the class of all $\mathcal{K}_0$-invariant tenses. I will construct, for an arbitrary set of predicate symbols and individual constants which contains at least one predicate symbol, a referentially complete interpretation $\mathcal{M} = \langle U, R, H, 0\rangle$ for $\mathcal{L}_{\mathcal{F}_0,Q}(N)$, relative to some member $\mathcal{T}$ of $\mathcal{K}_0$, such that

(i) if $\mathcal{L}_{\mathcal{F}_0,Q}$ contains a sentence λ satisfying the condition of Theorem 4, then there is a sentence ψ of $\mathcal{L}_{\mathcal{F}_0,Q}(N)$ which is true in $\mathcal{M}$ only at 0; and
(ii) every sentence φ of $\mathcal{L}_{\mathcal{F}_0,Q}$ is either true at no point, or else at infinitely many.

This will show that it is impossible to eliminate N from ψ in the sense of Theorem 2.

For simplicity I will assume that Q contains only one element, the 1-place predicate letter Q^1_1. It can be seen from the construction below—and will be argued after the construction has been given—that this assumption is not essential.

T will be equal to $T' \cup \{0\}$, where T' is a set of rational numbers different from 0, $\mathcal{M}$ the quadruple $\langle \omega, R, H, 0\rangle$; and for every $t \in T$, $R(Q^1_1)(t)$ will be an infinite, coinfinite subset S'_t of ω,[13] such that

(a) for every $t \in T$ there are $t' < t$ and $t'' > t$ such that $S_{t'} \cap S_t = S_{t''} \cap S_t = \emptyset$; and
(b) for every $t \in T$, $S_t \cap S_0 \neq \emptyset$.

H will be a 1–1 function mapping each c_i onto the (constant) function from T to i. In view of (a), (b) we have that

(1) the sentence $L(\exists v_0)(Q^1_1(v_0) \wedge NQ^1_1(v_0))$ is true in $\mathcal{M}$ at 0 but at no other moment.[f]

Moreover, the sets S_t for $t \in T'$ will be defined in such a way that

(2) each sentence of $\mathcal{L}_{\mathcal{F}_0,Q}$ is true either at all $t \in T$, or else at no t.

In order to describe the construction I will first make a few remarks on a certain kind of Boolean algebra, whose properties will play an important role in that construction, as well as in the argument which follows it.

Let U be an infinite set and $\mathcal{U}$ a subset of $\mathcal{P}(U)$. For any finite subsets K, L of U let $\mathcal{U}(K, L)$ be the set of all members V of $\mathcal{U}$ such that $K \subseteq V$ and $L \cap V = \emptyset$. Let $\mathcal{B}(U, \mathcal{U})$ be the Boolean algebra generated by the sets $\mathcal{U}(K, L)$ under the operations of finite set theoretic union, intersection, and complementation relative to $\mathcal{P}(U)$. One easily verifies that $\mathcal{B}(U, \mathcal{U})$ is generated by the sets of the forms $\mathcal{U}(\{u\}, \emptyset)$ and $\mathcal{U}(\emptyset, \{u\})$, and consequently that for every element V of $\mathcal{B}(U, \mathcal{U})$ there is a number $k \geq 1$ and a function B such that Dom $B = \{1, \ldots, k\}$ and for $i = 1, \ldots, k$ B_i is a pair of finite subsets B^1_i, B^2_i of U and $V = \cup^k_{i=1}\mathcal{U}(B^1_i, B^2_i)$. Such a function B will be called a *representation* of the element V. I will write $\mathcal{U}(B)$ for $\cup^k_{i=1}\mathcal{U}(B^1_i, B^2_i)$.

[13] A subset K of ω is *coinfinite* iff $\omega \setminus K$ is infinite. ω is the set of natural numbers.

[f] This statement is wrong. The sentence (i) $L(\exists v_0)(Q^1_1(v_0)$ & $NQ^1_1(v_0))$ is true at all times in the model $\mathcal{M} = \langle \omega, R, H, 0\rangle$. What should have been said is this: Let for any $t \in T$, $\mathcal{M}_t = \langle \omega, R, H, t\rangle$. Then (i) is true in $\mathcal{M}_0$ at all times in T but it will be true at no time in models $\mathcal{M}_t$ such that $t \neq 0$. On the other hand the sentences of $\mathcal{L}_{\mathcal{F}_0,Q}$ are all true at all times in all models $\mathcal{M}_t$ (i.e. whether $t = 0$ or not). This difference is enough to establish that $\mathcal{L}_{\mathcal{F}_0,Q}(N)$ cannot be reduced to $\mathcal{L}_{\mathcal{F}_0,Q}$. The following proof of this last claim is not affected.

Now let $\mathcal{S}$ be the set of all infinite, coinfinite subsets of ω. Then, whenever K, L are finite subsets of ω, and $K \cap L = \emptyset, \mathcal{S}(K, L) \neq \emptyset$; and thus every representation B such that for some $i \in \text{Dom } B, B_i^1 \cap B_i^2 = \emptyset$, represents an element of $\mathcal{B}(\omega, \mathcal{S})$ different from $\emptyset$. Exactly the same properties hold for the algebra $\mathcal{B}(V, \mathcal{V})$ where V is the set $\{\nu_0, \nu_1, \ldots\}$ and $\mathcal{V}$ is the set of all infinite, coinfinite subsets of V. The algebras $\mathcal{B}(\omega, \mathcal{S})$ and $\mathcal{B}(V, \mathcal{V})$ are of course isomorphic. In fact, let us understand by a *perfect assignment* a 1–1 function from V onto ω. Then for any perfect assignment h and finite subsets K, L of V, $\{h(W) : W \in \mathcal{V}(K, L)\} = \mathcal{S}(h(K), h(L))$—so that h induces an isomorphism between $\mathcal{B}(V, \mathcal{V})$ and $\mathcal{B}(\omega, \mathcal{S})$.

T' and the sets S_t for $t \in T'$ are constructed as follows. Let *Rat* be the set of rational numbers, and let $<$ be the natural ordering on *Rat*. Let $\{r_i\}_{i \in \omega}$ be an enumeration of $Rat \setminus \{0\}$ without repetitions. Let $\{J_k\}_{k \in \omega}$ be an enumeration of all pairs of mutually disjoint finite subsets of ω. Let W be an enumeration of all pairs $\langle i, -1\rangle$ and $\langle i, +1\rangle$ and all triples $\langle i, j, k\rangle$ where $i, j, k \in \omega$ and $i \neq j$. Let $\{E_\alpha\}$ be an enumeration of all members of $\mathcal{S}$.

For $n = 0, 1, 2, \ldots$ construct a pair $\langle i_n, S_n\rangle$ consisting of a natural number i_n and a member S_n of $\mathcal{S}$, and a subsequence W_n of W as follows:

(i) $\langle i_0, S_0\rangle = \langle 1, E_0\rangle; W_0 = W.$
(ii) Suppose that for $m \leq n$ $\langle i_m, S_m\rangle$ and W_m have been constructed.

Let w be the first element of W_n such that either

(a) $w = \langle i, +1\rangle$ and $i = i_m$ for some $m \leq n$; or
(b) $w = \langle i, -1\rangle$ and $i = i_m$ for some $m \leq n$; or
(c) $w = \langle i, j, k\rangle, i = i_m$ and $j = i_{m'}$ for some $m, m' \leq n$.

(It will be clear from the construction that W_n does indeed contain such an element w). In case (a) let i_{n+1} be the first natural number different from $i_0, \ldots, i_n$ such that $r_{i_{n+1}}$ is a rational number $> r_{i_m}$, and let S_{n+1} be the first member of $\{E_\alpha\}$ different from $S_0, \ldots, S_n$ which is disjoint from S_m. In case (b) let i_{n+1} be the first number different from $i_0, \ldots, i_n$ such that $r_{i_{n+1}}$ is a negative rational number $< r_{i_m}$; let S_{n+1} be as under (a). In case (c) let i_{n+1} be the first number different from $i_0, \ldots, i_n$ such that $r_{i_{n+1}}$ is between r_{i_m} and $r_{i_{m'}}$ and let S_{n+1} be the first member of $\{E_\alpha\}$ different from $S_0, \ldots, S_n$ which belongs to $\mathcal{S}(J_k)$. In all cases (a), (b), (c) let W_{n+1} be the sequence which we obtain when we eliminate w from W_n. Let $T = \{r_{i_n} : n \in \omega\} \cup \{0\}$, and let $\mathcal{T} = \langle T, <\rangle$, where $<$ is the natural ordering of T. For $t \in T, t \neq 0$ let $S_t = S_n$, where $t = r_{i_n}$; and let S_0 be a member of $\mathcal{S}$ which has at least one member in common with each S_t, where $t \in T, t \neq 0$. We show that the interpretation $\langle \omega, R, H, 0\rangle$ has indeed the properties (1) and (2). That (1) holds is clear. It is

somewhat less obvious that (2) is true. The next few pages will be concerned with showing that this is so.

Let $n_0, \ldots, n_k$ be a sequence of distinct natural numbers. For $K \subseteq \{\nu_0, \ldots, \nu_k\}$ we understand by '$K(n_0, \ldots, n_k)$' the set $h(K)$ $(= \{h(\nu_i) : \nu_i \in K\})$, where h is any perfect assignment such that for $i = 0, \ldots, k$, $h(\nu_i) = n_i$; for any representation B in the range of which occur only subsets of $\{\nu_0, \ldots, \nu_k\}$ understand by $B(n_0, \ldots, n_k)$ the function which assigns to each $i \in \mathrm{Dom}\ B$ the pair $\langle B_i^1(n_0, \ldots, n_k), B_i^2(n_0, \ldots, n_k)\rangle$; for any formula φ of $\mathcal{L}_{\mathcal{F}_0,Q'}$ all free variables of which are among $\nu_0, \ldots, \nu_k$, let $\varphi(n_0, \ldots, n_k)$ be the formula $[\varphi]c_{n_0}/\nu_0, \ldots, c_{n_k}/\nu_k$. For any sentence φ of $\mathcal{L}_{\mathcal{F}_0,Q'}$ let $T(\varphi)$ be the set of all $t \in T$ such that $[\varphi]^{\mathcal{F},\mathcal{T}}_{\mathcal{M},t} = 1$. For any representation B of an element of $\mathcal{B}(\omega, \mathcal{S})$ let $T(B) = \{t \in T : S_t \in \mathcal{S}(B)\}$.

It is a well-known fact of ordinary predicate logic with identity that for every formula φ there is a logically equivalent formula φ' satisfying the condition that

(3) every subformula of φ' which begins with a quantifier is either of the form $(\exists\nu_i)(\sim\nu_i = \nu_{i_1} \wedge \ldots \wedge \sim\nu_i = \nu_{i_n} \wedge \psi)$ or of the form $(\forall\nu_i)(\sim\nu_i = \nu_{i_1} \wedge \ldots \wedge \sim\nu_i = \nu_{i_n} \rightarrow \psi)$, where $\nu_{i_1}, \ldots, \nu_{i_n}$ are all the free variables of ψ.

One easily verifies that this is also true for the languages $\mathcal{L}_{\mathcal{F},Q}$ and $\mathcal{L}_{\mathcal{F},Q}(N)$, in the sense that for any formula φ of $\mathcal{L}_{\mathcal{F}}$, or $\mathcal{L}_{\mathcal{F},Q}(N)$ there is a formula φ' satisfying condition (3) such that $\varphi \leftrightarrow \varphi'$ is $\mathcal{K}$-valid (whatever $\mathcal{K}$ may be!). Formulae which satisfy (3) I will call *restricted*.

I will show that

(4) for each formula φ of $\mathcal{L}_{\mathcal{F},Q}$ the free variables of which are among $\{\nu_0, \ldots, \nu_k\}$ there is a representation $B(\varphi)$ in the Range of which occur only subsets of $\{\nu_0, \ldots, \nu_k\}$, such that for any choice of distinct natural numbers $n_0, \ldots, n_k$, $T(\varphi(n_0, \ldots, n_k)) = T(B(\varphi)(n_0, \ldots, n_k))$; either $B(\varphi)$ is $\{\langle 1, \langle\{\nu_0\}, \{\nu_0\}\rangle\rangle\}$, or else for each $i \in \mathrm{Dom}\ B$, $B_i^1 \cap B_i^2 = \emptyset$.

I will prove (4) by induction on the complexity of restricted formulae. In view of the equivalence of each formula with a restricted formula this will establish (4) for all formulae of $\mathcal{L}_{\mathcal{F}_0,Q'}$.

If φ is the formula q_i put $B(\varphi) = \{\langle 1, \langle\{\nu_0\}, \{\nu_0\}\rangle\rangle\}$; if φ is the formula $\nu_i = \nu_i$, put $B(\varphi) = \{\langle 1, \langle\{\nu_i\}, \emptyset\rangle\rangle, \langle 2, \langle\emptyset, \{\nu_i\}\rangle\rangle\}$; if φ is the formula $\nu_i = \nu_j$, where $i \neq j$, put $B(\varphi) = \{\langle 1, \langle\{\nu_0\}, \{\nu_0\}\rangle\rangle\}$; if φ is the formula $\mathcal{Q}_1^1(\nu_i)$, put $B(\varphi) = \{\langle 1, \langle\{\nu_i\}, \emptyset\rangle\rangle\}$.

Suppose that $\varphi = C_i^n(\varphi_1, \ldots, \varphi_n)$ and that for $i = 1, \ldots, n$ $B(\varphi_i)$ has been determined. Let $\varrho_1, \ldots, \varrho_{2^n}$ be all formulae of the form $\psi_1 \wedge \ldots \wedge \psi_n$, where for $i = 1, \ldots, n$ $\psi_i = \varphi_i$ or $\psi_i = \sim\varphi_i$. It follows from our remarks on the

Boolean algebras $\mathcal{B}(\omega, \mathcal{S})$ and $\mathcal{B}(V, \mathcal{V})$ that there are, for $s = 1, \ldots, 2^n$, representations B_s such that for any distinct natural numbers $n_0, \ldots, n_k$, $T(\varrho_s(n_0, \ldots, n_k)) = T(B_s(n_0, \ldots, n_k))$. Thus for each s it is the case that either for all choices of distinct numbers $n_0, \ldots, n_k$, $T(\varrho_s(n_0, \ldots, n_k))$ is empty, or else for all such choices $T(\varrho_s(n_0, \ldots, n_k))$ is dense in $\mathcal{T}$. Moreover, the non-empty sets $T(\varrho_s(n_0, \ldots, n_k))$ form, for any such choice, a partition of T.

Now let $t, t' \in T$ and let $n_0, \ldots, n_k$ and $m_0, \ldots, m_k$ be two choices of k distinct natural numbers such that for some $s \leq 2^n$, $t \in T(\varrho_s(n_0, \ldots, n_k))$ and $t' \in T(\varrho_s(m_0, \ldots, m_k))$. Then, by a simple generalization of Cantor's argument, there is an order-automorphism g of $\mathcal{T}$ such that $g(t) = t'$ and for $i = 1, \ldots, n$ $g(T(\varphi_i(n_0, \ldots, n_k))) = T(\varphi_i(m_0, \ldots, m_k))$. Since $\mathcal{F}_0(i)$ is $\{\mathcal{T}\}$-invariant it follows that $t \in T((C_i^n(\varphi_1, \ldots, \varphi_n))(n_0, \ldots, n_k))$ iff $t' \in T((C_i^n(\varphi_1, \ldots, \varphi_n))(m_0, \ldots, m_k))$. Let $s_1, \ldots, s_u$ be all those numbers $s \leq 2^n$ such that for any distinct $n_0, \ldots, n_k$ $T(\varrho_{s_j}(n_0, \ldots, n_k)) \subseteq T((C_i^n(\varphi_1, \ldots, \varphi_n))(n_0, \ldots, n_k))$ for $j = 1, \ldots, u$. Now let $B(\varphi)$ be a representation the Range of which is the union of the ranges of $B_{s_1}, \ldots, B_{s_u}$.[g] Then $B(\varphi)$ stands to φ in the relation required by (4).

Suppose that φ is a formula beginning with an existential quantifier, all free variables of which are among $v_0, \ldots, v_k$. For convenience let us assume that φ is the formula

$$(\exists v_{k+1})(\sim v_{k+1} = v_0 \wedge \ldots \wedge \sim v_{k+1} = v_k \wedge \psi).$$

Suppose first that $B(\psi) = \{\langle 1, \langle \{v_0\}, \{v_0\}\rangle\rangle\}$. Then for any choice of distinct $n_0, \ldots, n_k, n_{k+1}$, $T(\psi(n_0, \ldots, n_k, n_{k+1})) = \emptyset$. So $T(\varphi(n_0, \ldots, n_k)) = \emptyset$ for all distinct $n_0, \ldots, n_k$. So we may put $B(\varphi) = \{\langle 1, \langle \{v_0\}, \{v_0\}\rangle\rangle\}$. Now suppose instead that $B(\psi) \neq \{\langle 1, \langle \{v_0\}, \{v_0\}\rangle\rangle\}$. Assume that Dom $B(\psi) = \{1, \ldots, p\}$. For each $i \leq p$ let $C_i^1 = B_i^1 \setminus \{v_{k+1}\}$, $C_i^2 = B_i^2 \setminus \{v_{k+1}\}$. I claim that

(5) for any choice of distinct $n_0, \ldots, n_k$,
$T(\varphi(n_0, \ldots, n_k)) = \cup_{i \leq p} \mathcal{S}(C_i^1(n_0, \ldots, n_k), C_i^2(n_0, \ldots, n_k))$.

To show (5) first suppose that $t \in T(\varphi(n_0, \ldots, n_k))$. Then there is a number $n_{k+1} \neq n_0, \ldots, n_k$ such that $t \in T(\psi(n_0, \ldots, n_k, n_{k+1}))$. So $S_t \in \mathcal{S}(B_i^1(n_0, \ldots, n_k, n_{k+1}), B_i^2(n_0, \ldots, n_k, n_{k+1}))$ for some $i \leq p$. So $B_i^1(n_0, \ldots, n_k, n_{k+1}) \subseteq S_t$ and $B_i^2(n_0, \ldots, n_k, n_{k+1}) \cap S_t = \emptyset$. So $C_i^1(n_0, \ldots, n_k) \subseteq S_t$ and $C_i^2(n_0, \ldots, n_k) \cap S_t = \emptyset$. So $t \in \cup_{i \leq p} \mathcal{S}(C_i^1(n_0, \ldots, n_k), C_i^2(n_0, \ldots, n_k))$. Now suppose that $t \in$

[g] So the size of the Domain of $B(\phi)$ will be $\leq$ the sum of the sizes of the Domains of $B_{s_1}, \ldots, B_{s_u}$. In case $u = 0$ $B(\phi)$ will be the empty function. This means that $T(\phi) = \emptyset$. In this case we stipulate that $B(\phi)$ be instead the function $\{\langle 1, \langle \{v_0\}, \{v_0\}\rangle\rangle\}$.

$T(\mathcal{S}(C_i^1(n_0, \ldots, n_k), C_i^2(n_0, \ldots, n_k)))$ for some $i \leq p$. Since $B_i^1 \cap B_i^2 = \emptyset$, $\nu_{k+1} \notin B_i^1$ or $\nu_{k+1} \notin B_i^2$. Suppose that $\nu_{k+1} \notin B_i^1$. Let n_{k+1} be a number not in $S_t \cup \{n_0, \ldots, n_k\}$ (such a number exists as S_t is coinfinite). Then $B_i^1(n_0, \ldots, n_k, n_{k+1}) = C_i^1(n_0, \ldots, n_k)$ and $B_i^2(n_0, \ldots, n_k, n_{k+1}) \cap S_t = \emptyset$.[h] So $S_t \in \mathcal{S}(B_i^1(n_0, \ldots, n_k, n_{k+1}), B_i^2(n_0, \ldots, n_k, n_{k+1}))$. So $t \in T(\psi(n_0, \ldots, n_k, n_{k+1})) \subseteq T(\varphi(n_0, \ldots, n_k))$. The case where $\nu_{k+1} \notin B_i^2$ is treated similarly. Now let $B(\varphi) = \{\langle i, \langle C_i^1, C_i^2\rangle\rangle : i = 1, \ldots, p\}$. Then $B(\varphi)$ stands in the required relation to φ.

This completes the proof of (5). The case where φ begins with a universal, rather than an existential, quantifier is treated in exactly the same way. This completes the proof of (4).

Every set $T(B(\varphi)(n_0, \ldots, n_k))$ is either empty or else dense in $\mathcal{T}$; it follows from (4) that this is true also of the set of all $t \in T$ where a given sentence of $\mathcal{L}_{\mathcal{F}_0, Q'}$ is true in $\mathcal{M}$. In particular, since for every pair of disjoint finite subsets K, L of $\{\nu_0, \ldots, \nu_k\}$ and every $t \in T$ there are distinct numbers $n_0, \ldots, n_k$ such that $S_t \in \mathcal{S}(K(n_0, \ldots, n_k), L(n_0, \ldots, n_k))$ a sentence of $\mathcal{L}_{\mathcal{F}_0, Q'}$ will either be true in $\mathcal{M}$ at all t, or else false at all t. q.e.d.[i]

The assumption, made earlier, that Q contains only the predicate letter $\mathcal{Q}_1^1$, can be considerably weakened. In fact, all we have to assume about Q is that it contains at least one predicate letter of one or more places. That this assumption suffices can be seen as follows. Let us first assume that Q contains only predicate symbols. For any predicate letter $\mathcal{Q}_i^n \in Q$ let the interpretation $R(\mathcal{Q}_i^n)$ in the constructed interpretation $\mathcal{M}$ be defined by $R(\mathcal{Q}_i^n)(t) = \{\langle m_0, \ldots, m_n\rangle : m_0 \in S_t \text{ and } m_0 = m_1 = \ldots = m_n\}$. Then clearly the sentence $L(\exists\nu_0)(\mathcal{Q}_i^n(\nu_0, \ldots, \nu_0) \wedge N\mathcal{Q}_i^n(\nu_0, \ldots, \nu_0))$ is true only at 0. Associate with every formula $\mathcal{Q}_i^n(\nu_{i_1}, \ldots, \nu_{i_n})$, where for some $r, s \leq n$, $\nu_{i_r} \neq \nu_{i_s}$, the representation $\{\langle 1, \langle\{\nu_0\}, \{\nu_0\}\rangle\rangle\}$ and with every formula $\mathcal{Q}_i^n(\nu_i, \ldots, \nu_i)$ the representation $\{\langle 1, \langle\{\nu_0\}, \emptyset\rangle\rangle\}$. One easily verifies that in

---

[h] This statement is sloppy and strictly speaking it doens't mean anything. What should have been said: S_t satisfies the constraint imposed by $B_i^2(n_0, \ldots, n_k, n_{k+1})$ because (*i*) S_t satisfies the constraint imposed by $C_i^2(n_0, \ldots, n_k)$ and (*ii*) S_t doesn't cointain the number (n_{k+1}), so it will satisfy the additional constraint that B_i^2 imposes in case it cointains ν_{k+1}

[i] The point of this paragraph is not made as clearly as it could and should have been. The point—which is really simple enough—is that it follows from (4) that in the particular case where ϕ is a sentence, there are just two possibilities for $B(\phi)$:

(i) $B(\phi) = \{\langle 1, \langle\{\nu_0\}, \{\nu_0\}\rangle\rangle\}$.
(ii) The Range of $B(\phi)$ is $\{\langle\emptyset, \emptyset\rangle\}$, which comes to the same thing as saying that $B(\phi) = \{\langle 1, \langle\emptyset, \emptyset\rangle\rangle\}$.

In case (i) $T(\phi) = \emptyset$; in case (ii) $T(\phi) = T$.

either case the representation stands to the formula in the relation required by (4). This, together with the inductive steps proved above, establishes that (4) holds for all formulae of $\mathcal{L}_{\mathcal{F}_0,Q'}$ so that formulae of $\mathcal{L}_{\mathcal{F}_0,Q'}$ are again either true nowhere or else at a set of moments which is dense in $\mathcal{T}$.

Now suppose that Q contains individual constants as well. Let c_β be one of those. Let $\mathcal{T}$ be as before. The interpretation $\mathcal{M}$ will now have the form $\langle \omega \cup \{\omega\}, R, H', 0\rangle$, where $R(\mathcal{Q}_i^n)$ is defined as before; $R(c_\beta)(t) = \omega$ and $t \in T$; and H' is a 1–1 function from a set of constants not in Q onto the set of constant functions from T into $\omega \cup \{\omega\}$. We can again prove (4) by induction, now restricting attention to formulae in which every subformula beginning with $(\exists \nu_i)$ or $(\forall \nu_i)$ is of the form $(\exists \nu_i)(\sim \nu_i = \nu_{i_1} \wedge \ldots \wedge \sim \nu_i = \nu_{i_k} \wedge \sim \nu_i = c_\beta \wedge \psi)$ or $(\forall \nu_i)(\sim \nu_i = \nu_{i_1} \wedge \ldots \wedge \sim \nu_i = \nu_{i_k} \wedge \sim \nu_i = c_\beta \rightarrow \psi)$, respectively; and associating with every atomic formula $\mathcal{Q}_i^n(\beta_1, \ldots, \beta_n)$, where at least one β_i is a constant in Q, the representation $\{\langle 1, \langle \{\nu_0\}, \{\nu_0\}\rangle\rangle\}$.[j]

The above results can be summarized in the following

THEOREM 6 *Let $\mathcal{K}$ be a class of partial orderings which contains at least one dense linear ordering without endpoints. Let $\mathcal{F}$ be an indexed family of tenses which are $\mathcal{K}$-invariant. Let Q be any set of predicate letters and individual constants which contains at least one predicate letter. Then there is a sentence ψ of $\mathcal{L}_{\mathcal{F},Q}(N)$ such that for no sentence φ of $\mathcal{L}_{\mathcal{F},Q}$ is $\psi \leftrightarrow \varphi$ $\mathcal{K}$-valid.*

---

[j] This argument isn't quite right either, although, here too, the point it is trying to make is a simple and modest one. The argument should be this:

Suppose that $\mathcal{Q}$ contains individual constants. Then every formula ϕ of $\mathcal{L}_{\mathcal{F}_0,\mathcal{Q}}(N)$ may contain some of these constants, but it will contain only finitely many of them, say $c_1, \ldots, c_r$. We can once again define an interpretation relative to some time structure isomorphic to the rational numbers in which some sentence from $\mathcal{L}_{\mathcal{F}_0,\mathcal{Q}}(N)$ (the same one that was referred to in the proof above) is true only at 0, whereas all sentences of $\mathcal{L}_{\mathcal{F}_0,\mathcal{Q}}$ are either true nowhere or else true everywhere. The only difference with the earlier argument is that if ϕ is a formula of $\mathcal{L}_{\mathcal{F}_0,\mathcal{Q}}$, then there will be some finite number of individual constants $c_1, \ldots, c_r$ occurring in ϕ. Just as it was possible to rewrite ϕ equivalently as the formula in (3) earlier on, it is now possible to rewrite it as a formula in which each subformula beginning with an existential quantifier has the form in (2)

(2) $\quad (\exists \nu_i)(\sim \nu_i = \nu_{j_1} \wedge \ldots \wedge \sim \nu_i = \nu_{j_k} \wedge \nu_i = c_1 \wedge \ldots \wedge \sim \nu_i = c_r \wedge \psi)$

and analogously for subformulae beginning with a universal quantifier. By interpreting all individual constants in the interpretation as denoting some new element of the Domain of the model (such as, for instance $\{\omega\}$) that was not part of the Domain of the earlier interpretation and which does not occur in the extension of any predicate in $\mathcal{Q}$ at any time,—so that every atomic formula involving a predicate from $\mathcal{Q}$ will be false at any time for all assignments of values to its free variables—the inductive steps in the proof of (4) that deal with restricted quantified formulas can then be pushed through in much the same way as in the proof given in the body of the text.

The second theorem of this section is concerned with discrete orderings. I will construct an interpretation relative to the integers (regarded as a linearly ordered structure) in which again a sentence containing N is true only at 0, while all sentences not containing N will be true either nowhere or else at arbitrarily large positive and negative integers. This result will not be as general as Theorem 6, however, since it will only apply to languages $\mathcal{L}_{1,Q}$ and $\mathcal{L}_{1,Q}(N)$. For the construction I will assume that $Q = \{Q_1^1\}$. As in the previous proof this assumption can be weakened to the hypothesis that Q contains at least one predicate letter.

Let T' be the set of all integers different from 0, and let $T = T' \cup \{0\}$. The interpretation $\mathcal{M}$ will again be of the form $\langle \omega, R, H, 0\rangle$, where H is as before and $R(Q_1^1)(t)$ is a member S_t of $\mathcal{S}$. I will now describe the construction of the S_t. The idea is to construct successively S_1, S_{-1}, S_2, S_{-2}, etc., and finally S_0. Let $\{E_\alpha\}_{\alpha<2^\omega}$ and $\{J_k\}_{k<\omega}$ be the enumerations used in the previous proof. Let $\{i_j\}_{j<\omega}$ be the enumeration of the integers different from 0 which starts with $1, -1, 2, -2, \ldots$ and continues in the obvious way. Let W be an enumeration of all pairs $\langle i, -1\rangle$ and $\langle i, +1\rangle$, where i is an integer $\neq 0$, and all pairs $\langle k, -2\rangle$ and $\langle k, +2\rangle$, where k is a natural number. We construct for $n = 0, 1, 2, \ldots$ infinite, coinfinite subsets S'_n of ω and subsequences W_n of W as follows:

(1) Let $S'_0 = E_0$, $W_0 = W$.
(2) Suppose that S'_m, W_m have been constructed for $m \leq n$.
  (a) Suppose that i_{n+1} is positive. Let w be the first member of W_n such that either $w = \langle i, +1\rangle$ and $i = i_m$ for some $m \leq n$; or $w = \langle k, +2\rangle$. If $w = \langle i, +1\rangle$ let S'_{n+1} be the first E_α not yet used which is disjoint from S'_m; if $w = \langle k, +2\rangle$ let S'_{n+1} be the first E_α not yet used which belongs to $\mathcal{S}(J_k)$.
  (b) Suppose that i_{n+1} is negative. Let w be the first member of W_n such that either $w = \langle i, -1\rangle$ and $i = i_m$ for some $m \leq n$; or $w = \langle k, -2\rangle$; if $w = \langle i, -1\rangle$ let S'_{n+1} be the first E_α not yet used which is disjoint from S'_m; if $w = \langle k, -2\rangle$ let S'_{n+1} be the first E_α not yet used which belongs to $\mathcal{S}(J_k)$

In each of the cases under (a) and (b) let W_{n+1} be the sequence which results if we omit w from W_n. Let, for each $t \in T'$, $S_t = S'_n$, where $t = i_n$. Let S_0 be an infinite, coinfinite subset of ω which contains an element from each S_t such that $t \in T'$. One easily verifies that

(i) for each $t \in T'$ there are $t' < t$ and $t'' > t$ such that $S_{t'} \cap S_t = S_{t''} \cap S_t = \emptyset$; and
(ii) for any disjoint finite subsets K, L of ω and $t \in T$ there are $t' < t$ and $t'' > t$ such that $S_{t'}$ and $S_{t''}$ belong to $\mathcal{S}(K, L)$.

From (i) it follows that the sentence $L(\exists \nu_0)(Q_1^1(\nu_0) \wedge NQ_1^1(\nu_0))$ is true in $\mathcal{M}$ only at 0. (ii) enables us to prove that condition (4) holds (for formulae

of $\mathcal{L}_{1,Q}$). Again the proof is by induction on restricted formulae. It proceeds along the same lines as the argument given above, the only difference being that the inductive step where φ is of the form $C_i^n(\varphi_1, \ldots, \varphi_n)$ is now replaced by the two special cases where φ is $G\psi$ or $H\psi$. I will consider the case where φ is $G\psi$. Suppose that φ is $G\psi$ and that $B(\psi)$ has been defined. If $\mathcal{V}(B(\psi))$ consists of all infinite, coinfinite subsets of V then for every choice of distinct natural numbers $n_0, \ldots, n_k$ (where I assume that all sets which occur in the Range of $B(\psi)$ are included in $\{\nu_0, \ldots, \nu_k\}$), and every $t \in T(\psi(n_0, \ldots, n_k))$ is true at t. So $G\psi(n_0, \ldots, n_k)$ is true at all t, and we may therefore put $B(\varphi) = \{\langle 1, \langle \emptyset, \emptyset \rangle \rangle\}$. Now suppose that $\mathcal{V}(B(\psi))$ does not consist of all infinite, coinfinite sets of variables. Then, by our remarks on the Boolean algebras $\mathcal{B}(\omega, \mathcal{S})$ and $\mathcal{B}(V, \mathcal{V})$ there are finite sets K, L of variables such that $\mathcal{V}(K, L) \cap \mathcal{V}(B(\psi)) = \emptyset$. Let t be a member of $T, n_0, \ldots, n_k$ distinct natural numbers. By (ii) there is for every $t \in T$ some $t' > t$ such that $S_{t'} \in \mathcal{S}(K(n_0, \ldots, n_k), L(n_0, \ldots, n_k))$ and therefore $\psi(n_0, \ldots, n_k)$ is not true at t'. Thus for each choice of distinct $n_0, \ldots, n_k$ and each $t \in T$, $\varphi(n_0, \ldots, n_k)$ is false at t. Thus we may put $B(\varphi) = \{\langle 1, \langle \{\nu_0\}, \{\nu_0\} \rangle \rangle\}$.

The above argument yields the following

THEOREM 7 *Let Q be a set of predicate letters and individual constants which contains at least one predicate letter, and let $\mathcal{K}$ be a class of partial orderings which contains at least one ordering which is isomorphic with the integers. Then there are sentences ψ of $\mathcal{L}_{1,Q}(N)$ such that for no sentence φ of $\mathcal{L}_{1,Q}$, $\psi \leftrightarrow \varphi$ is $\mathcal{K}$-valid.*

REFERENCES

Bull, R., 1968. An algebraic study of tense logics with linear time. *The Journal of Symbolic Logic* **33**, 27–38.

Kalish, D., Montague, R., 1964. *Logic. Techniques of Formal Reasoning*. Harcourt, Brace & World, New York.

Kamp, H., 1968. *Tense Logic and the Theory of Linear Order*. Ph.D. dissertation, University of California, Los Angeles, CA.

Lemmon, E., Scott, D., 1966. *Intensional Logic*, preliminary draft of initial chapters by E. Lemmon, Stanford University. Published as: E. Lemmon, D. Scott, *An Introduction to Modal Logic: The 'Lemmon Notes'*, ed. by K. Segerberg. Basil Blackwell, Oxford, 1977.

Prior, A., 1968. "Now". *Noûs* 2, 101–119.

EVENTS, INSTANTS AND TEMPORAL REFERENCE[a,*]

Hans Kamp

I

The truth of my claim: "It is raining" depends on the time at which I made it. It was true if and only if it was indeed raining (in my vicinity) at that time. Similarly my assertion "it has been raining", made at time t, is true if there was a time t' preceding t such that it rained at t'.

But what is a time? In physics, where it has been standard practice to represent time as a structure isomorphic with the real numbers, we have become accustomed to thinking of times as either durationless instants or else collections of such instants.

The concept of a durationless temporal instant is a quite sophisticated one, the eventual response, after many centuries of experimentation, to Zeno's puzzle. It contains at least two distinct mathematical ideas, both of which are by no means obvious. In fact, they appear counterintuitive to many who encounter them for the first time. The first of these concerns the possibility of densely ordered structures, between any two elements of which there is a third; the second the possibility that a collection of elements each of which has zero size by itself, may nonetheless have a size bigger than 0. (Thus in particular, an interval may have finite duration even though each of the instants that constitute it has no duration whatever.)

* The research for this paper was done while the author was supported by the DFG-Project 'Die Beschreibung mit Hilfe der Zeitlogik von Zeitformen und Verbalformen im Französischen, Portugiesischen und Spanischen' at the University of Stuttgart.

[a] Editors' Note: This article was originally published in R. Bäuerle, U. Egli, & A. von Stechow (eds.), *Semantics from Different Points of View*, Springer, Berlin, 1979, 376–417. We would like to thank Springer for the kind permission to reprint the article. We carefully edited the article, removed some typos, added full bibliographical references and brought some formulas in line with now more common usage. We would like to thank Fritz Hamm and Ken Turner for careful proofreading and further comments. Hans Kamp provided some additional footnotes, indicated by Roman letters, which clarify some of the arguments in the original text.

© HANS KAMP, 2013 | DOI:10.1163/9789004252882_005
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

These principles tend to strike non-mathematicians as particularly puzzling when they are stated in the context of time; and one of the reasons for this is no doubt that the instant-concept which would satisfy them is so very far removed from the ideas that people ordinarily have of 'time'—however inarticulate these ideas may be.

Is time really like the real numbers? Or is it rather like the rationals, or like the integers, or has it some other structure? Before we can even make an attempt at answering such a question, we must first determine what sort of question it is. And the sort of question it is depends on what we take to be the nature of time. If time is, as Kant suggested, a mental category, then the question is about how, and to what extent, the structure of the mind determines the formal properties of this category. If we take—as, it seems, Poincaré advocated—the structure of time to be a matter of theoretical convention, with no more justification, and no more need of justification, then its utility to a highly successful complex of scientific theories, the question disappears. If we assume, with Leibniz, Einstein or Whitehead, that time is no more than the totality of temporal relations between the events and processes which constitute the history of our world, the question is about the actual structural relations between these events and processes. If we assume—as it appears Russell did at some stage—that statements about time, like other statements which appear to be about the world outside us, are in last analysis complex claims about actual and possible experiences, then the question is about the structure of these experiences.

Only the last two of these different conceptions are relevant to what I wish to say below. These two views, however different in their general philosophical perspectives, are nevertheless quite similar in an important formal respect: Each takes as primary certain entities—be they mental experiences or physical events—which are generally—in fact it is reasonable to assume *always*—of finite duration. I shall refer to such entities henceforth as *events*. If we can talk about durationless instants at all, these must, on either of the two views, be constructs, built from elements which are themselves not without duration.

II

There are various ways in which such a construction can proceed. The one which I shall present here goes back to Russell and Wiener.[1] This particular

[1] Cf. Russell (1936), Wiener (1914).

construction assumes that with the basic elements (be they mental or physical occurrences) are given two binary temporal relations, the relation of complete precedence, $\prec$, and that of temporal overlap, $\bigcirc$. We assume that these relations satisfy the following postulates:

(A1) $(\forall x \forall y)(x \prec y \rightarrow \neg y \prec x)$
(A2) $(\forall x \forall y \forall z)(x \prec y \wedge y \prec z \rightarrow x \prec z)$
(A3) $(\forall x) x \bigcirc x$
(A4) $(\forall x \forall y)(x \bigcirc y \rightarrow y \bigcirc x)$
(A5) $(\forall x \forall y)(x \prec y \rightarrow \neg x \bigcirc y)$
(A6) $(\forall x \forall y \forall z \forall t)(x \prec y \wedge y \bigcirc z \wedge z \prec t \rightarrow x \prec t)$
(A7) $(\forall x \forall y)(x \prec y \vee x \bigcirc y \vee y \prec x)$[2]

Suppose that $\tau = \langle T, < \rangle$ is a strict linear ordering, that E is a set of non-empty intervals of τ; and that for $e_1, e_2 \in E$

$e_1 \prec (\tau) e_2 \equiv_{df.} (\forall t_1 \in e_1)(\forall t_2 \in e_2)(t_1 < t_2)$; and
$e_1 \bigcirc (\tau) e_2 \equiv_{df.} (\exists t)(t \in e_1 \wedge t \in e_2)$.

Then the structure $\langle e, \prec (\tau), \bigcirc (\tau) \rangle$ satisfies each of A1–A7. By verifying that this is so, the reader may discover the intuitive motivation behind these postulates. But in view of the use we wish to make of them such an argument would be circular if presented as a justification for adopting them. Indeed the only possible justification for A1–A7 must derive directly from the intuitions regarding the events or experiences themselves and the temporal relations by which they are connected. Thus what justification can be given will depend largely on the metaphysical assumptions we can make about the nature of time; and it is here that what we might refer to as the Russellian and the Whiteheadian approaches part company.[3]

I am not much attracted either by Russell's phenomenalism or by Whitehead's metaphysics; and so I shall not attempt to justify the postuates A1–A7 in the context of either of these two frameworks. I shall, in due course, come up with an interpretation of my own. But first let me describe how the instants may be constructed from events structured in accordance with A1–A7. The construction is to provide us with instants such that for each of the events there is a certain set of instants at each of which the event is

[2] A2 is strictly speaking superfluous, as it can be deducted from A6 and A3.

[3] I should point out in this connection that Whitehead himself used a different construction of instants out of processes from the one I shall discuss here. However, the same fundamental issues that arise in connection with the present construction also arise in connection with his. A discussion of many of the same concerns to which this paper tries to address itself, but which uses a construction more akin to that of Whitehead is to be found in van Benthem (1980). See also Whitehead (1929).

"going on". This means that if i_1 is an instant at which e_1 goes on, i_2 is an instant at which e_2 goes on and $e_1 \prec e_2$ then i_1 must indeed be earlier than i_2. This indicates how we might define the earlier-later relation; but it does not show us how to construct the instants themselves.

When $e_1 \bigcirc e_2$, then, we would like to say, there ought to be some instant at which e_1 and e_2 go on simultaneously. Similarly, if e_1, e_2 and e_3 have a common overlap there ought to be an instant at which all three go on. Now, since $\bigcirc$ is a binary operator, the proposition that e_1, e_2 and e_3 have a common overlap cannot be expressed directly. However, we may convince ourselves that its content is given by the conjunction $e_1 \bigcirc e_2 \,\&\, e_1 \bigcirc e_3 \,\&\, e_2 \bigcirc e_3$. For, given that the events are, intuitively speaking, uninterrupted, how could the conjunction hold without there being common overlap?[4] Similarly, if x is a set of events such that for any two members e_1, e_2 of x $e_1 \bigcirc e_2$ then there ought to be an instant at which all events in x go on simultaneously.[5] Of course there might be several such instants. In fact there would have to be several such if there were events e_1 and e_2 such that $e_1 \prec e_2$ while both e_1 and e_2 overlap each of the events in x. For then there ought to be an instant at which all members of $x \cup \{e_1\}$ go on and an instant at which all members of $x \cup \{e_2\}$ go on, and of these the first must precede the second. This situation cannot arise, however, if x itself is *maximal*, i.e. if there is no event e which overlaps all the members of x and yet is not already in x. Even in this case we could assume without contradiction that there are distinct instants at which each of the members of x goes on; but the principle that any two which do *not* overlap do *not* share any instants does no longer force us to assume such multiplicity.

Indeed it is not difficult to show that if we identify instants with such maximal sets we get what we were looking for.

Formally:

Let $\varepsilon = \langle E, \prec, \bigcirc \rangle$ be a structure which satisfies the postulates A1–A7.

Def.

(a) An *instant of* ε is a maximal subset of E of pairwise overlapping events, i.e. i is an *instant of* ε iff (i) $i \subseteq E$; (ii) for any $e_1, e_2 \in i$ $e_1 \bigcirc e_2$; and (iii) for any $e_1 \in E \setminus i$ there is an $e_2 \in i$ such that $\neg e_1 \bigcirc e_2$

(b) Let $I(\varepsilon)$ be the set of instants of ε.

(c) For $i_1, i_2 \in I(\varepsilon)$, $i_1 < i_2$ iff there are $e_1 \in i_1$ and $e_2 \in i_2$ such that $e_1 \prec e_2$.

---

[4] It is easily seen that if e_1, e_2, e_3 are intervals of an ordering $\langle T, < \rangle$ and $\bigcirc$ is defined as above, then if $e_1 \bigcirc e_2$, $e_2 \bigcirc e_3$ and $e_1 \bigcirc e_3$ there is always a $t \in T$ such that $t \in e_1 \cap e_2 \cap e_3$.

[5] Actually there is a complication here when x is infinite. See also Section III below.

THEOREM Suppose $\varepsilon = \langle E, \prec, \bigcirc \rangle$ satisfies A1–A7. Then $\langle I(\varepsilon), <_\varepsilon \rangle$ is a strict linear ordering.

PROOF: We must show that ε satisfies the following conditions:

(i) $i_1 < i_2 \rightarrow \neg i_2 < i_1$;
(ii) $i_1 < i_2 \wedge i_2 < i_3 \rightarrow i_1 < i_3$;
(iii) $i_1 \neq i_2 \rightarrow i_1 < i_2 \vee i_2 < i_1$.

(i) Suppose $i_1 < i_2$, so there are $e_1 \in i_1$ and $e_2 \in i_2$ such that $e_1 \prec e_2$. Suppose now also that $i_2 < i_1$. Then there would also be $e_3 \in i_1$ and $e_4 \in i_2$ such that $e_4 \prec e_3$. Since e_3 and e_1 belonging to the same instant we have $e_3 \bigcirc e_1$. So by A6 we infer from $e_4 \prec e_3, e_3 \bigcirc e_1, e_1 \prec e_2$ that $e_4 \prec e_2$. But on the other hand e_4 and e_2 also belonging to the same instant i_2. So $e_4 \bigcirc e_2$. This contradics A5.

(ii) Suppose $i_1 < i_2$ and $i_2 < i_3$. So there is an $e_1 \in i_1$ and an $e_2 \in i_2$ such that $e_1 \prec e_2$. Also there is an $e_3 \in i_2$ and an $e_4 \in i_3$ such that $e_3 \prec e_4$. Finally since $e_2, e_3 \in i_2, e_2 \bigcirc e_3$. So by A6 $e_1 \prec e_4$.

(iii) Assume $i_1 \neq i_2$. So there is an e_1 such that either $e_1 \in i_1 \setminus i_2$ or $e_1 \in i_2 \setminus i_1$. Suppose $e_1 \in i_1 \setminus i_2$. Then there is an $e_2 \in i_2$ such that $\neg e_1 \bigcirc e_2$.[b] So by A7 either $e_1 \prec e_2$ or $e_2 \prec e_1$. If $e_1 \prec e_2$ then $i_1 < i_2$; in the other case $i_2 < i_1$. If $e_1 \in i_2 \setminus i_1$ the argument proceeds similarly. q.e.d.

I shall refer to the structure $\langle I(\varepsilon), <_\varepsilon \rangle$ also as '$\tau(\varepsilon)$'. Notice that if we assume that ε satisfies only A1–A6, then it is still possible to show that $<_\varepsilon$ satisfies (i) and (ii). But of course it is then no longer possible to show that $<_\varepsilon$ is connected.

Given the way we have defined instants the relation which holds between an event e and an instant i when intuitively we would wish to say that e "goes on" at i is simply the set theoretic relation of membership: "e goes on at i" can be defined as "$e \in i$". It is quite easy to show that for each $e \in E$ the set $\bar{e} = \{i \in I(\varepsilon) : e \in i\}$ is a non-empty interval of $\tau(\varepsilon)$.

For suppose $i_1, i_2 \in e$ and $i_1 <_\varepsilon i <_\varepsilon i_2$. We must show that $i \in \bar{e}$, or, in other words, that $e \in i$. Suppose not. Then there is an $e_0 \in i$ such that $\neg e_0 \bigcirc e$. So $e_0 \prec e$ or $e \prec e_0$. Suppose $e_0 \prec e$. There are $e_1 \in i_1$ and $e_2 \in i$ such that $e_1 \prec e_2$. Since $e_0, e_2 \in i, e_2 \bigcirc e_0$. So by A6 $e_1 \prec e$. But $e_1, e \in i$, so $e_1 \bigcirc e$: contradiction. The case where $e \prec e_0$ is treated analoguously.

Of course, in general not every interval of $\tau(\varepsilon)$ will correspond in this way to an event of ε. Nor can we be certain that an interval that does correspond to an event determines that event uniquely: there might be events which,

[b] The original text has '$e_1 \bigcirc e_2$' here. But this is obviously a mistake, it must correctly be $\neg e_1 \bigcirc e_2$.

although distinct, are yet perfectly simultaneous. Any two such events will stand, however, in exactly the same temporal relation to every other event of ε.

We have already seen how to define relations $\propto(\tau)$ and $\bigcirc(\tau)$ for intervals of some strict linear ordering τ. It is also easily verified that if $e_1 \propto e_2$ then $\bar{e}_1 \propto(\tau(\varepsilon))\ \bar{e}_2$, and if $e_1 \bigcirc e_2$ then $\bar{e}_1 \bigcirc(\tau(\varepsilon))\ \bar{e}_2$. Thus the function $e \Rightarrow \bar{e}$ is a homomorphic embedding of ε into the structure $\varepsilon(\tau(\varepsilon)) = \langle E(\tau(\varepsilon)), \langle \propto(\tau(\varepsilon)), \bigcirc(\tau(\varepsilon))\rangle\rangle$, where $E(\tau(\varepsilon))$ is the set of all intervals of $\tau(\varepsilon)$. To the structure $\varepsilon(\tau(\varepsilon))$ we can apply the Russell-Wiener construction once again. This yields a linear ordering, to which I shall for notational simplicity, refer here as $\tau' = \langle T', <'\rangle$. In general τ' will not be an isomorphic copy of $\tau(\varepsilon)$. To see this suppose that ε consists of two denumerable sequences of events $\{e_1, e_2, e_3, \ldots\}$ and $\{f_1, f_2, f_3, \ldots\}$ and that $e_1 \propto e_2 \propto e_3 \propto \ldots \propto f_3 \propto f_2 \propto f_1$. In this case each instant of $\tau(\varepsilon)$ will consist of precisely one event. Among the maximal sets of pairwise overlapping intervals of $\tau(\varepsilon)$, however, there is in particular one which contains (among others) for each $j = 1, 2, 3, \ldots$ both the interval (e_1, f_j) and the interval (e_j, f_1). This set of intervals determines an instant of τ' which precedes all the f_j's and is preceded by all the e_j's. This instant does therefore not correspond to any instant of $\tau(\varepsilon)$.

In general the structure τ' will not be what we should like it to be. Thus in the example just discussed τ' would contain besides the points already in $\tau(\varepsilon)$, in addition *two* new instants i_1 and i_2 such that $\{\bar{e}_1\} < \{\bar{e}_2\} < \ldots < i_1 < i_2 < \ldots < \{\bar{f}_2\} < \{\bar{f}_1\}$ where $\{\bar{e}_i\}$, $\{\bar{f}_i\}$ are the points of τ' corresponding to points $\{e_i\}$, $\{f_i\}$ of $\tau(\varepsilon)$. i_1 will be the maximal set containing all intervals of the form $\bigcup_{i>n}\{e_i\}$ where n is some natural number, while i_2 is the set which contains all intervals $\bigcup_{i>n}\{f_i\}$ for some n. This is displeasing—it would have been much more satisfactory had τ' been the Dedekind closure of $\tau(\varepsilon)$, producing, in particular, in the example under consideration exactly one instant for the gap between the points $\{e_i\}$ and the points $\{f_i\}$. To obtain the Dedekind closure we must restrict the family of intervals of $\tau(\varepsilon)$ which enter into the construction of the instants of τ' to all closed intervals which are bounded by instants of $\tau(\varepsilon)$, i.e. all intervals of the form $[i_1, i_2]$ where $i_1, i_2 \in I(\varepsilon)$.

This alternative has its awkward side too, however, in as much as some structures ε contain events e such that the corresponding interval $\bar{e}$ of $\tau(\varepsilon)$ is not of the form $[i_1, i_2]$. This would be the case for instance if the event structure ε of our example contained besides the e_i and f_i an event g which overlaps all of the e_i but none of the f_i. The same would be true for an event h which overlaps all of the events f_i but none of the e_i. As soon as we admit

the intervals $\bar{g}$ and $\bar{h}$ together with all the closed intervals of $\tau(\varepsilon)$ in the construction of τ' we obtain two instants i_1 and i_2, rather than one, to fill the gap of $\tau(\varepsilon)$. (i_1 would contain besides all intervals $[\{e_i, g\}, \{f_j, h\}]$ also $\bar{g}$ while i_2 would contain $\bar{h}$ as well as all these closed intervals.)

It is thus not true in general that τ' is the Dedekind closure of $\tau(\varepsilon)$. It should be added however that in all cases the difference between τ' and the Dedekind closure of $\tau(\varepsilon)$ is slight. It can consist only in that for certain gaps of $\tau(\varepsilon)\tau'$ has a pair of instants there where the Dedekind closure $\tau(\varepsilon)$ has got only one. I shall not investigate here the possibility of imposing additional conditions on ε which would warrant that τ' and the Dedekind closure of $\tau(\varepsilon)$ are indeed isomorphic.[6]

The event structure $\varepsilon(\tau')$ formed out of the intervals of τ' will be in general richer than $\varepsilon(\tau(\varepsilon))$. But after that we finally reach stability: $\tau(\varepsilon(\tau'))$ is isomorphic with τ'.

III

In section I I asked, by way of an introduction to the discussion which followed, the question how one might determine the structural properties of time. Let us assume that time, presented as a linear ordering of points, is derived from an underlying structure of events. Does this help us to answer that question?

The obvious strategy must now be (i) to discover additional postulates for event structures which allow us to derive additional structural conditions on the derived ordering; (ii) to find independent justification for these additional postulates. Russell made a valiant effort in this direction in his

---

[6] I should mention here an alternative method for constructing instants out of events which, it appears, was first studied by Whitrow (1961: 157 ff.). This method is based on the same idea that underlies Dedekind's construction of the real numbers from the rationals. Given an event-structure ε we can define a *cut of* ε to be a maximal pair (A, B) of subsets of E such that for each $e_1 \in A$ and $e_2 \in B$ $e_1 \prec e_2$—the pair is maximal in that for each event e not in $A \cup B$ there is an $e_1 \in A$ such that not $e_1 \prec e$ or an $e_2 \in B$ such that not $e \prec e_2$. It is easily shown that if we define instants as cuts and put $(A_1, B_1) < (A_2, B_2) \equiv_{df.} A_1 \subset A_2$, then we obtain again a linear ordering. This ordering moreover is already complete, so that τ' is in fact isomorphic with $\tau(\varepsilon)$ itself. It is easily verified that if (A, B) is a cut and $e \notin A \cup B$ then there are $e_1 \in A$ and $e_2 \in B$ such that $e_1 \bigcirc e$ and $e_2 \bigcirc e$. This makes it reasonable to define the events that go on at (A, B) as precisely those events that belong neither to A nor to B. The problem with this approach is that if e is an instantaneous event—i.e. if there are no e_1, e_2 such that $e_1 \bigcirc e$, $e_2 \bigcirc e$ and $e_1 \prec e_2$—then there will be no instant at all at which e_1 happens. It will become clear in section V why this makes the Walter approach ill-suited to the purpose of this paper.

work (Russell 1936). But his work seems to bear out the conclusion that this course is not as fruitful as one might have hoped. For the new postulates for ε tend to be just as plausible, or dubious, as the additional conditions on $\tau(\varepsilon)$ that can be derived with their help. To show that $\tau(\varepsilon)$ is dense, for instance, one has to adopt some such postulate as:

$$(\forall e_1)(\forall e_2)(e_1 \prec e_2 \rightarrow (\exists e_3)(\exists e_4)e_3 \prec e_2 \wedge e_1 \prec e_4 \wedge e_3 \bigcirc e_4).$$

It is evident that with this extra condition we can demonstrate the density of $<_\varepsilon$. But why should the postulate be accepted?

It is not for the sake of settling such problems however that I have discussed the Russell-Wiener construction. The two applications that I have in mind for it are in a quite different direction. The first of these applications is in fact quite distinct, I believe, from anything that the authors had contemplated themselves. And the second, though more commensurable, with their own intentions (as well as, for that matter, with those of Whitehead) is yet in important aspects different from what I expect either Russell or Whitehead would have accepted.

The first application is a purely linguistic one. I have become increasingly convinced that there are certain linguistic phenomena for which we can only account via a systematic analysis of the ways in which discourse participants process the information with which the discourse provides them. Among these phenomena there are some which concern reference to time, and of such phenomena I shall give some examples later on. To illustrate what I have in mind let us consider a very simple situation, in which one person, *S*, does all the talking, and another, *H*, all the listening. We shall assume, moreover, that *S* utters a sequence $s_1, \ldots, s_n$ of indicative sentences, as one would do when telling a story, or reporting on an event one has witnessed. As the discourse proceeds *H* will represent the information conveyed to him; and we shall suppose—another simplifying assumption—that he modifies his representation after each successive sentence that *S* utters. Thus *H* produces a series of representations $D_1, \ldots, D_n$. Each next representation D_{i+1} results from incorporating the new information contained in S_{i+1} into the structure D_i which represents all the information obtained from $s_1, \ldots, s_i$. Precisely how these D_i are structured will depend on many factors. Among these is the grammatical complexity of the sentences $s_1, \ldots, s_n$. I shall later on illustrate, for some exceedingly simple language, how the operation which produces D_{i+1} out of D_i and s_{i+1} might actually be defined. Suffice it for now to point out that in many types of discourse much of the information communicated is about certain events; when this is so the discourse will often make clear how the different events are temporally related. Such

information I propose to represent with the help of event structures of the kind discussed in the preceding section. The elements of E will be labeled events (where the label tells us what the event consists in); and the relations $<$ and $\bigcirc$ will contain the information about the temporal order of the events.

In general, however the information H receives will not specify the temporal order of the events completely.[7] This means that we must abandon postulate A7. I shall discuss the consequences of this in the next section. First, however, a preliminary sketch of the second application. As I already remarked, this application is more in accordance with the intentions of those who developed the theories of section II.

There is, I believe, a great deal of plausibility in the view that the nature of time is determined by the temporal connections of the events which constitute the history of our universe. I also believe that these temporal relations between events are objective—this is one of the important philosophical insights bequeathed by the theory of relativity. I do not believe, however, that a similar objectivity attaches to the concept of an event itself. What can count as an individual, separate, event depends on the means for singling it out; or for recognizing it as the presence, in some limited area of space and time, of a certain condition, or alternatively, as the change within some area of space-time in regard of some condition. But of conditions we can speak only where there is a conceptual scheme within which they can be expressed. And although conceptual schemes may well be partly determined by the nature of things, it would be unreasonable to assume that they are fully determined thereby. In particular, how rich the conceptual scheme is, (and therewith how rich the event structure it generates) depends not so much on intrinsic features of that in which the scheme is meant to create some order but rather on the needs and limitations of those who use it.

It would appear to be a consequence of the ways in which we learn and share concepts that many of them are vague. I do not have a conclusive argument that this is so. But, whether or not such an argument can be given, that many of the concepts we actually use are vague is an undeniable fact.

---

[7] This is particularly common in connection with eyewitness reports. The speaker tries to piece the events together as best he can. The various aspects of the case may come back to him more or less at random. Often he will be able to reconstruct the true temporal sequence of the occurrences he relates, and to communicate this additional information, only in the latter part of the discourse. Sometimes he won't arrive at a complete articulation of the temporal relations at all. In the former case some of the intermediate structures D_i, with $i < n$, will violate A7; in the latter case this will be true even of the final structure D_n.

Let us consider for simplicity just unary concepts. The vagueness of such a unary concept manifests itself in the presence of entities which are of the type to which the concept applies, but are nevertheless such that the criteria associated with the concept do not determine whether the entity falls under the concept or not. In some cases where the discovery of an entity which is neither clearly inside nor clearly outside the extension of a concept *c* confronts us with the vagueness of *c* we experience this as a deficiency; in such cases we will often try to sharpen *c* in such a way that the given entity *a*,—as well as, hopefully, many others whose relation to the concept had thus far been undecided—now fall clearly on one or the other side of the boundary of *c*'s extension; and our intention is that this modification of *c* be permanent. In other cases the encounter of such entities will not motivate us to do this, however. In fact it is arguable that some concepts,—such as e.g. *orange, clever, lonely, tiny, bald*,—are *essentially* vague; that in order to resolve the vagueness of these concepts we would have to alter them to a point where they would no longer fulfill their actual function: They would lose that flexibility which enables their users to adjust their extensions quite freely to the needs of the particular contexts in which they employ them. Although such concepts may be sharpened within particular contextual settings, the modification is meant to be provisional, and lapses when the context in question terminates. I have argued elsewhere[8] that vagueness is compatible with the laws of classical predicate logic, in as much as we can treat vague concepts as potentially sharp, although precisely *where* the boundary of the extension lies will be undetermined. In an earlier essay I may have given the impression that I believed that we must *always* take this attitude towards vague concepts. This would be an overstatement; and it is something which at that time I did not believe any more than I believe it now. It does seem to me to be correct, however, that we often treat vague concepts in this way. And that is all that is required for the following observation.

Suppose the sun is setting. First its colour is bright orange, then it turns into a bright red. At what time precisely does the sun cease to be orange, and become red? In as much as the borderline between orange and red is fuzzy this time cannot be determined with arbitrary precision. But does this mean that there is an *extended period p* during which the sun is turning red, a period such that up to the beginning of *p* the sun is orange, during *p* the sun is in transition and from the end of *p* the sun is red? I don't think

[8] See Kamp (1975).

that that is what it means. For with regard to p the same sort of question arises again: where precisely does p begin and where precisely does it end? It is not so much that the time of change is *extended*—it is rather that it is *undetermined*. Indeed, in any potential semantic extension of the language in which the borderline between red and orange is fixed perfectly the change will be instantaneous.

Suppose now that while the sun is setting the sky right above the tree to my right, which was blue at first, turns green. Once more, the colour changes gradually; but when the different shades of blue end and the different shades of green begin is again an issue that it will be impossible to answer exactly. Consider now the event of the sun's turning red, and that of the sky's turning green together. How are they temporally related? That too may be undetermined. In fact it will be undetermined if (i) it is semantically consistent to simultaneously fix the borderline between orange and red and that between blue and green in such a way that the first of these events occurs, according to this specification, before the second; and (ii) it is also consistent to fix the borderline so that the first event occurs later than the second.

Thus if we admit both the sun's turning red and the sky's turning green as elements e_1, e_2 of the structure ε of all events, then once again given the way the world is *and* the present status of our concepts we can assert neither $e_1 \prec e_2$ nor $e_1 \bigcirc e_2$ nor $e_2 \prec e_1$. So once again A7 must be abandoned.

It is natural to ask whether we could not save A7 by excluding such events as the sun's becoming red from ε altogether. I do not think, however, that that would be reasonable. For then we would eliminate far too much. Many, perhaps all, of those events would have to go which we identify as changes from a given condition into one of its opposites. What would remain? We often speak of an event where a certain condition pertains over a limited period of time; e.g. the nap I took two hours ago might be called an event. Such events, however, are always accompanied by a couple of changes, the beginning of the event—here the change from non-sleep to sleep—and its end—here the change from sleep into non-sleep. And precisely the same considerations that led us to the conclusion that changes may be incomparable leads therefore also to that conclusion in connection with events of the second type. How for example would one answer the following question: Did your nap completely precede mine, or was there a tiny overlap? Finally there are events that we might call *protracted changes*. Here we explicitly acknowledge that between the initial conditon c_1 and the final condition c_2 there are intermediate stages during which conditions obtain that are incompatible with both c_1 and c_2. With regard to such

protracted changes the same problem arises as with events of the second type: When for instance does the final condition c_2 obtain for the first time? Thus, even if we shun the potentially instantaneous changes, we still have no guarantee that A7 is satisfied.

IV

In each of the applications of the theory of events which I envisage the transformation of event structures into instant structures will play a central role. But, as we have just seen, we cannot, in the context of these applications, assume that the event structures satisfy A7. The instant structures into which the Russell-Wiener procedure translates them will therefore be in general only partially, but not linearly, ordered. This is unsatisfactory, for time is conceived as linear.[9]

There are various ways in which we can obtain a linear instant structure from the partial ordering that the Russell-Wiener construction gives us. In the first place, it is a familiar fact that each partial ordering $\langle T, < \rangle$ can be extended to a linear ordering $\langle T, <' \rangle$. This extension is of course not unique (except in the trivial case where $\langle T, < \rangle$ is already linear itself, and $\langle T, <' \rangle$ coincides with $\langle T, < \rangle$). No one of these extensions deserves, any more than its alternatives, to be regarded as giving *the correct* structure of time. Rather we should consider only those properties of temporal structure which all these extensions have in common.

Rather than first applying the Russell-Wiener method and then passing to extensions, we can also proceed in the reverse order: We first form an extension ε' of ε which satisfies A7, and then apply the Russell-Wiener transformation to ε'. Note in this connection that a model ε of A1–A6 can always be extended to a model ε' of A1–A7. For suppose $\tau' = \langle T, <' \rangle$ is a linear extension of the result of applying the Russell-Wiener transformation

[9] Relativity Theory has taught us that in one sense the relation of earlier and later is *not* connected: there may be two successive events e_1 and e_2 in one place neither of which is either earlier or later than a third event e_3 occurring at some distance from e_1 and e_2. But this circumstance arises only in connection with spatially distant events; and even in a case such as the one described connectivity fails only in the sense that e_1 nor e_2 is causally connected with e_3. If however an appropriate convention for measurement of time, viz. through synchronization of clocks at the respective place of the events, has been established in advance, the times of these events do become fully comparable, albeit in a way which is based on a partly arbitrary agreement. It would be wrong therefore to regard the theory of relativity as having proved the untenability of the traditional view that the order of time is a linear one.

to ε. Then τ' determines a model $\varepsilon' = \langle E, \prec', \bigcirc' \rangle$ of A1–A7 in the obvious way, viz. via the definitions: $e_1 \prec' e_2$ iff $\bar{e}_1$ lies entirely before $\bar{e}_2$; and $e_1 \bigcirc' e_2$ iff $\bar{e}_1$ and $\bar{e}_2$ overlap in ε'. We have, moreover, that $\tau' = \tau(\varepsilon)$. Thus each linear extension of $\tau(\varepsilon)$ can be obtained from a corresponding extension ε' of ε.

In fact the second method is more general than the first: there may be linear orderings of the form $\tau(\varepsilon')$, where ε' is an extension of ε, which cannot be obtained as extensions of the instant structure $\tau(\varepsilon)$ into which the Russell-Wiener construction transforms ε. The reason for this is that the instants of ε' need not be the same as the instants of ε. Suppose for example that i_1 and i_2 are incomparable instants of ε, i.e. for no $e_1 \in i_1$ and $e_2 \in i_2$ do we have either $e_1 \prec e_2$ or $e_2 \prec e_1$. Then there will be at least one extension ε' of ε in which for any $e_1 \in i_1$ and $e_2 \in i_2$ $e_1 \bigcirc e_2$. Thus the two distinct instants i_1 and i_2 of ε collapse into a single instant i of ε'.[10]

It should be intuitively clear that all the instant structures which can be obtained via completion of ε to a structure satisfying A7 represent linear time structures compatible with, and thus potentially contained in, the partial event structure ε itself. The second method is thus preferable to the first. It is questionable, however, whether the partial information about temporal order that can be derived either from ongoing discourse or from the (limited) accuracy with which our concepts mark the boundaries of the events they determine, is best represented in the form of structures $\langle E, \prec, \bigcirc \rangle$. For we may on occasion have information to the effect that e_1 is *not* completely before e_2: it may be that $e_1 \bigcirc e_2$, or it may be that $e_2 \prec e_1$ and yet be undetermined which of these disjuncts holds—certain being only that the third alternative is excluded. Similarly it may be definite that two events are not simultaneous, though not which preceded which. The best way to represent such information is to admit for each of the relations $\prec$ and $\bigcirc$, a *negative* as well as a *positive* extension. The positive extension of, say, $\prec$ will contain, as before, those pairs $\langle e_1, e_2 \rangle$ such that e_1 definitely precedes e_2; while its negative extension is to contain the pairs such that e_1 definitely does not precede e_2.

Talk of positive and negative extensions will remind some readers of certain fairly recent proposals for the model theoretic treatment of vague

[10] Of course i may absorb additional instants of ε as well—i.e. we may find that $i_1 \cup i_2 \subsetneq i$. In general the class of instant structures of the form $\tau(\varepsilon')$ corresponds not to the class of linear extensions of $\tau(\varepsilon)$, but to the class of those structures $\langle T', <' \rangle$ where the elements of T' are equivalence classes of elements of the structure $\tau(\varepsilon)$ under some equivalence relation which is also a congruence relation with respect to $<_\varepsilon$.

predicates.[11] Indeed it would in principle be possible to use the framework developed in these proposals to treat the relations $\propto$ and $\bigcirc$ as vague relations. It seemed desirable to me, however, to stick in the present study with a conventional metalanguage in which to state and discuss the conditions that should be imposed on event structures. I shall therefore represent the positive and negative extensions as extensions of distinct predicates. This means that we shall henceforth consider event structures of the form $\langle E, \propto^+, \propto^-, \bigcirc^+, \bigcirc^-\rangle$, where $\propto^+$ and $\propto^-$ represent the positive and negative extension, respectively, of the relation of complete precedence, and $\bigcirc^+$ and $\bigcirc^-$ the positive and negative extension of the relation of overlap.

In order that a structure of this form can be regarded as a representation of the temporal relations between the events in its domain it must, as before, satisfy certain conditions. Among these are, in particular, the condition that $\propto^+$ and $\propto^-$, and the condition that $\bigcirc^+$ and $\bigcirc^-$, are mutually exclusive. But there are others, which derive from the constraints that are imposed on the vague relations $\propto$ and $\bigcirc$ in virtue of their being subrelations of relations which jointly satisfy A1–A7. Which are these additional conditions?

If the structure $\langle E, \propto^+, \propto^-, \bigcirc^+, \bigcirc^-\rangle$ gives a complete specification of temporal relations, i.e. if $\propto^-$ is the complement of $\propto^+$ and $\bigcirc^-$ the complement of $\bigcirc^+$, then these conditions can be immediately obtained from those we postulated in section I: we simply replace in A1–A7 $\propto$ and $\bigcirc$ everywhere by $\propto^+$ and $\bigcirc^+$. But, of course, this is precisely the situation in which we are not particularly interested right now. In the general case, where $\propto^-$ and $\bigcirc^-$ are included in, but not identical with the complements of $\propto^+$ and $\bigcirc^+$, it is not so obvious what postulates the structure $\langle E, \propto^+, \propto^-, \bigcirc^+, \bigcirc^-\rangle$ must satisfy.

In fact this question is a special case of a much more general problem: Suppose that the analysis of a given number of concepts $c_1, \ldots, c_n$ leads to the conclusion that in principle these concepts jointly satisfy the set of postulates T—i.e. that ideally any formal representation of the domain(s) in which these concepts are operative should take the shape of a structure $\langle U, R_1, \ldots, R_n\rangle$ which is a model of T. Suppose also that in practice the concepts tend to be less than fully defined, so that a more correct formalization of what is actually determined about the extensions of the concepts must rather be in the form of a structure $\langle U, R_1^+, R_1^-, \ldots, R_n^+, R_n^-\rangle$, where R_i^+ and R_i^- represent the positive and negative extensions of c_i. This structure need not satisfy the postulates of T (or, more accurately, the results of replacing

[11] See, for instance Fine (1975), Kamp (1975).

in them the occurrences of R_i by R_1^+); but it must be capable of extension to a structure that satisfies these postulates. It is natural to ask:

(i) Is there a set of postulates T' such that $\langle U, R_1^+, R_1^-, \ldots, R_n^+, R_n^- \rangle$ satisfies T' if it can be so extended?
(ii) Is there a general method to derive what its postulates are from the original T?

The first of these questions does not always have a positive answer. It does have a positive answer however in a restricted class of cases which includes those in which we are specifically interested in the context of this paper. Moreover it is possible to give a quite simple characterization of T' in terms of T. This characterization however is not as effective as one might wish, as it does not provide us with an effective method for determining from an explicitly given finite set of postulates T an explicit list of the postulates of T'.

To discuss these matters properly, it is indispensible to introduce a certain amount of terminology. Some of this will already be familiar from the theory of models.[12]

Definition 1 Let $M = \langle U, R_1, \ldots, R_n \rangle$ and $M' = \langle U', R_1', \ldots, R_n' \rangle$ be relational structures of the same similarity type.

(i) We say that M is a *relational substructure* of M' if
  (a) $U \subseteq U'$; and
  (b) for $j = 1, \ldots, n, R_j \subseteq R_j'$.
(ii) M is a *strictly relational substructure* of M' if M is a relational substructure of M' and moreover $U = U'$.
(iii) For any first order theory T we shall understand by $RS(T)$ the first order theory of the class of all relational substructures of models of T; and by $SRS(T)$ the first order theory of the class of all strictly relational substructures of models of T.

Definition 2

(i) Let φ be a formula of first order logic which is in prenex disjunctive normal form. Then φ is called a *negative prenex formula* if each atomic subformula of φ which is not of the form $\sigma = \tau$ is preceded by a negation sign.
(ii) A formula is called *negative* if it is logically equivalent to a negative prenex formula, and *universal negative* iff it is equivalent to a negative prenex formula whose prenex consists of universal quantifiers only.
(iii) Let T be a first order theory. By $Neg(T)$ we understand the set of all consequences of the negative theorems of T; and by $Unneg(T)$ the set of all consequences of the universal negative theorems of T.

[12] See e.g. Chang & Keisler (1973).

Proposition

(i) M is a relational substructure of M' if M is a strictly relational substructure of some substructure of M'.
(ii) $Unneg(T) \subseteq RS(T)$.
(iii) $Neg(T) \subseteq SRS(T)$.

The proofs of these are trivial.

Theorem 2 Let T be any first order theory. Then

(i) $Unneg(T) = RS(T)$;
(ii) every model of $Unneg(T)$ is a relational substructure of a model of T;
(iii) $Neg(T) = SRS(T)$;
(iv) suppose T is a purely universal theory (i.e. the set of its theorems are derivable from some subset of purely universal theorems). Then every model of $Neg(T)$ is a strictly relational substructure of some model of T.

Note: Jonathan Stavi has recently pointed out to me that (iii), which I had only verified for the special case where T is purely universal, holds for arbitrary first order theories. The proof for the general case parallels that of Lyndon's theorem that the positive sentences are precisely the sentences preserved under homomorphisms. I shall here present only the much simpler proof for the case where T is assumed to be a universal theory.

Proof:

(i) That $Unneg(T) \subseteq RS(T)$ follows from Prop. (ii). To show that $RS(T) \subseteq Unneg(T)$ suppose $\varphi \in RS(T) \setminus Unneg(T)$. Let M be a model of $Unneg(T) \cup \{\neg\varphi\}$. Then M is not a relational submodel of a model of T. One easily verifies that this amounts to the statement that $D^+(M) \cup T$ is inconsistent, where $D^+(M)$, or the *positive diagram of* M, is the set of all formulae which have one of the following forms:
   (a) $P(\bar{u}_1, \ldots, \bar{u}_n)$, where P is an n-place predicate of the language of T, $\bar{u}_1, \ldots, \bar{u}_n$ are individual constants denoting the elements $u_1, \ldots, u_n$ of M and $\langle u_1, \ldots, u_n\rangle$ belongs to the extension of P in M;
   (b) $\bar{u}_1 = \bar{u}_2$ where $\bar{u}_1$ and $\bar{u}_2$ are names of the same element of M;
   (c) $\bar{u}_1 \neq \bar{u}_2$ where $\bar{u}_1$ and $\bar{u}_2$ name different elements of M.
Since $D^+(M) \cup T$ is inconsistent there are formulae $\alpha_1, \ldots, \alpha_k$ in $D^+(M)$ such that $T \vdash \neg(\alpha_1 \wedge \ldots \wedge \alpha_k)(\bar{u}_1, \ldots, \bar{u}_k)$, where $u_1, \ldots, u_k$ is an enumeration of all the constants which occur as arguments in any of the α_j. Since the $\bar{u}_i$ do not occur in T
   (1) $T \vdash (\forall x_1, \ldots, x_k)(\neg\alpha_1 \vee \ldots \vee \neg\alpha_n)(x_1, \ldots, x_k)$
where the x_i are variables not occurring in $\alpha_1, \ldots, \alpha_k$, $(\neg\alpha_1 \vee \ldots \vee \neg\alpha_n)(x_1, \ldots, x_k)$ is the result of replacing the $\bar{u}_i$ everywhere in $(\neg\alpha_1 \vee \ldots \vee \neg\alpha_n)$ by the x_i. The formula displayed in (1) is a universal negative sentence of T. So $Unneg(T) \vdash \neg(\alpha_1 \wedge \ldots \wedge \alpha_n)(\bar{u}_1, \ldots, \bar{u}_k)$. But this means that M is not a model of $Unneg(T)$, contrary to assumption. Therefore $RS(T) \subseteq Unneg(T)$.

(ii) A similar argument shows that if $M \models Unneg(T)$ then M is a relational substructure of a model of T. For if not then $D^+(M) \cup T$ would be inconsistent and the argument proceeds as under (i).

(iii) As I annouced in the Note I shall prove this part only for the special case where T is purely universal. $Neg(T) \subseteq SRS(T)$ holds in virtue of Prop. (iii). To prove the converse inclusion observe that if T is purely universal then

(2) M is a strictly relational substructure of a model of T if M is a strictly relational substructure of a substructure of a model of T (since T is preserved under taking substructures) if M is a relational substructure of a model of T (in view of Prop. (i)).

Suppose that $\varphi \in SRS(T) \setminus Neg(T)$. Then $Neg(T) \cup \{\neg\varphi\}$ has a model M which in virtue of (2) is not a relational substructure of a model of T. As under (i) we infer that for some atomic $\alpha_1, \ldots, \alpha_n$ in $D^+(M) Neg(T) \vdash (\forall x_1 \ldots x_k)(\neg\alpha_1 \vee \ldots \vee \neg\alpha_n)(x_1, \ldots, x_k)$. So M is not a model of $Neg(T)$, which once again contradicts our assumption.

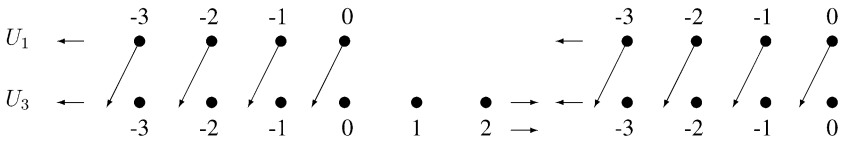
(iv) The argument is once more of the same form: Suppose M were a model of $Neg(T)$ but not a strictly relational substructure of a model of T. Then M would in view of (2) not be a relational substructure of a model of T. So $D^+(M) \cup T$ is inconsistent etc.

If T is not purely universal then it need not be the case that every model of $SRS(T)$ is a strictly relational substructure of a model of T. For example let T_1 be the theory of strict linear orderings which satisfy the additional condition that each element which has a successor has an immediate successor:

$$(\forall x)((\exists u)(x < u) \rightarrow (\exists u)(x < u \wedge \neg(\exists y)(x < y \wedge y < u)))$$

Consider two structures: $M_1 = \langle U_1, R_1 \rangle$ and $M_2 = \langle U_2, R_2 \rangle$, defined as follows:

Let $\langle U_1, S_1 \rangle$ be a linear ordering of order type $\omega^* + \omega^*$ and let $\langle U_3, S_3 \rangle$ be a linear ordering of order type $\omega^* + \omega + \omega^*$, and let $U_1 \cap U_3 = \emptyset$. R_1 is to be the successor relation of the ordering $\langle U_1, S_1 \rangle$, $U_2 = U_1 \cup U_3$ and R_2 is to be the union of the two successor relations on $\langle U_1, S_1 \rangle$ and $\langle U_3, S_3 \rangle$ respectively. Then M_2 is a strictly relational substructure of a model of T_1. For we can completely order its universe as indicated in the following picture:



(Thus we mesh the two parts of type ω^* of U_1 with the corresponding parts of U_3.)

M_1 on the other hand cannot be extended to a model of T_1 without adding new elements to its domain. For the only way in which we can extend R_1 to a transitive relation is to make it into the relation S_1 and then the last point of the first ω^* part of U_1 will be without an immediate successor.

It is easy to show that M_1 and M_2 are elementarily equivalent (see Appendix I). So, since $M_2 \models SRS(T_1)$ it follows that $M_1 \models SRS(T_1)$. So $SRS(T_1)$ has models which are not strictly relational substructures of any models of T_1. A similar counterexample can be constructed for the theory of dense linear orderings.

As both these counterexamples involve theories which are only slightly more complicated than purely universal theories—each is axiomatizable by a set of postulates all but one of which are universal while the remaining axiom is either of prenex form $\forall\exists$ or of prenex form $\forall\exists\forall$—one would suspect the failure of (iv) for non-universal theories to be a quite general phenomenon. However, as I already adumbrated, this circumstance does not affect the applications with which we are concerned in this paper, as these only deal with the theory T_{ev} whose axioms are:

(a) the axioms A1$^+$–A7$^+$, obtained from A1–A7 by replacing everywhere $\propto$ by $\propto^+$ and
(b) the sentences
(A8) $(\forall x,y)(x \propto^+ y \rightarrow \neg x <^- y)$ and
(A9) $(\forall x,y)(x \bigcirc^+ y \rightarrow \neg x \bigcirc^- y)$.

Even where T is universal there remains the question whether from a given axiomatization of T we can derive, in a more or less algorithmic manner, an explicit axiomatization for $Neg(T)$. It appears that the most satisfactory general solution to this problem is contained in the work of Barwise (1976). But this is something I cannot go into here. That the problem is not trivial is indicated by the fact that even where T is finitely axiomatizable this is in general not so for $Neg(T)$. An extremely simple example of such a theory T is the theory of strict partial orderings (i.e. the theory whose only axioms are $(\forall x,y)(x < y \rightarrow \neg y < x)$ and $(\forall x,y,z)(x < y \wedge y < z \rightarrow x < z)$). Another example is provided by T_{ev}. Axiomatizations of $Neg(T_2)$ and of $Neg(T_{ev})$ can be found in the Appendix II.

The reader may wonder how relevant this general model theoretic discussion is to the particular problem in the theory of events which led us to it. When considering event structures of the form $\langle E, \propto, \bigcirc\rangle$ I argued that the intuitively acceptable representations of event sets with partially defined temporal relations were the models of A1–A6. Here the subtheory that characterizes the class of admissible relational substructures is just as finitely axiomatizable as the original theory itself. And as it turns out,

something similar can be maintain in connection with the theory T_{ev}. For it appears plausible on intuitive grounds that the partial event structures $\langle E, \propto^+, \propto^-, \bigcirc^+, \bigcirc^- \rangle$ which represent incomplete information about the temporal relations between the events in their domains must nevertheless satisfy each of the following postulates:

A10 $(\forall x, y)(x \propto^+ y \rightarrow y \propto^- x)$
A2$^+$ $(\forall x, y, z)(x \propto^+ y \wedge y \propto^+ z \rightarrow x \propto^+ x)$
A3$^+$ $(\forall x((x \bigcirc^+ x)$
A4$^+$ $(\forall x, y)(x \bigcirc^+ y \rightarrow y \bigcirc^+ x)$
A7$^+$ $(\forall x, y, z, u)(x <^+ y \wedge y \bigcirc^+ z \wedge z <^+ u \rightarrow x <^+ u)$
A11 $(\forall x, y, z, u)(x <^+ y \wedge z <^- z \wedge z <^+ u \rightarrow x <^+ u)$
A12 $(\forall x_1, x_2, y_1, y_2, z_1, z_2)(x_1 \propto^+ y_1 \wedge y_1 \bigcirc^- y_2 \wedge y_2 \propto^+ z_1 \wedge$
$x_2 \propto^+ y_2 \wedge y_1 \propto^+ z_2 \rightarrow x_1 \propto^+ z_1 \vee x_2 \propto^+ z_2)$

It is easily checked that these formulae together with the principles

A8' $(\forall x, y)(x \propto^+ y \rightarrow \neg x \propto^- y)$ and
A9' $(\forall x, y)(x \bigcirc^+ y \rightarrow \neg x \bigcirc^- y)$

entail all axioms of $Neg(T_{ev})$ as given in Appendix II.

Thus, while the theory of relational and strictly relational substructures may have some interest in its own right, and while it may yet be found to have other, more subtle, applications to conceptual analysis, some of the complications which we have encountered during the preceding discussion do not directly matter to what is primarily at stake here.

One further observation before I conclude this section. In section II and III I created the impression that only strictly relational substructures of models of T_{ev} (the theory axiomatized by A1–A7) are of importance for the intended applications. As regards the application to discourse representation this is essentially correct. But in connection with questions concerning the structure of time as determined by the ensemble of events which a given conceptual scheme is capable of isolating from the flow of nature this is not strictly right. I already argued for the possibility that the time at which a certain change, say, from orange to red, takes place, cannot be exactly pinpointed because of the inherent vagueness in the concept used to define the change. For exactly the same reason that the time can in general not be perfectly determined, however, we should also allow for the possibility that it may be undetermined whether a change actually took place at all.[13] For

[13] My attention was drawn to this possibility by David Lewis.

suppose a certain object a, which is unambiguously orange at first, changes colour in the direction of red; but before it has reached a shade which is unambiguously red it reverts back to the original, indubitable, orange. The question whether here we have a change from orange to red followed by a change from red to orange, or rather no change at all would depend on where the borderline between red and orange is drawn. If it is drawn near the central instances of orange then according to that semantic decision there will have been two changes. But if the line is drawn near the central cases of red then, according to that stipulation, a would never, in the period under consideration, have ceased to be orange at all and thus no such change would have taken place. Thus while the partial event structure ε which incorporates only such information as is already definite in virtue of the semantic conventions as they stand, will not contain changes from orange to red and back, some of the event structures representing the information conveyed by particular ways of eliminating the truth value gaps of the underlying concepts will contain such changes.

It is a further question how the presence of such additional events in some of the completions of the partial structure might affect the structure of time as it is (partially) determined by ε. This is as it stands not a very precise question; but one way, consistent with what I have said before about this issue, of making it precise is the following. Let $\varphi(\varepsilon)$ be the class of all completions of ε.[c] Each such completion ε' determines a substructure $R(\varepsilon') = \langle E, \prec^{+''}, \prec^{-''}, \bigcirc^{+''}, \bigcirc^{-''} \rangle$ where the relations $\prec^{+''}, \prec^{-''}, \bigcirc^{+''}, \bigcirc^{-''}$ are the restrictions to E of the corresponding relations of ε'. (Thus, in the special case where the domain of ε' equals that of ε, $R(\varepsilon') = \varepsilon'$.) Let $\varphi'(\varepsilon)$ be the set of all these structures $R(\varepsilon')$ for ε' in $\varphi(\varepsilon)$. Furthermore let $\varkappa(\varepsilon)$ be the class of all instant structures $\tau(\varepsilon')$ where $\varepsilon' \in \varphi(\varepsilon)$ and let $\varkappa'(\varepsilon)$ be the class of all instant structures $\tau'(\varepsilon')$ for $\varepsilon' \in \varphi'(\varepsilon)$. We may then ask if the set of structural properties shared by all members of $\varkappa'(\varepsilon')$ does or does not coincide with the set of all properties common to the members of $\varkappa'(\varepsilon)$. This question is still not fully precise, for I haven't said what is to count as a "structural property". Here various options would seem reasonable. For instance, we could identify the structural properties with those properties which are first order definable in terms of $<$; or with the corresponding

[c] A potential source of unclarity in this discussion is the use of the term 'completion'. Intended is the following: $\varepsilon' = \langle E', \prec'^{+}, \prec'^{-}, O'^{+}, O'^{-} \rangle$ is a *completion* of $\varepsilon = \langle E, \prec^{+}, \prec^{-}, O^{+}, O^{-} \rangle$ iff (1) ε is a relational substructure of ε' and (2) $\prec'^{-}$ is the set-theoretic complement of $\prec'^{+}$ ($\prec'^{-} = E'^{2} \setminus \prec'^{+}$) and O^{-} is the set-theoretic complement of O^{+}.

second order properties; or with those definable by formulae of monadic second order logic, etc.

To answer the question arising from the adoption of any of these options would involve us in conceptual as well, probably, as mathematical complications whose subtlety would exceed what I am able to handle at the moment. But I should at least make clear why I have raised the issue at all.

The description I gave of the case where the object a, plainly orange at first, changes its hue in the direction of red and then reverts to its original state suggests that although it is indeterminate whether there have been changes e_1 from orange to red and e_2 from red to orange it is on the other hand definite that some other changes in colour took place at around that time. If those other changes are definable within the conceptual framework which provides the events of the partial structure ε they will already belong to ε itself. In view of this it is natural to ask whether the presence in a completion ε' of ε of the events e_1 and e_2, in addition to these other colour changes (which must belong to ε' since they already belong to ε) could have any impact on the structural features of $\tau(\varepsilon')$. Suppose that this were not so; and moreover, that it were not so not only in the particular case considered but in all other cases where the completion of ε contains events that do not occur in ε. In that case we would have to conclude that although some of the completions of ε have larger domains than ε, nevertheless the structure of time as determined by ε must be assessed via a class of structures of which ε is a strictly relational substructure. The completions ε' of which ε is a relational but not a strictly relational substructure would only be indirectly relevant, viz. as generators of the structures $R(\varepsilon')$ which enter into the assessment directly and of which ε is a strictly relational substructure. I hope to have made it sufficiently clear however that much careful analysis will be necessary of these issues before mere speculation can be replaced by informed opinion.

I should like to add one last remark. As the preceding discussions should already have revealed implicitly one of the distinctive features of the concepts of time and temporal order which underlies the formal developments of this paper is the thesis that the structure of time depends on the conceptual framework that is employed in the determination of the events which, according to the present theory, form the basis of that structure. This seems to introduce into the notion of temporal structure an important subjective element. It is an interesting, but extremely difficult, question how much room the present conception nonetheless leaves for an objective component to the notion as well. Essentially this comes down to the question what conceptual schemes are possible: can we carve up the "natural flow

of things" in just any way we like—and thus extract from it any temporal structure you might wish—or are the conceptual frames accessible to us restricted by an intrinsic order of nature, or by an intrinsic order in the interactions between nature and mind? On problems of this magnitude I would not dare to speculate. Suffice it to note their relevance for the philosophical implications of the view I have here advanced.

V

All natural languages contain devices of temporal reference. To describe these devices formally we need to represent, in one way or another, the structure of time. In the course of the fairly short history of the model theory of temporal reference the representation has been without exception, I believe, 'atomistic': time was modelled as a linearly ordered structure of instants. This was true, in particular, of the first formal work in the area, Cocchiarella's model theoretic analysis (Cocchiarella 1965) of Prior's P,F-calculus (or, to be more explicit, of a formal language which combines the familiar apparatus of first order predicate logic—predication, quantification and boolean connection—with Prior's operators P ("it was the case that") and F ("it will be the case that")). Cocchiarella's models take the form of triples (τ, U, F), where τ is a linear order $(T, <)$, U is a non-empty set (of possible individuals) and F is a function which assigns to each non-logical constant γ of the language an appropriate extension at each instant t of T.[14] The truth definition then associates also with each complex expression α an appropriate extension at each t of T; and a sentence, in particular, is in this way assigned at each t a truthvalue. Systems of tense logic such as those developed by Prior are structurally quite simple. This can be an advantage where we are concerned to clarify, by way of formalization, certain conceptual questions whose true nature the complexities of the ordinary vernacular tend to conceal. But when our concern is rather the precise analysis of the temporal devices which natural languages actually employ, this simplicity can be, and in fact has been, very misleading. The semantic description of many of these devices appears to require a number of modifications of the framework Cocchiarella introduced.

One of these, which has received much attention in recent work on tense and aspect, relates to the temporal elements with respect to which

[14] By 'appropriate' I mean here: a member of U of γ is an individual constant, a subset of U if γ is a 1-place predicate, etc.

such semantic notions as truth and satisfaction are (recursively) defined. For Cocchiarella these elements are simply instants; but in subsequent work it has been claimed that such semantic concepts must be analyzed as relations between expressions and intervals rather than relations between expressions and moments of time.[15]

The switch from instants to intervals has been motivated by various considerations. These considerations, which it would carry us too far to discuss individually with the care they deserve, fall, it seems to me, into two classes. Considerations of the first type concern the *recursive* part of the definition of truth, satisfaction (and, possibly, other semantic concepts). In the first place it has been argued that in order to arrive at a correct recursive characterization of the conditions under which assertoric utterances of certain complex sentences are true at the times at which they are made we must allow the recursion to pass through intermediate stages at which subexpressions of those sentences are evaluated with respect to intervals rather than to instants. Consider for instance the sentence:

(1) Fritz always writes an article in less than a month.

Suppose we analyze (1) as the syntactic result of combining the subject 'Fritz' with the complex predicate

(2) $(\lambda x)(x$ always writes an article in less than a month$)$

To arrive at the right truth conditions for (1) we must associate with (2) the right function from individuals to sets of instants (those instants at which the individual in question satisfies the predicate). Suppose further that this function is to be obtained from (functions associated with) underlying expressions

(3) $(\lambda x)(x$ writes an article$)$
(4) always
(5) in less than a month

The initial motivation for this requirement would be that (1) conveys that all occasions where Fritz writes an article are occasions of a certain kind viz. occasions on which the writing takes less than a month. What function f_3 ought to be associated with (3) so that from it we can, given appropriate interpretations for (4) and (5), correctly determine the function f_2 that must be associated with (2)?

---

[15] As far as I know the first explicit proposal of this sort is Bennett & Partee (1978). Since the first (informal) circulation of that paper there have been many other similar proposals, but I have decided to refrain from attempting to give a representative list of these.

Suppose f_3 gives us for each individual a set of instants at which it is true that a is (in the process of) writing an article. To derive f_2 from such an f_3 we would presumably have to make use of a clause to the effect that an instant t belongs to $f_2(a)$ if each continuous interval of instants all belonging to $f_3(a)$ has a duration of less than a month.[16] But this would yield the correct result only if a always writes only one article at a time. If on the contrary a starts one article on, say, the 1st of March and a second on the 15th of March, finishes the first on the 21st of that month and the second on the 7th of April, then the proposed clause would imply that he did not satisfy (2). Thus it is wrong. Nor does there appear to be any other clause that determines f_2 correctly from such a function f_3. A function f_3 for (3) from which f_2 can be reconstructed would, it seems, have to relate with each individual a, a set of *intervals*, the set which contains for each of the articles written by a the duration of the writing of that article. It should be intuitively clear how from a function of this sort we could define f_2.

Suppose next that we take the function f_3 to be determined by the semantic entities associated with, respectively, the 2-place predicate

(6) $(\lambda x)(\lambda y)(x \text{ writes } y)$

and with the quantifier phrase

(7) an article

It is natural to assume that this determination takes the form usually associated with the existential quantifier, viz. that $f_3(a)$ is the union of $f_6(a, b)$ where b ranges over all articles. In order that this operation should give the right value for $f_3(a)$ it is necessary that f_6 also associate with each suitable pair (a, b), again, a set of intervals. Note however that even if we insist that f_6 be of this form it might nonetheless seem possible for that function to be fully determined by an underlying function f_6' which maps pairs (a, b) to sets of instants—$f_6'(a, b)$ being the set of instants at which it is true that a is in the process of writing b. The difference between f_6 and f_3 lies therein that in the present case it might seem plausible that $f_6(a, b)$ consists of precisely those intervals which are maximal uninterrupted stretches of instants belonging to $f_6'(a, b)$—we saw that a similar assumption was unwarranted in connection with f_3.

---

[16] A general truthdefinition should of course contain, rather, a clause with two parameters of which the clause referred to in the text would result as a special instance if we were to substitute for these parameters the particular interpretations which the theory associates with, respectively, (4) and (5).

It seems therefore that the function f_6 is reducible to a function which assigns instants; and thus the example suggests the following picture: at the basis of the recursion we have certain relations between individuals and instants; but at some point in the recursion we must switch from such relations to relations between individuals and intervals. Eventually however there is a return to notions involving instants, such as in particular that of the truth of a complete sentence at the instant it is used.

Of course the discussion of the example is very far from a conclusive argument that this is indeed the right picture. In the first place I have done nothing to show that (1) must be analyzed into components in precisely the way I proposed. To argue for this persuasively is possible only via careful semantic and syntactic considerations of all sentences of a substantial language fragment to which (1) belongs, and this would go not only beyond the scope of this paper but also beyond my present capacities.

As a matter of fact I do not really believe that the difficulties to which the discussion drew attention could really be circumvented via an alternative analysis of (1) into basic components. There was however another point in the discussion that is genuinely dubious, and which deserves more careful attention than I have given it so far. This is the point where I remarked—with appropriate diffidence—that "it might seem plausible" that f_6 is determined by an underlying function f_6' which assigns sets of instants to pairs (a, b).

Is this really plausible? An interval which counts as the period during which a wrote a particular article b will as a rule not consist exclusively of instants at which a is actually writing, or even engaged in directly related activities such as reading up on the literature of the subject, looking up references in the library etc. Even the most devoted scholar will have to take out time for sleep and food unless his project—that of writing b—can be accomplished within less than, say, 24 hours.

What can be said about an instant which belongs to the period during which the article is written but at which the author actually was asleep? Earlier, when considering the reducibility of f_6 to f_6' I tried to dodge this issue by describing the relevant instants as instants at which the author is "in the process of writing" the article. But how is one to tell that an instant at which he is actually asleep is also an instant at which he is in the process of writing? Only, surely, by evaluating the position of that instant within the stretch of time which contains also many instants when he is engaged in what we might coin "writing his article in the strict sense" (i.e. writing, checking related papers and such like). This strongly suggests that in some cases the judgement whether the author is in the process of writing at a particular

instant reduces to a conceptually prior judgement regarding whether he is writing *during* some interval to which the instant belongs.

From all I have said it would appear that this latter judgement depends in turn on judgements about the instants at which the author is writing in the strict sense. It is the distribution of these instants throughout the interval i which determines whether i counts as a period during which the author is in the process of writing. So could we not construct a semantic theory in which the truth conditions for (1) are spelled out along the lines sketched above, but in which f_6 is determined by an underlying function f_6' that assigns to a pair (a, b) the set of instants at which a is writing b in the strict sense?

This is a difficult question to which I do not know the answer. It is difficult for two, related reasons. First it may be very hard to specify with any precision what it is for a to be writing in the strict sense. Which activities are to be taken as directly part of the actual writing? If looking up a reference in the library counts as such, should being on the way to the library in order to check a reference also count among these activities? And if that counts, what about having a nap so as to be able to carry on more efficiently with the writing afterwards?

Secondly, even if we assume that "writing in the strict sense" could be defined with sufficient precision, which distributions within i of instants at which a is writing in this sense make i an interval during which a is writing b? If you stop to think for a moment about a *general* clause which would correctly define this class of intervals in terms of the corresponding distributions of instants, you will soon realize how very complicated such a clause would have to be. In fact it might well have to be so complex that we would not wish to incorporate it into a semantic theory whose purpose is to elucidate semantic relations—a purpose hampered by the surfeit of detail which the clause would introduce into the theory; in which case it would be better to treat f_6, and similar functions that involve intervals rather than instants, as primitives.

But apart from these methodological considerations there may be an even more conclusive reason for treating f_6 as primitive: It is conceivable that f_6 is not uniformly reducible to notions involving instants in any way whatever. For even if it is true that the judgement that a was writing b during i is based on what went on at the various instants that constitute i, the integration of information concerning the various instants which yields this judgement might occur at a level of mental processing which—unlike the semantic relations between grammatically complex expressions and their parts, to which formal methods of analysis have thus far been applied with some measure of success—defy description with the help of the discrete

combinatorial techniques which have been the essence of all formal semantics to date.

It may seem that the example we have here considered—the relation "*a* writes *b*"—is a rather special one. Nevertheless the question to which it gave rise, viz. whether judgements about intervals are reducible to judgements about instants, is a perfectly general one, which can, and must be raised in connection with any predicate. Note however that what lends these questions their appearance of urgency is the particular temporal framework within which the theory of temporal reference has been operating: If we choose to represent time as an instant structure the choice between the use of instants and the use of intervals may look like the choice between greater theoretical simplicity on the one and greater empirical adequacy on the other hand. But from the perspective that we have developed in the preceding sections this dilemma simply disappears. For what, according to the advocates of the use of intervals in the better theory, and possibly the only theory that can do justice to the facts of natural discourse, now also appears as the theoretically simpler description, in that it evaluates expressions relative to what according to this new perspective are the basic constituents of time.

VI

Let us briefly consider what such a semantic theory might be like. For the sake of simplcity we shall restrict our attention to languages L whose primitive expressions are singular terms and predicates and which combine formulae into larger ones with the help of certain logical operators, about which more below. As before I shall assume that time is given in the form of a partial event structure $\varepsilon = (E, \prec^+, \prec^-, \bigcirc^+, \bigcirc^-)$. Furthermore we shall assume that an interpretation for L relative to ε takes the form of a function A which associates with each n-place predicate of L a function that assigns to any tuple $(a_1, \ldots, a_n)$ of appropriate arguments a subset of the event set E. Intuitively $A(Q)(a)$, where Q is a 1-place predicate and a an individual, is to consist of those events e during which a has Q.[d]

Note that the absence of e from $A(Q)(a)$ should not be interpreted as conveying that throughout ea fails to have Q. For instance e may be the event of a's washing b; it temporally includes the event e' of a's rinsing b; thus e'

[d] The phrasing in the text—'events e during which a has Q'—is inaccurate and potentially misleading. What should have been said is: 'events e consisting in a satisfying Q'.

will belong to $A(\text{'rinse'})(a, b)$, while e will not. It is clear in this case that the condition 'a rinses b' does not fail throughout e. Thus the satisfaction relation relative to events is not bivalent; and the distinction between events during which an atomic sentence fails and events during which it partly fails and partly holds, which evidently is essential to the truth conditions of the negations of simple sentences, must somehow be made explicit by the interpretation. We shall do this by assuming that A assigns to Q and $a_1, \ldots, a_n$ not a single set $A(Q)(a_1, \ldots, a_n)$ but a pair of sets $A^+(Q)(a_1, \ldots, a_n)$ and $A^-(Q)(a_1, \ldots, a_n)$ consisting, respectively, of the events during which $a_1, \ldots, a_n$ stand in the relation Q and the events during which $a_1, \ldots, a_n$ do not stand in that relation.

Suppose ε' is a model of T_{ev} which is a strictly relational extension of ε, and let t be an instant of $\tau(\varepsilon')$. Under what conditions would it be appropriate to say that a has Q at t? Our first guess might be that

(1) a has Q at t if there is an $e \in E$ such that $e \in A^+(Q)(a)$ and $e \in t$.

However, our earlier discussion of the phrase 'a writes b' has already shown how problematic this clause is: If we allow for the possibility that e belongs to the set assigned by A to Q and (a, b) in virtue of e's being a 'period when a was writing b' then being an instant at which e occurs is not a sufficient condition for it being correct to say that at that instant a is writing b. (1) could therefore only be correct if we excluded from $A(Q)(a, b)$ all those events which contain parts during which strictly speaking a does not stand in the relation Q to b. But to do so would be to eliminate precisely that aspect of the new theory that renders it more promising than theories which build their semantic recursions on the basis of assignments that involve instants. It thus appears that (1) cannot be right. What we need instead is the following clause

(2) a has Q at t if for each $e \in t$ there is an $e' \in E$ such that $e' \subseteq e$ and $e' \in A^+(Q)(a)$
(Here "$e' \subseteq e$" is short for: $(\forall e'')(e'' \bigcirc^+ e' \rightarrow e'' \bigcirc^+ e)$. This last formula is easily seen to be equivalent to $(\forall t' \in \tau(\varepsilon'))(e' \in t' \rightarrow e \in t')$.)

According to (2) a has Q at t if no matter how short an event e occurring at t we consider there is an event e' temporally included in e during which a has Q. Thus (2) deals with the difficulty which we observed in connection with "a writes b"; and I cannot see any other reason why it would be inadequate. Let us adopt it.

With the help of (2) we can convert the interpretation function A into a function A' which tells us for each n-place predicate Q, individuals $a_1, \ldots, a_n$ and instant t whether $a_1, \ldots, a_n$ satisfy Q at t. In this way we can pass from

the pair (ε, A) to a structure $(\tau(\varepsilon'), A')$ which much resembles the models for tense logic which were first introduced by Cocchiarella.

There is however an important difference which the existing similarities threaten to conceal. Cocchiarella's assignment functions are total: for the sentence $\neg Q(a)$ to be true at t in one of his models $M = (\tau, A)$ it is (necessary and) sufficient that t does not belong to the set of instants which the interpretation function A associates with Q and a. But it is far from obvious that we can deal with negation in the same simple fashion when we work with the models $(\tau(\varepsilon'), A')$ which are the products of our interpretations relative to event structures. We already noticed that the truth relation between statements and events is not bivalent: besides the possibilities: (i) s is true during e, and (ii) s is false during e, there is a third, viz. that s is true during some parts of e and false during others. By itself this does not necessarily produce failure of bivalence in the derived instant model. For nothing we have said so far excludes the possibility that for each s and each t there either is an event surrounding t during which s is true or else an event surrounding t during which s is false. There is however another reason why we should not expect, and indeed not *want*, bivalence even at the level of instants. In section IV I argued that among the events that are crucial to the determination of temporal structure there are in particular those which I termed changes. Suppose that e is an event of this kind, and that e is identified as an event of a's turning from not-Q into Q. Let us assume moreover that this change is ideally instantaneous, so that in $\tau(\varepsilon')$ there is an unique $t(e)$ to which e belongs. Does a satisfy Q at $t(e)$? There are, it seens to me, in general no good grounds either for confirmig this or for claiming that a does satisfy not-Q at $t(e)$. $t(e)$ is a time at which a is in transition in regard to Q, and this is a situation which is distinct both from that where a is in the state, or engaged in the activity, denoted by Q and from that where a is subject to the condition of not being in the state (or not being engaged in the activity).

Indeed we might find it tempting to represent this third kind of situation by a third 'truthvalue', distinct from 1, which the statement that a has Q should receive in a situation where a is in the state or engaged in the activity, and 0, which the statement would have in situations where a is subject to the opposite condition. If we decide to do this then we should introduce of course also a fourth value to represent those situations in which a is turning from Q into not-Q.

This would give rise to a type of four valued logic. But whether a viable logic emerges will depend, as always, on the possibility of stating the values of complex expressions in terms of the values of their components. And, as usual, the possibility is severely limited. Indeed the languages for which

this analysis seems at all workable are all very impoverished. An example would be the language L whose atomic formulae are of the form $Q(\alpha_1, \ldots, \alpha_n)$ where Q is an n-place predicate and $\alpha_1, \ldots, \alpha_n$ are singular terms; and in which complex formulae are formed only with the help of the connectives 'and', 'or' and 'not'. Moreover we must assume that any change from satisfaction of a predicatc Q by individuals $a_1, \ldots, a_n$ to non-satisfaction of Q by these individuals—and similarly any change from non-satisfaction to satisfaction of Q—is ideally instantaneous; and finally—but this hardly seems a restriction—that a change from not-p to p, where p stands for a condition expressed by any atomic sentence of L, is flanked by an event during which not-p holds on the left and by another event during which p holds on the right. Thus if p stands, say, for the condition expressed by the atomic sentence $Q(\alpha)$ and e is a change from not-p to p then e should satisfy in the structure (ε,A) the condition:

(3) $(\exists e_1)(e_1 \prec^+ e \,\&\, e_1 \in A^-(Q)(a) \,\&\, (\forall e')(e' \prec^+ e \rightarrow e_1 \prec^- e'))\&$
$(\exists e_2)(e \prec^+ e_2 \,\&\, e_2 \in A^+(Q)(a) \,\&\, (\forall e')(e \prec^+ e' \rightarrow e' \prec^- e_2))$.

Similarly a change from p to not-p should satisfy:

(4) $(\exists e_1)(e_1 \prec^+ e \,\&\, e_1 \in A^+(Q)(a) \,\&\, (\forall e')(e' \prec^+ e \rightarrow e_1 \prec^- e'))\&$
$(\exists e_2)(e \prec^+ e_2 \,\&\, e_2 \in A^-(Q)(a) \,\&\, (\forall e')(e \prec^+ e' \rightarrow e' \prec^- e_2))$.

Under these conditions the truth values of complex sentences are, it seems, correctly determined by the following four-valued truthtable:

$\neg$

p	$\neg p$
0	1
1	0
2	3
3	2

$\&$

p \ q	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	2	0
3	0	3	0	3

$\vee$

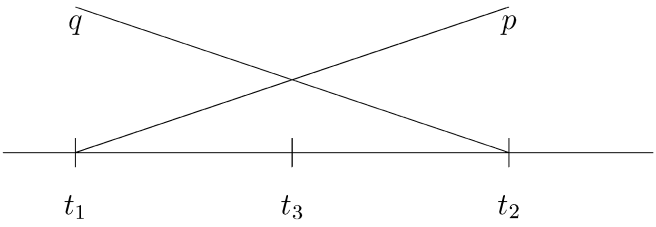
p \ q	0	1	2	3
0	0	1	2	3
1	1	1	1	1
2	2	1	2	1
3	3	1	1	3

The joker hides in the table for 'and' and 'or'. His hiding places are the entries for $(p = 2, q = 3)$ and $(p = 3, q = 2)$. That the value of $(p \;\&\; q)$ must be o in these cases can be argued as follows. By assumption the event e at which p has, say, the value 2 and q the value 3 is instantaneous. So p is turning from false to true at the very instant when q turns from true to false. So the conjunction $(p \,\&\, q)$ which was false just before the change on account of p is also false just after the change on account of q. Consequently there is no change in truthvalue for $(p \,\&\, q)$ in the immediate vicinity of the instant of change. It would therefore seem reasonable to stipulate that $(p \,\&\, q)$ have that same value also at the instant of change itself. The same considerations

lead us to postulate the value 1 for $(p \vee q)$ in the case where $p = 2$ and $q = 3$ and that where $p = 3$ and $q = 2$.

The logical calculus which we obtain via these truthtables is perhaps mildly amusing. But the ideas behind it get us nowhere in general. In particular, as soon as we admit into the language L an operator, T say, which turns a sentence p into a sentence Tp which says that it is becoming the case that p, the plausibility of the truthtables vanishes. For clearly if e is a change from not-p to p then the value of Tp at $t(e)$, if it is to be any one of the four values available to the present semantics, must be 1. This implies however that, since p itself has the value 2 with respect to e, $(p \,\&\, Tp)$ gets, according to the truthtable for 'and', the value 2. But this is strongly counterintuitive. The conjunction says that it both is the case that p and that it is becoming the case that p. This seems to be a contradiction and thus should receive the value 0.[17]

The four-valued system also loses its plausibility in the absence of the assumption that the changes are ideally instantaneous. For suppose the event e which constitutes a change from not-p to p and also a change from q to not-q corresponds to the interval (t_1, t_2) of $\tau(\varepsilon')$. Graphically we might represent this situation in some such way as the following:



Here t_3 is some point interval included within the interval (t_1, t_2). The diagram suggests, and I think rightly, that there is no good justification for giving $(p \,\&\, q)$ the value 0 at t_3, especially not if we give on the other hand $(p \vee q)$ the value 1 at t_3. Nor would either of the values 2 and 3 be appropriate. For these indicate change; but neither $(p \,\&\, q)$ nor $(p \vee q)$ does change in the vicinity of t_3, at least not in the sense in which we have understood change so far, i.e. as change either from definite truth to definite falsehood or as change from definite falsehood to definite truth.[18]

---

[17] I had at first overlooked how soon the truthtables here presented become inadequate. Barbara Partee protected me from the embarrassment that would have resulted from my failure to observe this within the present, more generally accessible version of the paper.

[18] One might be inclined to assign to $(p \,\&\, q)$ at t_3 some 'intermediate truthvalue'; but

The upshot of all this is that in general we cannot hope for a complete specification of the values of complex sentences in terms of the values of their components even if we adopt the additional values 2 and 3. The best which it seems we have reason to expect is a partial truth definition. Lest it be thought that this partial-valuedness is a drawback of the present approach when compared with more familiar treatments of temporal reference I want to stress my belief that this is on the contrary a virtue.

It is a mistake to adopt a framework which makes it appear that there always ought to be an answer to the question: Is p true at time t? If t is a time of change for p this question has in general no clear answer. Only in some particular cases specific considerations may persuade us that one of the two possible answers is more suitable than its opposite. Thus if p expresses the sentence "object a is at rest relative to the object b" and t is a time where this sentence turns from true to false we may want to say that p is still true at t. The reason for this would be that, according to the received theory of motion, a has at times shortly after t a non-zero velocity relative to b while at times shortly before t the relative velocity will be zero. Moreover, the limit of the velocity at t', where t' approaches t from the right, is zero too. And, given the usual definition of instantaneous velocity this limit will in fact be equal to the velocity at t itself. Thus t shares with its temporal predecessors the property that a has zero velocity and differs in this respect from its near successors. This may, though it need not, sway us to count t as one of the instants of rest.

But even in cases such as this one the decision does not alter the status of t as an instant of change for p. Indeed, it is this feature of models for temporal reference based on event structures, i.e. that times of change are identified as such *directly*, and not as transition points between times of no change, which makes them very natural for the analysis of the logical properties of 'becoming'. In another paper[19] I shall analyze a number of formal languages which contain besides certain devices of temporal reference, such as tense operators, also the operator T ("it is becoming the case that") mentioned earlier in this section.

There are a number of different options between which one must choose when formulating the semantics for such languages. One of these I should

---

not only would we thereby abandon the four-valued scheme that is under scrutiny; the introduction of such intermediate values leads in any case to notorious difficulties. See e.g. my 'Two Theories about Adjectives', in Kamp (1975).

[19] See my contribution to the proceedings of the Conference on Quantifiers and Tense, Stuttgart, April 1979 [=Kamp (1980)].

at least mention here, as it is directly relevant to the discussion of section V. Suppose we are given a pair (ε,A) where ε is a partial event structure and A an interpretation relative to that structure. There are then three possible models with respect to which we could try and formulate the truth definition:

(a) The model (ε,A) itself; or, alternatively some model (ε',A) where ε' is some strictly relational extention of ε.
(b) The model $(\tau(\varepsilon),A')$ where $\tau(\varepsilon')$ is the instant structure determined by some strictly relational extention ε' of ε and A' is obtained from A via (2) on p. 80, in the manner indicated in the paragraph below (2).
(c) The structure $(\varepsilon(\tau(\varepsilon')),A'')$ where $(\varepsilon(\tau(\varepsilon'))$ is the interval structure determined by $\tau(\varepsilon')$ and A'' is determined by A.[20,e]

Which of these different models will provide the most suitable framework for the formulation of the truth definition of any particular language L will depend on the specific features of L. These questions too are better left to the later paper.

VII[21]

In section III I briefly stated my intention to apply the theory of events to the analysis of discourse. In this last section I wish to say a little more about such applications. The reader will have to be content with a mere outline of the possibilities which arise in this connection and some discussion of certain methodological and conceptual implications. I hope to provide further details in the already mentioned Kamp (1980).

We shall continue to limit our attention to the forms of discourse where *A* does all the talking (or writing) and *B* all the listening (or reading); where

[20] Precisely how A'' is to depend on A is not quite so simple a question. Clearly intervals of the form $\bar{e}$, where $e \in E$, will have to belong to $A''^{\pm}(Q)(a_1, \ldots, a_n)$ if e belongs to $A^{\pm}(Q)(a_1, \ldots, a_n)$. But for intervals which are not of this form such questions are complicated by the problems which we encountered in our discussion of '*a* writes *b*'. I must leave this question to the other above mentioned paper.

[21] The material presented in the present section owes much to the help and advice from members of the DFG project on tense and aspect in Stuttgart, in particular Franz Guenthner, Jaap Hoepelman and Christian Rohrer.

[e] The original text has '$\iota(\tau(\varepsilon'))$' here. But what is meant is nothing other than what is referred to in Section II, p. 58 as '$\varepsilon(\tau(\varepsilon'))$'.

the discourse consists of a succession of indicative sentences $s_1, \ldots, s_n$; and where *B* adjusts his representation of the information he has obtained after each successive sentence. Moreover I shall limit myself here to discourse that is exclusively in the past tense. Discourse that satisfies all these constraints one finds typically

(i) in written and oral narrative prose;
(ii) in reports about past events, such as those given, for instance, by eyewitnesses under interrogation.

Even in the presence of these severe restrictions one encounters phenomena whose significance can be properly understood only, I believe, within the framework that has been developed in the earlier sections of this paper. Central among these phenomena is the 'directive' function of certain devices that occur in natural language, of which I shall discuss a couple of examples.

The first of these is strictly speaking a combination of two contrasting devices, represented in French by, respectively, the Passé Simple and the Imparfait. Quite a bit has been written about the difference between these two tenses, but mostly by grammarians working in a non-formal framework, or at any rate in one very different from that of modeltheoretic semantics. It has been suggested: that the Passé Simple is used when one wants to emphasize the punctual character of the reported event, while the Imparfait stresses its durative aspect; that the Imparfait is used to provide the background to a certain event or sequence of events, whereas the succession of events which unfolds against this background is reported with the help of the Passé Simple; that the use of the Imparfait places the addressee as it were inside the locus of action, where the use of the Passé Simple would have kept him at what might be called an outside observation post.

Each of these contrasts contains an important element of truth; and no one can be captured, it would seem, within a theory of truth conditions: It does not seem possible to articulate those contrasts as differences in truth conditions between sentences in respectively the Passé Simple and the Imparfait. Exactly the same event can be correctly reported by the sentence

"Il y a deux ans la Compagnie acheta un navire de 100.000 tonnes."

and by the sentence

"Il y a deux ans la Compagnie achetait un navire de 100.000 tonnes."

Which tense is appropriate depends not on features of the event as such (e.g. the time that it took to conclude the deal) but rather on the role that is

attributed to it, or the angle from which it is viewed in the context in which it is mentioned.[22]

The claim I wish to make is that what distinguishes Imparfait and Passé Simple does not so much lie in the contribution they make to the truth conditions of the sentences in which they occur, but rather in the different directives they convey to the addressee concerning how he is to represent to himself the contents of the sentences which these tenses mark.

As a first approximation we can say, reiterating one of the theses referred to above, that the Passé Simple indicates that the reported event is to be viewed as punctual, or temporally undivided. Let us assume that the discourse representations $D_1, \ldots, D_n$ which B constructs in response to A's utterances of $s_1, \ldots, s_n$ are partial interpretations of the form (ε, A) which we described in the previous section. Let (ε_k, A_k) be the structure B has constructed at the point where A has uttered $s_1, \ldots, s_k$. Suppose s_{k+1} is a sentence in the Passé Simple, and moreover a particularly simple sentence of this sort, e.g.

> "Jean ôta sa veste."

If, as would normally be the case, this sentence gives new information, not yet conveyed by $s_1, \ldots, s_k$, then B is to add to the event set E_k a new event e; the Passé Simple indicates that this event is to be treated as *punctual within* the representation of the discourse which B is in the process of building up. I shall interpret this formally as the condition that in $\varepsilon_{k+1}e$ must satisfy the condition that it is not partitioned by other events:

$$(\forall e_1, e_2)(e_1 \bigcirc^+ e \,\&\, e_2 \bigcirc^+ e \rightarrow e_1 \prec^- e_2)$$

It is natural to ask whether this condition should be regarded as remaining in force in the course of the subsequent modification, which yield, in succession, $D_{k+2}, \ldots, D_n$. It seems that this is not always so. Consider for instance the following passage:

> "Hamilcar continua seul à pied, sans escorte, car les réunions des Anciens étaient, dans les circonstances extraordinaires, toujours secrètes, et l'on s'y

---

[22] I do not want to go so far as to say that there are no events, processes or states which can be correctly reported with the help of only one of these tenses. For example it seems doubtful that one could ever be justified in the use of the Passé Simple in a sentence such as: "Francois fut le fils du Directeur et Mme Delahaye". Even of such a sentence however it is by no means evident that it is unacceptable because it is (necessarily) false. It is rather that certain events, processes or states cannot be viewed in a certain way or attributed a certain role, because the aspect that the viewpoint emphasizes or that is required for the fulfillment of the role, is entirely absent from them.

> rendait mystérieusement. D'abord il longea la face orientale de l'Acropole, passa ensuite par le Marché-aux-herbes, les galeries de Kinsido, le Faubourg-des-Parfumeurs. ..." (Flaubert, Salammbô)

Here the first event is reported in the Passé Simple; and yet the passage continues to recount various stages of that event. There is a natural temptation to explain texts of this type by saying that a certain thematic 'break' occurs between the introduction of the event which is subsequently subdivided and the following sentences which effect the subdivision. Unfortunately it is not easy to formulate the concept of thematic 'break' in a way which protects such an explanation against circularity. I shall leave the problem of accounting for texts of this sort for a later occasion, and ignore this complication in the present section. Thus we shall make here the simplifying assumption that once a condition of the form (1) has been introduced into D_{k+1} it remains in force in all the subsequent $D_{k+2}, \ldots, D_n$.[23]

Connected with the principle that the Passé Simple indicates punctuality appears to be the rule of narrative discourse that a succession of sentences in the Passé Simple is to be taken as reporting a similarly ordered sequence of successive events ("The Passé Simple pushes the action forward"). Indeed, although I do not think that there is any clear sense in which the principle can be said to *imply* the rule the following considerations may suggest how the former could have contributed to the development which established the latter.

Suppose e_1 and e_2 are events reported by two successive Passé Simple sentences s_1 and s_2. Let t_1 and t_2 be the time points these events determine. There are in principle three possibilities for their relative location:

(i) t_2 after t_1;
(ii) t_2 coincident with t_1; and
(iii) t_2 before t_1.

Why should possibilities (ii) and (iii) be ruled out? To begin with we should observe that neither is always precluded. The second possibility for instance represents a genuine interpretation of a two sentence sequence such as:

> "Pierre épousa Francine. Eduard épousa Chantale."

---

[23] It follows from these remarks that strictly speaking a discourse representation D_i involves, besides a structure (ε_i, A_i) also a set C_i of conditions which stipulate the status of some of the members of E_i, and which are carried over to the next structure D_{i+1}. In most of what follows, however, I shall ignore this and speak of the D_i as if they consisted simply of the pairs (ε_i, A_i).

The cases where a consecutive interpretation of the two events appears strongly preferred are typically those where their simultaneity implies that at least one of them must have lasted for some time. This arises in particular in contexts where the sentences s_1 and s_2 report actions by (or, more generally, events involving) one and the same individual. In some such cases it is down right impossible to be involved in both events at the same time (getting married and getting pregnant are arguably events of this sort, at least if we understand the former in the legal sense of the term); in which case coincidence of t_2 with t_1 is ruled out trivially. But more often it is rather that though the actions could be performed, or the events could happen, simultaneously, this is only possible if the individual involved in both 'meshes', so to speak, his involvements: he has to perform the second action, say, while engaged in, and perhaps briefly suspending, the first action. The interpretation of the two as simultaneous would then impose upon at least one of them a durative aspect which the use of the Passé Simple precludes.

The preference of (i) over (iii) has, I believe, a quite different origin. There is a natural preference for listing a collection of events in the temporal order in which they occurred rather than in some other order. This is so because only by following the actual order of the events in his account does the speaker enable his audience to 'live through' the events in their true succession. It is very plausible moreover that as soon as one way of recounting a series of events has some over-all preferability over its potential competitors this one way becomes established as a convention. For clearly it is a practical advantage to be able to convey the actual order of events simply through the order of enunciation, rather than having to add explicit verbal information for this purpose. And, given that it is advantageous to have such a convention it should be expected that the one to establish itself is that which provides the most natural correlation between order of enunciation and order of occurrence.

It is important to notice in this connection that such a convention is only operative where the speaker can be assumed to know what the true order of events is—this is so in particular in fiction, where the order is, as much as any other aspect of the story, of the speaker's own making. In situations, on the other hand, where the speaker cannot be assumed to have complete knowledge as regards order, such as when a witness tries to recall the various events relevant to the case in connection with which he has been called, the convention does not hold: the witness may recall the events in the wrong order; and if later on in his account he makes an explicit statement about the temporal order of two events which he introduced at some earlier point, he thereby invalidates his earlier statements no more if this order goes against

that in which the events were introduced than it would if the two orders agreed.

While the Passé Simple is typically used to report the successive elements of the main course of action of a story, the Imparfait serves to present the setting in which the action is taking place.

Of course, as the story progresses the background may change as well. Therefore a sentence in the Imparfait, enunciated at a particular point in the narration, will convey that the condition to which it refers obtains at the time to which the action has progressed at that point of the story. It may, but need not, obtain also at earlier or later parts of the course of events. Roughly therefore narrative prose is governed by the following principles.

(1) The Passé Simple conveys punctuality of the event reported.
(2) In successions of sentences the Passé Simple conveys a similar temporal succession of the reported events.
(3) A sentence in the Imparfait conveys that at the point to which the story has proceeded a certain condition obtains.

(3) implies that a series of successive sentences in the Imparfait conveys that the corresponding conditions all obtain at the same time. If the discourse then continues with a sentence in the Passé Simple which introduces a new 'punctual' event these conditions may, but do not have to, hold also at this next time. There is one exception to this, viz. when the narrative starts with a number of sentences in the Imparfait which set the stage for the beginning of the action. Here it seems to be understood that the first event does fall in the time span where all the above mentioned conditions obtain. But the events introduced by Passé Simple sentences following this first one are again free of this constraint: at the times they represent the conditions mentioned at the start of the story need no longer be in force.

Evidently these rules cover only a relatively small part of a very special form of verbal communication. But they provide a simple framework within which it will be possible to discuss the few general points to which this section will be confined.

First I must say a little more about the sense in which the Passé Simple indicates 'punctuality'. I already suggested that this indication takes the form of a directive to the hearer and that formally this amounts to his adopting in the passage from D_k to D_{k+1} an instance of (1) involving the new event e which is introduced at that point. If e does satisfy (1) in ε_{k+1} then $\tau(\varepsilon_{k+1})$ will contain a single instant at which e goes on; so that in precisely this sense e is indeed instantaneous in $\tau(\varepsilon_{k+1})$. The same is true of the structures $\tau(\varepsilon')$ where ε' is a strictly relational extension of ε_{k+1} satisfying A1–A7, provided e also satisfies (1) in ε'. (Theorem 2 of section IV implies

that a strictly relational extension satisfying this extra constraint will in fact always exist.)

The fact that e is instantaneous in $\tau(\varepsilon')$ does of course not mean that e is instantaneous in any absolute sense. To be a little more precise let us assume that the discourse consisting of $s_1, \ldots, s_n$ is non-fictional. In that case the question arises whether what was said was indeed true. Formally this should mean that in a model which represents the 'real world' as it develops through time the sequence of sentences is true at the time they were uttered. For the present discussion it is not so important how this real world model is precisely set up. Let us assume, to start, that it has the familiar form of a pair (τ, A) where τ is a strict linear ordering, e.g. an ordering isomorphic to that of the real numbers, and A is an interpretation function relative to τ of the primitive non-logical expressions of the language in question. I shall make the idealized, but harmless, assumption that the entire sequence $s_1, \ldots, s_n$ is asserted at one single instant t_0.

Under these assumptions the question of truth is the question whether the sentences $s_1, \ldots, s_n$ are all true in the model (τ, A) at the time t_0. Suppose that one of these sentences, s_k say, reports an event in the Passé Simple, e.g.

"Jean ferma la porte."

If this sentence is true then there must be one or more instants of τ before t_0 at which the relevant event goes on, i.e. at which John and the door stand in the relevant relation (let us leave aside for now the question which relation this is). Now from a physical point of view events such as the closing of a door always take time. So, if (τ, A) describes the world from a physical standpoint it will contain an entire interval of points at which the relation holds. This is moreover something of which speakers are of course perfectly well aware, even when they treat the event as punctual: what emerges as an instant in the instant structure induced by the discourse representations D_i may nevertheless be a temporally extended event from another point of view; and that other point of view may moreover be the one that generally passes for the 'objective' point of view.[24]

In fact there ought to be, in those cases where A speaks truthfully, a natural correspondence between the discourse structures D_1 which B constructs in response to his words and the real world model (τ, A). In order to give an idea of the formal nature of this correspondence let us assume that the

[24] From what was said in earlier sections of this paper it follows that there is no such thing as *the objective* point of view in some absolute, uniquely determined sense.

sentences $s_1, \ldots, s_n$ are drawn from a very simple language L, all of whose sentences are of either one of two forms: $P_i q_j$ or $P_s q_j$. These sentences result from applying the operators P_i and P_s to underlying sentence symbols q_j. Intuitively the sentences $P_i q_j$ and $P_s q_j$ stand for certain simple French sentences in, respectively, the Passé Simple and the Imparfait such as

> "Jean ferma la porte" and
> "Francine se brossait les cheveux";

and the sentence symbols q_j which occur as parts of them can roughly be viewed as standing for the corresponding present tense sentences ('Jean ferme la porte', 'Francine se brosse les cheveux').[25]

The informal rules for the interpretation of narrative discourse now come to this: We distinguish two cases

(a) s_1 is of the form $P_s q_j$; and
(b) s_1 is of the form $P_i q_j$.

We first consider case (a).

(1) D_1 is to be the pair (ε_1, A_1), where E_1 is to contain two events, e_s—the event of utterance of the discourse $s_1, \ldots, s_n$—and e_1, the event reported by s_1. ε_1 incorporates moreover the information that e_1 precedes e_s; and e_1 is put into $A^+(q_j)$

(2) Suppose D_k has already been constructed. s_{k+1} will be either of the form $P_s q_j$ or of the form $P_i q_j$. Suppose first that it is of the form $P_s q_j$. In this case we add a new event e to E_k and include the ε_{k+1} the information that it precedes e_s and is itself preceded by the event reported by the last Passé Simple sentence preceding s_{k+1}. Moreover e is put into $A^+(q_j)$. Suppose next that s_{k+1} is of the form $P_i q_j$. Then we also introduce a new event into E_k and into $A^+(q_j)$; but now we specify in that this event overlaps with the last event introduced by a Passé Simple sentence.

In case (b) the construction is to proceed in almost the same way. The only difference concerns the beginning of the construction. Suppose that the first

[25] It has been one of the basic assumptions of tense logics of the sort briefly alluded to in the opening paragraphs of section V that sentences in the simple past tense can be analyzed as resulting from applying a 'past tense operator' to an underlying sentence, often identified with the corresponding present tense version of the past tense sentence in question. This identification is notoriously problematic, especially in English. However, whatever the precise relation may be between the past tense and the present tense sentence, it usually makes perfectly good sense to ask whether the conditions referred to in the sentence—whether it be in the past or in the present tense—obtain at a given time, e.g. if at a given time John and the door stand in the relation which obtains whenever John is closing the door. This is sufficient justification for the introduction of the sentence symbols q_j.

Passé Simple sentence of the discourse is s_{r+1}. Then each of the sentences $s_1, \ldots, s_r$ is to be taken as introducing an event which overlaps each of the others, and moreover does not lie entirely after e_s. The event introduced in response to s_{r+1} is also to overlap all of these first r events as well as to completely precede e_s. After that the construction follows the same rules as under (a).

We assume that A spoke truthfully. But what precisely does this mean? It means, at the very least, that each of the sentences $s_1, \ldots, s_n$ is true in $(\tau, \mathcal{A})$ at t_0. However it might be argued that this is not by itself quite enough for A to have spoken the full truth; that in order for that to be so the times at which the underlying 'present tense sentences' q_j are true in $(\tau, \mathcal{A})$ must moreover be in the temporal order which the sequential order of the sentences $s_1, \ldots, s_n$ conveys.

I shall come back to this question soon. But for the moment let us assume that the order in which the mentioned events happened is indeed in accordance with that conveyed by the discourse.

To investigate the relation that obtains in this case between the discourse structures D_i and the real world model $\langle \tau, \mathcal{A} \rangle$ let us concentrate on the last structure D_n. As before I shall assume that D_n has the form $\langle \varepsilon_n, \mathcal{A}_n \rangle$, where for each q_j $\mathcal{A}_n^+(q_j)$ and $\mathcal{A}_n^-(q_j)$ are disjoint subsets of E_n.[26] The function $\mathcal{A}$ on the other hand will be assumed to assign to certain combinations of a sentence q_j and a time $t \in T$ a truth value (1 or 0).

Let us first suppose that for each of the q_j that occur in any of the sentences $s_1, \ldots, s_n$ there is an unique interval i_j before t_0 consisting of all and only those points $t \in T$ such that $\mathcal{A}(q_j, t) = 1$. Let $\iota = \langle I, \prec, \bigcirc \rangle$ be the structure consisting of precisely these intervals, together with the one point interval $[t_0, t_0]$, where the relation $\prec$ and $\bigcirc$ are defined in the obvious way (see section II). It is easily verified that since $s_1, \ldots, s_n$ report the events in their true order, ι corresponds to a strictly relational extension of ε_n; i.e. there is such an extension $\varepsilon' = \langle E_n, \prec_n^{+'}, \prec_n^{-'}, \bigcirc_n^{+'}, \bigcirc_n^{-'} \rangle$ such that $\langle I, \prec, \bigcirc \rangle$ is isomorphic to $\langle E_n, \prec_n^{+'}, \bigcirc_n^{+'} \rangle$. This isomorphism preserves moreover the semantic connection with the q_j: $e \in \mathcal{A}^+(q_j)$ iff the corresponding interval i consists of all the points at which q_j is true according to $\mathcal{A}$.

Each instant of $\tau(\iota)$ will be a finite set of overlapping intervals. Let f be the function which maps each t that belongs to the union of all the

---

[26] In discourse structures constructed in response to the special type of discourse we are considering now many, and possibly all, of these sets are either empty or singletons; moreover, as we have formulated the language L, the sets $\mathcal{A}_n^-(q_j)$ will all remain empty. But in general these special conditions need not be satisfied.

intervals of I to that instant of $\tau(\iota)$ all members of which contain t. Then f is a homomorphism of a substructure of τ (that substructure which is determined by I) to the structure $\tau(\iota)$. f induces therefore a homomorphism g from that substructure of τ to a strucure $\tau(\varepsilon')$ where ε' is some strictly relational extension of ε_n. It should be evident that this homomorphism also preserves truth, in the sense that if q_j is true at t according to A then there is an $e \in E_n$ such that $e \in g(t)$ and $e \in A^+(q_j)$, and conversely.

In general the sets of times at which the q_j are true according to A need not constitute single, uninterrupted intervals. There are at least two reasons why this need not be so. In the first place the actual past may contain several events of the type of q_j, only one of which the discourse in question reports. In the second place, as we observed in section IV, a certain period may qualify as the duration of a protracted event or process without it being strictly true that the conditions in terms of which the event or process is identified obtain at every single instant of that period.

Let us first assume that only the former of these two reasons is operative. In that case it remains possible to single out a set I of intervals of τ such that the structure $\langle I, \prec, \bigcirc \rangle$ determines a strictly relational extension of ε_n in the way described, and which is such that if the interval i of I corresponds to the event e of E_n then q_j is true, according to A, at each $t \in i$ iff $e \in A^+(q_j)$. But it will in general not be possible, as it was in the simpler case considered earlier, to *define* the members of I explicitly in terms of the information A provides about the q_j.

If the second reason is operative also then it need no longer be true that the interval of τ which, intuitively speaking, ought to correspond to the event e, introduced into ε_n in response to, say, the sentence $P_s q_j$ consists exclusively of instants at which q is true according to A. The correspondence between ι and ε_n will then no longer verify (2). An adequate description of the relation between D_n and the real world would in such cases seems to require that we represent the actual past also in the form of a model $M = (\varepsilon, A)$, where ε is a partial event structure. Each of D_n and M then induces its own instant structure and the second of these will be a 'refinement' of the first—as ε will contain so many more events than ε_n.

These extra complications do not alter the moral, however, that is implicit in our discussion of the simple case. The moral is this. There is a perfectly legitimate sense in which we can, in a given context, treat, or look upon, certain events as punctual. Such an attitude does not imply any commitment to the belief that these events are indeed instantaneous absolutely. But it will determine (and manifest itself in) particular ways of talking about them, the

interpretation of which leads to temporally ordered structures in which they do *play* the role of instants. There where discourse concerns real events, and reports them correctly, this structure can be regarded as containing, among other things, a possibly incomplete representation of part of the real world. This representation involves as a rule some sort of *contraction* of what are in physical terms protracted happenings to indivisible temporal units.

At two earlier points I already touched upon the question whether the difference between Passé Simple and Imparfait has anything to do with the contributions these respective tenses make to the truth conditions of the sentences in which they occur. I first expressed myself as if this were not so. However, subsequent discussions seemed to indicate that that was too rash a denial. More careful reflection upon the matter shows that what we should say here ultimately depends on the role we wish to attribute to the concept of truth.

What might appear to lend the denial some plausibility is the assumption—implicit in so much work within formal semantics—that truth conditions can be assigned to sentences in isolation. For, as already remarked, there is little if anything in the respective truth conditions of two sentences which differ only in that the first has a Passé Simple where the second has an Imparfait to set them apart. The assumption that sentences determine their truth conditions individually, and that a bit of non-fictional assertoric language is true as a whole if and only if each of the sentences uttered as part of it is true separately arises naturally, but surreptitiously, from a conception of semantics that is too simple-mindedly compositional.

Yet the implausibility of that assumption flies in the face of anybody who stops to think about, for instance, the phenomenon of intersentence anaphora. As it has been realized for some time how close the parallels are between tenses and pronouns, and in particular how important anaphora is not only in connection with pronouns and demonstratives but also in connection with tenses,[27] it should be no surprise that the dubious assumption proves equally untenable within the present domain of enquiry—indeed the mechanisms of which I tried to give a description on page 99—I might well be said to be themselves mechanisms of anaphora; for they allow us to infer the time of a reported event, or at any rate some information about the temporal position of that event, from preceding discourse.

Once we have accepted that it is only to the discourse as a whole that truth conditions can be assigned in a systematic way, we may find it natural to

[27] See in particular Partee (1973).

explicate the difference between Passé Simple and Imparfait at least partly in terms of their distinctive contributions to the truth conditions of the discourse. For as we saw the particular temporal order that the discourse conveys depends on where it uses the Passé Simple and where the Imparfait.

But is this order really part of the truth conditions of the discourse? After all the mechanisms that convey the order are operative only in certain types of discourse; and the question whether a particular discourse is of such a type may depend on such typically pragmatic factors as the speaker's knowledge about the order of the events which he recounts. There has been a strong tendency in the past to keep such pragmatic factors and their implications separate from typically 'semantic' notions such as truth conditions, a strategy which, as Grice was the first to point out, often has very striking advantages.

As a matter of fact the separation of pragmatic from semantic issues leads in some cases to considerable technical difficulties.[28] I do not know if such difficulties are to be feared within the context of the present phenomenon also. But let us assume for the sake of argument that they do not arise. The question whether the conveyed order of events is part of the truth conditions of the discourse is then, it seems to me, a question about the *status* of the mechanism for conveying it; and this status may differ from one case to the next. Thus we may become persuaded that with certain forms of language, in particular literary forms, the mechanism is associated in a strictly conventional manner, while in other cases the association is based on the consensus among the discourse participants that the necessary pragmatic preconditions for it obtain. In the former cases it would seem reasonable to regard the order as a matter of truth conditions; in the latter probably not.

Some of the issues to which I have tried to draw attention in this section also arise in connection with the second of the two examples which I promised to discuss. This is the English expression 'and next'.[f] 'And next' has commanded a certain amount of attention within philosophy, primarily through the work of von Wright (1965), in whose theory of action and obligation it plays an important part. Von Wright analyzes 'and next' as a 2-place sentential operator, which combines sentences p and q into a

[28] See Kamp (1978). Admittedly the issues there discussed are quite different from those which concern us here.

[f] Actually no such promise is made in the version of the paper that got published. (Or perhaps the author had made such a promise just to himself.)

compound sentence *pNq* which is stipulated to be true at a time *t* iff *p* is true at *t* and *q* is true at the next time after *t*. According to this stipulation *N* can be used non-trivially only in a context where time is discrete, i.e. where moments have immediate successors. Indeed it is only time of this type that von Wright considers.

This strikes me as unsatisfactory—at least if things are left at just that. For most of us do not really believe that time is discrete. Yet we all use the expression 'and next' often, and apparently without effort or strain.

The explanation of this must, I think, once again be sought in the *directive* function that attaches to the expression: it instructs the addressee to treat the events e_1 and e_2 reported by the first and the second, respectively, of the two sentences which it combines as *consecutive*. By using 'and next' the speaker conveys to the hearer that within the context none of the events which from an 'objective' point of view do come between e_1 and e_2 are to be taken into account. It is precisely the element of stipulation that makes it often so difficult to fault a speaker on this score: if the event e_2 was indeed after e_1 how is one to make a convincing case that what the speaker said was wrong because other events, which he did not mention, intervened?

Nevertheless there are situations in which such an attack seems justified. Suppose you and I went to town this morning to do the shopping and have just returned to the cottage with the groceries. I say to my wife: "First we went to the butcher and next we went to the bakery". Whereupon you interrupt me saying: "No, that it is not quite right. After the butcher we went to the greengrocer's, and only then to the bakery". Here it seems that if things were indeed as you say my account was in fact false. The reason is that the context *already* determines quite clearly which events are to count as significant. An enumeration which claims two of them to be consecutive while there is a third from this very same class which came in between therefore strikes us as being simply not true to the facts. But what appears in this case as part of the truth conditions will in others, where the class of significant events has not been fixed in advance, only be understood as a contribution to the definition of that class; as such it might have an effect on the truth conditions of subsequent sentences; but it cannot be itself attacked as being in breach of the truth.

Both examples we have discussed indicate how difficult it is to draw the fashionable distinction between 'semantic' and 'pragmatic' aspects of language. The difficulties we have encountered are twofold. In the first place it looks as if the effects of what is in each case essentially the same mechanism (the mechanism for conveying order of occurrence through order of enunciation) should sometimes be classified as pragmatic and sometimes

as semantic. But to insist on such a classification is unhelpful. Not only is it likely to lead to quite arbitrary cuts in a continuous spectrum of types of discourse (stretching from those where, according to existing criteria for the distinction, the contributions are clearly pragmatic to those where the pragmatic element has become completely ossifed, and the contributions would count as semantic); also by attributing too much importance to the distinction we risk concealing from view the unifying principle that is in each case responsible for these contributions.

In the second place there is the status of the directive functions themselves. In as much as they are directives to the hearer they belong in the general theory of speaker-hearer interaction and thus might seem to qualify as pragmatic. At the same time however they have an important semantic aspect, as they serve the hearer in his effort to construct his representations of what is said; and what, if it isn't that, could be identified as his assigning a *meaning* to what he has been told?

Indeed it seems to me that the rules for the construction of discourse representations have at least as good a claim to being constitutive of meaning as the clauses which make up the definition of truth. Meaning after all attaches to fiction no less than to factual discourse. What distinguishes the two is that the latter, but not the former, can be (literally) true or false. But both induce the formation of structures that mirror their contents; and the rules that guide such formation are, we have seen, the same in either case.

Appendix I

Let $M_1 = (U_1, R_1)$ and $M_2 = (U_2, R_2)$ be as in Section IV.

By the *distance in* M_1 *from* the element u_1 of U_1 *to* the element u_2 of U_2 we understand:

(1) the unique positive integer n such that u_2 is the n-th successor of u_1, provided there is such an integer;
(2) the unique number $-n$ such that u_1 is the n-th successor of u_2, provided there is such an n;
(3) ∞ if neither (i) nor (ii) applies.

We write '$d(u_1, u_2)$' for this distance. In the same way we define the *distance in* M_2 *from* an element v_1 of U_2 *to* another element v_2 of U_2. We shall refer to the elements of M_1 and M_2 which have no immediate successors as "zero's" (of M_1 or M_2, respectively). We say that a k-tuple $(u_1, \ldots, u_k)$ of elements of U_1 and a k-tuple $(v_1, \ldots, v_k)$ of elements of U_2 are *similar mod* n (in symbols: $(u_1, \ldots, u_k) \approx_n (v_1, \ldots, v_k)$) iff

(i) for each $i, i \leq k$ either $d(u_i, u_j) = d(v_i, v_j)$ or else both $|d(u_i, u_j)|$ and $|d(v_i, v_j)| > n$ and
(ii) for $i \leq k u_i$ and v_i either are at the same distance from a zero or each is from every zero at a distance whose absolute value is greater than n.

We associate with each formula φ of the language of M_1 and M_2 a number $n(\varphi)$ such that, assuming φ has k free variables, (1) if $(u_1, \ldots, u_k) \underset{n(\varphi)}{\sim} (v_1, \ldots, v_k)$ then $M_1 \vDash \varphi[u_1, \ldots, u_k]$ iff $M_2 \vDash \varphi[v_1, \ldots, v_k]$.

$n(\varphi)$ is defined by induction on the complexity of φ:

(i) if φ is atomic then $n(\varphi) = 1$;
(ii) if φ is $\neg\psi$ then $n(\varphi) = n(\psi)$;
(iii) if $\varphi = \psi_1 \,\&\, \psi_2$ then $n(\varphi) = max(n(\psi_1), n(\psi_2))$;
(iv) if $\varphi = (\exists x_k)\psi(x_1, \ldots, x_{k-1}), x_k)$, then $n(\varphi) = 2.\, n(\psi) + 1$.

(1) is shown by induction on φ. The only interesting case is that where φ has the form $(\exists x_k)\psi(x_1, \ldots, x_{k-1}), x_k)$. Assume $(u_1, \ldots, u_{k-1}) \underset{n(\varphi)}{\sim} (v_1, \ldots, v_{k-1})$. Suppose $M_1 \vDash \varphi[u_1, \ldots, u_{k-1}]$. Then there is a $u_k \in U_1$ so that $M_1 \vDash \psi[u_1, \ldots, u_{k-1}, u_k]$. We distinguish the following cases:

(i) $d(u_i, u_k) \leq n(\psi)$ for some $i < k$. In this case choose v_k at distance $d(u_i, u_k)$ from v_i. Note that this is possible even when $d(u_i, u_k) > 0$. For if v_i were less than $d(u_i, u_k)$ points from a zero of M_2 then u_i would have been that same number of points from a zero of M_1 and so the distance from u_i to u_k could not have been as much as $d(u_i, u_k)$. It is easy to see that $(u_1, \ldots, u_{k-1}, u_k) \underset{n(\psi)}{\sim} (v_1, \ldots, v_{k-1}, v_k)$. Suppose e.g. that $|d(u_k, u_j)| \leq n(\psi)$ for some $j \neq i$. Then $|d(u_i, u_j)| \leq n(\varphi)$. So $d(v_i, v_j) = d(u_i, u_j)$, and so, since also $d(v_i, v_k) = d(u_i, u_k)$, $d(v_k, v_j) = d(u_k, u_j)$. In the same way one argues that if $|d(v_k, v_j)| \leq n(\psi)$ for some $j \neq i$, then $d(v_k, v_j) = d(u_k, u_j)$, and also that if either u_k or v_k is at some distance from a zero which is less than or equal to $n(\psi)$ absolutely, then the other point is at the same distance from a zero too.
(ii) For no $i < k |d(u_i, u_k)| \leq n(\psi)$, but $|d(u_k, z)| \leq n(\psi)$ for some zero z of M_1. In that case we choose v_k to be a point of U_2 which has that same distance to some zero of M_2. Clearly $|d(v_i, v_k)| > n(\psi)$ for all $i < k$. For if for some $id(v_i, v_k) \leq n(\psi)$ then v_i would be at a distance $\leq n(\varphi)$ from a zero of M_2 and so u_i would have been at that distance from z. But then $d(u_i, u_k) = d(v_i, v_k)$ and so $|d(u_i, u_k)| \leq n(\psi)$, contrary to assumption. Thus we have again that $(u_1, \ldots, u_{k-1}, u_k) \underset{n(\psi)}{\sim} (v_1, \ldots, v_{k-1}, v_k)$.
(iii) For no $i < k |d(u_i, u_k)| \leq n(\psi)$ and for no zero $z d(u_k, z) \leq n(\psi)$. In this case we choose a v_k which has distances $> n(\psi)$ to all v_i as well as to all zero's of M_2. Clearly $(u_1, \ldots, u_{k-1}, u_k) \underset{n(\varphi)}{\sim} (v_1, \ldots, v_{k-1}, v_k)$.

Since in each of the cases (i), (ii), (iii) $(u_1, \ldots, u_{k-1}, u_k) \underset{n(\varphi)}{\sim} (v_1, \ldots, v_{k-1}, v_k)$ we may infer by the inductive hypothesis that $M_2 \vDash \psi[v_1, \ldots, v_k]$. So $M_2 \vDash (\exists x_k)\psi[v_1, \ldots, v_{k-1}]$. Analogously, if $M_2 \vDash \psi[v_1, \ldots, v_{k-1}]$, then $M_1 \vDash \psi[u_1, \ldots, u_{k-1}]$. This completes the induction.

It follows in particular that if φ is a sentence then $M_1 \models \varphi$ iff $M_2 \models \varphi$. So M_1 and M_2 are elementarily equivalent.

Appendix II

Let T_0 be the theory of strict partial orderings. (Thus T_0 is axiomatized by the two axioms $(\forall x, y, z)(x < y \,\&\, y < z \rightarrow x < z)$ and $(\forall x, y)(x < y \rightarrow \neg y < x)$). Then $RS(T_0) = SRS(T_0) = Neg(T_0) = Unneg(T_0)$ is axiomatized by the set of all sentences φ_n of the form

$$(\forall x_1, \ldots, x_n)\neg(x_1 < x_2 \,\&\, x_2 < x_3 \,\&\, \ldots \,\&\, x_{n-1} < x_n \,\&\, x_n < x_1)$$

where $x_1, \ldots, x_n$ are n distinct variables. (Intuitively such a sentence says that there is no cycle of n elements). It is easy to see that this set of axioms cannot be replaced by any finite set. The axiomatization of $RS(T_{ev}) = SRS(T_{ev})$ is a little more complicated. Indeed, the statement of the axioms that I shall give here involves an element of recursion which did not enter into the axiomatization of $RS(T_0)$; and I do not know that this element could be eliminated from a perspicuous formulation of axioms for $RS(T_{ev})$.

The axioms will be the universal closures of disjunctions of negated 'basic' formulae (where these basic formulae are either atomic formulae or conjunctions of two atomic formulae). However, it will facilitate the subsequent description to take as axioms rather the equivalent forms in which the matrix of the sentence is the negation of a conjunction of basic formulae, and to identify any such axiom simply in terms of that conjunction. Thus $x <^+ y \,\&\, y <^+ z \,\&\, z \bigcirc^+ x$ represents the axiom $(\forall x, y, z)(x <^+ y \,\&\, y <^+ z \,\&\, z \bigcirc^+ x)$. It is easily verified that all sentences represented, in the way just explained, by the following conjunctions are theorems of $RS(T_{ev})$:

(C1) $R_1(x_1, x_2) \,\&\, R_2(x_2, x_3) \,\&\, \ldots \,\&\, R_{n-1}(x_{n-1}, x_n) \,\&\, R_n(x_n, x_1)$

where for $i = 1, \ldots, n$ each formula $R_i(x_i, x_{i+1})$ has one of the forms:

(i) $x_i <^+ x_{i+1}$
(ii) $x_i \bigcirc^- x_{i+1} \,\&\, x_{i+1} <^- x_i$
(iii) $x_{i+1} \bigcirc^- x_i \,\&\, x_{i+1} <^- x_i$
(iv) $x_i \bigcirc^+ x_{i+1}$
(v) $x_{i+1} \bigcirc^+ x_i$
(vi) $x_{i+1} <^- x_i$

and where in the cycle $\{1, 2, \ldots, n, 1\}$ any two formulae $R_i(x_i, x_{i+1}), R_j(x_j, x_{n+1})$ each of which has one of the forms (iv), (v), (vi) are separated by a formula $R_k(x_k, x_{k+1})$ which has one of the forms (i), (ii), (iii).

N.B. we shall henceforth refer to formulae of the forms (i), (ii) and (iii) as *strong conjuncts* and to those of forms (iv), (v) and (vi) as *weak conjuncts*.

We shall indeed adopt all formulae of the form (1) as axioms. But that is not enough. It cannot be enough since none of them contains any basic conjunct $x_i \bigcirc^- x_{i+1}$. ($\bigcirc^-$ can only occur in combination with $<^-$!) There are however negative sentences which are valid in T_{ev} and in which $\bigcirc^-$ occurs by itself e.g. $(\forall x, y)(x \bigcirc^+ y \rightarrow \neg x \bigcirc^- y)$. The difficulty with the relation $\bigcirc^-$ is that $x \bigcirc^- y$ means intuitively that either $x <^+ y$ or $y <^+ x$, but which of these two is the case the formula does not specify. We cannot therefore have any 'cyclical' axioms of the by now familiar form, which contain a single occurrence of a conjunct $x \bigcirc^- y$. For example, although the sentence represented by $x <^+ y \,\&\, y <^+ z \,\&\, z <^+ x$ is valid, this is not the case for that represented by $x \bigcirc^- y \,\&\, y <^+ z \wedge z <^+ x$. For $x \bigcirc^- y$ *might* be true because y lies before x, and not x before y. Similarly the sentence represented by $x \bigcirc^- y \& x <^+ u \& u <^+ y$ is not valid. However the combined conjunction $x \bigcirc^- y \& y <^+ z \& z <^+ x \& x <^+ u \& u <^+ y$ does represent a valid sentence, which can be seen more clearly if we write that sentence in the form:

(2) $\quad (\forall x, y, z, u)[(x \bigcirc^- y \& y <^+ z \rightarrow \neg z <^+ x) \vee (x \bigcirc^- y \& x <^+ u \rightarrow \neg u <^+ y)]$

Since for any values u_1, u_2 for x, y we must have either $\bar{u}_1 <^+ \bar{u}_2$ or $\bar{u}_2 <^+ \bar{u}_1$, one of the conditionals in the matrix of (2) must be true, irrespective of what z and v are.

Generalizing we can observe the following:

(3) Suppose the conjunctions (C3) $\varphi_1 \,\&\, \ldots \,\&\, \varphi_n$ and (C4) $\sigma_1 \,\&\, \ldots \,\&\, \sigma_m$

each represent a theorem of $RS(T_{ev})$; (and in paricular that each of the $\varphi_1 \,\&\, \ldots \,\&\, \varphi_n, \sigma_1 \,\&\, \ldots \,\&\, \sigma_m$ has one of the forms (i)–(vi) mentioned above). Suppose that v_1, v_2 are two variables so that there is a formula φ_i which has the form (a) $v_1 <^+ v_2$ and a formula σ_j which has the form (b) $v_2 <^+ v_1$. Let (C5) be the conjunction of (i) the result of replacing in (C3) all conjuncts φ_i which are occurrences of (a) by $v_1 \bigcirc^- v_2$; and (ii) the result of replacing in (C4) all conjuncts which are occurrences of (b) by $v_1 \bigcirc^- v_2$. Then (C5) represents again a theorem of $RS(T_{ev})$. Similarly we get the representation of a theorem by forming (C6) the conjunction of (i) the result of replacing in (C3) all instances of (a) by $v_2 \bigcirc^- v_1$, and (ii) substituting $v_2 \bigcirc^- v_1$ for all instances of (b) in (C4).

(3) offers a means of constructing new axioms out of pairs of sentences which have already been adopted as members of the axiom system.

It turns out that we only need to close the class of sentences represented by the conjunctions of form (C1) under the construction step described under (3) to obtain a complete axiomatization for $RS(T_{ev})$.

Summarizing, we put $A_5^0 =$ the set of all sentences represented by conjunctions of the form (C1); $A_5^{n+1} = A_5^n$ together with all sentences which can be obtained from pairs of sentences in A_5^n by applying the step described under (3). $RS(T_{ev})$ is then axiomatized by $A_5 = \cup_n A_5^n$.

It should be noted that neither of the theories $RS(T)$ we have considered is finitely axiomatizable even though the axiomatization of T was in both cases finite and in fact, one would feel, extremely simple. That $RS(T_{par})$ is not finitely axiomatizable is seen very easily. For define for $n = l, 2, \ldots \psi_n = \&_{i \leq n} \varphi_i$. Then for $m > n \vdash \psi_m \rightarrow \psi_n$. On the other hand $\nvdash \psi_n \rightarrow \psi_m$. For let $M = (U, <)$ where $U = \{1, 2, \ldots, n + l\}$ and $<= \{(i, i + 1) : i \leq n\} \cup \{(n + 1, 1)\}$. Then $M \vDash \psi_n$ but $M \nvDash \varphi_m$. So the ψ_n form a strictly increasing set of axioms for $RS(T_{par})$. This implies that the theory cannot be axiomatized finitely. Similar, though somewhat more involved, arguments show that $RS(T_{ev})$ cannot be finitely axiomatized. The proofs that the axiomatizations given here are indeed complete are not difficult but long and tedious. I therefore decided not to include them.

References

Barwise, J., 1976. Some applications of Henkin quantifiers. *Israel Journal of Mathematics* 25, 47–63.

Bennett, M., Partee, B.H., 1978. *Toward the Logic of Tense and Aspect in English*. Indiana University Linguistics Club, Bloomington, IL.

van Benthem, J., 1980. Points and periods. In: C. Rohrer (ed.), *Time, Tense and Quantifiers. Proceedings of the Stuttgart Workshop on the Logic of Tense and Quantification*. Niemeyer, Tübingen, 39–58.

Chang, C.C., Keisler, H.J., 1973. *Model Theory*. North-Holland Publishing Company, Amsterdam.

Cocchiarella, N., 1965. *Tense and Modal Logic. A Study in the Topology of temporal Reference*. Ph.D. dissertation, University of California, Los Angeles, CA.

Fine, K., 1975. Vagueness, truth and logic. *Synthese* 30, 265–300.

Kamp, H., 1975. Two theories about adjectives. In: E. Keenan (ed.), *Formal Semantics of Natural Language*. Cambridge University Press, Cambridge, 123–155. Reprinted in: K. von Heusinger, A. ter Meulen (eds.), *Meaning and the Dynamics of Interpretation. Selected Papers of Hans Kamp*. Brill, Leiden, 2013, 225–261.

Kamp, H., 1978. Semantics vs. pragmatics. In: F. Guenthner, S. Schmidt (eds.), *Formal Semantics and Pragmatics of Natural Languages*. Reidel, Dordrecht, 255–287. Reprinted in: K. von Heusinger, A. ter Meulen (eds.), *Meaning and the Dynamics of Interpretation. Selected Papers of Hans Kamp*. Brill, Leiden, 2013, 185–217.

Kamp, H., 1980. Some remarks on the logic of change, part I. In: C. Rohrer (ed.), *Time, Tense and Quantifiers. Proceedings of the Stuttgart Workshop on the Logic of Tense and Quantification*. Niemeyer, Tübingen, 135–180.

Partee, B.H., 1973. Some structural analogies between tenses and pronouns in English. *The Journal of Philosophy* 70, 601–609. Reprinted in: B.H. Partee, *Compositionality in Formal Sematics: Selected Papers by Barbara H. Partee*. Blackwell, Oxford, 2003, 50–58.

Russell, B., 1936. On order in time. *Mathematical Proceedings of the Cambridge Philosophical Society* 32, 216–228. Reprinted in: B. Russell, *Logic and Knowledge. Essays 1901–1950*, ed. by R. Marsh. Allen & Unwin, London, 1956, 345–363.

Whitehead, A., 1929. *Process and Reality*. Free Press, New York.

Whitrow, J., 1961. *The Natural Philosophy of Time*. Harper, London.

Wiener, N., 1914. A contribution to the theory of relative position. *Proceedings of the Cambridge Philosophical Society* 27, 441–449.

von Wright, G., 1965. 'And next'. *Acta Philosophica Fennica* 18, 293–304.

DEIXIS IN DISCOURSE. REICHENBACH ON TEMPORAL REFERENCE[a]

Hans Kamp

1. Introduction

It is a great honor for me to be allowed to give this lecture, in the place where, more than thirty years ago, I made the transition (painful as it is in most cases, I guess) from a "mere" student to someone trying to be a genuine member of his field by contributing to it, to the best of his ability, things of his own making or finding. At that time Reichenbach, about whom I had heard a good deal as a student in the Netherlands before I came to UCLA, was already part of its Philosophy Department's illustrious history, and I should confess that his work did, as that point, not have an immediately identifiable influence on the education which UCLA gave me—what influence there was was already part of a general heritage of philosophical insights and logical techniques. The visible importance Reichenbach has had, for me personally and more generally for that part of the field which has absorbed an increasing proportion of my attention, came—as far as I was concerned—later. This occurred at a time when, in part because of Reichenbach's work on temporal reference, I felt I had to abandon some of the scientific convictions I had formed during the three years of my stay at this university. My outlook on much that is central to the problems which have preoccupied me both before and after—the way in which language functions and its implications

[a] Editors' note: This paper is the slightly revised version of the Reichenbach Lecture at UCLA, which Hans Kamp gave in 1999. The Reichenbach Lectures are dedicated to the life and work of Hans Reichenbach (1891–1953), who worked at UCLA from 1938 to 1953. The original manuscript was never published, though it circulated as a copy since then. It consists of the lecture and an appendix that elaborates on critical questions after the talk. We have included the lecture and the main part of the appendix. We carefully edited the manuscript, corrected typos and updated the bibliographical references. We would like to thank Daniel Altshuler for proofreading and constructive comments.

© K. von Heusinger & A. ter Meulen (eds.), *Meaning and the Dynamics of Interpretation. Selected Papers of Hans Kamp*, Brill, Leiden, 2013.

© hans kamp, 2013 | doi:10.1163/9789004252882_006
This is an open access chapter distributed under the terms of the cc by-nc-nd 4.0 license.

for logic and for our understanding of the human mind—was thereby quite radically transformed.

The opportunity I have been given to address this audience here today—on an occasion dedicated to the furtherance of the scientific and philosophical ideals Reichenbach pursued, but also, I presume, to reflect on the lasting impact of his own work—has made me think back, in an attempt to recapture the changes which then occurred, to the field as a whole but, as a side-effect, also to me personally. That made it necessary to try and recall, as vividly as was still possible after so many years, what it was people thought in the days before those changes, those when I was a student here, and—an even harder task—the times before I was a UCLA graduate student. And so the experience which culminates in my standing here before you became something like the closing of a long and fairly tortuous loop.

Reichenbach should probably be characterized as first and foremost a philosopher of science. Even a cursory perusal of his bibliography reveals a clear predominance of works in the philosophy of physics and on related issues of scientific and logical methodology. (In fact it was in this context that I first became familiar with his work at the university of Amsterdam, and turned into a firm afficionado.) In comparison, his contribution to the philosophy of language may appear much less prominent—though admittedly the line between philosophy of language, philosophy of logic and scientific methodology is, in his case no less than that of many of his contemporaries, not so easy to draw. Yet, in this domain too Reichenbach has been extraordinarily influential. And probably the largest part of that influence is due to one section—of no more than ten pages—in the chapter he devoted to the analysis of natural language in his book "Elements of Symbolic Logic". This is the section on the tenses of the verb (Reichenbach 1947: pp. 289–298). Over the past 25 years it has given rise to an unending flow of publications, most of them within linguistics. And although much has happened in this domain since the first papers appeared that were inspired by it, it is still often quoted—and all of us who are working on temporal reference in natural language know that Reichenbach is the principal source from which our work received its ultimate inspiration.

This is an unusual feat, and it deserves some reflection why this comparatively short section should have had such an enormous impact. In brief the answer is something like this: (i) the section reveals what we can now recognize as at least two distinct, extremely important insights; and (ii) these were insights that, as far as I can guess, went very much against the spirit of the times—or, more correctly perhaps, against the spirit of the logical tradition to which Reichenbach belonged and which he himself had helped so

much to establish. In fact, I think it is fair to say that the section does not fit very comfortably within the chapter of which it is a part; it is quite different from all the other sections, and it is hard to see it as an instance of what Reichenbach states as his aims in the opening paragraph of the chapter to which it belongs: "In the present chapter we shall attempt to apply the methods of symbolic logic to an analysis of conversational language."[1] Unlike the other sections this one does not make any use of the symbolic tools provided by formal logic; if it is to be counted as an application of the "methods of symbolic logic", then only in the attenuated sense that what it has to say is evidently the product of a logical mind—the kind that it would be nice to think training in formal logic could create or improve. But it is hardly credible that someone with the mental powers of a Reichenbach should have needed a training in symbolic logic to be able to arrive at the conclusions which this section contains. On the contrary, one might rather be inclined to think that what is so remarkable about this section is that Reichenbach's knowledge of symbolic logic as it then existed did not deflect him from insights which are so clearly at loggerheads with the then prevalent notions of how logic could and ought to be applied.

2. Reichenbach on Tense

In order to make clearer to you what it is that makes the section on tense so remarkable I must remind you of what it actually has to say. Its central thesis, I think it is fair to hold, is the one often referred to as the "two-dimensional" analysis of tense. What is meant by this is best illustrated by the example which Reichenbach himself takes as his point of departure, the English Past Perfect. The Past Perfect, he notes, presents the information it conveys—the occurrence of some event or the obtaining of one state of affairs—as lying in the past of a certain time which in its turn lies in the past of the Speech time. And this second time is more often than not made available by the context within which the past tense sentence appears. Significantly, Reichenbach illustrates this point by two quotations of *bits of text*, each consisting of several sentences. The first of these I repeat here:

(1) But Philip ceased to think of her a moment after he had settled down in his carriage. He thought only of the future. He had written to Mrs. Otter, the

[1] *Elements of Symbolic Logic*, Ch. 7, opening sentence (p. 251); here and elsewhere all references are to *the Free Press* paperback edition of 1966, which has the same pagination as the original version from 1947.

> *massière* to whom Hayward had given him an introduction, and had in his pocket an invitation to tea on the following day.
>
> (From: Somerset Maugham, *Of Human Bondage*)

I have underlined the Past Perfects of this passage. Each of the three Past Perfect clauses of the passage illustrates the two relevant points: (i) the information conveyed is presented as lying in the past of some past "Reference time", and (ii) this Reference time is provided by the surrounding context. For the first Past Perfect clause the Reference time is provided by the clause "he ceased to think of her"; for the other two it is provided by the sentence "He only thought of the future."[2]

Before we come to Reichenbach's generalization of these two points to other English tenses, let me round off my earlier remark on the novelty of his section on tense. First, there is the interest in the logical functioning of natural language as such. This of course is the concern of the entire chapter, and not something particular to the section on tense. But it deserves to be mentioned here, because so much of the logical climate of the day pointed, to the exent I can tell at this distance in time, away from such interests. The success of symbolic logic since the discovery of the predicate calculus by Frege (and Peirce), and the conviction that clinging too closely to the apparent structure of natural language sentences had been responsible for the failure to arrive at a satisfactory logic for centuries, seems to have led to a general sense that natural language was beyond logical analysis—that the carrying through of such an analysis would be very hard, if not impossible, and that in any case the enterprise would hardly be worth the effort. Evidently this is an attitude which Reichenbach did not share—that is clear from the entire chapter, which seems intended as a comprehensive survey of the different logical and semantic devices which natural languages employ. What is especially remarkable about the section on tense, however, goes beyond this. Not only do we find in this section the observation that tenses like the Past Perfect are logically complex in ways unfamiliar (I presume) from work in formal logic up to that time; even more importantly, it is also noted that they have an *anaphoric* dimension, in virtue of which they can establish semantic links between the sentences of which they are part and

---

[2] The text suggests that these two reference times more or less coincide; so, using the original reference time also for the second and third Past Perfect would have yielded essentially the same interpretation. Although I am assuming here that the Reference time for these Past Perfects is the time of Philip's thinking of the future, I actually do not think there are clear principles governing the interpretation of the Past Perfect which force us to take this as the Reference time.

other sentences which co-occur with these in a given discourse or text. It is especially because of this that Reichenbach's theory of tense has had such an impact on subsequent work, an impact that continues up to the present day.[3]

But let us return to the actual content of the section. After his observations on the Past Perfect, Reichenbach goes on to argue that what is comparatively easy to see in the case of this tense also holds true for the others: Each tense involves a "Reference time" as well as an "Event time"—the time at which, roughly speaking, the verb is said to hold of its arguments. And the tenses differ from each other in that they have different things to say about (i) the temporal relation between Event time E and Reference time R and (ii) the temporal relation between Reference time R and the utterance time, or "Speech time", S. Here is Reichenbach's table for the tenses of English:

	Relation betw. E and R	Relation betw. R and S
Past Perfect (*had slept*)	$<$	$<$
Simple Past (*slept*)	$=$	$<$
Future of the Past (*would sleep*)	$<$	$>$
Present Perfect (*has slept*)	$<$	$=$
Present (*sleeps*)	$=$	$=$
Simple Future (*will sleep*)	(a) $>$ (b) $=$	$=$ $>$
Future Perfect (*will have slept*)	$<$	$>$

It is remarkable how well this table has weathered the more than five decades since it was first conceived. Even today, after the very large amounts

[3] The section also differs from the other sections of the chapter in that there is no attempt to formalize the analysis which it informally describes. Indeed, formalization would have been impossible with the logical tools available in the mid forties; and, as we now know, much time had to pass and much preparatory work had to be done before such tools could be developed. The remarkable achievement of Reichenbach in this section lies in large part, it seems to me, therein that he was not deflected from the insights the section records by his realization that these insights could not be formalized with the means he had.

of work that have since been done on the temporal systems of English and many other languages, the table still stands largely unscathed. However, it raised, and still raises, the difficult question exactly what its role should be within a comprehensive, logically precise theory of the meaning of sentences and discourses in English and other natural languages. In fact, much of the work it has provoked concerns this question. For a theory in which the table could be integrated did not exist at the time, and it took many years to arrive at one.

3. Interlude: Montague and Prior

To appreciate some of the difficulties that one faces when trying to make Reichenbach's table part of a systematic account of the semantics of natural language, we must allow ourselves another foray into the history of the logical analysis of time and its role in language.[4] First a brief remark about Richard Montague's work on the logic and semantics of natural languages. There is a wide, perhaps universal consensus nowadays that Montague's work must be seen as the starting point of a new discipline, the one now usually referred to as "formal semantics of natural language"; in fact, there could hardly be a better proof of this than the widespread tendency to use the term "Montague Grammar" as an alternative denotation for this discipline. Montague accomplished what many of his colleagues at that time (the second half of the sixties) probably thought impossible—if, that is, they gave the problem any thought at all: To apply the model-theoretic methods that had been developed, largely by his teacher Alfred Tarski, for the artificial languages of formal logic effectively to the analysis of a significant fragment of some natural language; to specify, that is, the semantic values that the syntactic constituents of the sentences of such a fragment determine in any model, and to do this in such a way that comparatively simple, syntactically triggered rules for combining them into values for larger constituents assign the sentences the truth conditions which they intuitively seem to have (and thus to provide a formally precise characterization of entailment and synonymy relations between those sentences which are consistent with our intuitions).

[4] Again we touch upon a chapter in the history of logic which is intimately connected with the history of UCLA's Philosophy Department; lest you begin to think this is all a fawning ploy set up for the express purpose of this lecture, it really is not.

For what I want to say later on it is important to note that Montague's work on natural language semantics consists of two distinct achievements. First, it involved the delevopment of a new logic, known as Higher Order Intensional Logic. This formalism, which nowadays is standard fare for linguistics students who want to do work in formal semantics (and one would wish the same was true for philosophy students who want to work on problems relating to language and logic), constituted a major breakthrough at the time. It offered a solution to a number of awkward puzzles which had developed around the work of Carnap and Church on the logic of "Sinn" or intension, providing for the first time a formal logic for intensional contexts with which it was comparatively easy to work and which seemed to provide more or less what the analysis of natural language required.[5] Without HOIL only a small part of Montague's natural language semantics would have been possible.

Montague's second achievement had to do with the way in which he defined the semantic values of certain grammatical constituents of English sentences. The most striking and most often quoted example of this is his treatment of noun phrases—*Socrates, every man, the King of Denmark*, etc. By interpreting NPs across the board as being of the logical type of a quantifier Montague regained conformity between syntax and semantics precisely there where logicians since Frege had believed that the gulf between grammatical and logical form in natural language had been shown to be unbridgeable once and for all. The uniform treatment of all NPs as quantifiers is just one application of a method (sometimes referred to as "generalization to the worst case") of which Montague made use repeatedly and on which the success of his natural language semantics crucially depended.

It is hardly possible to overestimate Montague's achievements, both in the domain of formal logic and in that of the systematic analysis of English. And yet, as we will see, the architecture of his theory is still too narrow, too reminiscent of that of the syntax and model theory of classical predicate logic, to accommodate some of the most important and ubiquitous aspects of natural language meaning. In particular, it is an architecture that is fundamentally unsuited to the incorporation of Reichenbach's insights about tense.

---

[5] In the meantime we have come to realize that the use of intensional logic in the analysis of many "intensional" terms of natural language, among them most notably the propositional attitude verbs, is highly problematic, but that is another story, not to be told today.

The second development I want to mention temporally overlapped with the one just described, though its beginnings lie a few years farther back. I am referring to the invention of *tense logic*, in the form in which it was developed by, first and foremost, Arthur Prior.[6] When Prior, in the fall of 1965, joined the philosophy department of UCLA for a one semester visiting appointment, the syntax and proof theory of the system for which he is still best remembered (the so-called "*P*, *F*-calculus") were already in place. It was during that semester that he became acquainted with the model-theoretic semantics for a similar system that had been developed by Montague and his student Cocchiarella; and in subsequent work, starting that semester, the connections between the two approaches were cleared up. At the time these results led many of us to believe that the combination of Prior's results (Prior 1967) with those of Montague and Cocchiarella provided a powerful basis for the logical analysis of temporal reference of any sort, and for some years after that tense logic was seen by many as the tool to use in the analysis of tense and other devices for referring to time that are found in natural languages. In fact, it took a fair bit of time before we came to realize, after a series of unsuccessful attempts to make such applications work, that our optimistic ideas about how to crack the problem of temporal reference in natural language had been altogether on the wrong track. If one had paid closer attention to what Reichenbach had said twenty years before, this realization might have come a good deal earlier, for his analysis of the tenses is so very clearly at variance with some the general assumptions that are intrinsic to the tense-logical approach.[7]

The principal incompatibilities are these. Priorean tense logics are designed after the model of modal logic, with its sentential operators "it is necessary that" and "it is possible that". In analogy to these the (simple) past tense, for instance, is represented as the sentential operator *P*. According to the standard semantics for this operator (another assumption taken over from modal logic) this operator works as a quantifier—the truth conditions

[6] Witness the often used term "Priorean tense logic" to denote the still best-known system of tense logic, Prior's *P*, *F*-calculus as well as similarly structured systems.

[7] It may be questioned, though, whether this would have been in the interest of logic as a whole. Tense logics of the Priorean kind, some of them developed in the sincere belief that they would further our understanding of natural language, have become an important tool within computer science, where they have been used in the formalization of the theory of execution sequences, with applications in, among other things, program verification, certain types of theorem proving, etc. Without the reigning misconception about the relation between natural language and Prior-type tense logics these systems might not have been developed until much later.

of a formula *Pq* (with sentence letter *q*) are: *Pq* is true at a time *t* if there exists a time t' before *t* such that *q* is true at t'. Other tenses get analogous quantificational analyses. These analyses ignore the anaphoric dimension of the tenses to which Reichenbach had drawn attention.[8]

Some of the insights that came from the work on tense logic found their way into Montague Grammar.[9] However, the attention devoted to questions of temporal reference in the treatment of those fragments is rather "basic", certainly in comparison with the so much more detailed analyses it devotes to other aspects of meaning.[10]

4. Temporal Reference as Meaning-in-Context

What is needed, rather, if we want to do justice to those observations, is a systematic account of *meaning-in-context*. The basic implications of this desideratum are by now reasonably well understood. First and foremost, what is wanted is a formally precise notion of *context*, satisfying the following desiderata:

1. the context must be able to guide the interpretation of the anaphoric elements of any sentence *s*;

   and

2. the interpretation of *s* in the given context must then be able to trigger a suitable update of this context, so that the resulting update can serve as context for the interpretation of the sentence following *s*.

---

[8] From the perspective of natural language there is another fundamental design fault with the modal approach. This one has to do with the fact that tenses are formalized as modal *operators*. Such a formalization (i) falsely predicts the possibility of arbitrary iterations of tenses, with semantic effects that simply are not to be found in natural languages; and (ii) it forces not only the tenses but also other devices of temporal reference, such as temporal adverbs and prepositional phrases, into the mould of modal operators, and thereby imposes a scope order between any two of them. It is an old observation, which we will review later on (see Section 5.2), that often such scope orderings are precisely what is *not* wanted in an analysis of natural language. As a rule natural language tenses and adverbs in the same clause work in *conjunction*, not in series.

[9] See in particular 'The Proper Treatment of Quantification in Ordinary English' in: R. Montague, *Formal Philosophy: Selected Papers of Richard Montague*, ed. by R. Thomason (Montague 1974).

[10] In a way this is not surprising, given that Montague Grammar is unsuited to provide a faithful modeling of how the tenses of English and other natural languages actually work—and not because of some contingent features at its periphery, which could easily be changed, but because of the very character of its general architecture and the concept of semantics which lies behind it. For Montague Grammar—I noted this already—is a theory of the semantics of *sentences*, not of coherent discourse.

Semantic theories which involve such a two-way interaction between context and current sentence have been around for about twenty years. (The most common current name for the entire family of these theories is "Dynamic Semantics".[11]) All members of that family with which I am familiar are based on the principle that the context induced by a given piece of discourse represents the full content of that part of the discourse that has already been interpreted; and when a new sentence gets interpreted in such a context, the result of the interpretation is an update of the context in which the semantic contribution made by this new sentence is represented as well.

For certain purposes, however, it is not necessary to assume that the context captures the full content of what has been interpreted already. In particular, if we are concerned only with the anaphoric properties of temporal expressions, then a type of context will suffice in which only the temporal elements are represented which may be needed as antecedents of subsequent tenses and temporal adverbials. Thus, for our present purposes we can make do with contexts of the following sort:

(2) A *context* is a structure $< T, <, \subseteq, \supset\subset, \ldots, SAL >$, where

- (i) T is a set of "times" (intuitively intervals of time as well as instants)
- (ii) $<, \subseteq, \supset\subset, \ldots$ are relations of complete precedence, temporal inclusion, abutment and ...[12]
- (iii) *SAL* is a salience ranking among the elements of T (represented in the form of a stack)
- (iv) one of the elements of T is S (the Speech time)

---

[11] I am using "Dynamic Semantics" in a broad sense here, according to which it includes not only systems for which this term is sometimes employed exclusively (in particular the systems of Groenendijk and Stokhof known as "Dynamic Predicate logic" (Groenendijk & Stokhof 1991) and "Dynamic Montague Grammar" (Groenendijk & Stokhof 1990)) but also "File Change Semantics" (Heim 1982) and "Discourse Representation Theory" (Kamp 1981, Kamp & Reyle 1993). By now the family includes a good many more systems besides the four just mentioned, among them versions of Dynamic Semantics (in the narrow sense of the term) due to Dekker (1993), Muskens (1996), Visser & Vermeulen (1997), and also various extensions of DRT, such as the two level DRT-variant of van der Sandt (1992) and van der Sandt & Geurts (1991) (now the one most widely used), *lambda*-DRT (Pinkal 1996), U(nderspecified) DRT (Reyle 1993) and S(egmented) DRT (Asher 1993, 1996).

[12] The dots are to indicate that depending on what temporal expressions the theory means to account for, additional structural predicates may be needed to structure the set of context times. In particular, a systematic account of the meaning of temporal prepositional phrases involving "calendar terms"—like *in an hour, last week, next Wednesday, in the fall of 1997, on the first of November, at five to eleven, on Christmas eve*, to name but a few—the context times will have to be represented as satisfying certain such predicates too. For instance, it will be important that the context time t_1 is a calendar week, t_2 a calendar day, that t_3 is the day immediately following t_4, and so forth. I do not want to commit myself here to exactly what structural predicates should be used.

Interpretation of a tensed sentence relative to a context C of this kind will make use of, among other things, the interpretation instructions which Reichenbach's table specifies for its tense. As an example, suppose that the first two sentences of the passage from *Of Human Bondage* have given rise to the context

$$C =< \{S, t_1, t_2, t_3\}, t_2 < t_1 \leq t_3 < n, ..., \{t_3\} >,$$

where t_1 is the time of the ceasing to think, t_2 the time of the settling down and t_3 that of the thinking of the future. (The last component is to be understood as saying that there is only one salient time and that that is t_3). Interpretation of the clause "He had written to Mrs. Otter" now involves, in accordance with the Reichenbachian characterization of the Past Perfect, the choice of a "past" time from the context—that is: an element of T_C which precedes S—and then locates the event it relates to a time earlier than that time.[13] If we assume that the principle includes a default preference for the most salient time preceding S, then the choice will be t_3 and the interpretation of the clause that Philip wrote to Mrs. Otter at some time before t_3. Moreover, this new time—that at which the writing occurred—must be incorporated into the context so as to be available for the interpretation of the next clause. Calling that time t_4, we extend the context to a new context C' as follows:

$$C' =< \{S, t_1, t_2, t_3, t_4\}, t_2 < t_1 \leq t_3 < n, t_4 < t_3, ..., \{t_4, t_3\} >,$$

with the last component now signifying that t_4 has been put at the top of the salience order. We can exploit this new addition to the context in the interpretation of the next clause, "[the *massière*] to whom Hayward had given him an introduction", by choosing t_4 as Reference time for the Past Perfect of this clause. This interprets the event of Hayward giving Philip an

---

[13] I am assuming here a certain way of converting the entries of Reichenbach's table into rules for using and updating the context. There is a substantial body of work in which these rules have been spelled out explicitly. The extant implementations do not agree with each other on all details, and all of them involve certain decisions that could be questioned. (More about that later, see in particular the Appendix.) On the interpretation of Past Perfects like the ones in (1), however, all the explicit treatments I know of agree with the implementation of Reichenbach's entry of which I am making an informal use in this paper.

N.B. The implementations I am thinking of do not involve an explicit formalization of a notion of salience. That such a notion does play an important part in anaphora of any kind is undeniable, though how exactly it should be formalized is still a matter of much contention and uncertainty. I have operated here on the assumption that salience is just a matter of recency of introduction. But in general this is much too simple. Again, for a little more information the reader should consult the Appendix.

introduction to Mrs. Otter as preceding the time of writing. (It is hard to tell, in this case, whether the Reference time for this last Past Perfect is t_3 or t_4. In the latter case the processing of the tense itself would leave it open whether the introduction time is before, simultaneous with or after the time t_4. The conclusion that it preceded t_4 might nevertheless be forthcoming, however, but now as the result of a further inference, based on the contents of the two clauses and the reader's general world knowledge.)

5. What is Reference Time?

One of the questions which Reichenbach's proposal raises is: What exactly is a Reference time? What is the intuitive conception behind it? And how exactly does the notion work in the actual interpretation of tense (and other sentence constituents which contribute to the temporal dimension of interpretation)? In earlier work[14] I have criticized Reichenbach's concept of Reference time as one which is not able to sustain the multiplicity of demands placed upon it, and I proposed to split it into two notions, that of the *R*(*eference*) time and that of the *T*(*emporal*) *P*(*erspective*) time.

5.1. *Narrative Progression*

The reasoning which led to this counterproposal went as follows. On the one hand there is the role which Reichenbach's Reference time plays in the analysis of the Past Perfect. Here, the intuition seems to be that of a past time "from which the given information is viewed", in other words, a time identifying a view point or perspective. On the other hand there is the notion of Reference time as it has been used in accounts of the dynamic properties of narrative. It should be said right away that this use of the Reference time is not found in *Elements of Symbolic Logic*. But Reichenbach's rule for the English Simple Past, according to which for this tense Reference time and Event time coincide, has certainly encouraged this use of the term, and there is nothing in *Elements of Symbolic Logic* that suggests he would not have foreseen this additional use, or at any rate accepted it if the issue had been put to him. These accounts address a phenomenon which has in some form or other been a concern of students of discourse and literary texts for a very long time: The sentences which make up past tense narrative discourse

[14] Cf. Kamp & Rohrer (1984/86); Kamp & Reyle (1993). The criticism was mine.

can be distinguished into those which "drive the action forwards" and those which do not. In English, Simple Past tense sentences of non-stative verbs tend to do this, whereas Past Progressives and Simple Pasts of stative verbs almost invariably do not. The point is illustrated by the following pair

(3.i) Philip settled down in his carriage. He took a note pad from his briefcase and started writing.
(3.ii) Philip settled down in his carriage. He felt hot and tired.

In (3.i) the taking out of the pad and the beginning of the writing are understood as *following* the settling down, whereas in (3.ii) Philip's feeling hot and tired is naturally understood as being the case *while* he settled down. Usually this difference (cf. Partee (1973, 1984); Hinrichs (1986)) is described by saying that in (3.i) the textual transition from the first to the second sentence has the effect that the Reference time shifts forward, whereas in (3.ii) the effect is that the Reference time stays where it was.

In the literature which takes its clues from Reichenbach we find two ways of working out the details of this. The first strategy is to maintain that the Reference time introduced by the first sentence is the same as its Event time—as the time, in other words, at which Philip settled down; the second sentence is then interpreted as locating its Event time after the Reference time it finds in the context (and then incorporating this event time as a new Reference time into the context for subsequent use). The second account, favoured by Hinrichs and Partee themselves, has it that a non-stative sentence like the first sentence of (3.i) and (3.ii) leaves behind in the context a Reference time that is itself located (just) after the completion of the described event, i.e. after the settling down. This Reference time can then be identified with the Event time of the second sentence. This second sentence introduces a new Reference time which is distinct from its Event time and is located immediately after the completion of the event. Stative sentences, on the other hand, do not introduce new Reference times of their own account; they simply exploit the Reference times that are already there.

The Hinrichs-Partee strategy has the advantage that it can, without additional assumptions, account for cases like (4), where the state reported in the second sentence begins only after the event mentioned in the first sentence has taken place:

(4) Mary lit the lamp. In the murky light a person was visible lying on the bed in the far corner.

In the absence of special contextual information which might suggest otherwise, the natural understanding of this passage is that the murky light comes from the lamp, and thus that the state of the person being visible in it came

into effect as a result of the lamp being lit—in other words, it did not yet obtain when the lighting process started. By identifying the time of the state with the Reference time previously introduced by the first sentence of (4), which, according to the second strategy, is located just after the moment at which the lamp had been lit, we get just the interpretation that it seems is wanted here.

For cases like (3.ii), on the other hand, this strategy has to say something more, viz. that it is a natural inference that the state described in the second sentence obtained not only at some time immediately after the end of the event mentioned in the first sentence, but already at the time when this event was in progress. This inference has the plausible justification that the event in question could not have been responsible for bringing the state about; so if the state obtained immediately after the event, then presumably it should already have been there before.

Examples like (4), on the other hand, are a problem for the first strategy, which identifies the Reference time with the Event time. It is forced to say about such cases that here, because the first sentence introduces not just an event but one which effects a relevant change, the first sentence introduces an additional time, viz. that of this resulting effect state. This time is then available as an alternative possible Reference time for the next sentence. The choice between this Reference time and the one which coincides with the first event is then to be seen as one that is based on inference; and the considerations which justify the inference needed in the second strategy for dealing with a discourse like (3.ii) now have to be invoked in making this choice.

From today's vantage point there is not much to choose between these two strategies. For we now know that the general problem of interpreting sequences of past tense sentences is much more complicated than either strategy acknowledges. The three cases illustrated by (3.i), (3.ii) and (4) do not exhaust the full repertoire of possibilities by a long shot; and the choice from this wider repertoire that each interpretation must make is influenced by a number of factors other than the morphology of tense. (Especially important are the rhetorical relations in which each new sentence must be seen as related to one or more other sentences of the discourse, in order that the discourse as a whole can be understood as coherent.) When the effects of these other factors are integrated into the theory of interpretation, the differences between the two strategies we have just mentioned get lost in the general overhaul. Within the scope of this lecture it is not possible to go into the details of these further complications. In view of the particular role,

however, that these arguments are playing in the general argumentation structure of this lecture, this is not a grievous omission.[15]

There seems to be no obvious sense in which narrative progression can be regarded as a matter of perspective, and the role that Reference time is made to play in the accounts of narrative progression that we have just touched upon appears to be of a very different kind than the one it plays in Reichenbach's account of the Past Perfect. This second role, we already noticed, has the character of identifying something like a *vantage point*. And this, it seemed to me, is something altogether different from the other role, that of recording, in some way or other, to what point the narrative, or some sub-plot of it, has progressed. In fact, the two roles seem to be so different that it would be surprising if one and the same concept could consistently play both roles, and in particular that, in cases where that might be required, it could play both at the same time. And I thought I had found a conclusive proof that it couldn't, in the form of a type of discourse in which both roles are apparently needed simultaneously. This type of discourse is known as an *extended flashback*. In extended flashbacks a sentence in the Simple Past is followed by a number of sentences in the Past Perfect, which may form "sub-discourses", in which we can observe essentially the same dynamic narrative effects that are found in mini-discourses like (3.i), (3.ii) and (4). (5) is an example.

(5) Shortly before noon we reached the top. We *had got up* at six, *had had* a light but sustaining breakfast, and *had gone* on our way. We *had reached* the last cabin at about eight thirty. There we *had taken* a brief rest and then we *had set off* again.

In (5) the first sentence introduces a time which can then be exploited as Reference time in the interpretation of all the Past Perfect clauses which follow. But these Past Perfect clauses form what presents itself as a little narrative in its own right, concerning a number of events each one of which we perceive as coming after the last one, irrespective of whether the clause contains a temporal adverbial (such as "at about eight thirty") to make the progress in time explicit. To make sense of this intuition along the lines already drawn, we would, to interpret each of the Past Perfect clauses in (5) except for the first, seem to need two kinds of Reference time. Each of those Past Perfects requires a Reference time according to the rule associated with the Past Perfect entry in Reichenbach's table; and since this Reference time has to be situated between the Event time and the Speech time, the only

[15] See in particular the work of Asher (1993, 1996) and Lascarides & Asher (1993).

possible choice for it is the time of the arrival at the top of the mountain. But in addition we also seem to need a Reference time in the sense of our discussion of (3) and (4) for the purpose of situating the event described by the clause in relation to the Past Perfect clause preceding it. This Reference time must precede the event of the new clause and thus be distinct from the first one.

Examples of this sort led me to distinguish between two types of "Reference time" that are needed in the interpretation of tensed clauses, the *R*(*eference*) time and the *T*(*emporal*) *P*(*erspective*) *time*. *TP*-time was to play the role of Reichenbach's Reference time in the sense of the entry of the Past Perfect and the other tenses and *R*-time was to act as the Reference time in the account of narrative progression. Each of these is to be recovered from the context in which the tensed clause appears. (If a viable *TP*-time or *R*-time is not forthcoming from the context as is, then it will have to be, if at all possible, accommodated.) The rules for recovering (and accommodating) *TP*-time and *R*-time are not quite the same; but the details do not really matter here and there is no need to go into them.

5.2. *Temporal Perspective and Temporal Adverbs*

There was an additional type of consideration which seemed to provide independent support for this distinction. It concerns the interaction between tenses and certain temporal adverbs. An example is the interaction between the past tenses and the adverb *now*. It is well-known that in narrative prose the adverb *now*, which has a primary use as a means of (normally emphatic or contrastive) reference to the utterance time, can shift its reference to some time in the past, but that this is possible only under certain conditions. We find the same with the counterparts of *now* in other languages, among them the French *maintenant*. Uses of *maintenant* and *now* which refer to past times are known to be restricted in roughly similar ways. But in French, with its two tenses corresponding to the English Simple Past, the Passé Simple and the Imparfait, the restrictions manifest themselves in an unmistakably "grammatical" way: Past-directed references of *maintenant* can co-occur with the Imparfait, but not with the PS.

Before we look at the explanation of these facts, there is a matter which I mentioned briefly in footnote 8 of Section 3, but to which I devote a few more words at this point. This is the way in which tenses and temporal adverbials interact in general. In footnote 8 I noted that one of the problems of the Priorean approach to temporal reference is that all constituents contributing to the location of the described event or state in time are to be

treated as operators. This entails that any two such constituents occurring in the same sentence (so, in particular, any pair consisting of tense and a temporal adverb belonging to the same clause) must stand in a definite scope relation. That this leads to irritating problems, and sometimes to the wrong truth conditions, has long been known. As an example consider the following sentence:

(6) Bill will arrive in two days.

In a Priorean framework this sentence can be represented either as

(6.i) IN TWO DAYS (FUT (Bill arrive))

or as

(6.ii) FUT (IN TWO DAYS (Bill arrive))

If FUT is the future tense operator F of Prior's (P, F)-calculus and IN TWO DAYS(p) is interpreted as true at t iff p is true at some time within the second day after the day of t, then neither of these formulas gives us truth conditions for (6) that are even approximately right. (Both allow Bill's arrival to take place arbitrarily far in the future.) To be precise, the truth conditions of (6.i) are those in (6.i')

(6.i') $(\exists d)(d = d(n) + 2 \,\&\, (\exists t)(t \in d \,\&\, (\exists t')(t' > t \,\&\, \text{Bill arrive at } t')))$

and those of (6.ii) are the ones in (6.ii')

(6.ii') $(\exists t')(t' > n \,\&\, (\exists d)(d = d(t') + 2 \,\&\, (\exists t)(t \in d \,\&\, \text{Bill arrive at } t)))$

Actually it is not all that easy to find examples to just this effect, i.e. where both scope orderings are intuitively meaningful, but both lead to the wrong truth conditions. But that does not mean that a Priorean analysis will yield the wrong result only in exceptional cases. The more common situation, illustrated by a sentence like

(7) Bill arrived yesterday.

is that the adverb has indexical properties which put it, for purposes of interpretation, outside the scope of the original Priorean theory to begin with. This limitation can be formally overcome by the device of multiple indexing, which enables the truth definition to retain access to, for instance, the utterance time when it is applied to subformulas within the scope of other temporal operators. (In the original "single-index" semantics for the simple Priorean logics this access is lost.) This method makes it possible to fix, through liberal multiplication of indices, any number of such scope problems. But such repairs tend to be progressively ad hoc. In fact, the

history of this particular branch of the semantics of temporal reference has made plain that the method is devoid of any sound linguistic foundation; the formalisms to which it has given rise got ever farther removed from the way in which natural language actually seemed to function, instead of getting nearer to it.

The right and simple answer to the scope problems which the tense-logical approach generates is that devices of temporal reference just do not interact in the manner of logical operators. Tenses and temporal adverbs from the same natural language clause do not take scope one over the other. Rather, they act in conjunction, like constraints from logic programming or conditions in systems based on unification. In this respect the interaction of tense and temporal adverb is not fundamentally different from that of adjective and noun in a simple noun phrase like *a blue car*, which describes an object which is (i) a car and (ii) blue. Conjunction is one, and arguably the most basic, interpretation of the syntactic operation of adjunction.

This is another case where paying closer attention to what Reichenbach had had to say nearly two decades earlier would have made it easier to see that this wasn't the right way to think about temporal reference in natural language. For in ESL he noted that temporal adverbs acted as predicates of times. More precisely, he claimed that a temporal adverb occurring in a tensed clause had to be treated as a predicate of the Reference time needed in the interpretation of the tense of the clause. In last analysis this claim does not seem to be quite correct: Temporal adverbs act as predicates of the Event time, not the Reference time.[16] But the larger issue here is whether such adverbs are operators or predicates. And on that issue Reichenbach was clearly right.

Let us, after this excursus, return to the distributional possibilities of past-directed *maintenant*. We can account for the observations we have made on the basis of the following four assumptions:

(i) temporal locating adverbs—these include terms denoting dates and terms for clock times (*Wednesday, the third of March, in 1999, at five past eleven*, etc.) as well as terms like *in an hour, tomorrow, the following Sunday, now* and so

[16] One argument against the claim that a temporal adverbial always acts as a predicate of the Reference time is that in a discourse like "We arrived at 11.30. We had left the house at 6.15." the second sentence is naturally understood as saying that the leaving took place at 6.15, while the Reference time relevant to the interpretation of this sentence is, by Reichenbach's own lights, the time of arrival, that is 11.30. This does show that such adverbials always act as predicates of the event time. There is plenty of further evidence to support the general claim.

forth—always act as characterizations of the Event time. (As observed in the preceding paragraph.)

(ii) *now* always refers to the *TP* time;

(iii) the Imparfait allows for an interpretation according to which the *TP* time lies in the past of the Speech time and the Event time coincides with the *TP* time (this is formally just the entry for the English Simple Past in Reichenbach's table);

(iv) the PS cannot be interpreted in this way, but only permits the interpretation according to which the *TP* time coincides with the Speech time and the Event time lies before the *TP* time.

These assumptions entail that it is possible to combine *now* with the Imparfait, but not with the PS. For in order that *now* combines with either of these tenses Event time and TP time must coincide (assumptions (i) and (ii)); and this possibility is allowed by the Imparfait (assumption (iii)) but excluded by the PS (assumption (iv)).

In isolation this may not seem much of an explanation. But the assumptions (i), (iii) and (iv) do a good job accounting for a number of similar interactions (when combined with suitable analogues to (ii) for the interacting adverbs). And if this and similar explanations are indeed correct, then they provide additional evidence for the thesis that *TP* time and *R* time must be distinguished. For French PS sentences will, when they occur as part of naratives, normally require, just like English non-stative sentences in the Simple Past, past *R*-times to locate the described event, and these *R*-times will have to be distinct from the *TP*-time, which is assumed to coincide with the Speech time.

6. Must we Distinguish Between R-Time and TP-Time?

For many years I supposed that these arguments provided, when taken together, powerful evidence for the separation of Reference time into *R*-time and *TP*-time. Recently, however, the package has been challenged by an alternative account of these phenomena, which has made me reconsider this opinion; and as this lecture is first and foremost about the influence that Reichenbach's theory has had on subsequent theorizing, it would seem proper to include a brief presentation of this alternative (or, more accurately, of my understanding of it). Also, this alternative account will naturally lead us to our next topic.

One arguably unsatisfactory feature of the account just sketched is that it treats the different past tenses involved in the argument we have considered as showing distinctive behaviour with regard to the choice of the *TP*-time.

This may not appear as objectionable for a language like French, where IMP, PS and PC are morphologically distinct tenses. But it seems awkward for a language like English, where we have to postulate different "Reichenbach entries" for the Simple Past tenses of stative verbs and for those of non-stative verbs. Here one cannot help feeling that a needed distinction is being drawn in the wrong place.

An alternative account, urged upon me by Antje Roßdeutscher,[17] developed out of a critique of this unsatisfactory aspect of the account; more specifically, Roßdeutscher objects to the explanation I just sketched of the distributive possibilities of *now/maintenant*, in which this differentiation between the Simple pasts of different verbs was an essential ingredient. Her own account makes use of two notions, which did not play an explicit part in the one which it is meant to replace. The first of these I will refer as *View Point Aspect*.[18] The second notion is that of a *cut*, a conceptual operation which is a necessary part of the interpretation of some (and perhaps all) tensed sentences and clauses. The View Point Aspect of a clause or sentence is intuitively a matter of whether the information it carries is presented as seen "from the inside"—so-called *I*(*nternal*) *V*(*iew*) *P*(*oint*) *A*(*spect*)—or as seen from an external perspective—the case of *E*(*xternal*) *V*(*iew*) *P*(*oint*) *A*(*spect*). EVPA is to be divided further into *R*(*etrospective*)*VPA*, in which the presented information is viewed with hindsight, and *P*(*rospective*)*VPA*, where the information is viewed as pertaining to the future. For the most part in this lecture we will be concerned with the RVPA variety of EVPA.

The means of marking VPA vary from language to language. In English, marking results from a combination of lexical information—verbs are lexically classified as *stative* or *non-stative* (more about this below), and in addition, as first explicitly recognized and analyzed in Verkuyl (1972), there are rules for determining the aspectual class of complex VPs and, eventually, of full "tenseless" clauses from the lexical classification of the main verb and relevant properties of the argument phrases, complements and adjuncts which are combined with it to form the VP or clause. This determines the VPA of "simply tensed" uses of such clauses (i.e. uses in which the tense form is not a progressive): they have IVPA when the tenseless clause is stative, EVPA when it is non-stative. The aspect of a tenseless clause also

[17] Roßdeutscher, p.c.; see also Roßdeutscher (2000).

[18] Emphasis on the importance of the distinction between View Point Aspect and another notion of Aspect, which I will call *Objective Aspect*, as well as extensive discussion of the difference between the two notions and the relations between them, are prominent in the work of Carlota Smith. See in particular Smith (1991).

determines whether the clause allows for the Progressive—it does if and only if it is non-stative. And in that case the application of the Progressive always leads to IVPA. In French IMP always conveys IVPA, whereas PS (as well as that use of the PC which has replaced the PS in spoken and in certain kinds of written contemporary language) determine the VPA as External.[19] This statement I now see as too strong. There are cases involving the Past Progressive in English which have External rather than Internal View Point; and the same appears to be true fort he IMP in French. The future tenses and the present tense in French, on the other hand, do not determine VPA either way; however, normal[20] uses of the present tense impose IVPA, and with future tense uses of stative clauses there also is a strong tendency to interpret them as having IVPA. In languages like German or Dutch (my own) no tense form fixes VPA. Here the assessment of VPA is always left to other factors. In the Slavonic languages, the study of which was the original source of "Aspect" as a concept used in theoretical linguistics, the story is a different one yet again.[21] And so on.

Cut enters into the interpretation of tensed clauses in the following way. It is part of the interpretation of a tensed clause that a certain temporal period that includes the eventuality[22] which the clause describes comes with a "cut". Where this cut is situated in relation to the eventuality itself depends on the VPA of the given clause. When the clause has IVPA, the cut must fall within the described eventuality itself, in case it has EVPA the cut lies to the right of it, cutting the result state to which the eventuality has given rise.

This accounts for the observations about the distribution of *now* and *maintenant* if we assume that Reichenbach's entry for the English Simple Past extends to all the "simple past tenses" of English and French, i.e. that it is the entry for both the Simple Past and the Past Progressive of English and also for both the IMP and the PS of French. For the question whether *now/maintenant* is admissible has now turned into the question whether the

---

[19] This statement I now see as too strong. There are cases involving the Past Progressive in English which have External rather than Internal View Point; and the same appears to be true for the IMP in French.

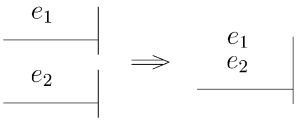
[20] By "normal" uses of the Present Tense I mean those where the speaker describes what is currently the case. Excluded are (i) the future-oriented use of the Present known as the "time table" use; (ii) the "reportive" use, as one gets it in live reports of athletic events; (iii) the "historical present", a past-directed use of the Present, used as a stylistic device to convey greater directness or vividness in narrations of past events. For more details see footnote 25.

[21] Especially interesting in this regard are Bulgarian and Macedonian, with their tense-and-aspect systems that contain elements inherited from their Slavic ancestry on the one hand and from a heavy influence of Greek on the other.

[22] I follow the by now standard practice of using the term "eventuality" as a cover term for events, process and states.

Reference time qualifies as Event time, that is, whether it is a time at which the described eventuality occurs, obtains or is going on. This will depend on whether the cut places the reference time within the eventuality or outside it and that, we stipulated, depends on the VPA of the sentence. And given our stipulations as to how VPA is determined in the two languages, it follows that *now* is possible with Simple Past stative and Past Progressive non-stative clauses, but not with Simple Pasts of non-stative clauses, and that in French past-directed *maintenant* goes with the IMP but not with PS or PC.

It should be added that Roßdeutscher's cut concept does not, in and of itself, solve all problems about temporal reference which we have so far encountered. This remains true even after we adopt a couple of further assumptions which Roßdeutscher assumes in connection with this concept. These assumptions are most easily explained for the case of two successive past tense sentences S_1 and S_2, with the first one providing the context for the interpretation of the second. In fact, let us confine the case even further by assuming that both sentences have External VPA and that the first is in the Simple Past. In that case the first sentence provides, according to these additional assumptions, a context of interpretation consisting of the event e_1 it describes together with an (external) cut—a time lying within the result state produced by, and thus located to the right of, e_1. Interpretation of the second sentence S_2, relative to the given context, is a two-step process, which involves, first, a preliminary interpretation of S_2 on its own and, second, an integration of this interpretation into the given context. In case S_2 is also in the Simple Past, its preliminary interpretation has the same form as the one we just described for S_1. In this case (as in others) integration of the preliminary interpretation within the context is assumed to involve aligning its cut with that of the context. This however leaves the temporal relation between e_1 and e_2 still undecided, as shown in the diagram below:



Roßdeutscher has argued that this is as it should be, and that the temporal relation which we perceive to hold between e_1 and e_2 is to be accounted for on the strength of additional properties of the combination of S_1 and S_2. Her arguments are, it seems to me, plausible enough and they are in line with more widely accepted views on the nature of temporal relations in discourse. Alas, by pursuing this matter further with the degree of detail

which it would require I would prevent myself from addressing a number of other issues which I see as more pressing in the context of this lecture.[23]

The few problems that we have looked at in our discussion of the two accounts constitute—this ought to be stressed—but a tiny sample from a large empirical domain. On the basis of such a limited discussion, a choice between the account to which I did for so many years subscribe and Roßdeutscher's alternative could hardly be perceived as definitive. Nevertheless, on the strength of what the two accounts have to say about these few problems (which seemed to me to provide adequate support for my own account at the time when I first proposed it), there would seem to be a strong presumption that Roßdeutscher's account should be preferred. And with that account we are back to Reichenbach's old analysis of the English Simple Past.

Roßdeutscher's account, then, is, among other things, a vindication of Reichenbach's table.[24] But it raises, through the new notions of which it makes use—View Point Aspect and Cut—a number of questions of its own. In particular, we need a clarification of the notion of View Point Aspect. Such a clarification touches upon two concepts about which many things (and in fact: many different things) have been said in the literature, *Aspect*—as we find it e.g. in the often encountered phrase "(the theory of) tense and aspect"—and *Perspective*. It is with these two notions that the remaining sections of this lecture will be concerned. Here again our discussion can only be schematic, and moreover, we will have to confine ourselves to just

---

[23] This and a number of other issues connected with Roßdeutscher's proposal are discussed at length in the Appendix.

[24] In fact, there is one other point where I have, in what has been said up to now, deviated from Reichenbach's principles. Reichenbach maintains that locating adverbs like *two hours ago* or *on the first of November*, etc. always concern the Reference time. In contrast I have maintained (see (iv) above) that they always concern the Event time. Only some adverbs, like e.g. *now*, are Reference time oriented in addition, but this is then a special property of the particular adverb, having to do with its "indexical" character—see below. This is a difference of opinion with Reichenbach on which I continue to stand firm. One argument in favour of the position I wish to uphold is that with Past Perfect sentences the locating adverb often goes with the Event time, not the Reference time, cf. the first and fourth Past Perfect clauses of (5). There are admittedly also cases where the adverb goes with the Reichenbachian Reference time of the Past Perfect sentence, e.g. in "At seven I (already) had dressed, breakfasted and answered my mail." On my story these are cases where the Past Perfect is literally what the word says, viz. the past tense of the perfect, an aspectual operator which turns an event verb into the description of a corresponding result state. For such Past Perfects the "event" is the result state, so Event time and Reference time coincide, as with the Simple Past, and it may look once again as if it is the Reference time to which the adverbial *at seven* applies. For details see Kamp & Reyle (1993).

some out of a larger number of issues. We will begin by saying something about perspective that is relevant to View Point Aspect, then give a very quick summary of the notion of *Objective Aspect*[25] as I understand that notion and add a few words on the relationship between it and that of View Point Aspect. In the final section we will look at another dimension to the notion of perspective, that of perspectival shift. As we could see already to a certain extent in Section 5.2, this topic is closely bound up with the interaction between tenses and certain temporal adverbials, and—this will be the main concern of that section—with the bearing perspectival shift has on the concept of indexicality as we know it from the philosophical literature, principally through the work of Kaplan (1989).

7. Aspect

7.1. *Internal vs. External View Point*

Information can be seen, or presented, from more than one perspective. In particular, it can be presented from an external perspective or from an internal one. First some observations about the internal perspective.

The various things I can think about the world in which I am, naturally divide into three kinds: thoughts about what is irrevocably past, thoughts about what is still to come and thoughts about what is the case, or is going on, right now. Consider thoughts of the third category, thoughts about what is going on at present. Some of these thoughts concern prevailing conditions: it is 73°F in the room in which I am sitting (as I just saw on the thermometer near the door), I am hungry, across the road there is a tree I can see as well as a tall brick building, and so on; some of these states are more permanent than others, but as far as I can tell, they are all obtaining right now. Also, certain things are happening: a bird is singing outside my window, in the bathroom water is dripping from the tap. And finally there are events in progress: they are putting up a new building in a vacant lot across the street, I am writing this paper, the cleaning lady is doing the bath room, in Rambouillet a bewildering assortment of diplomats are negotiating an agreement for the Kosovo. All of these "eventualities" are either states of affairs or processes. Of the processes some (the last four) are activities which have a goal, a natural conclusion which will be reached if all goes according to plan. But there is no knowing whether the conclusion will in

[25] See footnote 18.

fact be reached. From my internal, "present tense" View Point it cannot be more than conjecture or presumption that they will. All that I can "see" right now is that a certain activity is going on—in this regard, the building across the road is like the dripping of the tap.

In English this Internal View Point is reflected by the grammatical rule that goings on must be described using the Present Progressive, rather than the Simple Present. This principle holds for all "non-stative" verbs, verbs that can be used to describe processes (whether goal-directed or not); only stative verbs, verbs that describe states such as that of it being 73°F or of my being hungry, are, in "normal" uses of the present tense, felicitous when appearing in the Simple Present.[26] The English Progressive is thus a linguistic device for conveying that what is being described is presented from an Internal View Point; it is a means of focussing on the process (which may culminate in the intended result when all goes well, but need not) rather than on the entire, successfully completed event as a whole, with culmination included. Other languages—the Romance and Germanic ones among them—are less particular in this respect. They permit the use of the Simple Present tense with non-stative as well as with stative verbs. But the interpretation of present tense sentences in these languages is nevertheless the same as that of Present Progressive sentences in English. In "normal" uses of the present tense in those languages, the verbal predicate gets "coerced", as the current terminology has it, into an interpretation which the English Progressive makes formally explicit. Irrespective of what individual language we look at, what is presented from the Internal View Point is always a process or state.[27] So much for the Internal View Point.

---

[26] There are certain uses of the English Simple Present which admit non-stative as well as stative verbs. One is the future-oriented "time table" use, exemplified by sentences like "The train leaves at seven.", in which the present tense is used to say something about some time in the future. A second one is the so-called "reportive present", found in on-the-spot "real time" accounts of ongoing events, such as e.g., a soccer match. ("Moore passes to Jones. Jones moves quickly to the center, he gets past Fletcher, he gets past Carswell, he shoots. Bingo!") A third use, related to the second, is the "historical present", known to many of us from that perennial of high school Latin, Caesar's *De Bello Gallico*. A fourth one, also related to the second is given by the Simple Present uses of perception verbs in sentences like "I see an owl sitting on the top of the chimney." I am using "normal" use of the present tense in a sense in which it excludes each of these cases. "Normal" uses of the present tense are those in which one 'now' is saying something about what is going on at the time of speech, but where the time of speech is assumed to remain the same throughout the discourse, rather than moving along as one sentential utterance follows another.

[27] Indeed, some have argued (see Vlach 1981) that the Progressive always has the effect

The External View Point comes in two forms: Viewing things from a time before they take place—the *Prospective* View Point—and viewing from from a time afterwards—the *Retrospective* View Point. The conceptual difference between these two view points has also found its reflection at the level of linguistic form, it shows up for instance in the requirement, in English and many other languages, that prospective information must be marked by the use of a tense form that is morphologically future—and this irrespective of whether the time of the described event or state is before or after the time of speech. Thus (8.i), uttered on the 15th of February, is in accordance with the rules of grammar, whereas (8.ii) is not:

(8.i) Exactly two weeks ago Mary expected she would be in Paris by November 7th.
(8.ii) Exactly two weeks ago Mary expected she was in Paris by November 7th.

View Point Aspect and the relations that are constitutive of Reichenbach's table (that between R and S and that between R and E) are independent of each other. But they do interact; and the interactions that matter in the present discussion are those which occur when the tense of the sentence is one according to which the relation between Event time and Reference time is "$E = R$". For then, as we have seen, the Reference time is to be used as cut, and VPA determines where the cut is to end up.

We have introduced VPA as a notion which concerns the angle from which information is viewed. Thus it has something to do with perspective—with the perspective from which the presented information is or can be viewed. But as we will see in the final section it is a perspectival concept only in a very weak sense, and the kind of perspective it identifies is by no means the only one that is linguistically relevant.

View Point Aspect is defined as concerning the way in which information is presented or viewed. As such it belongs at the level of conceptualization of what goes on, not to that of what actually goes on as such. Nevertheless, as the above discussion revealed, the distinction between Internal and External View Point also has a bearing on what it is that can be presented in either way. Specifically, the IVPA limits the kinds of eventualities that can be presented from this View Point; they must be either processes or states. This is the "objective correlate" of the IVPA-EVPA distinction, and it is in this way that View Point Aspect is related to the complex notion of objective Aspect.

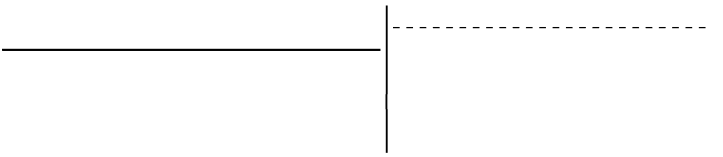
---

of turning a verbal predicate into a predicate of states. If we accept this analysis, and take English as our clue, then everything that is viewed "from within" has a state-like character.

7.2. *Objective Aspect*

To explain what we understand by Objective Aspect it is best to start with an example. Following Moens & Steedman (1988), we can represent the event structure described by a typical telic verb, like *paint* or *eat*, as we find them in "Mary painted the cupboard." or "Fred ate his meal." in the following schematic way:

(9)



According to this schema, known as the "(aspectual) nucleus", the various uses of these verbs will involve one or more of the following components:

(i) the activity leading up to the intended result (preparatory phase)
(ii) the reaching of that intended result (culmination point)
(iii) the state that is brought about by reaching the culmination (result state)

The Objective Aspect of a given verb phrase which contains such a verb as main verb concerns the question which part or parts of the nucleus it describes. The simple sentences just quoted differ in this regard from progressive sentences such as "Mary was painting the cupboard." or sentences in the perfect such as "Fred has eaten his meal." The first of these only involves the preparatory phase, while the second only concerns the result state. The Progressive and the Perfect can thus be seen as *aspectual operators*; they turn the verbs they are applied to into new verb forms, which describe just those parts of the nucleus that the operator selects.

Aspectual operators thus create out of given verb forms that differ from each other with regard to the parts they select from the nucleus. But there is more variety than just that which is produced by these operators when they are applied to verbs like *paint* or *eat*. For lexical verbs also differ from each other in what we might call their *basic nuclear properties*. These differences manifest themselves in two different ways. First, verbs may differ with regard to the parts they use of the nucleus given in (9) when they occur in their "basic" forms (such as, in English, verbs in the Simple Past). This is the difference between the so-called *accomplishment verbs*, illustrated by *paint* and (transitive) *eat*, and the achievement verbs such as *arrive* or *reach* (*the*

top of the mountain). The latter describe, when used in their basic capacity, only the culmination part of (9), whereas the former describe, as we saw, culmination part together with preparatory phase. Second, there are verbs which have a different type of "nucleus" to begin with. Among them we find the *activity verbs*, such as *sleep* or *walk*, whose nucleus is best assumed to consist of a single stretch of activity, and the stative verbs, such as *love* or *have*, whose nucleus should be thought of as just a stretch of time during which the state obtains. The aspectual operators Progressive and Perfect are not always applicable to such verbs (more precisely, progressives of stative verbs aren't really possible; when they occur, they seem to require a reinterpretation of the underlying verb as describing some kind of activity); or else the effect may be somewhat different from what it is with the verbs so far considered.

We thus arrive at the following picture. The verbs of English (and other natural languages) each belong to one of a number of different aspectual classes—the so-called *Aktionsarten*—according to the type of nucleus they determine and those parts of those nuclei that are included in the denotations of their basic forms. In addition there are aspectual operators, which can be applied to some or all of the Aktionsarten and whose application produces verb forms that denote complexes drawn from the associated nucleus; usually these are different from the complexes denoted by the basic verb forms.[28]

But this is still not the end of the story. For most (and perhaps all) natural languages have means for turning verbs of one Aktionsart into verbs belonging to another, often with a systematically related meaning. Well-known is the formation of verbs of the "second aspectual class" through prefixation in the Slavonic languages[29] or the Hif'il system in Hebrew and its analogues in the other Semitic languages.[30]

---

[28] For English, a delicate case, which I will not pursue here, are the progressive and non-progressive uses of activity verbs. See Kamp & Reyle (1993), Ch. 5.2 for some discussion.

[29] Creation of new verbs through semantically systematic and often productive prefixation operations is found in many languages, among them those spoken in eastern and central Europe. (This is a phenomenon that seems to cross-cut the principles of language typology; for instance we find it in the Slavonic languages, in Hungarian and, albeit somewhat less systematically, in the Germanic ones.)

[30] Many of these kinds of verb forming operators carry us outside the domain of the purely aspectual, as the prefix will alter the meaning of the verb to which it applies also in other than purely aspectual ways. Here we touch upon a vast area of linguistic research, belonging largely within the domain of lexical semantics, very little of which has to my knowledge been systematically explored (at least within the setting of a logically informed methodology which I am presupposing).

It should be clear from what I have said about aspect that Objective Aspect and View Point Aspect are, conceptually speaking, very different notions; one has to do with the event-like structures that various verbs or verb phrases describe, the other with the way in which information is viewed. Yet there is nevertheless a close connection between the two, for the Internal View Point can "see" only event structures of a certain kind. As a consequence, if a clause or verb phrase is to qualify as a linguistic representation of the Internal View Point, then it must satisfy the corresponding constraints on its Objective Aspect.

8. Perspectival Shifts and Indexicality

In this last section of this paper we return to the notion of perspective. We already noted that certain temporal adverbs—our only examples so far were *now* and its French counterpart *maintenant*—impose restrictions on the tenses with which they can be combined. These observations have implications for the concept of *indexicality*, as it has been defined and treated in the philosophical and linguistic literature; for *now* and its counterparts have generally been portrayed as "indexicals" par excellence. In fact, the observations we have already made about past-directed *now*, and some others we will get to presently, appear a challenge to the most popular and most elegant account of indexicality, that which is part of Kaplan's "three-level" theory of meaning, with its levels of extension, intension and character.[31]

Before we proceed to a general reflection on the implications of such facts for the three-level theory we must first look at a few more data. Our earlier observations about *maintenant* and *now* suggest the following gloss: In order that these adverbs can be used to refer to some time in the past, there must be some kind of "perspectival shift" from the utterance time—which plays the general role of a kind of "default perspective time" to that earlier time to which the adverb refers. (At a minimum, what happens in such cases according to the account we have sketched is that this past time is the Reference time and that the sentence in which the adverb occurs has Internal VPA; suggesting that the information it carries is seen from the vantage point of the past Reference time.) *Now* and *maintenant* need such "perspectival shifts" to earlier times in order to be able to refer to those times, and in this they are not alone.

---

[31] See in particular Kaplan's "Demonstratives", which first appeared in print in Almog et al. (1989) (eds.) *Themes from Kaplan*.

In fact, the phenomena pertaining to perspectival shift are more complicated than our observations so far could reveal. A look at the some other "shift-requiring" adverbials will make this clear. In French a temporal prepositional phrase of the form "*dans* + NP"—e.g. *dans deux jours* (*in two days*)—must refer to the second day after the day of the Reference time, whether that Reference time be the utterance time or some other time, something illustrated by the following pair of sentences.[32]

(10.i) Mercredi, le premier Septembre Jean arriva à Paris.
✓ Dans deux jours il partirait pour Berlin.

(Wednesday, the first of September Jean arrived in Paris. In two days he would leave for Berlin.)

(10.ii) Mercredi, le premier Septembre Jean arriva à Paris.
*Dans deux jours il partit pour Berlin.

(Wednesday, the first of September Jean arrived in Paris. In two days he left for Berlin.[33])

In this respect *dans deux jours* differs from both *deux jours après* ("two days after") and *deux jours plus tard* ("two days later"), which go happily with the tenses of both (10.i) and (10.ii):

(11.i) Mercredi, le premier Septembre Jean arriva à Paris.
✓ Deux jours après/plus tard il partirait pour Berlin.

(Wednesday, the first of September Jean arrived in Paris. Two days after/later he would leave for Berlin.)

(11.ii) Mercredi, le premier Septembre Jean arriva à Paris.
✓ Deux jours après/plus tard il partit pour Berlin.

(Wednesday, the first of September Jean arrived in Paris. Two days after/later he left for Berlin.)

In comparison with *deux jours après* and *deux jours plus tard*, the use of *dans deux jours* thus appears to be more restricted.[34] But on the other hand

[32] During the lecture Dominique Sportiche pointed out that *dans deux jours* can also be used to refer to the day after the day of the Speech time in cases where the Reference time of the clause which contains the phrase is distinct from the Speech time.

[33] Some speakers find the English translation of (10.ii) acceptable. But the French original is consistently judged as bad.

[34] Actually the uses of *deux jours après/plus tard* do not constitute a simple superset of those of *dans deux jours*. *Deux jours après/plus tard* is "allocentric", in the sense that its "origin of computation"—the time from which the one to which the phrase refers is "computed"; for *deux jours après* this is the time such that the referent of *deux jours après* is the second day after that—cannot be the time of Speech. *Dans deux jours* is not allocentric. Thus as a way

we find that *dans deux jours* appears to be *more* flexible than certain other adverbs, including *après-demain* (*the day after tomorrow*), which nevertheless denotes the same time as *dans deux jours* in those contexts where both can appear (just as we saw for *dans deux jours* and *deux jours après/plus tard*). This is shown by the following pair:

(12.i) Les habitants continuèrent a déblayer les débris.
Tout de même ils se rendaient bien compte:

✓ Après-demain
✓ Dans deux jours
} les bombardements allaient recommencer.

(The inhabitants carried on with sweeping up the rubble. And yet they realized full well. The day after tomorrow/ in two days the bombing would resume.)

(12.ii) Les habitants continuèrent a déblayer les débris.
Ils ne se rendaient pas compte,

*qu'après-demain
✓ que dans deux jours
} les bombardements allaient recommencer.

(The inhabitants carried on with sweeping up the rubble. They didn't realize that the day after tomorrow / in two days the bombing would resume.)

The point of this pair of mini-discourses is that the uses of *après-demain* are more restricted than those of *dans deux jours*. *Après-demain*—and the same is true of similar adverbs—*demain* ("tomorrow"), *hier* ("yesterday"), *aujourd'hui* ("today") and their counterparts in other languages such as English *yesterday, today, tomorrow*; as well as (according to most speakers) expressions containing *last, next* and *ago*, as we find them in *an hour ago, last week, next Wednesday*, seem to require awareness on the part of someone 'on the scene'. The examples suggest that the kind of perspectival shift which such adverbs seem to need in order to be felicitous (and of which their presence therefore is a signal) is one that is found (exclusively or primarily) in "free indirect discourse". This is a literary device in which, one might want to say, the perspective is shifted from the external utterance time to a story-internal time, and where the information is presented as seen through the eyes of one or more protagonists from the story at that time. An example is

---

of saying that Jean will arrive the day after tomorrow (12.i') is grammatical whereas (12.ii') is not:

(12.i') Jean arrivera dans deux jours.
(12.ii') Jean arrivera deux jours après.

the last sentence of (12.i), a new sentence which presents its material as a content of thought on the part of the inhabitants at the time. This is a more radical kind of pespective shift than that which is required by past-directed references of *now* and *maintenant*. Perhaps it is a kind which is possible only in narrative fiction, where it is one of the conventions of the genre that the external utterance time, that at which the story is supposed to have been written (or, perhaps, that at which it is read), has been "bracketed out": Interpretation of the story does not require that one relates the narrated events temporally to the utterance time—or, if at all, only in an extenuated sense.[35]

Adverbs thus seem to differ as regards the kind of perspectival shift which they need to be felicitous. I do not know whether the full repertoire of possibilities is exhausted by the two kinds of perspective shift our examples indicate or whether there are further kinds of shift, required by other expressions, grammatical constructions or stylistic devices—in English, French or in any of the innumerable other natural languages which lie outside my very limited linguistic horizon—which have not yet been investigated from this particular View Point.[36]

However, even the distinctions exemplified by our triple *deux jours après/ plus tard, dans deux jours* and *après-demain* present enough of a challenge to the most attractive and justly popular framework for dealing with meaning in natural language (including, especially the meaning of "indexicals" like *now* or *tomorrow*) which has guided thought about natural language semantics for nearly thirty years, Kaplan's three level theory. According to this theory the general form of the semantic value of an expression is that of a *character*, a function which maps contexts of use to intensions, i.e. to functions which map "points of evaluation" to suitable extensions. In the case of a temporal adverb the suitable extensions are times (if we stick to

---

[35] The examples (12.i,ii) have been chosen so as to maximize the contrast that interests us—I believe few French speakers would quarrel with the acceptability of the first version of (12.i) or with the unacceptability of the first version of (12.ii). But the examples do not clinch the question where the border line between the acceptable and the unacceptable cases lies, they leave open what happens when we present the claim that the inhabitants knew about the resumption of the bombing into normal indirect discourse, or when their ignorance is presented in the free indirect discourse form. My own impression is that both of these additional versions are bad, though presumably for somewhat different reasons. Should this assessment be correct, this would tend to show that an account of the acceptable use of adverbials will have to pay attention to a number of distinct linguistic and (possibly) conceptual factors.

[36] The little that has come to my attention about the different kinds of "logophoricity" one finds in various languages governing the behaviour of pronouns and reflexives inclines me to think this might well be so.

the assumption on which much of the foregoing is based, viz. that such adverbs act as predicates of times). So the value of such an expression should be a function from contexts of use to functions from points of evaluation to times. With this general scheme one often combines the ancillary thesis that the semantic values of lexical items either have constant character or constant intensions.[37] According to this additional, lexical principle a context-sensitive adverb—like *now*, say—whose character is not constant, has constant intensions as values which its character yields in different possible contexts of use. Indeed, the original proposal within this theory for what the value of *now* should be conform to is this: The value of *now* maps each context C to the constant intension which assigns to every point of evaluation one and the same time, viz. the ("utterance") time of *C*.

But what about the uses of *now* where it refers to some time in the past? Clearly the semantic value just described won't do in this case if we assume that the relevant context is the speech time in such cases. Or, rather, it won't do if we assume that the relevant context time is in these cases—just as it is in others—the utterance time of the entire sentence or discourse. If we want to save the lexical principle, then we have to assume that in such cases the context is not what it normally is (e.g. for an utterance such as "Right now it is five to one", aimed at telling somebody the time of day. And so we will have to allow for the possibility of context *shifting*: What we have thus far described as "perspectival shift" should be recast as a shift of context.

If this is the way we try to adjust the three level theory to the demands placed upon it by past-directed *now*, *tomorrow* and the likes, the question must be asked how much of the original structure of the theory, the intuitions which motivated it and the general theses that went with it survives and what will fall by the wayside. Without further details of what such a more comprehensive modified three level theory will look like it is difficult to guess.

So what might such a more comprehensive theory look like? I cannot go into this question here and now in the depth necessary for a well-founded judgement of how much of the original three level conception can be seen as having been preserved. But let me, in just a very few words, sketch one way in which the features of context-sensitive adverbs that appear to be alien to the three-level theory as I understand it might be accounted for in

[37] "Constant" in the sense in which a function is constant iff it returns the same value for all its arguments.

such a revised theory.[38] In the theory I have in mind, which I will think of as implemented within Discourse Representation Theory,[39] the "meanings" of the tenses and temporal adverbs are given (in part at least) in the form of rules for the construction of semantic representations of the sentences and discourses in which tokens of these expressions and morphemes occur. However, the representations for sentences and discourses that are to be constructed with the help of these rules are not quite like the representations (so-called "Discourse Representation Structures" or "DRSs") familiar from the generally accessible literature on DRT; for they may contain DRS-like substructures which represent bits of information that the represented sentence or discourse presents as some person's thought or, alternatively, as a content presented from the perspective of some time other than the utterance time but without interpreting it as the content of any particular individual's or individuals' mental content. These substructures determine their own "propositions",[40] and when the part of the sentence or discourse that gives rise to the substructure contains an "indexical" adverb, then it will make to this "proposition" a contribution defined by the shifted context of that substructure. For instance, when the part giving rise to the substructure contains the word *now*, then the construction rule associated with that word will insure that its representation in the given substructure is interpreted as referring directly to the time of the (possibly shifted) context of the substructure.

Along these lines[41] we can, I think and hope, hold on to the basic insights about indexicality which found such an elegant culmination in the three level theory, while at the same time doing justice to the phenomena of "perspectival shift". In fact, it should be possible in this way to account also for the different *kinds* of perspectival shift which appear to go hand-in-hand with different types of perspectivally sensitive adverbs and different stylistic forms.

How much of the spirit of the original three level theory will be preserved in this way, everyone will have to decide for himself—and of course one can't

---

[38] What I am about to say must remain in the realm of speculation until the details of the theory are spread out before us. In fact, quite a number of the details have been worked out already, though here I can do no more than give the merest hint. Those familiar with Discourse Representation Theory (see in particular Kamp & Reyle 1993) should be able to make some sense of the following remarks in relation to that theory.

[39] Also known as "DRT", the abbreviation I will use in the few remaining paragraphs which are still to follow. For an introduction see in particular Kamp & Reyle (1993).

[40] More correctly: "information states"; see Kamp (1999); Kamp et al. (2011).

[41] See Kamp & Reyle (1993).

be expected to do this until the details of the revised theory are properly in place. At this point all I can do is to promise that I will provide these before long. But whatever our verdict on this matter is and whatever the future time when we will be in a position to give it, there seems no denying that the ways in which natural language allows us to manipulate perspective show that its mechanisms for conveying meaning go well beyond what the original tree level theory was intended to deal with and what, in its original form, it enables us to say.

9. Conclusion

This brings me to the closing words of this lecture. From the few linguistic problems we have looked at it should have become clear that no very simple architecture for a semantic theory of a natural language, which follows the model of syntax and semantics of predicate logic as closely as, for instance, standard Montague Grammar, will ultimately do. Let me summarize what we have found to be some minimal requirements on an architecture which might do and some of the most important tasks that must be accomplished by a semantic theory with such an architecture:

1. First, as I mentioned early on—and this is in my view the single most important difference between natural language and the predicate calculus—natural languages have means for exploiting the possibilities of cohesive discourse, possibilities which have no equivalent in predicate logic as we normally state and use it: In natural language a multi-sentence discourse can express what in predicate logic can be stated only with the help of a single (often distressingly long) formula.
2. To make this exploitation of discourse possible, many words and morphemes of natural language have meaning components that point towards the context to which the sentence or clause containing them belongs. The tenses, we have seen, and quite a few of the temporal adverbs as well, have such a meaning component; but in fact, there are many more such items in English (and, I am pretty sure, in most if not in all other natural languages). In fact, one of the most important challenges for formal semantics as it exists today is to correctly classify and explicitly state the meaning contributions made by such morphemes and words.
3. In addition there is an issue which we touched upon in passing, but which deserves a more explicit articulation than I have thus far given it. The insight that in natural language multi-sentence discourse often accomplishes what in predicate logic can be accomplished only by a single formula is compatible with a predicate-logic-based concept of logical form—a concept of logical form, according to which each of the logical forms made available by the semantic representations of discourse one adopts is logically equivalent to

> expressions of some well-established version of predicate logic. (This, is the case of DRT in its original form, all of whose representations (DRSs) are equivalent to formulas of the first order predicate calculus.) The study of temporal reference, however, and in particular its perspectival dimension, shows that more is needed than just an account of how single logical forms can be expressed by cohesive sequences of sentences in a natural language. For as we have seen, the effect of the different perspectivally sensitive adverbs is to convey that the parts of the discourse containing them have a *distinctive status* within the discourse as a whole. A good discourse representation must account for this dimension of meaning: for the different kinds of status a bit of information may have within the information structure as a whole, and not just for what that bit of information is (i.e. for its truth conditions). And the semantics for the representation must do justice to the logical relation between the status of the given bit and the status of other components of the representation. This is a problem that is akin to that of describing complex attitudinal states consisting partly of beliefs, partly of desires, partly of intentions and so forth. But there are important differences as well; a perspectival orientation is not the same thing as a propositional attitude (although, as we have seen, perspective often involves some person's attitudes at some time); and the logical relations between different perspectives are, similarly, of a different character than, say, that between an intention and a belief. While considerable progress has, I think, been made on the subject of complex attitudinal states, corresponding work on the logic of discourse representations, which assign distinct perspectival status to their different parts and provide a satisfactory analysis of their logical interaction, is, to my knowledge, still rudimentary at best.

These are problems for which, I suspect, nobody at this time has as much as an outline of a solution. But they are problems that must be addressed if we are to capture what is surely one of the most important aspects of information as we find it in thought and language: the ways in which the mind organizes information as a network of connected "information nodes", in which the nodes are alike in that they all carry propositional content, but in which nevertheless each node has a separate role to play within the structure overall; and the ways in which languages reflect this aspect of information structure and the devices they have for making certain aspects of it explicit. On the other hand, I am convinced that by addressing these problems semantics should be able—more in this way perhaps than in any other—to make new contributions to the philosophy of language, logic and mind, and ones no less fundamental than those it has made already in the course of this century; and at the same time it is by work in this domain that we should, at long last, succeed in building a solid bridge between "formal" semantics and the systematic study of both conversational language and literary devices and forms.

If it is this far that the road starting from Reichenbach's observations about the behaviour of tense has led us, it would seem to be a fair question how much of that road, of the part that we have been able to cover in the half century since *Elements of Symbolic Logic* appeared and the part that is still in front of us, he himself foresaw. Probably not all of it, but on the basis of the terse discussion in the section from *Elements of Symbolic Logic* which has been our point of departure it is difficult to draw firm conclusions. What is clear beyond a suspicion of doubt is that in this section Reichenbach was evidently very much ahead of his time.

Appendix

Most of the points that were raised during the discussion after the Reichenbach lecture concerned the details of how we determine temporal relations between the eventualities described by successive past tense sentences in short bits of cohesive discourse. The audience drew attention to a number of complications which the simple-minded accounts I had presented during the lecture are unable to deal with as they stand. In fact the arguments I presented have long been known to fall short by a long stretch of what is needed when we want to account for the full spectrum of interactions that are revealed by the judgements of human readers and listeners about temporal order of described events and states. The more recent literature—starting roughly in the mid-eighties—has gone well beyond these simple-minded accounts, and my only excuse, if any, for presenting matters in the oversimplified a way I chose was that within the temporal limits presented by the occasion a more refined analysis would have been infeasible.[42]

As I see it myself, the most serious oversimplification in the lecture was the analysis I presented of narrative progression. That this is a much more complicated issue than the analysis suggested was already clear in the late seventies; already at that time certain counterexamples to this analysis were being discussed. Among these counterexamples there were in particular felicitous sequences of simple past tense non-stative sentences where the order of the reporting sentences does not mimic the temporal order of the

[42] I wish I had been able to find a form of presentation which would have given the details their due while allowing the non-expert to follow the gist of the arguments which take their departure from Reichenbach and lead to the kind of semantic theory that we now know to be necessary for dealing with the phenomena to which he was the first to draw attention within a context of formal logic.

reported sentences: the principle of isomorphism between order of events and the succession of the sentences reporting them in the discourse, is at best a default principle that is often, and easily, overruled.

That the order in which the events are introduced in the discourse is only one among many clues about the temporal order of their actual occurrence (or, more acurately, the order of occurrence presumed by the narrator) is one of the prominent aspects of the work on discourse of Nicholas Asher (1993). A central feature of Asher's theory, which he first developed in the second half of the eighties and which has been steadily expanded and improved since then, is the emphasis it puts on the connections between temporal order and rhetorical relations. The importance of rhetorical relations for the interpretation of discourse is summed up by the following general principle:

> (PRC) (Principle of Rhetorical Connectedness)[43]
>
> Each new sentence or clause in a discourse or text must be interpreted as standing to one or more other clauses or sentences in the text in a certain rhetorical relation.
>
> (If interpretation does not succeed in establishing such a relation, then the interpreter has the feeling that the current sentence or clause has not been properly understood and the discourse appears to lack coherence.)

PRC embodies a very important insight into the ways in which we interpret discourse. But it is not easy to make it precise and there continues to be much disagreement about how this should be done. One of the main controversies concerns the repertoire of rhetorical relations from which each sentence- or clause-interpretation must make a choice. In the extant literature on rhetorical relations one finds quite a diversity of proposals as to what this repertoire might be. The relations which have been most prominent in the work of Asher and others on the interpretation of temporal relations—and these are the ones which I will be using also in this Appendix.[44]

1. *Explanation*

This is the relation which holds between the current sentence or clause and an earlier one if the first serves to give a kind of causal explanation of the latter. An often discussed example is:

---

[43] The literature on rhetorical structure and rhetorical relations is by now quite substantial. Work, which is still of general importance, and which has been a significant anchor point in Asher's writings on this issue, is that of Mann and Thompson (see Mann & Thompson 1987).

[44] Here I am relying heavily on the work of Asher and his various coauthors.

(13) Bill fell. John pushed him.

According to most speakers, this sentence pair has an interpretation according to which the second sentence presents its event e_2 (the pushing) as what caused the event e_1 described in the first (the falling). As a consequence e_2 cannot be understood as following e_1, as the cause of an event cannot begin later than the event itself. The choice between the remaining options—in particular that between partial overlap between e_1 and e_2 and that of e_2 completely preceding e_1—the interpreter will make on the basis of other factors, e.g. on the basis of what is known about how events of the types described could be causally connected. If the information required to make such further decisions is absent, they need not be made; in that case the only constraint that the interpretation imposes on the temporal relation between them is the condition that e_2 starts no later than e_1. We will encounter other cases of such temporal "underspecification" below.

2. *Elaboration*

Here the current sentence or clause is interpreted as elaborating the eventuality described by some other one. An example is

(14.i) John glued the heel back on his shoe. He used Evostick.

Here the second sentence has the function of telling us more, viz. what kind of glue was used in the glueing process, about the event e_1 described in the first sentence. An event e_2 described in a sentence S_2 which is interpreted as standing in the Elaboration relation to some other sentence S_1 is typically understood as temporally included in the eventuality e_1 described by S_1. There are examples, however—and in fact such examples have been salient in the literature going back to the early eighties—which may seem to put this general principle in doubt. One instance is given in (14.ii):

(14.ii) They finally built a bridge across the estuary. A Finnish architect drew up the plans.

It has been suggested that here we have an exception to the temporal inclusion principle, inasmuch as the drawing up of the plans must have preceded the building of the bridge. But this is only one way of looking at the complex event that (14.ii) speaks of. It seems just as possible to see the first sentence of (14.ii) as describing the building of the bridge in a wide sense, according to which the architect's design is part of it too. For instance, if the decision to build the bridge included or was followed by the decision to commission the Finnish architect to make the plans, or to have

a competition for the design which the Finnish architect proceeded to win, then this would seem a natural interpretation. It is my impression that in those cases where S_2 is to be construed as providing an Elaboration of S_1, the eventuality described by S_1 can always be seen as comprehensive enough to include the one described by S_2.[45]

According to this stipulation Elaboration excludes the possibility that e_2 precedes e_1. (In fact, we may want to adopt a stronger principle according to which e_2 must be included in e_1 in this case. See the remarks below about "Background".) Nevertheless, as in the case of Explanation, the temporal relation between e_2 and e_1 is not completely fixed thereby. For temporal inclusion of e_2 within e_1 is compatible with four more specific relations: e_2 properly included in e_1, temporal coincidence of e_2 and e_1, e_2 starts at the same time as e_1 but ends before it, and e_2 starts after e_1 but the two end at the same time. As in the case of Explanation, a choice between these four, will, if it is forthcoming at all, have to depend on other factors, that have to do with lexical meaning and world knowledge.

There is however a special case of Elaboration in which the choice between these four possibilities comes about automatically. This is the case where the event e_2 described by the elaborating sentence S_2 is identical with the event e_1 of the sentence S_1. Arguably (14.i) itself is of this kind; its second sentence is naturally interpreted as saying of the event E mentioned in the first sentence—the glueing back of the heel—that it (i.e. that very E) involved Evostick as the glue. Another example of this sort is (15):

(15) Ford wrote a review of Lennox' latest novel. He tore the book to shreds.

---

[45] To distinguish between Elaboration and Explanation on intuitive grounds is not all that easy. For in a sense saying something about how a certain event or state came about is a case of "saying more about it", and thus a way of "elaborating" it, too. In such cases S_2 seems to qualify intuitively as an Elaboration of S_1 *in virtue of* qualifying as standing to it in the Explanation relation. Thus the principle that Elaboration always entails temporal overlap between e_2 and e_1 (and, more specifically, inclusion of e_2 in e_1 when e_2 is an event) is to be seen as in part a matter of stipulation, in which cases of "elaboration in the form of causal explanation" are excluded from the Elaboration relation. For the interpretation of (14.ii) just discussed this decision entails that we must understand the "building" event of the first sentence as including the architect's design. Compare this case with the one in which the architect made a design for a bridge across a river at some particular place. As a result, the council decided to build a bridge of that design. In this case it would seem more natural to see the building as not including the design and the latter as something which led to it. In other words, in the light of this additional information about the described events the more natural analysis would now be that according to which S_2 stands in the Explanation relation to S_1 and e_2 precedes e1 rather than being temporally included in it.

It is *in* writing the review, as we might put it, that Ford tore Lennox' novel to shreds.

(14.ii), on the other hand, seems clearly not of this kind. Even if we count the drawing up of the plans for the bridge as part of building it, it surely must have been a proper part. A perhaps even more straighforward example of this second pattern, where the event of the first and that of the second sentence are distinct, is the following dinner description, a variation on a theme that has been among the literature's stock-in-trade examples since the early eighties:

(16) The meal that Mary had prepared delighted her guests. The hors d'oeuvre, a crab cocktail with a cointreau-based dressing of her own invention, was much appreciated. The guests were even more impressed by the swordfish which followed.

Here the events of the second and third sentence, that of appreciating the hors d'oeuvre and the impression made by the swordfish, are distinct parts of the delight produced by the meal as a whole, just as the two courses in question, the hors d'oeuvre and the fish are distinct and therefore proper parts of the entire meal, which must therefore be distinct from it.

3. *Background*

A relation which appears closely related to that of Elaboration and that could, in a different way of slicing up the rhetorical cake, have been treated as a special case of it, is the one we will refer to as "Background". Background holds between sentences S_1 and S_2 if S_2 is stative and describes a condition which obtained when the eventuality described by S_1 took place, or while it lasted. An example is (17):

(17) After dinner they took a stroll through the grounds. The air was warm and fragrant, more so than one had any right to expect from a night in early April.

Here the second sentence "places" the event described by the first sentence (the protagonists' stroll) within a context—or "against a background"—of a weather condition which we understand as holding throughout the time when the stroll occurred. The temporal relation which accompanies the relation of Background is inclusion in the direction opposite to that which corresponds to Elaboration as defined under *2*: In the case of Background it is e_1 that is included in e_2.

4. *Narration*

Finally we come to the relation which, I said, is sometimes treated as the rhetorical relation par excellence and thus as the default relation—the one which the interpreter will assume unless special considerations lead him to prefer one of the others. We adopt the name used by, among others, Asher for this relation, that of *Narration*. Narration is the rhetorical relation which is assumed to hold between non-stative sentences S_1 and S_2 when the event e_2 can be seen as being a natural follow-up to e_1. Here are two examples:

(18) John fetched the newspaper from the doorstep. He sat down, took out the business section and checked the performance of the stocks his father-in-law had told him to buy the week before.

(19) He told her that as far as he was concerned she could get lost. She replied that getting lost was just the thing she intended to do.

In both cases each next sentence is naturally understood as describing an event that can be seen as the next one in a "causally"[46] connected succession. In (18) we can interpret the successive events as parts of a single plan, which was "pre-programmed" in John's mind even before he went to fetch the paper. But it is not necessary to impose such a strongly intentional gloss on what (18) reports. It could have been that John went to get the paper out of sheer habit, that the idea of looking at the business section only came to him after he sat down and that he checked on the stock he had been advised to buy only when he reached the page with the stockmarket quotations. But even if the events are as loosely connected as this second scenario makes them appear, they still qualify as forming a sequence in which each event naturally leads to the next one, notwithstanding the fact that the second one wasn't planned at the time when the first one occurred. Moreover, it is neither necessary that the events are seen as caused by their predecessors, nor need the predecessor be viewed as providing a precondition necessary for the occurrence of its successor. All that it required is that the first event leads to a situation which fulfills the necessary preconditions for the occurrence of the next event and in which this next event can be seen as a natural continuation of the sequence of events narrated up to that point.[47]

---

[46] See footnote 46.

[47] Roßdeutscher (2000) argues that the relation that is typically perceived as holding between the events expressed by successive sentences in narrative discourse are what she calls "non-accidental dependency". This is a relation which certainly holds between e_1 and e_2 if e_1 is the cause of e_2, but does not require this; in order that the relation holds it suffices that e_1 creates the possibility for an event of the type of e_2 to take place, or, in the domain of

(19), in which the event of the first sentence is an utterance made by one person and the event of the second sentence the reaction to that first event by another, instantiates an important, special case of this. The first event is neither a necessary nor a sufficient condition for the second—the woman referred to as "she" could have said something different from what she did say, or she could have chosen not to react at all. But this does not alter the fact that we see the second utterance reported in (19) as a *reaction* to the first. (In fact this is what is entailed by the semantics of the verb *reply*.) The exact nature of the possible connection(s) between two events which is implied by the rhetorical relation of Narration between the sentences reporting them is not easy to describe; certainly the concept of causation as we find it in the philosophical literature is not what we need here. Causation between e_1 and e_2 in the sense in which causation is normally understood is *one* of the possible connections between two events which is compatible with Narration, but it is not the only one. The metaphysical counterpart of Narration is either a set of possible relations which contains causation as only one of several members (or, if one wants to distinguish between several different causation relations, which includes the set of those relations as a proper subset); or else it is a single relation which is entailed or presupposed by causation but does not require it. It seems to me that a better understanding of what this relation is (or what these relations are) is an important task, not just as part of the foundation for the analyis of discourse but more generally for metaphysics and the philosophy of logic as a whole. But it is a task that I believe hasn't been accomplished yet, and at this point I do not know how to it should be accomplished.

From what we have been saying it should be clear that when S_1 and S_2 stand in the relation of Narration, then e_2 must follow e_1. It is important, however, to stress that whatever the event relation or relations may be which correspond(s) to the Narration relation, Narration is *not just* a matter of temporal succession between the reported events. In this respect Narration is like the other rhetorical relations we have discussed: It has consequences for the temporal relation between the described eventualities, but it is more

---

human action and reaction, when e_1 can be seen as having provoked e_2 as response, even if the response could also have been another one. Exactly what the general characterization should be or, for that matter, whether a single uniform account of non-accidental dependency is possible, is not clear to me at this point. But that some such relation, including causation as usually understood but much more comprehensive, is crucial to our understanding of discourse coherence, appears to be an important and correct insight.

than a matter of temporal relations alone. We will return to this point below, when we address problems of the interaction between tenses and adverbs.

This list is not complete. Missing in particular is the relation of *Concession*, the relation which holds between S_1 and S_2 iff what the one sentence says makes what is said by the other less likely than it would have been otherwise.[48] Concession is uninformative about the temporal relationship between the eventualities which S_1 and S_2 describe. Within the context of this Appendix it is therefore of no direct interest.

I leave it open whether a theory of rhetorical relations should distinguish still further relations. For the observations I want to make in this Appendix the four I have discussed suffice.

5. *Multiple Rhetorical Relations*

There is a further aspect to the Elaboration relation which we need to mention. Often an elaboration extends over several sentences, and when that is so, the later sentences of the elaboration are sometimes understood as not only standing in the Elaboration relation to some earlier sentence, which describes the event about which they all provide further details, but as standing also in certain rhetorical relations to each other. For instance, there may be several sentences elaborating a sentence S_1 which mentions some complex event (such as a dinner) which it is possible to see as pairwise connected by the relation of Narration. (20) is an example of this.

(20) Frankie got the five year old daughter out of the kindergarten. He drove up to the gate and parked his car in one of the two remaining free spaces. He went into the building, found the class room where they kept the children until they would be collected after school, told the teacher that Mrs. Van Bentick had sent him. She said that it was all right and he took the girl by one hand and her belongings in the other. They went to the car, he opened the door, put her inside and drove off.

You may read this as being about a kidnapping or as a friendly neighbour helping out at a time of domestic crisis. But in either case all clauses following the first sentence are naturally interpreted as part of one long elaboration of a complex event which the first sentence describes in very general terms. And in addition each of the subevents described by the second conjunct of the second sentence and from then on is understood as "the next

[48] For a careful analysis of the relevant notions of conditional and non-conditional likelihood see the work of Merin, in particular the publication listed in the bibliography (cf. Merin 1996).

thing to happen" after the event described by the immediately preceding clause; these pairs of successive clauses appear to be candidates for the Narration relation that are as good as any.

Cases of this kind, where a sentence or clause is interpeted as rhetorically connected to two sentences or clauses at once, are quite common. Typically (in fact I am not aware of other possibilities) they are of the following pattern: There is one relation R_1 in which each of the sentences of a sequence $\mathcal{S}$ stands to some fixed sentence S_1 outside that sequence (normally preceding it), whereas the sentences within the passage also stand in certain rhetorical relations R_i to each other. Such cases could be alternatively described as involving (i) the relation R_1 as holding between S_1 and $\mathcal{S}$ and (ii) the relations R_i between the relevant pairs of members of $\mathcal{S}$. In this way we may be able to do away with the idea that new clauses or sentences may have to be rhetorically related to the antecedent discourse by more than one rhetorical relation; but of course the real complexity is thereby not eliminated; each member S_1 of the sequence $\mathcal{S}$ will have to be recognized as belonging to a sequence which as a whole stands in the Elaboration relation to the first sentence, as well as being rhetorically related by the relation R_i to its immediate predecessors in $\mathcal{S}$.

The most prominent cases of this phenomenon—i.e. of double rhetorical attachments of a new clause to the interpretation structure already associated with the text or discourse up to that point—may well be those where the "principal" relation is Elaboration. But Explanation seems to allow for such possibilities too, witness the following variation of the well-worn example (13):

(21) Bill fell. John pushed him and he stumbled over a root sticking out from the path.

Here the two conjuncts of the second sentence are naturally understood as jointly explaining why Bill fell. They form a sequence, one might say, which, as a whole, stands in the Explanation relation to the first sentence, while the two clauses which make up the sequence appear to be related by Narration.

If cases like (21), in which one sentence in the Simple Past is followed by two or more sentences or clauses also in the Simple Past, which jointly provide an explanation for the eventuality of the first sentence and moreover stand in the Narration relation to each other, are not all that common, this is probably connected with a matter which we discussed at some length in the lecture (Section 5.1). There we drew attention to extended flashbacks, like (5). The eventualities described by all but the first clause or sentence of an extended flashback, we said, each involved two temporal relations, one

to the eventuality which, in Reichenbach's terminology, provides the Reference time of the different Past Perfects that the members of the flashback contain, and a second temporal relation to the eventuality of the immediately preceding clause. In the light of our observations about double rhetorical relations it seems natural to wonder whether extended flashbacks might not be on a par with the rhetorically connected sequences of sentences and clauses of which we have just spoken. More specifically, aren't extended flashbacks just one more type of discourse involving the relation of Explanation?

Although the example we have just considered might suggest this, I think that in general the answer to the question should be "no". For there are cases in which a flashback gives us what might be considered some "past background" to the eventuality which provides the Reichenbachian Reference time, but where this background does not in any way *explain* that eventuality. An example is (22):

(22) When Fred took the bar exam in April, he made the highest score of all two hundred fifty six candidates. The year before he had not even come close to passing.

Here the information that the year before Fred failed the exam dismally may be seen as providing relevant background information to the event mentioned in the first sentence because of the striking contrast between a hopeless performance on the first occasion and a brilliant one on the second. But surely the earlier disaster does not explain the subsequent success; rather, it makes that success appear so remarkable because it would have sugested the opposite. Here the rhetorical relation between the two sentences is, if anything, that of Concession.

Extended flashbacks serve well as a point of entry into the general question which discourse situations may lead to double rhetorical relations for the new sentence or clause, and how. First, there are the flashback cases we already considered, where a first sentence in the Simple Past is followed by a series of event sentences in the Past Perfect, for all of which this first sentence provides the Reference time in the sense of Reichenbach's Past Perfect table entry, but where in addition the successive Past Perfect sentences appear to be standing in the Narration relation to each other. In these cases it is hard to tell what comes first in the intepretation of all but the first of the Past Perfect sentences, the choice of the Event time of the first sentence as reference time for the new sentence or the rhetorical relation (Narration) between the new sentence and its immediate Past Perfect predecessor. If the new sentence is perceived as standing in the Narration relation to its

predecessor, then the use of the Event time of the first sentence as Reference time for its Past Perfect migh be regarded as a consequence. Not only will, in the simple examples we have considered, no other Reference time be available (for the new event will be interpreted as following all the events introduced by preceding Past Perfect sentences; so none of the Event times of those events will do as Reference time for the new Past Perfect); it is also plausible that interpreting the new sentence as standing in the Narration relation to its predecessor is tantamount to seeing it as part of the flashback to which that predecessor has already been assigned; as a part of the flashback, the new sentence should use the Event time of the initial sentence as the Reference time for its Past Perfect tense.

In principle there is also another order in which these decisions could be made: First, the Event time of the initial sentence is chosen as Reference time, then the new sentence is related to its predecessor, thereby becoming unequivocally part of the same flashback. Note, however, that in general neither the first nor the second of these two steps is obligatory. For the choice of Reference time the Event time of the immediate predecessor might have served too, thus giving rise to a "flashback within a flashback".[49] Also, once the Event time of the initial sentence has been chosen as the Reference time for the new Past Perfect sentence, it is not necessary to relate it to any of the earlier Past Perfect sentences. An example is

(23) Fred and Susannah met in Paris. He had flown in from Brussels.
She had made her way on bicycle from Ostrava.

Here both the second and the third sentence are naturally interpreted as flashbacks from the time when Fred and Susannah met. But there would seem to be no further rhetorical relationship between them.[50]

In the case of the third sentence of (23) the choice of the Reichenbachian Reference time must come first, for the simple reason that no further rhetorical relation is forthcoming. A case where the relation between the new sentence and its immediate predecessor appears to have priority are those

---

[49] This possibility has in fact been suggested as the right analysis of discourses like

(i) Getty bought the Chardin from Karmousian. Karmousian had acquired it ✓ from Lord Kilmarnock, who had inherited it from his great-aunt, the Viscountess Hardwick.

where the main clause of the second sentence can be taken as supplying the Reference time for its relative clause.

[50] At least not one of any of the types we have considered. Intuitively there is an element of contrast between the two.

where the new sentence is not in the Past Perfect but in the Past Progressive (in English) or in the Imparfait (in French):

(24) (i) We reached the top shortly before noon. We had got up at half past five. Already the sun was shining brightly …
(24) (ii) Nous avons atteint le sommet peu de temps avant midi. Nous nous étions levés a cinq heures et demie. Déjà le soleil brillait à pleine force.

The intended interpretation here is that the sun was shining when the climbing party got up, not when it reached the top. Sentences with Internal View Point Aspect can be interpreted as involving as Reference time the Event time of a preceding sentence in the Past Perfect. In fact, the tenses of the third sentences in (24) are the natural ones to use if this is the intended interpretation. (In the third sentence of (24.i) a Past Perfect ("the sun had been shining") would have been just about possible, but the Past Progressive is clearly preferred. Past Perfect in the third sentence of the French example ("le soleil avait brillé") would be virtually ungrammatical.) In these examples the interpretation of the third sentence as rhetorically related to its immediate predecessor (in the sense of Background) would seem to be the first and decisive step, and counting it as part of the flashback an automatic effect of that.

These different cases suggest the following picture. Each new sentence must be related to some preceding sentence.[51] The relationship established between the new sentence S_{n+1} and the chosen "anchor point" S_i must be consistent with the requirements imposed by the new sentence's tense form and its View Point Aspect. In certain cases the first relation may then give rise to a second one, between the new sentence or clause and some third sentence or clause S_j. In those cases where two relations get established it seems to be difficult to say which of them has priority in the process of interpretation. What matters in these cases—and this is certainly so from the limited perspective of this Appendix, where our ultimate concern is with temporal relations—is that the combination of relations satisfies a number of simultaneous constraints imposed by linguistic properties of the current clause or sentence on the one hand and the general demands imposed by coherence on the other.

[51] Presumably one not too far back, though I do not know by what rules sentences or clauses that lie too far back are excluded, nor whether this is a phenomenon of gradual deterioration, or one with sharp boundaries, in the sense that anything lying farther back in the discourse is categorically excluded as rhetorical anchor point.

To my knowledge the patterns of multiple rhetorical relations which we observed in connection with the Past Perfect cover the range of possible cases. First, as far as I know there are no cases which involve three or more such relations—two is the limit. Second, the cases where interpretation apparently leads to the assumption of two relations conform to what we have just seen: One relation is necessary as a minimal fulfillment of the requirements of coherence; the second relation either (i) is a natural consequence of the first or (ii) results from applying an interpretational strategy which also leads rhetorical relations in cases where we get only one relation.

An example of (i) is that where a new sentence in the Simple Past is interpreted as part of the elaboration of an event introduced by some initial sentence S_1. This sentence may then be interpreted in addition as the next element in a narrative sequence, which provides this elaboration and to which its immediate predecessor also belongs. An example is the "dinner" example (16). Analogous cases arise when the first relation is Explanation. Thus, in (21), the second conjunct of the second sentence is interpreted as part of the Explanation started in the first conjunct and the two conjuncts are naturally seen as standing in the relation of Narration. In these cases there is, I believe, no telling whether it is first the Elaboration or Explanation connection which is established, or whether the interpretation first "sees" the narrative connection with the immediately preceding sentence or clause. But, as we noted in connection with our Past Perfect examples, this question is of no importance to what emerges as interpretation in the end. In particular it is of no consequence for what the interpretation has to say about temporal relations.

With both Elaboration and Explanation it is possible for the new sentence or clause to be understood as standing in this relation to S_1 without standing in any rhetorical relation to the sentence or clause that immediately precedes it and that has already been interpreted as standing to S_1 in the same relation (i.e. Elaboration or Explanation).

An example of the first sort in which the relation to S_1 is Elaboration is given in (25). (26) is a similar example, but with Explanation instead of Elaboration.

(25) After that time our paths divided. Mary went to a university in Strasbourg, I got a job in a bank in our home town.

(26) The company went brankrupt. Its products were outdated and unattractive. And its main competitor launched an aggressive advertising campaign, which it was in no financial position to match.

In (25) the second and third sentence both elaborate what is put forward by the first, the coming apart of the two protagonists' ways of life. But

there is no further rhetorical relation connecting them in some additional way. (26) is similar. Bankruptcies (and many other complex events) often have a number of distinct causes, of which they are cumulative effects. The second and third sentence of (26) give two causal factors in the company's collapse. Here too there need be no further connection between these two (although, of course (26) does not exclude the possibility that the competitor's campaign was wholly or partly provoked by the inadequate product line, which made the company vulnerable to attack).

These are only the beginnings of a theory of discourse relations and much is needed to turn it into a theory that approximates the degree of formal clarity and precision which has become the standard in formal semantics. Nevertheless, some of the implications that rhetorical relations carry for our understanding of the temporal relations between the eventualies of which a discourse speaks can already be discerned from the little that has been said here.

Central to this last point, which is the one of direct concern here, is the following observation. It is generally important for discourse interpretation that any new sentence can be seen as standing to what came before it in some meaningful rhetorical relationship, and in general this relationship goes hand in hand with a certain temporal relation between the eventualities involved. But the type of the rhetorical relation is decisive for the temporal relation only in those cases where the entry for the tense of the new sentence or clause in Reichenbach's table contains the specification that "$E = R$"; in that case the precise rhetorical relation decides how precisely the events and/or states are temporally related; in fact when "$E = R$", any temporal relationship is still possible as we have seen, and we need our understanding of the rhetorical connection between the two sentences or clauses if we want to decide what the exact temporal relationship is.

One implication of this observation is that the condition "$E = R$", which we find in so many entries of Reichenbach's table, requires some elucidation. If the table is made part of a theory of discourse interpretation along the lines sketched in the lecture and in this Appendix, is it then still possible to interpret "$E = R$" literally as identity? To see more clearly what depends on this issue, let us consider once more the case of two consecutive sentences in the Simple Past, where the first, S_1, is taken to supply the Reference time for the second, S_2. In that case for both S_1 and S_2 Event time and Reference time will, according to Reichenbach's table, "coincide"—$E_1 = R_1$ and $E_2 = R_2$. Suppose, moreover, that the first sentence provides the Reference time for the second. As we have seen, all this is still compatible with any one of the

logically possible topological temporal relations between the eventualities that S_1 and S_2 describe; what temporal relation we take them to stand in will depend on the rhetorical relation we perceive to hold between the two sentences.

What does this imply for the interpretation of "$E = R$"? If this condition is to be understood as saying literally that Event time and Reference time coincide, then evidently we cannot insist that the reference time R_2, which we are assuming the first sentence must somehow supply for the interpretation of the second, must be always identical with either E_1 or R_1. Moreover, how R_2 is determined in relation to the information that S_1 contains can't be a matter just of the interpretation of S_1 as such, if the condition "$E_2 = R_2$" forces R_2 to coincide with the Event time of the new eventuality. For exactly where the new eventuality will be situated relative to that of S_1 depends, as we have seen on the rhetorical relation between S_1 and S_2; and that in turn depends not only on what S_1 says, but equally on the content of S_2.

Neither of the two strategies I mentioned in Section 5.1 of the lecture for dealing with narrative progression in a (quasi)-Reichenbachian fashion—the one which makes the time of the event e_1 into the Reference time R_2 for the next sentence and the one which places R_2 immediately after e_1—satisfy these requirements, for neither allows for the different temporal relations between e_1 and e_2, which depend on how S_1 and S_2 are seen as rhetorically connected.

In fact, if we want to use Reichenbach's entry for the Simple Past as an interpretation rule for Simple Past tense clauses and sentences in discourse, and use it as a rule which operates directly on the context as it has been established through interpretation of the sentences and clauses preceding the one containing the tense in question, then "Reference time" must be understood in a somewhat different way than we have done so far. The best option, it seems to me, would be to regard the Reference time provided by a given eventuality as an interval surrounding this eventuality, but possibly extending somewhat beyond it in either direction. In the cases where, according to Reichenbach's table, "$E = R$", the eventuality e_2 described by the currently interpreted sentence must *overlap* with the Reference time R, but that condition is now compatible with any temporal relation between e_2 and the eventuality e_1 that is described by the sentence which is supplying R. What this relationship between e_2 and e_1 precisely is must then, we have seen, be decided by other factors (including matters of aspect and rhetorical relations). Only when, according to the table, the relationship between Event time and Reference time is "$E < R$" or "$E > R$", will we get

a verdict on the temporal relationship between e_2 and e_1 from the interpretation of tense alone.[52]

Let us summarize the principal points of our discussion so far. They are:

1. To determine the temporal relation of the eventuality described by a new sentence or clause of a discourse relative to the eventualities introduced by earlier sentences or clauses, we need more than the information provided by (a) tense and (b) View Point Aspect. This is so in particular when the tense is one for which Reichenbach's table specifies the condition "$E = R$".
2. Only when the tense of the new sentence or clause is one for which the table gives one of the conditions "$E < R$" or "$E > R$", will the temporal relationship between the new eventuality and that of the sentence which provides R be settled in general. But even in such cases, a supplementary rhetorical relation to some third sentence may carry information about the temporal relationship between the new eventuality and the one described by that sentence.
3. All in all the temporal dimension of discourse interpretation cannot be seen in isolation from other aspects of interpretation.

In what I have said in this Appendix temporal adverbs have been systematically ignored. (I have gone out of my way to keep such adverbs out of the examples we have been looking at, and have allowed them in only in a few places, where their presence was not directly relevant to the points the examples were used to illustrate.) I have done this in full awareness of the fact that a crucial dimension of the problem of temporal reference in natural language has thereby been left out. In fact, when temporal adverbs are taken into account as well, most of what has been said here has to be qualified or revised. First, temporal "locating" adverbs (see Section 5.2 of the lecture) impose absolute constraints on the eventualities described by the clauses in which they occur—absolute in the sense that they cannot be overwritten by other conflicting considerations. More specifically, the presence of such an adverb suspends the need to relate it in any specific rhetorical way to an earlier sentence or clause. (The discourse will still seem strange to the interpreter if he can see no relevance of the new clause or sentence to what the discourse has transacted up to that point, but this is different

[52] If we adopt this option, however, then the version I gave in the Lecture of how Roßdeutscher's theory of View Point Aspect and cut applies to successions of past tense sentences, where the second sentence may be either stative or non-stative, must be adjusted. Now it can no longer be the new Reference time which produces the cut; rather, the cut must be made by a more specific contextual temporal element. As far as I can see, the time occupied by the eventuality e_1 of the sentence which supplies the Reference time for the current sentence or clause should give us the intuitively right results.

from an obligation to relate it in terms of some specific rhetorical relation to some particular recent sentence or clause.)

On the other hand there are certain adverbials which have a rhetorical as well as a temporal dimension. Well-known in this respect are the adverb *then* and subordinate clauses beginning with *when*.[53] *When*-clauses have been argued to do more than just locate the main clause eventuality in time.[54] Whether the semantic contributions made by *then* and various other temporal adverbials are in the first instance temporal or rhetorical, and how their rhetorical or temporal contributions can then be accounted for as secondary effects produced by pragmatic mechanisms, is a matter for further investigation.[55]

Bibliography

Almog, J., Perry, J., Wettstein, H. (eds.), 1989. *Themes from Kaplan*. Oxford University Press, Oxford.

Asher, N., 1993. *Reference to Abstract Objects in Discourse*. Kluwer, Dordrecht.

Asher, N., 1996. Mathematical treatments of discourse contexts. In: P. Dekker, M. Stokhof (eds.), *Proceedings of the 10th Amsterdam Colloquium on Formal Semantics*. Vol. 1. ILLC Publications, Amsterdam, 21–40.

Dekker, P., 1993. *Transsentential Mediations, Ups and Downs in Dynamic Semantics*. Ph.D. dissertation, University of Amsterdam.

Groenendijk, J., Stokhof, M., 1990. Dynamic Montague Grammar. In: L. Kalman, L. Pólos (eds.), *Papers from the Second Symposion on Logic and Language, Budapest*. Eotvos Lorand University Press, Akademiai Kiadeo, 3–48.

Groenendijk, J., Stokhof, M., 1991. Dynamic predicate logic. *Linguistics and Philosophy* 14, 39–100.

Heim, I., 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. dissertation, University of Massachusetts, Amherst, MA.

Hinrichs, E., 1986. Temporal anaphora in discourses of English. *Linguistics and Philosophy* 9, 63–82.

Kamp, H., 1981. Événements, représentations discursives et référence temporelle. *Langages* 64, 39–64.

Kamp, H., 1999. *A DRT-based Formalism for the Representation of Attitudinal States and Indirect Discourse*, unpublished manuscript, University of California, Los Angeles, CA. Available online: http://www.ims.uni-stuttgart.de/archiv/kamp/.

---

[53] Such subordiate clauses also act as additional specifiers of the eventualities described by the main clauses to which they are adjoined, and as such they are on a par with adverbs and adverbially used prepositional phrases.

[54] See the references to Steedman (1997) and Moens & Steedman (1988).

[55] In this connection too, the reader is recommended to consult the forthcoming work by Roßdeutscher cited in the Bibliography and further works cited there.

Kamp, H., van Genabith, J., Reyle, U., 2011. Discourse Representation Theory. In: D. Gabbay, F. Guenthner (eds.), *Handbook of Philosophical Logic*, 2nd Edition. Vol. 15. Springer, Berlin, 123–394.

Kamp, H., Reyle, U., 1993. *From Discourse to Logic. Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer, Dordrecht.

Kamp, H., Rohrer, C., 1984/86. *Temporal Reference in French*, unpublished manuscript, Bedford College, London/CSLI, Stanford University, Stanford, CA/University of Stuttgart, Stuttgart. To be published with CSLI Publications, Stanford, CA, 2013/14. Available online: http://www.ims.uni-stuttgart.de/archiv/kamp/.

Kaplan, D., 1989. Demonstratives. An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. In: J. Almog, J. Perry, H. Wettstein (eds.), *Themes from Kaplan*. Oxford University Press, Oxford, 481–563.

Lascarides, A., Asher, N., 1993. Temporal interpretations, discourse relations and commonsense entailment. *Linguistics and Philosophy* 16, 437–493.

Mann, W., Thompson, S., 1987. Rhetorical structure theory: Description and construction of text structures. In: G. Kempen (ed.), *Natural Language Generation. New Results in Artificial Intelligence, Psychology, and Linguistics*. Martinus Nijhoff Publishers, Dordrecht, 85–95.

Merin, A., 1996. *Die Relevanz der Relevanz: Fallstudie zur formalen Semantik der englischen Konjunktion 'but'*. Habilitation thesis, University of Stuttgart. Reprinted as Vol. 142 of Arbeitspapiere des Sonderforschungsbereichs 340, Sprachtheoretische Grundlagen der Computerlinguistik, University of Stuttgart, Stuttgart, 1999.

Moens, M., Steedman, M., 1988. Temporal ontology and temporal reference. *Computational Linguistics* 14, 15–28.

Montague, R., 1974. The proper treatment of quantification in ordinary English. In: R. Montague, *Formal Philosophy: Selected Papers by Richard Montague*, ed. by R. Thomason. Yale University Press, New Haven, CT, 247–270.

Muskens, R., 1996. Combining Montague Semantics and Discourse Representation. *Linguistics and Philosophy* 19, 143–186.

Partee, B.H., 1984. Nominal and temporal anaphora. *Linguistics and Philosophy* 7, 243–286.

Partee, B.H., 1973. Some structural analogies between tenses and pronouns in English. *The Journal of Philosophy* 70, 601–609. Reprinted in: B.H. Partee, *Compositionality in Formal Sematics: Selected Papers by Barbara H. Partee*. Blackwell, Oxford, 2003, 50–58.

Pinkal, M., 1996. Radical underspecification. In: P. Dekker, M. Stokhof (eds.), *Proceedings of the 10th Amsterdam Colloquium on Formal Semantics*. ILLC Publications, Amsterdam, 587–606.

Prior, A., 1967. *Past, Present and Future*. Oxford University Press, Oxford.

Reichenbach, H., 1947. *Elements of Symbolic Logic*. Macmillan Co, New York.

Reyle, U., 1993. Dealing with ambiguities by underspecification: Construction, representation and deduction. *Journal of Semantics* 10, 123–179.

Roßdeutscher, A., 2000. *Lexikalisch gestützte formale Textinterpretation*. Arbeitspapiere des Sonderforschungsbereichs 340, Sprachtheoretische Grundlagen der Computerlinguistik, Vol. 157. University of Stuttgart, Stuttgart.

van der Sandt, R., 1992. Presupposition projection as anaphora resolution. *Journal of Semantics* 9, 333–377.

van der Sandt, R., Geurts, B., 1991. Presupposition, anaphora, and lexical content. In: O. Herzog, C.-R. Rollinger (eds.), *Text Understanding in LILOG Integrating Computational Linguistics and Artificial Intelligence Final Report on the IBM Germany LILOG-Project*. Springer, Berlin, 259–296.

Smith, C., 1991. *The Parameter of Aspect*. Kluwer, Dordrecht.

Steedman, M., 1997. Temporality. In: J. van Benthem, A. ter Meulen (eds.), *Handbook of Logic and Language*. Elsevier, Amsterdam, 893–938.

Verkuyl, H., 1972. On the compositional nature of the aspects. *Foundations of Language, Suppl. Series* 15, 40–109.

Visser, A., Vermeulen, C., 1997. Dynamic bracketing and discourse representation. *Notre Dame Journal of Formal Logic* 37, 321–365.

Vlach, F., 1981. The semantics of the progressive. In: P. Tedeschi, A. Zaenen (eds.), *Syntax and Semantics 14: Tense and Aspect*. Academic Press, New York, 271–292.

PART TWO

SEMANTICS AND PRAGMATICS

INTRODUCTION TO SEMANTICS AND PRAGMATICS

Hans Kamp

The majority of papers in this collection touches on the relations between semantics and pragmatics in some way. This is true in particular of all those that deal with aspects of the development and applications of DRT, but it also applies to some of the other papers, besides the two that constitute this part. The central underlying issues are what mechanisms are involved in obtaining the intended interpretations of utterances, what information these mechanisms draw on and what the logical order is in which they must be applied. Considerations of this general sort shape the notions that linguists and philosophers entertain of what divides semantics and pragmatics. The classical classification, if I am right, is still that proposed by Morris in the nineteen thirties, according to which semantics deals with linguistic meaning and pragmatics with the use of language. But it is not clear that what the papers in this section are concerned with and many of the other papers in this collection also make contact with in one way or another is quite that distinction. The question is not easy to answer because there appears to be a good deal of unclarity about the relation between *linguistic meaning* and *language interpretation*. Problems pertaining to that relation are highlighted by the notion of Gricean implicature, to take just one instance. Are Gricean implicatures ever part of the linguistic meaning of the sentential utterances that give rise to them? I suppose that the standard answer, and the one that Paul Grice himself endorsed, was '*no*'. Recent discussions of the conventionalisation of certain implicatures might suggest a different answer, or a more differentiated one. But what matters for the discussion right now is the question what can be a reason for wanting to say that some aspect of utterance meaning is pragmatic, or rather that it is semantic.

Here is my answer to that question: an aspect of the interpretation of an utterance belongs to pragmatics if it becomes part of the interpretation via a mechanism that requires some aspect of the use of the uttered expression, and does not just operate on its properties as a mere expression (i.e. as an abstract morpho-syntactic object). That, I think, is the guiding principle for drawing the line between semantic and pragmatic aspects of utterance interpretation that most people concerned with meaning in natural

© HANS KAMP, 2013 | DOI:10.1163/9789004252882_007
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

language would nowadays go by or that they would acknowledge as a guiding principle, when the question is put to them.

One consequence of this way of making the distinction between semantics and pragmatics is a striking asymmetry. Semantic mechanisms are those that can be defined without reference to any kind of feature of language use. That makes 'semantic' a unified (if negatively characterized) concept. On the other hand, uses of language have many different kinds of properties. So it is possible to distinguish at least as many different kinds of pragmatics and that is, in fact, what we see when we look at how the field of linguistic pragmatics is partitioned into different subfields. There are good reasons for such divisions. For among the things we have learned about natural language interpretation in the course of the past fifty years is that different mechanisms, and the different expressions and constructions that activate them, exploit different kinds of contextual, utterance-related information. This state of affairs reinforces a perception that I think I share with many, viz. that pragmatics naturally divides into different 'chapters', concerned with different kinds of utterance-related or contextual information, with the mechanisms that exploit those kinds of information, and with the parts or aspects of expressions that trigger those mechanisms. But for the most part such pragmatic interpretation mechanisms do not only make use of certain kinds of contextual information. They typically also build on aspects of the interpretation that have already been established through the application of other mechanisms to the linguistic input that the interpreter receives. This imposes a certain logical order on the application of the different mechanisms that the interpretation of any given utterance calls upon. Since it is part of the characteristics of particular mechanisms that they require inputs of certain specific sorts, this logical order is a relation imposed on the mechanisms themselves, and not only on their individual applications in the interpretations of particular utterances for which they are jointly required. There is a certain partial ordering between interpretation mechanisms and every application of those mechanisms in any complex utterance interpretation event will have to be consistent with that order. Among the mechanisms involved in interpretation there are in particular those that do not involve contextual information of any kind, hence qualify as (purely) semantic mechanisms. For mainstream formal semanticists the paradigm of such a mechanism is a truth definition that operates directly on the syntactic structure of the uttered sentence, yielding a set of truth conditions, or a semantically expressed proposition. This output can then serve as input to further mechanisms of a pragmatic nature.

Although the papers in this part antedate Discourse Representation Theory, I will allow myself to say a few words at this point about DRT in relation to this general discussion about what semantics and pragmatics are and how they might be thought to interact. A partial excuse is that the second paper, *Semantics versus Pragmatics*, was written at a time when the problems it overtly addresses—how are we to draw the line between semantics and pragmatics, if we want to draw it at all—were much on my mind and when DRT emerged only a little later, it seemed to carry with it at least a partial answer to this question. DRT provides a different picture of the relation between syntax and semantics than the one I hinted at above. As far as semantics is concerned, it offers what may look a little more like a mechanism than does a compositional truth definition, viz. its various DRS construction algorithms. Such an algorithm delivers, in conjunction with the truth definition for the DRS-language to which the constructed DRS belongs, the truth conditions for the represented utterance, possibly as a basis for further pragmatic processing. But the line-up between DRT and what I have just called the mainstream conception is, notoriously, not straightforward. The DRS construction algorithm of classical DRT includes the resolution of anaphoric pronouns to sentence-external, as well as sentence-internal antecedents. According to the mainstream view this is an intrusion of pragmatics into semantics, certainly for the cases of trans-sentential anaphora and, according to what seems the dominant view today, also for sentence-internal cases of donkey-type anaphora. For the trans-sentential cases this view follows from the principle that the unit of semantic analysis is the sentence, since on that view context, in the form of the sentence containing the antecedent, is needed to interpret the pronoun. DRT is in a position to classify such cases as semantic, because part of the general view it represents is that the unit of semantic analysis is not the single sentence, but the (typically multi-sentence) discourse. And indeed, DRT has often been described as innovative in redrawing the boundaries between semantics and pragmatics, incorporating into semantics what was previously held (and, as I just said, is still held by many today), to belong to pragmatics.

For presuppositional DRT, in which preliminary DRSs are constructed for single (uttered) sentences and then, during the second stage of the processing, the presuppositions of this preliminary representation are resolved within the given context, the picture is different from what it was for earlier versions (such as that of ***From Discourse to Logic***). As far as the semantics-pragmatics distinction goes, the new architecture allows for a straightforward and clean interpretation: the first stage—construction of the preliminary DRS—constitutes the syntax-semantics interface, and the second—

converting the preliminary DRS into a completed DRS through presupposition justification—is the interface between semantics and pragmatics. After the complete DRS has been constructed, it becomes available for the application of further pragmatic mechanisms; integration of this DRS with the context DRS should best be thought of as involving the application of such further pragmatic mechanisms, at least in those cases where the integration doesn't happen automatically, as a kind of default.

But there is at least one salient possible exception to this principle of semantic priority, according to which the application of semantic mechanisms must precede the applications of pragmatic ones. It has to do with the role of referring terms and is perhaps most clearly visible on the Strawsonian view that making a statement involves a combination of two linguistic acts, one of referring and one of predicating. On this simple view of what it is to make a statement or assertion, pragmatics may play its part in determining what the referring term refers to, before the presumably semantic operation of predicating something of that referent is possible. In such cases at least some operations involving pragmatic principles may precede some semantic operations. This same question about the order in which semantics and pragmatics enter into utterance interpretation is also raised, and for much the same reason, by what Keith Donnellan has to say about the referential use of definite descriptions. Here too, pragmatics comes into play, possibly in conjunction with semantics, in determining the referent. Only afterwards the referent can be used in an application of the presumably semantic predication operation. David Kaplan's suggestion that the identification of the referents of indexicals like *I* and *now*, requiring the pragmatic input of the utterance context, comes first and evaluation of the simple and complex predications that those referents are involved in follows, also raises this same issue.

This question about the interleaving of what are arguably semantic and pragmatic aspects of interpretation is the central concern of the two papers in this part, *Free Choice Permission* and *Semantics versus Pragmatics*. These papers are more closely connected than any other pair in the present collection. *Semantics versus Pragmatics* is an attempt to explore the methodological implications of its predecessor *Free Choice Permission* for the ways in which semantics and pragmatics could be related. The particular issue that *Free Choice Permission* raises is of a type that wasn't touched upon above, since the pragmatic notions involved—those having to do with Speech Act Theory—played no part in the discussion. But the point that needs making here can be made regardless: the interleaving implicit in the semantic proposal made in *Free Choice Permission* is of an unusually severe sort. In order

to compute the content of a disjunctive permission statement like *You may go to the beach or go to the cinema*, one first has to compute the sets of options made available to the addressee by the disjuncts *go to the beach* and *go to the cinema*, respectively, and then apply the semantic operation expressed by *or*, which, the paper assumes, is always set-theoretic union to this pair of sets. And the result of this procedure may then be in its turn part of the interpretation of some more complex sentences, as when the addressee says to himself in reaction to what he has just been told: *if I may go to the beach or to the cinema, I will go to the cinema*.

Especially the first of these two papers has attracted some attention in recent years, in the context of a renewed or intensified interest in the interactions between semantics and speech act type determination, and also in relation to new attempts to arrive at a general account of when disjunctions may have the apparent force of conjunctions. But in none of the more recent work that I have seen has the specific issue that the proposal of *Free Choice Permission* brings into play and that *Semantics versus Pragmatics* tries to come to grips with, been frontally addressed: the idea that what is uniform about disjunction and its cognates in other languages is that it always expresses set union, and that what varies are the roles played by the sets. The more general idea, of which this is one manifestation, is that Boolean structure, as conveyed in English by the words *not*, *and* and *or* (and some other words and constructions that are semantically definable with the help of these) constitutes a semantic foundation for language that pervades all the different uses we can make of it in performing different types of speech acts. *Free Choice Permission* can be taken as making the more general claim that this can take forms as dramatic as the account it suggests for the disjunctive permission statements it discusses. Perhaps this claim is wrong; but I still haven't seen a convincing argument that it must be.

FREE CHOICE PERMISSION[a,*]

Hans Kamp

§1

(1) 'You may go to the beach or go to the cinema',
I almost told my son Michael. But I thought better of it, and said:
(2) 'You may go to the beach.'
Boys shouldn't spend their afternoons in the stuffy dark of a cinema, especially not with such lovely weather as today's.

Thus what I did in fact permit was less than what I first intended to permit. We might even be inclined to say that the permission I contemplated, entailed, but was not entailed by, the permission I gave.

This indeed has been the verdict of most people to whom I have put the question whether there is a relation of entailment between these two permission statements.

But what exactly could we mean by saying that there is a relation of entailment between statements which—at least at first view—are not statements of fact? Logical analysis has provided us with a basically sound understanding of entailment between assertions.[1] Very roughly, one assertion *entails* another if and only if in all possible states of affairs (or possible situations, or possible worlds) in which the first assertion would be true the second would be true also. What makes this characterization compelling is the comprehensive and precise logical theory into which it is incorporated; but I will not go into the details of that theory now.

---

[*] Meeting of the Aristotelian Society at 5/7 Tavistock Place, London, WC1, on Monday, 26th November, 1973, at 7.30 p.m.

[1] I will use the expressions 'statement of fact', 'assertion' and 'assertive statement' interchangeably.

---

[a] Editors' note: This article was originally published in the *Proceedings of the Aristotelian Society*, 74, 1973, 57–74. It is reprinted here by courtesy of the Editor of the *Aristotelian Society* © 1973. We carefully edited the article, corrected some typos, added full bibliographical references and slightly adjusted the layout. We would like to thank Tyler Wilbers for careful proofreading and helpful comments.

© HANS KAMP, 2013 | DOI:10.1163/9789004252882_008
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

No doubt the success of this analysis of entailment between assertions has been at least partly responsible for the tendency to try to understand the logical structure of non-assertive discourse by reducing it to assertions. Such a reduction has been suggested, in particular, in connection with permissions and commands. The latter are to be reduced to assertions of obligation and the former to assertions of permission—statements which, rather than creating permissions, claim that certain permissions are in force. The logical relations between these various assertive statements are then assumed to hold also between their non-assertive originals.

According to this recipe (1) and (2), in particular, reduce to the assertive statements:

(3) 'It is permitted (to Michael) to go to the beach or to go to the cinema'; and
(4) 'It is permitted (to Michael) to go to the beach'.

Besides the logical features which statements of obligation and permission share with all other assertions—those properties which we may reveal through formalization into, say, the predicate calculus—there are others of a more specific nature. These depend essentially on the meaning of such expressions as 'has the obligation to', 'ought', 'it is permitted', 'it is forbidden', and others:—those expressions, whose logical properties are analyzed by what is currently called *deontic logic*.

§ 2

The past twenty years have seen the rise of a variety of formal systems of deontic logic—and the fall of some, but by no means enough of them. I will discuss here only one system, perhaps the best known, and probably of the least controversial. It is also one of the oldest. It was developed by von Wright in the early 50's. The version I will present here does not entirely coincide with von Wright's formulation. I will refer to the present version as $\mathcal{D}$.

$\mathcal{D}$ is a simple extension of the ordinary sentential calculus. Let the vocabulary of the latter consist of the sentence letters $p, q, r, \ldots$ and the sentential connectives $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$ (with their usual meanings); its *sentences* are defined in the obvious, traditional way. $\mathcal{D}$ is obtained by adding to this language one new symbol, O. O is a one-place sentential operator. Thus if ϕ is a sentence of $\mathcal{D}$, then so is $O\phi$. $O\phi$ should be read as 'it ought to be the case that ϕ', or alternatively, as 'it is obligatory to see to it that ϕ' or 'x has the obligation to see to it that ϕ' (for some fixed person x), depending on the particular use to which we want to put the system. Entailment between sentences of $\mathcal{D}$

is characterized in terms of derivations: we say that a sentence ϕ is *entailed* by the sentences $\phi_1, \ldots, \phi_n$ just in case there is a derivation of the sentence $(\phi_1 \wedge \ldots \wedge \phi_n) \rightarrow \phi$; as usual a derivation of a sentence is a list of sentences each of which either is an axiom or comes from one or more preceding lines by an application of a rule of inference.

Axioms and rules of $\mathcal{D}$ include a complete set for the classical sentential calculus; in particular, $\mathcal{D}$ has the rules of *Modus Ponens* and *Substitution*; and its axioms comprise a set which together with these two rules gives all classical tautologies. Besides these rules and axioms there are some whose acceptability depends essentially on the intended meaning of O. It is precisely through such additional axioms and rules that systems of deontic logic become interesting. But it is also there precisely that they usually become dubious; for hardly any of the various proposed axioms and rules are acceptable for all the intended applications.

There is only one principle which appears to be sound on any plausible understanding of 'ought' or 'obligatory'. It is the principle that if ϕ and ψ are provably equivalent then so are $O\phi$ and $O\psi$. This principle, i.e., the inference rule

$$(5) \quad \frac{\phi \leftrightarrow \psi}{O\phi \leftrightarrow O\psi}$$

is indeed a rule of $\mathcal{D}$. In addition I adopt for $\mathcal{D}$ one new axiom, the sentence

$$(6) \quad O(p \wedge q) \rightarrow Op,$$

which, ostensibly, says that if the conjunction of p and q is obligatory then so is p by itself.

This principle is certainly not valid in all of the intended interpretations of $\mathcal{D}$. It is not valid, for example, if we read O as 'it ought to be the case that', and understand 'ought' as 'ought, no matter what'. For it could be the case that independently of whatever else may happen the realization of both p and q would bring about a better state of affairs than would ensue if the conjunction were not realized, while the realization of p in the absence of q would mean disaster. Perhaps the man standing under the first floor window of the burning house ought, no matter what, to tell the frightened girl behind it to jump, and catch her in his arms when she does. But if he cannot catch her, or doesn't want to risk his bones, he certainly ought *not* to tell her to jump.

Formally, p ought to be the case, no matter what, if every possible world in which p holds is better than any world in which p is false but which in all other respects (or as many as possible, given that p is false) is the same as the first. Clearly this condition may hold for $p \wedge q$ without holding for p.

There are systems in which, precisely for this reason, the axiom is not valid.[2] Here, however, we need a slightly different interpretation of $\mathcal{D}$, in which Op is understood as 'x has the obligation to see to it that p'—i.e., that p's realization is among x's various obligations. If x fails to observe certain other obligations his realizing p may do more harm than good; and perhaps he deserves *blame* for the thoughtlessness of his realizing p under those conditions, over and above the blame he already deserves for his failure to comply with these other obligations; but for all that, p is no less among x's obligations.

Formally, p is an obligation of x (in the here intended sense) iff p is true in all those worlds in which x fulfills all his obligations. So if $p \wedge q$ is among x's obligations then $p \wedge q$ is true in all the worlds in which he fulfills all his duties, and so, consequently, is p. Thus p is an obligation of x as well.

Within $\mathcal{D}$ we can represent certain deontic concepts other than what is rendered by the operator O itself. We can for example supplement $\mathcal{D}$ with the sentential operator P, with intended reading 'it is permitted that', or '(x) has permission to make it true that' and add the sentence

(7) $\quad Pp \leftrightarrow \neg O \neg p$

to the axioms of $\mathcal{D}$. This axiom expresses the view that for any sentence p, it is permitted that p be (made) true just in case it is not obligatory that its negation be (made) true.

This view seems sound enough. Notice however that in the extended system we can derive

(8) $\quad Pp \rightarrow P(p \vee q)$;

on the other hand the formula

(9) $\quad P(p \vee q) \rightarrow Pp$;

is not derivable. Moreover we cannot really make (9) valid by fiat, viz., by adding it as axiom, or by strengthening the axiom system of $\mathcal{D}$ in such a way that (9) becomes derivable. For from (9) we could derive with the help of (7) and the rules and axioms already present in $\mathcal{D}$

(10) $\quad Op \rightarrow O(p \wedge q)$

and this last formula is clearly unacceptable. For among the intended interpretations of $\mathcal{D}$ there must be some in which x has the obligation to see to it

[2] Cf. David Lewis, *Counterfactuals*, Oxford, 1973.

that p be true but not that both p and q be true (p may stand for something good, and q for something evil!)

§ 3

I have given this somewhat lengthy exposé of the system $\mathcal{D}$ not so much because of any intrinsic interest in the system itself, but in order to show that its axioms and rules are valid on the interpretation which I said is here intended; so that when presently we will arrive at unpalatable results when applying $\mathcal{D}$ to our problem concerning the permission statements (1) and (2) we will be sure that we cannot blame these results on an inadequacy of $\mathcal{D}$ with respect to its originally intended application to assertions of permission and obligation.

What can $\mathcal{D}$ tell us about the logical relations between (1) and (2)?—Or, rather, between the assertions (3) and (4) to which according to the method under discussion, they can be reduced? This depends on how (3) and (4) should be formally represented in $\mathcal{D}$. As is common in applications of formal languages to the analysis of natural discourse there is no general and precisely defined canon for translating the ordinary vernacular into $\mathcal{D}$. However, established practice would have it, apparently, that they may be represented as, say,

(11) $P(p \vee q)$; and
(12) Pq,

respectively, where p is short for 'Michael goes to the beach' and q for 'Michael goes to the cinema'.

But, as we have seen, in $\mathcal{D}$ it is (12) which entails (11) and not (11) that entails (12). So if we accept the above symbolization we are led to the conclusion that, similarly, (4) entails (3) but not conversely; finally by the proposed reduction method, we are made to conclude that (2) entails (1) and that (1) does not entail (2).

This is in direct opposition to the intuition concerning (1) and (2) which I stated in the very beginning of this paper. This conflict between our intuition and the verdicts of Deontic Logic is well-known. Von Wright[3] calls it the '*Paradox* of Free Choice Permission'. None of the solutions that he himself offers for the paradox, nor any other that I have come across, satisfies me.

---

[3] G. H. von Wright, *An Essay on Deontic Logic and the Theory of Action*, Amsterdam 1969. Chapter I, Section 7.

The failure of these solutions indicates in my opinion that the very attempt to grasp the logical properties of permission statements by reducing them to assertions is misguided.

I will, instead, give an analysis of the function of permission statements which will explain why and in what way our intuition concerning (1) and (2) is legitimate.

§ 4

What exactly is the function of permission statements? It will be useful to concentrate on a particular situation in which such a statement can be used successfully. Let us elaborate on the context of (1) and (2). That situation involves two persons, the speaker (me) and the addressee, Michael. What enables me to extend a permission to Michael is a certain authority I have over him. This relation of authority is characterized by a variety of standing prohibitions—things which Michael is not supposed to do—not at least unless I tell him. It is only with reference to these prohibitions that my permission statement can fulfill the function it has: to remove some of these prohibitions from the list—for a short while, or, sometimes, for good. Thus a permission statement, when it is successful, moves a certain class of actions from the realm of the prohibited into that of the permitted.

But how is this class determined? What, for example, are the actions rendered permissible by my statement (2)? Surely not *every* way of making the statement 'Michael goes to the beach' true. He is not to go by taxi; he is not to go to that section of the beach where there is no swimming guard. He is not to put on the coat which his mother bought him last month and which he is only to wear on special occasions and with explicit permission. And of course, as always he is to keep out or mischief and not to engage in behaviour that is offensive to others. All these ways of going to the beach remain proscribed; for they would either violate a standing prohibition which my permission statement has not lifted or else some of the laws, moral rules or codes of conduct which govern the lives of all of us—as citizens, social beings and persons; those laws, codes and rules I do not even have the power to modify.

What I have just said about this particular situation applies in essence to all natural contexts for permission statements. Let me try to give a preliminary schematic account of these essential features.

We will concentrate on situations involving only two persons, *A* and *B*. They are members of a society and as such subject to a variety of prohibi-

tions of either legal or moral nature. We moreover assume that A has over B certain authority. This relation involves some further prohibitions which B must observe. Some, and perhaps all, of these A has the power to rescind.

But what are prohibitions? And what is it to rescind, or *lift* a prohibition? The answer I would like to give to these questions are these: Prohibitions are generic actions, or action types. Besides generic actions there are individual actions; these stand to generic actions as material objects stand to properties of material objects. A generic action P is a prohibition for a person B if B may not perform any individual action, which falls under P; an individual action is prohibited to B if it falls under a prohibition for B; and permissible iff it is not prohibited. To lift a prohibition is to render permissible a class of individual actions which fall under the prohibition—and thus were previously prohibited.

However, these definitions will serve our purpose only if we countenance among individual actions not only those which are in fact performed but also actions which are merely possible, i.e., never actually performed at all. Though I regard it as conceivable that such an ontology could be put on a solid footing, I have no foundation to offer now. And even if I had it would probably take too much time to expound it here.

The doctrine which I will offer instead has perhaps a less intuitive ring to it. But it avoids the tangles of the general theory of action; and for our present purposes it will do very well indeed. According to this doctrine prohibitions are propositions. The prohibition for Michael to go to the beach, e.g., is the proposition 'Michael goes to the beach'. Of course not every proposition is a plausible candidate for being a prohibition, cf. 'the earth is round'. The propositions we should think of in this connexion are primarily those which express that the addressee engages in a certain action or activity. But it is not necessary for our purpose to give an explicit characterization of the class of those propositions which are possible prohibitions. That the proposition p is a prohibition for B (at t) means that B is prohibited from realizing any possible world in which p is true. These possible worlds will then also be called prohibited; while the remaining worlds are called permitted. To lift p is to render permissible a class of worlds in which p is true and which, thus, were previously prohibited.

In the sequel I will sometimes for convenience's sake talk about generic and individual actions, rather than about propositions and possible worlds. However, there will never be any problem in converting to the other terminology.

Let $p_1^S, \ldots, p_m^S$ be the social, legal and moral prohibitions to which B is subject and $p_1^A, \ldots, p_n^A$ those that are specific to the authority relation

between A and B. Let us first assume that p_i^A is independent of all the other p_j^A and all the p_j^S in that there are individual actions which fall under p_i^A but not under any of the remaining prohibitions and which are *open* to B in the given situation, i.e., it is possible for B to perform them. This assumption, though it is, as we will soon see, not always satisfied, holds true for the most straightforward cases of permission giving and considerably simplifies the analysis. Under this hypothesis, lifting the prohibition p_i^A will have the effect of rendering permissible the class of all those previously prohibited worlds in which p_i^A is true at some time following the statement and in which none of the remaining prohibitions is true at any such time.[4] Let us denote this class as $C(p_i^A)$.

In the context of the permission statements (1) and (2) one of the standing prohibitions was against going to the beach and another against going to the cinema. Let us assume these prohibitions to be p_1^A and p_2^A, respectively. My statement (2) will then have the effect of moving from the realm of the prohibited into that of the permitted class $C(p_1^A)$. But which class would have been moved by (1)? According to our first intuition it should be a larger class than $C(p_1^A)$, presumably something like the union of $C(p_1^A)$ and $C(p_2^A)$. Why should that be so? Well, this is precisely what we should expect, given the meaning of the word 'or'. For, as logicians have realized with regard to assertions for at least a century, 'or' stands essentially for set-theoretic union: The set of possible situations in which a disjunctive assertion is true is the union of the sets of possible situations which realize its disjuncts.

'Or', I maintain, *always* stands for set-theoretic union. Previously we could not see that it does have this function also in disjunctive permission statements because we lacked the appropriate understanding of what permission statements do. And so we saw ourselves forced into the implausible and totally *ad hoc* doctrine, that in such statements 'or'—usually at least—plays the same rôle that 'and' plays in assertions.

We can now also understand the nature of our intuition concerning the entailment between (1) and (2). (1) would have moved a larger class of actions than did (2); thus it would have had a greater effect; it is, in plausible terms, a stronger permission than (2). Indeed, (1) is *logically* stronger than (or better, *logically at least as strong as*) (2) in so far as in any permission context the set of actions moved by it includes the set of actions moved by

[4] This characterization is of course only correct on the assumption that B is not subject to the authority of any individuals besides A, an assumption that will be in force throughout this paper.

(2). Thus, if we introduce the term '*p*(ermission)–entails' for this relation, we may put for arbitrary sentences *s* and *s′*: *s p*–entails *s′* iff in in any situation in which both *s* and *s′* can be used to grant permissions the class of actions which would be moved by a permission statement involving *s* would include the class of actions moved by a permission statement involving *s′*.[5]

Thus if *s p*–entails *s′*, then whatever is rendered permissible by the use of *s′* is necessarily also rendered permissible by the use of *s*.

Just as the significance of *p*–entailment derives from the nature of permission giving, so does the importance of entailment between assertions—as characterized in section 1—derive from the nature of assertion. To assert *s* is to claim that the actual world is among the various possible situations compatible with *s*; thus the smaller the class of compatible situations the stronger the assertion—for it is *a priori* less likely that the actual world will belong to the class. And, furthermore if the class of those situations compatible with *s* is included in the class of those compatible with *s′* then if the assertion of *s* is correct, i.e., the actual world does belong to the first class, then it must also belong to the second and thus the assertion of *s′* must be correct too.

§ 5

But matters aren't quite as simple as the last section may have suggested. Indeed, pursuing the analysis I there began, one encounters a great many difficulties. Of these, and of the new insights that dealing with them has given me, I can, in the space left to me not discuss more than a minute portion. Nor will I have the opportunity to integrate these various observations into a comprehensive formal semantics for a fragment of English which contains a substantial variety of those sentences commonly used in permission statements. I hope to give a detailed presentation of this formal theory on some other occasion.

The first problem is this. Not all disjunctive permission statements have the force of (1). Consider a king telling his vassal:

(13) 'You may pillage city *X* or city *Y*. But first take counsel with my secretary.'

The second part of this statement makes it clear that the vassal should not infer from the first part that he may make his own choice of city. Which

[5] This definition is clumsy and sloppy. A formally precise version requires a technical framework which I cannot develop here.

one he may loot ultimately depends on the secretary's advice, the tenour of which—we may assume—is at this point unknown to king and vassal alike. Thus the first part of (13) sets, as it were, the outer limits of the permission without giving *carte blanche* for everything lying within these boundaries.

I believe that the proper way to look at the contrast between (13) and (1) is this. (13) does not, in the manner of (1), remove both its disjuncts from the list of standing prohibitions; rather it removes the disjunction as a whole. This removal will be vacuous however if prohibitions against the separate disjuncts remain in force. (For there is no world in which the disjunction is true but in which both disjuncts are false.) Thus if a permission statement such as (13) is to have any effect at all it must at least partially lift the prohibition against one of the disjuncts. The contextual factors that determine which disjunct is lifted are complex. I will not go into this analysis. Suffice it to remark that in many cases the prohibitions against the disjuncts have different weight: one is *stricter* than the other (e.g., enforced by means of more severe penalties). In such a situation it is in general the weaker of the two which the disjunctive statement lifts.

It should be noticed that this example shows the limitations of our earlier assumption concerning the independence of the prohibitions p_i^A. For the disjunction removed by (13) is *not* independent of its two disjuncts; while all three are standing prohibitions. Indeed, the assumption seems not to hold in most complex cases. It fails, in particular, in all examples yet to be discussed.

The contrast between (1) and (13) has an interesting parallel among quantified permission statements. Consider:

(14) 'You may pick some flowers',
(15) 'You may pick any flowers',
(16) 'You may pick all flowers'.

I claim that all three are different in force. The difference between (14) and (15) can be brought out as follows: Suppose that the authority relation between Mary and her mother is characterized by a prohibition against picking whatever flowers the garden contains, together with a stricter prohibition against picking roses. In such a situation (14) will normally be understood as lifting the prohibition against picking flowers, but not that against picking roses, while (15) will be understood as permitting the picking of no matter which flowers, roses included. (15), on the other hand, is also different from (16). This becomes clear if we consider a situation in which besides a weak prohibition against picking a few flowers there is a strong prohibition against picking the whole lot. (16), in what is in my opinion its more natural

sense, then removes the strong, as well as the weak, prohibition, while (15) only removes the latter.

'Any' is a word that has puzzled many—linguists and logicians alike. The puzzlement arises at least in part from the fact that the word sometimes appears to behave as an existential and sometimes as a universal quantifier. In many occurrences it could be either. Thus in

(17) 'If anyone will solve this problem I will'; and
(18) 'I haven't solved any problem',

'any' may be regarded either as (i) an existential quantifier with narrow, or (ii) as a universal quantifier with wide, scope.

There are also cases, such as

(19) 'Anyone can solve this problem'; and
(20) 'Anyone who has seen Naples has seen enough',

which appear to show unambiguously that 'any' is a universal quantifier. And I presume it is these examples which convinced certain logicians that the second view is the correct one. Yet on the other hand we have

(21) 'Has anyone come?',

which means the same as 'Has someone come?', and not as 'Has everyone come?'; and

(22) 'I wonder if anyone has come',

which means neither 'I wonder if everyone has come' nor 'It is true of everyone that I wonder if he has come', but rather 'I wonder if someone has come'.

Another curious fact about 'any' is that its distribution shows striking parallels with that of 'or'; and that in most statements containing 'or' which have natural counterparts with 'any', the 'or' can be regarded as having the force of 'and'. Consider for example:

(23) 'If John or Bill will solve this problem, I will.'
(24) 'I haven't solved either this problem or that one',
(25) 'John or Bill can solve this problem',
(26) 'Did John or Bill come?'

We have seen that there are at least some statements, such as (1), in which 'or' appears to have the force of 'and' and yet for which we can give a semantic account even when we accept my claim that 'or' always stands for set-theoretic union. I now want to claim in a similar vein that 'any' also stands invariably for set-theoretic union—but union over a family of sets, rather than between two; further that this is also true of 'some', so that in

this respect 'any' and 'some' are alike; and finally that 'and' stands always for binary intersection, and 'every' and 'all' for intersection over families of classes.

My claim about 'any' will be puzzling in virtue of (19) and (20). In the next section I will briefly indicate why they are not the counter-examples to my view they seem to be.

But let us return to (14) and (15). The contrast between these is essentially the same as that between (1) and (13). And on the view that 'any' is basically an existential quantifier it is just as difficult to explain.

The account I want to propose may not be ultimate. But in any case, I cannot expound it here with the coherence which it acquires when incorporated into the formal theory which, I said earlier, is beyond the scope of this paper. But let me try.

Permission statements contain occurrences of a certain one-place existential operator which are often concealed in the surface structure; I represent this operator by the letter '*F*'. I call *F* the 'focus-operator' as it is its function to 'move' the sub-formula which stands within its scope into focus and thereby subject it (as a whole) to the particular semantic-pragmatic operation which is associated with the mode, or pragmatic function, of the speech act in question. The difference between (1) and (13), in particular becomes then manifest as that between the formulae

(27) $Fp \vee Fq$; and
(28) $F(r \vee s)$.

The effect of using (28) in a permission statement will be to lift the prohibition against $r \vee s$—and what that amounts to precisely depends, as we have seen, on certain aspects of the situation. The effect of using (27) is to rescind the prohibitions p and q individually and to make permissible the union of the two classes of worlds which these separate removals render permissible.

Similarly we may, abbreviating 'x is a flower' as $Q(x)$ and 'Mary picks x' as $P(m, x)$, represent the difference between (14), (15) and (16) as that between

(29) $F(\exists x)(Q(x) \wedge P(m, x))$
(30) $(\exists x)(Q(x) \wedge FP(m, x))$
(31) $F(\forall x)(Q(x) \rightarrow P(m, x))$

respectively. (29) removes the prohibition against the proposition that Mary picks some flower, while (30) renders permissible the union over the family of all classes C_x of worlds which are rendered permissible by rescinding the various propositions of the form 'Mary picks x', where x is a particular flower in the garden. (31) simply rescinds the prohibition against picking all at once, and there with the weaker prohibition against picking a few flowers.

It should be noted that while the formal distinction between (27) and (28) does not manifest itself at all in the English sentences (1) and (13) they stand for, that between (29) and (30) manifests itself in their English counterparts (14) and (15) as the difference between the words 'any' and 'some'. It is this, and the evident similarity of the semantic contrast between (1) and (13) and that between (14) and (15) which support the use of *F* in the symbolization of the latter two.

§ 6

Let me conclude with some very brief remarks concerning (19) and (20). The reason why they do not refute my view of 'any' is connected with a fact about natural language which itself is of much wider importance than the analysis of this particular—and, it seems, specifically English—lexical item: It has been a tacit assumption behind most formal analyses of natural language that at least all assertions have the same pragmatic function. So far I have not challenged that assumption but only attempted to show that it is pointless to try to reduce statements which are not assertions to statements which are. What I want to claim now is that even among what we call assertions we find considerable pragmatic diversity, and that only if we understand the different functions of various types of asserting statement, we may hope to arrive at a correct analysis of the semantics of assertive discourse.

Here are a few examples of what I have in mind:

(I) Statements of possibility, such as (19), have the primary function of moving into the purview of the hearer a class of worlds, all those *plausible* worlds in which what follows the possibility operator is true. Thus (10) should be understood as moving into the hearer's purview the union of all those classes C_x of plausible worlds in which 'ν solves this problem' is true for particular persons x. Contrast this with

(32) 'Someone can solve this problem'

which can be understood as moving into purview the class of plausible worlds in which 'someone solves this problem' holds (and this will be a smaller set if it is plausible that x should solve the problem, while vastly unlikely that y would succeed in doing so.[6])

[6] There is still another analysis of (32) which gives a probably more natural reading of it but which I cannot discuss here.

(II) Conditionals have the primary function to inform the hearer that in any of the plausible worlds in which the antecedent holds the consequent holds as well. Thus they offer, as it were, for consideration a class of worlds—the plausible worlds in which the antecedent holds—and assert that each of these satisfies the consequent. In this light (17) offers for consideration the union of the classes C_x of worlds in which 'x solves this problem' is true and asserts of each of the members of this union that I solve the problem in it.

We can see how a similar analysis leads to the intuitively correct truth conditions of:

(33) 'If everyone will solve this problem, then I will'; and
(34) 'If someone will solve this problem then I will'.

(III) Denials. The function of a denial is to remove from consideration the class of all (plausible) worlds in which what is being denied is true. Thus (18) removes from consideration the union of those classes which are removed by the denials of the individual propositions 'I have solved x' where x is a particular problem.

It should by now be clear how to give parallel analyses of (23)–(25) which explain why 'or' seems to have, in each of them, the force of 'and'.

Of the three remaining examples concerning 'any', viz., (20), (21) and (22), we should expect the last two to be unproblematic; for in them 'any' functions ostensibly as an existential quantifier. But let us have a look at them anyhow. (21) is a question. What is the function of a question, more specifically a yes-no question? It is to offer to the addressee a certain class of possible worlds for the purpose of finding out whether the actual world is among them. To ask (21) is to offer the addressee the union of the various classes C_x of worlds in which the statement 'x has come' are true.

(22), though an assertion, must be understood, I believe, in terms of our analysis of (21). It expresses the fact that I am asking myself if the answer to (21) is 'yes', i.e., it expresses that I have a particular attitude towards a certain class of possible worlds; 'any' helps to determine that class in its usual manner.

There remains (20). This example appears troublesome, partly because it has no obvious counterpart containing 'or'. The solution to its analysis must lie, I believe, in the functional similarity it bears to a certain conditional, viz.,

(35) 'If anyone has seen Naples he has seen enough.'

This suggestion is somewhat supported by the fact that statements of the same form as (20) tend to be unnatural when 'any' is followed by a sortal—,

rather than by a noun-phrase which is normally used to assert or deny a property of individuals previously identified under some other sortal:

(36) 'Any tree lost its leaves last winter',

sounds odd to me—although I could, if need be, make the sense of it which I would guess it had been meant to have, if it had been meant to have any sense at all: that which we associate with the statement:

(37) 'Take any tree you like; what I am telling you is, it lost its leaves last winter.'

But can't we convert each and every categorical statement with a subject of the form 'any so-and-so' into a conditional? Well, indeed we can; this is what logicians have done ever since they began symbolizing sentences of natural language into predicate logic, whether the subjects of these sentences contained 'any', 'every', 'all' or 'each'.

But then isn't our doctrine about 'any' ultimately just the same as the traditional view that 'any' is a universal quantifier with wide scope? I do not think so. Rather, the present line of thought would show that universal quantification can be expressed by two quite different mechanisms; the first uses such operators as 'every' and 'all' which stand for set-theoretical intersection over classes of possible realizations; while the second is a more roundabout process which presents the union over a family of classes of situations with one characteristic and claims of all these that they have a certain other characteristic as well.

But why does (36) mean—if anything—the same as

(38) 'Every tree lost its leaves last winter'

and not the same as 'Some tree lost its leaves last winter'? Doesn't our account of 'any' suggest the latter reading and not the former? No, not quite. For 'any', though similar to 'some' in that both represent set-theoretic union, differs from it in being associated—while 'some' is not—with certain pragmatic operations, each of which has the task to call before the mind families of possible worlds. In the formal theory I mentioned—in which the operator *F* is always an indication that such an operation applies—this difference manifests itself, roughly, in that 'any' always stands outside the scope of *F*, while 'some' does not. In the most natural analysis of sentences of the form '… lost its leaves last year' there is no *F* at all. That is why 'any' is there out of place. If we want to make sense of (36) nonetheless then we have to impose an analysis in which *F* does occur; and this gives us something like (37), which endows (36) with the sense of (38).

What I have said in this last section is meant as food for thought, not as an argument. But even if only half of what I have said is only half true

the semantical analysis of natural language is of fundamentally greater complexity than is imagined by those who believe that we can describe the semantics of at least the declarative parts of natural languages by theories which are based on no other concepts than those of satisfaction and truth. I should want to believe that too. So I wish all this were wrong. Yet I am afraid there may be enough truth in it to render the belief absurd.

SEMANTICS VERSUS PRAGMATICS[a,1]

Hans Kamp

I.

Consider the sentences

(1) You may take an apple,
(2) You may take a pear,

and

(3) You may take an apple or take a pear.

In one sense of 'may' these convey permissions and thus would seem to fall within the province of deontic logic.[2] According to deontic logic the logical forms of (1) and (2) must be

[1] The history of this paper is long and rather unfortunate. In the summer of 1973 I produced a draft, part of which I subsequently incorporated into Kamp (1973). Since then I have given several talks on the material that had remained unpublished, with increasing diffidence and a corresponding inability to persuade the audience. I should mention in particular presentations at the conference in Bad Homburg which led to these proceedings, and to the Oxford Philosophy Society in the winter of 1976. I am grateful for the many criticisms that I have received on those occasions, in particular from Richmond Thomason in Bad Homburg and from Michael Woods in Oxford; they have helped me to face up to my own growing doubts about the solution to the problem of disjunctive permission sentences of which I gave a sketch in Kamp (1973) and about the elaborate theory I had built upon that solution.

[2] I assume that the 'may' of permission and the 'may' of possibility are distinct lexical items. Unless the contrary is stated explicitly (which will happen only once, towards the very end of the paper) the occurrences of 'may' in the examples we shall discuss will always be instances of the first of these two 'may's.

[a] Editors' note: This article was originally published in F. Guenthner & S. Schmidt (eds.), *Formal Semantics and Pragmatics for Natural Languages*, Reidel, Dordrecht, 1978, 255–287. We would like to thank Reidel (now Springer) for its kind permission to reprint the article. We carefully edited the article, corrected some typos, added full bibliographical references and brought some formulas in line with now more common usage. We would like to thank Tyler Wilbers for careful proofreading.

© HANS KAMP, 2013 | DOI:10.1163/9789004252882_009
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

(1′) It is permitted that (*x* take an apple),

and

(2′) It is permitted that (*x* take a pear)

while that of (3) would have to be either

(3′) It is permitted that (*x* take an apple or *x* take a pear),

or alternatively

(3″) It is permitted that (*x* take an apple) or it is permitted that (*x* take a pear)

Many systems of deontic logic postulate that both (3′) and (3″) are logical consequences of both or (1′) and (2′), while no deontic system countenances (1′) and (2′) as consequences of either (3′) or (3″). This appears to be in conflict with the common intuition that it is (3) which 'entails' (1) and (2), rather than the other way round.

Some years ago I sketched a solution to this problem in Kamp (1973). The ideas on which that solution is based are closely related to an unpublished paper by David Lewis (1970), which gives a model-theoretic analysis of the illocutionary force of permissions and commands. Lewis observes that the pragmatic effect of a command is to restrict the options of the addressee. Similarly a permission statement tends to effect a broadening of his options. The question which Lewis asks, and partly answers, is: By how much do particular permissions and commands extend or restrict the addressee's options for action?

Lewis's explicatory model is simple and persuasive. Consider a community of language users consisting, we shall here assume for simplicity, of two persons, *A* and *B*, and in which *A* has authority over *B*. We study the development in time of the options for action open to *B*, and in particular the ways in which these options are affected by permissions and commands that are issued from time to time by *A*.

The effect of a command is easily stated. Suppose that *A* gives *B* at time *t* the command; 'Polish my shoes!' In so doing he limits the options of *B* to those futures in which, at some time after *t*, *B* polishes *A*'s shoes.

By the *options* of an agent, at time *t*, we shall understand here, as the preceding sentence may already have indicated, those possible continuations of the world after *t* in which the agent fulfills all his obligations and forbears doing the things from which he is prohibited. I shall represent these possible continuations in a manner which I have found convenient in another study (Kamp 1978) where I adopted the convention that each possible continuation of world *w* at time *t* is a possible world which spans the full course of

time—from the most distant past to the most distant future—but which, in as much as it is a possible continuation of w after t, coincides with w in all respects throughout the period which precedes t. Among the possible continuations of w after t will of course be in particular w itself; w will be the only such continuation just in case w is completely deterministic from t onwards.

We shall denote the set of possible continuations of w after t as $Pos(w, t)$. Those continuations in which B honors his obligations and does not transgress any prohibitions by which he is constrained form a subset of $Pos(w, t)$, which we shall refer to as $Per(w, t, B)$. *Pos* and *Per* change with time. One source of change is the occurrence of events which are not fully determined by what went before: If such an event e occurs in the world w at t then the worlds w' in which e does not occur or in which it has different features from those it has in w, and which belong to $Pos(w, t')$ for times t' preceding t no longer belong to $Pos(w, t')$ for t' later than t. Indeed, for all we know, at all times something is going on that is not completely determined by the past, in which case *Pos* is shrinking continuously. These changes in *Pos* will in general produce corresponding changes in *Per*.

Besides this probably continuous change in *Per* there are the modifications effected by A's permissions and commands. We will assume that the performance of these speech acts is instantaneous and that the effect they have on *Per* manifests itself first at the hypothetical instant of performance. Thus if A tells B in w at t: "Polish my shoes!" then the effect of this shows up in $Per(w, t, B)$, but not in $Per(w, t', B)$ for any t' before t. To state what the effect of the command on *Per* is we must refer to what $Per(w, t, B)$ would have been had the command been absent. This set we may identify with the limit of the sets $Per(w, t', B)$ for t' approaching t from the past. Now it seems reasonable to assume that there must be some open interval (t_0, t), with t as end-point, during which no commands or permissions are issued. Then the limit of $Per(w, t', B)$ will be equal to $\cap_{t' \in (t_0, t)} Per(w, t', B)$. We shall denote this set as $Per^{-}(w, r, B)$.

The effect of A's command to B at t now comes to this: Let S be the set of all worlds in $Pos(w, t)$ in which B polishes A's shoes at some time later than t. Then

(4) $\quad Per(w, t, B) = Per(w, t, B) \cap S$[3]

---

[3] One could object to (4) on the ground that commands often require compliance within a limited time span. Sometimes the limit is given explicitly, such as when A says: 'Polish my shoes before tomorrow morning'. But just as often the command itself contains no such specific reference to time, and in these cases the limit is essentially vague in as much as

A command has thus the effect of modifying the set *Per* into a new set which results from intersecting the old set with a set determined by the content of the sentence used. A permission on the other hand tends to enlarge the set *Per*. The question is: by how much? This is Lewis' problem about permission. The solution to this problem is by no means simple. Kamp (1973) contains an elementary suggestion, which appears to be correct for at least some cases. Suppose that *B*'s options at time t are restricted only by a number of standing prohibitions $P_1, \ldots, P_k$ and that each prohibition P_i is represented as an intransitive verb phrase φ_i, which P_i prohibits B from satisfying at any time after t. Suppose moreover that A gives B at t the permission 'You may φ_i' where φ_i is the verb phrase that represents P_i. According to the proposal of Kamp (1973) the effect of *A*'s utterance is to lift the prohibition P_i while the remaining prohibitions are all upheld. In other words, the set that is added to $Per(w, t, B)$ consists of those worlds in which B satisfies, at some future time, the verb phrase φ_i, but satisfies none of the verb phrases representing the other prohibitions at any time later than t.[4]

Of course this solution is adequate only in those cases where the permission sentence used corresponds unambiguously to one particular standing prohibition, and where moreover it is possible to violate that prohibition without violating any of the others. These conditions are by no means always fulfilled.

Robert Stalnaker ([no date]), in a comment on Lewis' paper, suggested a more general analysis of the effect of permissions. His proposal assumes that the members of $Pos(w, t)$ can be compared with respect to the extent to which B departs in them from the standard of irreproachable conduct.

---

there is no time t' after t such that the addressee has unambiguously obeyed if he performs the required action before t', but has failed to do so if he performs it at any time after t'. With permission we find a similar phenomenon: many permissions are to be understood as licensing a certain action only for a restricted period; if the action hasn't been performed by the end of that period then the permission lapses. As this aspect of permission and commands is irrelevant to the main issue of this paper I have decided to ignore it.

Another, and related, problem with commands is that many are to be understood as 'standing orders'. This is typically the case when the imperative uttered contains a negation, as in 'Do not lean out of the window', or 'Do not covet thy neighbour's wife'. But there are also negation-free commands which are thus intended, e.g. the 'Keep left' signs which admonish the continental driver upon his entry into Britain. Again there is a parallel complication in the analysis of permission sentences: 'You may smoke' can be used not only to convey that the addressee may have one cigarette, but also to lift the prohibition against smoking for an extended period of time, possibly even for good. This complication will also be ignored here.

[4] Usually the intention of the permission is that the verb phrase which represents the cancelled prohibition may be satisfied only once, although this is not always so. Compare note 3. As I already said there, this type of complication we shall ignore.

Formally this comparability can be represented (as is also one in Lewis' theory of the counterfactual conditional; see Kaplan 1989) as a nest of sets whose union is $Pos(w, t)$, whose smallest number is $Per(w, t, B)$, and which moreover is such that if w_1 belongs to a member of the nest to which w_2 does not belong then w_2 is a world in which B behaves on the whole more reprehensibly than he does in w_1. The effect of a permission of the form 'You may φ' is to add to the set *Per* all those worlds in which B satisfies φ at some time after t and which belong to the smallest member of the nest in which there are any such worlds. I shall refer to the nest of subsets of $Pos(w, t)$ as $Pos(w, t, B)$.

The proposal of Kamp (1973) becomes a special case of Stalnaker's theory on the reasonable assumption that $Pos(w, t, B)$ contains a set in which there are worlds in which B violates the prohibition that the permission he has been given cancels, but no such worlds in which he violates some of the remaining prohibitions as well. But, as I already intimated, there are many cases to which the theory of Kamp (1973) cannot be applied, at least not the simple version of it which I have presented so far here. Suppose e.g. that the standing prohibitions include the prohibition against taking an apple and the prohibition against taking a pear, and that A gives B a permission by uttering (3). There is then no obvious correspondence between the permission sentence[5] used and any one particular member of the set of standing prohibitions (the two prohibitions which I just mentioned being equally good, and thus equally bad, candidates). Stalnaker's theory on the other hand applies to this case as well. And it might seem at first that it handles it correctly.

In particular if we assume that the prohibition against taking an apple and the prohibition against taking a pear are equally strong then the theory makes exactly the right prediction. For in this case any set in $Pos(w, t, B)$ that contains a world in which B satisfies the verb phrase 'take an apple' should also contain worlds in which B satisfies the verb phrase 'take a pear', and conversely. So the set *Per* will be extended with worlds in which B takes an apple and also with worlds in which he takes a pear. And this is just as it should be.

Not in all cases does the theory give us the correct answer, however. Consider for example a situation in which there are again separate prohibitions

[5] By a *permission sentence* I will understand in this paper any sentence of the form 'You may φ' where φ is an intransitive verb phrase, and 'may' is the 'may' of permission. As sentences of this form are by no means the only ones to be used for the purpose of granting permissions, the term 'permission sentence' could be misleading. It is intended strictly to facilitate the discussions of the present article.

against, respectively, taking an apple and taking a pear, but in which the prohibition against taking an apple is much stronger than that against taking a pear (e.g. because the apples have to be bought, while the pears come from the garden). Stalnaker's theory predicts that the set of worlds added to $Per(w, t, B)$ by a performative utterance of (3) in this situation consists exclusively of worlds in which B takes a pear. Yet even here, it would seem intuitively, (3) conveys that taking an apple is all right too.

This appears to refute the theory. But we should not be too hasty to reject it. In particular we should first consider whether we could not apply the theory of conversation to the counterintuitive prediction it makes for the effect of (3) in this last situation. We might expect this to be possible, as Stalnaker's proposal claims that in a situation of this kind the effect of (3) is identical with that of the simpler sentence (2). Thus something like the maxim of brevity would require the use of (2) rather than (3) if this were indeed the effect intended. We might hope to find additional conversational principles which justify the further inference that if A does use the sentence (3) his intention must have been to permit not only worlds in which B takes a pear but also some in which he takes an apple.

We will explore this possibility later, in section III. Before that I wish to reexamine, in the next section, the account of (3) which I sketched in Kamp (1973) and which does give the correct prediction for the effect of (3) in either of the two situations we have considered above.

To conclude this section let me reiterate the observation, already made at some length in Kamp (1973), that the analysis of permission sentences we have discussed in this section offers a way to explain the apparent puzzle about the logical relations between (1), (2) and (3): (3) is 'stronger' than e.g. (1) in as much as the set or worlds which a performative utterance of (3) adds to the options of the addressee includes, but is not necessarily included in, the set added through a performative use of (1). Indeed it is tempting to introduce, as was proposed in Kamp (1973), a relation of 'entailment' specific to permission sentences by putting:

> Permission sentence φ (*permission*)–*entails* permission sentence ψ iff in every situation the set of worlds added to the options of the addressee through the use of φ includes the set of worlds added to the set of options through the use of ψ.

We have just seen that as it stands (3) does not permission–entail (1) according to Stalnaker's theory. (3) does entail (1) according to the proposal which we shall investigate in the next section. That investigation will lead us, however, to questions of a much more general and theoretical nature. The entailment relation defined above will not be of any further concern.

II.

It will be convenient to generalize somewhat the framework which I sketched when expounding the Lewis-Stalnaker account. Henceforth I shall always assume that with each utterance is associated a situation, or *context, of use*. Moreover we shall also consider contexts of use in which no utterance actually takes place, e.g. to contemplate what the effect of uttering a certain sentence *would* have been in that context *if* it had been made. For simplicity I shall continue to consider only contexts in which there is one speaker, A, and one addressee, B; A will always be assumed to have the authority over B which is necessary to make his permission statements to B effective. Each context c determines a world w_c, the world in which that context is situated, and a time t_c, the time at which the utterances we consider in connection with B are, or would have been, made. We retain the functions *Pos* and *Per*, writing '$Pos(c)$' for '$Pos(w_c, t_c)$', '$Per(c, B)$' for '$Per(w_c, t_c, B)$', etc.

We denote as $[\![\varphi]\!]_c^{Per}$ the sets of those worlds which are added to the options of the addressee through the permission granting utterance of the permission sentence φ in the context c. In cases where such an utterance actually takes place in c this set can be expressed as $Per(c, B) \setminus Per^{-}(c, B)$. It is reasonable to assume, however, that this set is fully determined by (i) the sentence φ; and (ii) contextual features whose presence in the context c is independent of whether the utterance is actually made. (Both proposals for the computation of $Per(c, B)$ from $Per^{-}(c, B)$ which we considered in Section I. exemplify this principle). On this assumption it is meaningful to speak about the set $[\![\varphi]\!]_c$ also when it is not uttered in c.

In Section I. we considered two situations in which A utters (3). In each of these, I maintained, the 'net effect' of this speech act, $[\![(3)]\!]_c^{Per}$, is equal to the union of the net effects produced by utterances of, respectively, (1) and (2):

(5) $\quad [\![(3)]\!]_c^{Per} = [\![(1)]\!]_c^{Per} \cup [\![(2)]\!]_c^{Per}$.

In Kamp (1973) I claimed that this equation must always hold, i.e.

(6) $\quad [\![\text{you may } \varphi \text{ or } \psi]\!]_c^{Per} = [\![\text{you may } \varphi]\!]_c^{Per} \cup [\![\text{you may } \psi]\!]_c^{Per}$

for any intransitive verb phrases φ and ψ; and I formulated a general hypothesis about the uniform character of logical operators such as 'or' and about their function in sentences that are typically used in performative utterances, which was intended to explain why equation (6) holds. I shall briefly restate that hypothesis and retrace the line of thought that led up to it.

One of the central assumptions of modern semantics is that a theory of truth conditions for a language must include a component which has the form of a recursive definition (usually a definition of truth or of satisfaction). There is in general no fully fixed format for this component. Even the truth definition for so simple a language as the classical propositional calculus can be written down in a number of different though essentially equivalent ways. Among these there is one which makes that aspect of the sentential operator 'or' that is crucial here explicit. I shall give the essentials of this particular form of the truth definition for propositional logic first.

I assume that the assertoric use of a sentence φ of propositional calculus in the context c determines—in ways which we need not analyze right now—the *propositions* expressed in c by the atomic components of φ. For each atomic component q_i of φ we identify the proposition expressed by q_i in c, $[\![q_i]\!]_c$, with the set of all those worlds w' in which q_i is true at t_c. The propositions expressed in c by arbitrary compounds built up from these components by means of the sentential connectives are then given by a recursive definition with such clauses as:

(7) $\neg[\![\phi]\!]_c = W \setminus [\![\phi]\!]$ (where W is the set of all possible worlds);
(8) $[\![\varphi \,\&\, \psi]\!]_c = [\![\varphi]\!]_c \cap [\![\psi]\!]_c$;
(9) $[\![\varphi \vee \psi]\!]_c = [\![\varphi]\!]_c \cup [\![\psi]\!]_c$.

Such definitions invite the view that the meanings of the sentential connectives are given by the corresponding clauses of the definition.

Now observe the similarity between (5) and (9). It was this similarity which led me to the following three-part hypothesis:

(10) (i) logical particles such as 'not', 'and', and 'or', represent semantic operations of a simple set-theoretic nature (operations such as complementation, intersection, and union);
(ii) they represent these operations *uniformly*, i.e. for all occasions of use; and
(iii) the role played by the sets on which these operations are performed, as well as the sets they produce, depends on the type of speech act that is performed by uttering the sentence in which the operator occurs.

According to this hypothesis 'or', in particular, has the same semantic function in (3) as it has in (11)

(11) Jones has taken an apple or has taken a pear.

The difference between (3) and (11) is that the sets $[\![(1)]\!]_c^{Per}$, $[\![(2)]\!]_c^{Per}$ and $[\![(3)]\!]_c^{Per}$ represent new options for the addressee, while the sets $[\![$Jones has taken an apple$]\!]_c$, and $[\![(11)]\!]_c$ are *truth sets*—sets of which the corresponding assertoric utterances claim that the actual world is among them.

It is an appealing feature of this hypothesis that it avoids the need to postulate for such occurrences of 'or' as we find in (3) a different meaning or function from the one it is supposed to have in the linguistic environments which have traditionally been the primary concern of formal logic. To postulate such a separate function for occurrences of 'or' without further explanation of how they are connected with the function of 'or' in sentences such as (11) is particularly unsatisfactory in view of the fact that precisely the same puzzle about disjunctive permission arises in many languages other than English. This is evidently a systematic phenomenon that cannot be dismissed as an idiomatic curiosity of one particular vernacular. Note however that the Lewis-Stalnaker theory avoids such *ad hoc* assumptions too. The choice between that theory and the analysis we are now considering must therefore be made on different grounds.

According to (5) the set $[\![(3)]\!]_c^{Per}$ is determined as the union of the sets $[\![(1)]\!]_c^{Per}$ and $[\![(2)]\!]_c^{Per}$. If the operation that forms $[\![(3)]\!]_c^{Per}$ out of $[\![(1)]\!]_c^{Per}$ and $[\![(2)]\!]_c^{Per}$ is indeed the last step of the communication of the effect of *A*'s permission granting utterance, what are the other steps of this computation like?

One of the factors, which determine the set $[\![\text{You may } \varphi]\!]_c^{Per}$ is the sentence 'You may φ' itself. It should be clear, moreover—and the specific proposals for the computation of the sets $Per(w, t, B)$ which we considered in the previous section bear this out—that it is more specifically the satisfaction conditions for the verb phrase[6] which are needed for this computation. In particular the computations of $[\![(1)]\!]_c^{Per}$ and $[\![(2)]\!]_c^{Per}$ require the truth conditions in *c* of the verb phrases

(12) take an apple,

and

(13) take a pear.

I will take it for granted that the theory must specify the satisfaction conditions of complex verb phrases via the type of recursion on the syntactic complexity of these expressions which is by now familiar from the work of, in particular, Montague and his school. If we assume for example—the exact

[6] I am assuming that intransitive verb phrases are semantically treated as one place predicates of individuals; and I will take this to imply that the satisfaction conditions for an intransitive verb phrase φ in the context *c* are represented by a function which for each pair $\langle w, t\rangle$ gives the set of individuals satisfying the verb phrase *w* at *t*. I will denote this function as $[\![\varphi]\!]_c$.

details are insignificant for the present discussion—that (12) is constructed by first forming the singular term 'an apple' by combining the noun 'apple' with the indefinite article, and then forming the verb phrase by combining this term with the verb 'take'; and if we assume moreover—as is usual in Montague Grammar—that the computation of the satisfaction conditions of (12) follows this grammatical construction process step by step, then the computation of (12) (and similarly that of (13)) will involve at least two recursion steps.

The picture which emerges from these considerations is the following. To compute the set $[\![(3)]\!]_c^{Per}$, we first have to perform a number of steps of the kind familiar from extant work in formal semantics to obtain the satisfaction conditions for (12) and (13); from these we compute by what would appear to be typically pragmatic considerations the sets $[\![(1)]\!]_c^{Per}$ and $[\![(2)]\!]_c^{Per}$: and from these we then finally obtain $[\![(3)]\!]_c^{Per}$, by a step which according to our hypothesis (10) constitutes an application of the very same recursive clause that enables us to compute e.g. the truth conditions of (11) in *c* from the truth conditions in *c* of 'Jones has taken an apple' and 'Jones has taken a pear'.

Permission sentences are used not only for the purpose of granting permissions but also for making assertions. Suppose e.g. that *A* has decided to give *B* the permission to draw from his, i.e. *A*'s, bank account, and that he has already made the formal arrangements which this entails (such as depositing a statement to this effect with the bank). *B* then visits the bank and the teller says to him: 'You may draw from *A*'s account.' Clearly this utterance must be taken as a report that a certain permission has been given; it is not a performative which creates a permission.

There are moreover many cases where it is difficult to say whether the utterance is a performative or an assertion. Suppose e.g. that *A* himself says to *B*: 'You may draw from my account' after having made the arrangement. Is this utterance to be taken as an act of granting the permission or of communicating that it has been granted? Another problematic case is that where *A* wants to give *B* a certain permission; say, the permission to take an apple, and instructs *C* to tell *B* this in case *B* turns up at a time when he himself is absent. If then *B* does turn up in *A*'s absence and *C* says to *B*: 'You may take an apple', has *C* given a permission to *B* or only *reported A*'s permission?[7]

[7] Our uncertainty about such questions is reflected in the practical uncertainty we feel in situations where *B* has acted in a way which is acceptable only if the permission is in force, but without having been told this explicitly. Suppose e.g. that *B* has taken an apple before *C* has had the opportunity to say to him 'You may take an apple', but after *A* has instructed *C* to tell *B* this. Has *B* been guilty of a transgression?

It is worth noting that it is not only difficult to decide for some such cases whether the utterance was an assertion or a performative; from the point of view of the addressee it is usually also quite unnecessary. For as long as he has good reason to believe that the utterance is *appropriate*, then, whether he interprets it as a performative or as an assertion the practical consequences will be precisely the same. For either the utterance is a performative and as such creates a certain number of new options, or else it is an assertion; but then if it really is appropriate it must be true; and its truth then guarantees that these very same options already exist.

I am making this observation because it brings out a general principle that correlates the assertoric and the performative uses of permission sentences. This principle can be roughly stated as:

(14) An assertoric utterance of a permission sentence φ in the context c is true iff all those worlds already belong to the options of the addressee that a performative use of φ would have added to the set of his options if they had not already belonged to it.

(This formulation leaves something to be desired as there are obvious problems with the interpretation of the counterfactual construction; it is unnecessary in the present context to eliminate this element of vagueness.)

It is important for us to realize that this principle holds not only for simple sentences such as (1) but also for more complicated permission sentences. Assertoric utterances of (3) for example often have the force of reports that both taking an apple and taking a pear are permissible acts. Notice that this is in conflict with what one is led to expect by the familiar semantic accounts of disjunctive declarative sentences. This discrepancy requires an explanation no less than the problem relating to performative uses of (3) with which we set out. One possible explanation is the following: The performative use is the *primary* use of permission sentences; it is primary in the sense that the truth conditions governing the assertoric use of these sentences reflect in the manner indicated by (14) the pragmatic effects produced by the corresponding performative uses.[8] Given what we have said

[8] Very much would have to be said about the general framework within which an explanation of this form would naturally fit. This would require among other things a systematic classification of utterances into speech act types, and an elaboration of the idea that certain types of speech acts are the primary speech acts associated with certain grammatically characterized types of sentences. I cannot even begin to do this here. For some ideas along these lines see Searle (1975).

about permission sentences earlier such an account seems natural enough, and we will pursue it a little further.

Its formal implementation requires that we add to the recursive component a further clause which determines the truth conditions of a permission sentence φ in the context c in terms of $[\![\varphi]\!]^{Per}_c$, for example:

(15) $[\![\varphi]\!]_c = \{\langle w, t\rangle : (\exists c')(w = w_{c'} \;\&\; t = t_{c'} \;\&\; [\![\varphi]\!]^{Per}_c \subseteq Per(c', B_c))\}$
(where B_c is the addressee of c)[9,10]

The recursive definition will then determine the set $[\![(3)]\!]_c$ via a computation in which the last computation step of $[\![(3)]\!]^{Per}_c$ is followed by a step which leads us back to another value of the same function $(\lambda\varphi)[\![\varphi]\!]_c$ with which the first steps of the computation are also concerned.

I stress the order in which these various computation steps follow each other because it seems to me to conflict with a widespread, though not often articulated, view about the general form that a comprehensive theory of language should take. In much of the recent formally oriented thinking about language there has been, I believe, the implicit assumption that the semantic and pragmatic components of such a theory are *separable*, in a sense which I shall try to explain presently.

I have already referred to the generally recognized view that any language theory which is to account for, at least, the truth conditions of declarative sentences of unbounded complexity must incorporate a component which has the form of a recursive definition. If we accept this view, the implicit assumption to which I just alluded can be formulated as consisting of the following two parts:

(SSP) (i) The concept, or concepts, characterized by the recursive component of the theory belong(s) to semantics.
(ii) Moreover no notion to which this component refers belongs to pragmatics.

The distinction between semantic and pragmatic notions which this principle presupposes goes back at least to the work of C.S. Morris (1955), who

---

[9] An exact version of the recursive definition would have to make explicit that (15) applies only to those φ for which $[\![\varphi]\!]^{Per}_c$ is defined, and for which moreover the permission granting use is primary. This class should include all permission sentences as defined in note 5. Whether it should include other expressions as well is a question I shall not consider here.

[10] As we have defined them here the sets $[\![\varphi]\!]^{Per}_c$ and $[\![\varphi]\!]_c$ show an asymmetry that may seem awkward: the former is a set of worlds while the latter consists of world-time pairs. It would of course be possible to eliminate this asymmetry by redefining $[\![\varphi]\!]^{Per}_c$ as a set which also consists of pairs $\langle w, t\rangle$. I can see no real merit in doing so however.

introduced the tripartite division of the theory of language into syntax, semantics and pragmatics.[11]

Although the literature which discusses the relations between semantics and pragmatics is voluminous what is supposed to distinguish the two categories is only rarely made explicit; consequently the terms are often applied with less consistency than one might wish. It may well be impossible to give a coherent account of the distinction that does justice to the entire spectrum of intuitions associated with it, even were one to take into account only what has been said about semantics and pragmatics during the last decade. What I have to say about the subject should not be taken as an attempt to do this, although I will repeatedly appeal to certain intuitions about the nature of pragmatics and semantics which I believe to be quite general and deeply entrenched.

Morris' original distinction, according to which semantics is concerned with the relations between signs and the objects to which they refer, while pragmatics has for its subject the interaction between signs and their users, still leaves considerable room for manoeuvre to the theorist who wants to convert this characterization into a formally precise definition. By way of a first step in the direction of a precise distinction we shall consider the following principle, which seems to accord well with Morris' intentions:

> (DSP) Semantic (as well as syntactic) concepts can be applied to linguistic expressions irrespective of how and when they are used; pragmatic concepts on the other hand can be properly applied only to individual utterances of expressions, as the question whether such a concept applies depends essentially on features of the context in which, and on the purpose for which, the expression is used.

It should be pointed out at once that DSP has certain implications which may at first seem counterintuitive. For example the truth value of a statement

(16) The sun is shining

is according to DSP a pragmatic concept. For what the truth value is depends on when and where the statement is made. Notice however that on the other

---

[11] The distinction between syntax and semantics is irrelevant to the present paper, and I won't say anything about it. In fact I may occasionally use the term 'semantic' rather broadly, including under it concepts which can be taken to belong to syntax rather than to semantics. No significance should be attached to these real or apparent violations of established terminology.

hand the truth *conditions* of (16), i.e. the conditions which specify in what circumstances an assertoric utterance of (16) would be true, are classified by DSP as semantic. Similarly the truth value of an assertion of

(17) I am thirsty.

is a property which can be ascribed to the utterance, but not to (17) in isolation; but again, if by the truth conditions of (17) we are prepared to understand conditions that specify in what circumstances, including information about the time of utterance and the identity of the speaker, an assertion of (17) is true, then these truth conditions are part of semantics. As these two examples indicate, the properties of truth and falsehood are, at least with respect to a language as rich as English, to be regarded as pragmatic properties. Yet there remains a sense—viz. that of truth conditions—in which the 'concept of truth' may still be said to be part of semantics. This last observation should render the conclusion that the properties of truth and falsehood belong to pragmatics less unpalatable.

The examples we have just considered illustrate at the same time something else: To the property of truth, which is as we have seen a property of utterance, there corresponds a semantic functional concept which assigns to a sentence a function that determines for each context of use whether an assertoric utterance of the sentence in that context is true or false. This correspondence is evidently of a general nature. Indeed, suppose that P is any property of utterances such that the question whether an utterance of sentence φ in context c has P is fully determined by (i) the syntactic form of φ and (ii) certain contextual features which c has independently of whether φ is asserted in c (i.e. if c has, respectively lacks, any of these features then it would have done so equally if φ had not been asserted). Then it is always possible to introduce a concept P' which assigns to each sentence φ that function $P'(\varphi)$ from contexts to truth values such that for any context c $P'(\varphi)(c) = 1$ iff an assertoric utterance of φ in c would have P. P' is a sentence concept and thus qualifies according to DSP as a semantic notion.

This possibility of converting pragmatic concepts into corresponding semantic notions might make us wonder if the distinction which DSP draws is a genuine distinction. For it might seem that many of the sentence concepts which result from these conversions are classified as semantic and yet incorporate essentially pragmatic information.

Before we dismiss DSP on these grounds, however, it is worthwhile to devote a few remarks to a recent paper by David Kaplan (1978), in which he draws a distinction quite similar to that marked by DSP. To explain the

here relevant aspects of Kaplan's paper it is best to summarize what he has to say about simple indexicals such as e.g. the word 'I'. On the one hand the reference of 'I' varies with the context of use so that we cannot associate any particular referent with the word as such. On the other hand there is a simple *rule* which tells us for any context of use *c* what the referent is of 'I' in *c* (viz. the speaker of *c*). This rule defines a function from contexts to individuals, which Kaplan calls the *character* of the word 'I'. Kaplan argues persuasively that it is natural to identify this character with the meaning of 'I'. For it is this function, or alternatively the rule of which it is the crystallization, which a speaker may be supposed to have assimilated when he can be said to have mastered the use of the word.

Generalizing from examples such as this Kaplan argues for a theory in which each meaningful expression has a character, i.e. a function which associates with each context of use *c* a representation of what might be roughly described as the 'semantic content' of the expression in that context.[12] The argument that characters can be regarded as meanings also generalizes naturally.

If we accept this identification of characters with meanings then it becomes almost imperative that we regard the *character concept*, i.e. the concept which assigns to each expression its character, as a semantic concept. For what could semantics be if it is not the theory of meaning?

The preceding paragraphs indicate that it is reasonable to classify certain sentence concepts as semantic even though they may be felt to incorporate pragmatic aspects. There is however an important distinction that this observation threatens to obscure. Compare e.g. the sentences (17) and

(18) The Ohio is a tributary of the Mississippi.

(17) contains an indexical element while (18) does not. This difference is reflected by the characters of the two sentences: the character of (18) is a constant function on contexts; the character of (17) is not. This difference between the two characters arises because the computation of $[\![(17)]\!]_c$ involves a step which imports into it information specific to *c* (the identity of the speaker of *c*) while the computation of $[\![(18)]\!]_c$ involves no such step. Indeed we might want to say that the pragmatic nature of $[\![(17)]\!]_c$ is a consequence of the fact that some step of its computation is an application of a clause which refers to pragmatic notions. This however won't help us in drawing the line between pragmatic and semantic concepts. For how are we

[12] I am of course using the word 'semantic' here in a loose, pretheoretic sense.

to decide which of the notions referred to by the clauses of the definition are pragmatic notions? As long as we confine our attention to characters it may appear that we can circumvent this circularity by capitalizing on the difference between constant and non-constant context-functions. Let us try this and call a character concept for a language *purely semantic* if it assigns to each expression of the language a constant character.

The recursive definition which we have just been discussing thus characterizes a concept which is not purely semantic in this sense. It should be noticed however that the theory of indexicals and demonstratives can be reformulated in such a manner that its recursive component does define a purely semantic character concept. All one needs to do is to first associate with each expression that contains indexical or demonstrative elements an abstract representation in which the indexicals and demonstratives are replaced by variables. The satisfaction conditions for these representations will then be independent of context; and so all representations will have constant character. The pragmatic component of the theory can evidently extract all the required information from these satisfaction conditions. It will specify e.g. that an assertoric utterance of the sentence φ in c is true iff the representation associated with φ is satisfied in w_c at t_c by any sequence which assigns to the variables those objects which the context c provides as references for the indexicals and demonstratives corresponding to these variables.[13] I mention the possibility of this reformulation to make it clear that at least as far as indexicals are concerned we can design the theory in a fashion that insures that the recursive component generates a purely semantic character concept.

Let us now return to the computation of $[\![(3)]\!]_c$, and the recursive definition which yields that computation. It is my intuition—but I think it is one I share with many others—that the steps which compute the sets $[\![(1)]\!]_c^{Per}$ and $[\![(2)]\!]_c^{Per}$ out of $[\![(12)]\!]_c$ and $[\![(13)]\!]_c$ introduce a pragmatic element into

[13] Actually the matter is a little more complicated than I make it appear. For there are indexical expressions such as 'now', of which it is not clear whether they could be replaced by variables. If we take 'now' to be a shorthand for 'at the present time' then of course we can replace it by 'at t', where t is a variable ranging over moments. It has been proposed however (Gabbay 1976; Kamp 1971; Prior 1968) that 'now' should be treated as sentential operator, on a par with the operators 'it was the case that' and 'it will be the case that' which are traditionally considered in tense logics of the type first introduced by Prior. This proposal is, at least in the form in which it is available in the literature, incompatible with the replacements that the revised form of a theory requires. I have reached the conviction however, that nothing compels us to a view of 'now' on which the substitution of 'at t' is inadmissible.

the computation, and that by the same token the clause[14] of which they are applications introduces a pragmatic aspect into the recursive definition.

The formal considerations of the preceding pages confirm this intuition. For, of the two concepts $(\lambda\varphi)(\lambda c)[\![\varphi]\!]_c$ and $(\lambda\varphi)(\lambda c)[\![\varphi]\!]_c^{Per}$[15] that the recursive definition which I have sketched simultaneously defines, the second assigns context functions that are not constant and so is not purely semantic. As a matter of fact the concept $(\lambda\varphi)(\lambda c)[\![\varphi]\!]_c$ also assigns nonconstant functions to certain expressions; e.g. $(\lambda c)[\![(3)]\!]_c$ depends on c because the right hand side of (15) asks for the addressee of c. We have seen however that the context dependencies created by such simple indexicals as 'I' and 'You' can be eliminated from the recursive component of the theory if we associate with each expression that contains such terms a representation in which they are replaced by variables.[16] When thus modified the definition will assign to each meaningful expression φ a constant function $(\lambda c)[\![\varphi]\!]_c$, *even* when the computation of this function involves intermediate steps which compute values of $(\lambda\varphi)(\lambda c)[\![\varphi]\!]_c^{Per}$.

However the definition which results from this modification is still in violation of SSP, for one of the concepts it defines viz. $(\lambda\varphi)(\lambda c)[\![\varphi]\!]_c^{Per}$, will assign non-constant context functions no less than it did according to the unmodified definition.

One might want to look upon the recursive component in a somewhat different perspective, and take it as a definition of just the concept $(\lambda\varphi)(\lambda c)[\![\varphi]\!]_c$. The pragmatic component of the theory will then account for the effect of performative utterances of permission sentences on the basis of the satisfaction conditions for the representations of the sentences

---

[14] I haven't yet stated this clause explicitly. From the discussion in Section I. it should be clear, however, that it should have some such form as:

> $[\![$You may $\varphi]\!]_c^{Per} = \{w \in Pos(c) : (\exists t > t_c)(B_c \in [\![\varphi]\!]_c(\langle w, t\rangle))$ & w belongs to the smallest member of $Pos(c, B_c)$ which contains any worlds w' such that $(\exists t > t_c)(B_c \in [\![\varphi]\!]_c(\langle w', t\rangle))\}$

(here B_c is the addressee in c).

[15] This notation is a little misleading in that it conceals the fact that $[\![\varphi]\!]_c^{Per}$ is defined on only a small subset of the domain of $(\lambda\varphi)(\lambda c)[\![\varphi]\!]_c$.

[16] This requires in particular that we modify (15) in such a way that if

(19) x may take an apple or take a pear.

is the representation of (3) then for any assignment s (of objects to variables)

(15′) $[\![(19)]\!]_c(s) = \{\langle w, t\rangle : (\exists c')(w_{c'} = w$ & $t_{c'} = t$ & the addressee of $c' = s(x)$ & $[\![(19)]\!]_c^{Per} \subseteq Per(c', s(x)))\}$.

used. This would presumably involve an appeal to (14) in the 'opposite direction' (more about this in the next section). One could then consider the concept $(\lambda\varphi)(\lambda c)[\![\varphi]\!]_c^{Per}$ as no more than a side product of the recursive definition, needed as an auxiliary in the characterization of $(\lambda\varphi)(\lambda c)[\![\varphi]\!]_c$, but not among the concepts that the first part of SSP refers to. However, even when conceived in this way the new definition remains in conflict with SSP, in virtue of SSP.ii; for some clauses, such as e.g. (15′) (cf. footnote 16), their right hand sides will refer to a concept that is not purely semantic.

It must be asked at this point whether this really matters. Does SSP express a conception about the form of linguistic explanation that should be taken at all seriously? In order to speak to this question I must say a little more about the considerations that motivated me to state SSP in the form I gave here, within the context of our discussion of permission sentences. The conception of which I tried to capture the relevant consequences in SSP goes back at least to Frege. Frege assumed[17] that the content of an expression is the same no matter in what type of speech act it is used. This speech act neutral content is determined from the contents of the smallest constituents of the expression by a recursive process the successive steps of which are dictated by the syntactic structure of the expression. The contribution which the expression makes to the speech act in which it is used is completely accounted for by its content.

Linked with this picture is the notion that the part of the theory which deals with the contents of expressions is its semantic component, while the parts which deal with the properties of utterances are to count as pragmatic. This accords well with what has been said earlier about the difference between semantics and pragmatics. For on the one hand we have already mentioned that the analysis of the properties of speech acts belongs to the province of pragmatics. I might add to this that those parts which are specifically concerned with the illocutionary effects of performative utterances seem to have a further claim to being pragmatic in as much as these effects are instances of the interaction between signs and their users which is the proper subject of pragmatics according to Morris' original definition. On the other hand there is good reason to believe that the systematic analysis of content is the one and only possible form for a theory of meaning.[18] So in

[17] See e.g. Frege (1960, 1967).

[18] This assumption is fundamental to both model theoretic semantics, as it was established by Montague, and to truth theory as advocated by Davidson and his school. See in particular Davidson (1967).

as much as semantics is the theory of meaning it is natural to identify the component of the theory which deals with content as its semantic component.

I must confess that I can see no good reason why semantics and pragmatics should be separated in the sense that SSP tries to articulate. In particular, if we think of the recursive component of the theory as representing, in some way, the means of processing expressions that are available to the speakers, there are no a priori grounds, it seems to me, that would favor a theory which agrees with SSP over one that violates it: computation processes that SSP prohibits appear no less accessible than those which it allows. This is an extremely delicate matter however, to which we shall briefly return in the last section.

I must now mention some other objections to the theory which we have considered in this section, and which are more specifically concerned with the details of the analysis of disjunctive permission sentences it proposes.

First, it must be conceded that assertoric utterances of (3) need not be interpreted as saying that both taking an apple and taking a pear are permitted actions. There is nothing wrong, say, with *C*'s responding to *B*'s inquiry about what permission *A* has given, by saying:

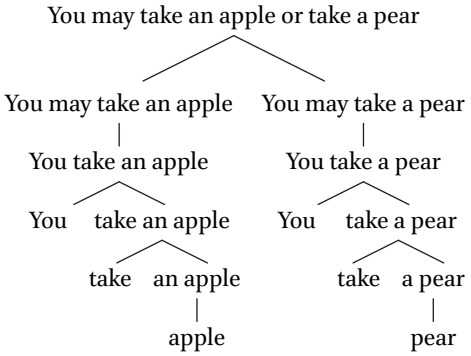
(20) You may take an apple or take a pear; but I don't know which.

It appears, moreover, that in certain fairly infrequent situations even performative utterances of disjunctive permission sentences can be taken in this weaker sense. Suppose e.g. that we have just arrived in the town where we shall spend our holiday, and that in the immediate environment of this town there are two swimming areas, the first a part of Shoal Creek, the second a part of Shingle Creek. We have heard that one of these areas is dangerous and that in fact only a week earlier some boy drowned there. But I don't know which of these two it is. Suppose further my son comes to me and asks if he may join tomorrow some of his newly acquired friends for a swim. I can then reply: 'You may go to Shoal Creek or go to Shingle Creek. But stay away from the dangerous one.' I think it is right to hold that in this case a permission, viz. the permission to go for a swim somewhere, has been issued; but it is only the permission to realize the disjunctive verb phrase 'go to Shoal Creek or go to Shingle Creek'. Nothing has been said about the disjuncts separately.

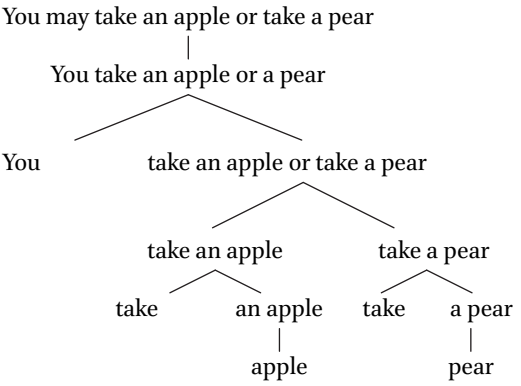
In Kamp (1973) I already gave a brief indication of how we might accommodate both interpretations of disjunctive permission sentences within the framework there advocated. I suggested that a sentence such as (3) has two distinct analyses which induce different computations. Besides the

analysis we have already discussed there is another in which the union operation is applied *before* the transitions from $[\![\ ]\!]_c$ to $[\![\ ]\!]_c^{Per}$ take place: it operates on $[\![(12)]\!]_c$ and $[\![(13)]\!]_c$ rather than on $[\![(1)]\!]_c^{Per}$ and $[\![(2)]\!]_c^{Per}$. From the union so obtained we then compute the set $[\![(3)]\!]_c^{Per}$, as indicated earlier.[19] It will be helpful to represent these two analyses as trees:

(21)



(22)



If we look at these diagrams we cannot help noticing that from a purely grammatical point of view (21) is decidedly less plausible than (22). For according to (21) (3) results through combining the sentences (1) and (2) by means of 'or'. And this operation must have the syntactic effect of not only inserting 'or' between the two constituents but moreover of deleting

[19] See footnote 14.

Once we accept that there are two distinct computations for $[\![(3)]\!]_c^{Per}$ and $[\![(3)]\!]_c$, the notation used so far becomes strictly speaking ambiguous. To proceed with full propriety we should make the notation dependent on the analysis. But let that pass.

the second occurrence of 'may'. This is certainly not the way to obtain (3) which purely syntactic considerations would suggest. Of course we could try to defend the analysis by claiming that (3) can be derived through verb phrase deletion from

(23) You may take an apple or you may take a pear

which can be used for the very same purposes (both assertoric and performative) that are served by (3) itself. This defense is less than fully satisfactory, however. For sentence (23), though it *can* be used to license simultaneously both the act of taking an apple and the act of taking a pear, sounds somewhat strained when used in this way, whereas (3) doesn't at all. (23) is a much more natural choice of expression, it seems, when we want to report that a certain permission is in force but are unsure which it is.

A second syntactic problem for our theory already arises with (1) and (2). Our present analysis of (1) assumes that the sentence is formed from combining 'may' with the sentence 'You take an apple'. From a purely syntactic point of view it seems more natural, however, to assume that 'may' is combined with the verb phrase 'take an apple' to give 'may take an apple'. This second verb phrase is then combined with 'You' to yield (1). If we were to adopt such a syntactic analysis for (1), we could either (i) retain the semantic account given earlier but thereby give up the strict correspondence between syntactic and semantic recursion steps which we have observed so far (it is a difficult question how important this is; but I will continue to take it here that there always must be *good* grounds to justify a departure from the principle); or (ii) revise the theory in such a way that $[\![\text{take an apple}]\!]_c$ yields an appropriate interpretation for 'may take an apple', and that from this interpretation we can compute in turn $[\![(1)]\!]_c^{Per}$. This is possible but makes the theory rather more awkward. (I will spare the reader the uninteresting details of this modification.)

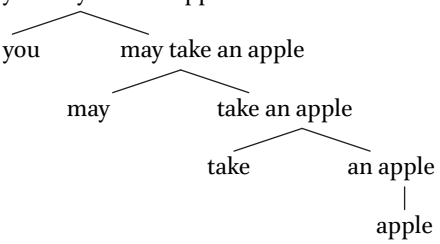
A third source of criticism is the fact that permission sentences are, unlike imperatives, indicative sentences. The view that indicatives are primarily used for the purpose of asserting, and that in all cases where they appear in speech acts of a different type there must be a special account of why this should be possible, is as problematic as it is old. And the theory of indirect speech acts—which as far as I know is the only serious effort to make this thesis work—has only sharpened our awareness of all the difficulties that a defense of it must overcome. Nonetheless it might be maintained that the prima facie evidence for such a view is so strong that a theory which conflicts with it can be accepted only after we have tried in earnest to deal with its problems in a manner with which the view is compatible.

III.

So let us try to account for the assertoric use of permission sentences while ignoring, for the time being, that they are also used in performatives, and to deal with their performative uses subsequently. Our primary task is now to state the truth conditions of permission sentences. We shall begin by considering this question for simple sentences such as (1) and (2).

Again we might consider a number of different syntactic analyses for these sentences which would dictate different computations; but, as the reader will see, this time there is no good reason to consider any but the intuitively most natural analysis. For (1) the analysis we adopt is given by the following tree:

(27)[b] you may take an apple



The crucial step in the computation of $[\![(1)]\!]_c$ is now the one that corresponds to the formation of

(28) may take an apple

out of

(12) take an apple.

As (12) and (28) are both intransitive verb phrases they ought to be assigned the same type of semantic entity. So if we stay with our earlier proposal for $[\![(12)]\!]_c$ (see p. 194), then $[\![(28)]\!]_c$ must be a function from world-time pairs to sets of individuals.

But which function? Well, the function, surely, which assigns to the pair $\langle w, t\rangle$ the set of individuals u such that taking an apple belongs to the permissible options of u in w at t. Generalizing from this particular case

[b] Editors' note: In the original article example numbers 24–26 are missing. We follow the original numbering.

we may conclude that the computation of $[\![$ may $\varphi]\!]_c$ from $[\![\varphi]\!]_c$ requires the following recursion clause:

(29) $[\![\text{may } \varphi]\!]_c(\langle w, t\rangle) = \{u : (\exists w')(w' \in Per(w, t, u)) \;\&\; (\exists t')(t' > t \;\&\; u \in [\![\varphi]\!]_c(\langle w, t\rangle))\}$

The other steps in the computation of $[\![(1)]\!]_c$ need, after all that has been said on the subject in the previous section, no further comment.

Before we try to deal with the truth conditions of more complex permission sentences such as (3), let us first consider the question how we might account, on the basis of these truth conditions, for the fact that utterances of a sentence such as (1) are so often taken to be performatives, whose function it is to grant a permission. Towards an explanation of this phenomenon we might consider the following argument. Suppose that *A* has authority over *B* and that this fact is common knowledge shared between *A* and *B*. Then *B* may be expected to react to *A*'s utterance: 'You may take an apple' with the reflection: 'It is up to *A* whether I may take an apple or not. Therefore he knows whether what he says is true or false. It may be assumed moreover that he is not saying what he knows to be false, as this would go against established principles of conversational propriety. So I may conclude that I have the permission to take an apple.'

The effect of *A*'s utterance is thus to make *B* realize that he has been given the permission.

The problem with this explanation is that it doesn't go quite far enough. For it fails to establish that there is an actual *institution* of giving permissions by uttering permission sentences.[20] Suppose *A* says to *B* 'You may take an apple', *B* then takes an apple, whereupon *A* berates him for doing so, claiming he had no permission to take an apple. In such a situation it is not just that *B* can excuse himself by pointing out that he was *misled* by *A*'s utterance. No, *B* can justly claim that he *had* the permission, in virtue of what *A* said to him. There are situations where *A* just cannot mislead *B* simply because his utterance constitutes the granting of the permission. The hypothesis that the addressee of an utterance of, say, (1) tends to respond with the kind of reasoning I have just sketched can at best provide an intuitive explanation of how this institution could have developed in a language which already allows for the assertoric use of permission sentences. It does not show, however, that there *must* be such an institution. A theory which is to deal with

[20] This is a problem generally with explanations in the theory of indirect speech acts, to which the present argument belongs. For the theory of indirect speech acts see e.g. Searle (1975).

the various aspects of language from a purely synchronic perspective must therefore postulate the performative use of permission sentences without the possibility of justifying this claim conclusively by an appeal to conversational rules.

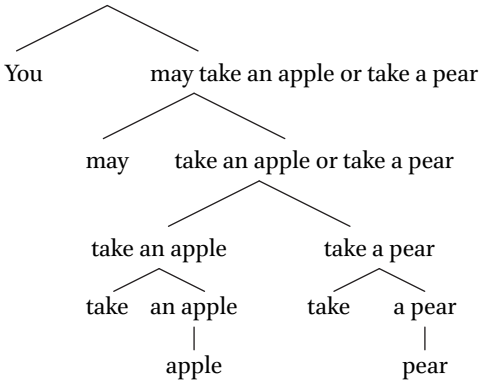
Of course the theory must not only imply that permission sentences can be used to give permissions; it must also specify what effects such sentences produce when they are so used. Once again the theory of conversation goes some way towards predicting these effects. For we may suppose the addressee *B* to continue the reflections which *A*'s utterance of (1) provokes roughly as follows: 'If what *A* says is true then there must be some futures in which I take an apple and to which he does not object. For all I knew there were no such futures. However, some of the futures in which I take an apple are less objectionable than others. Now it is reasonable to assume that if future w is now permitted and w' is no more objectionable than w then w' is permitted too. So if any such futures are permitted then surely all those are which from what I knew are as unobjectionable as could be given that in each of them I will take an apple.'

The value of this argument is as limited as that of the previous appeal to conversational principles. We may feel that it articulates a valuable insight into the coherence between, and possibly the origin of, certain linguistic conventions; but I do not see how it could be so formulated that the existing conventions regarding the power of permission granting utterances becomes strictly derivable from the truth conditions of the permission sentences used in these utterances together with conversational principles which are general enough to be supported by other evidence than their power to correctly predict the effects of performative utterances of sentences such as (1).

If indeed there is no hope of actually *deriving* from the assumption that permission sentences are used for assertoric purposes and from the conditions which specify when an assertoric utterance of such a sentence is true either the fact that permission sentences are also used to give permissions or the effects that their performative uses produce; then the difference between an account of such sentences which follows the strategy we have thus far outlined in this section on the one hand and on the other the theory sketched in the preceding section is not really that big. In fact the only difference so far is that while the earlier theory derives the truth conditions of permission sentences from the effects they produce when used performatively, the present theory treats truth conditions and performative effects separately, allowing only for a partial, and thus far purely informal, explanation of how the performative use could have developed out of the assertoric use.

The crucial question however is, as before, the treatment of (3), and we must now consider how the new theory might cope with this sentence. If we analyze (3) so as to give 'or' wide scope over 'may' then there is no hope of obtaining the strong reading for (3), at least not if we are to use no more than the resources that the present theory offers. (This is something which I will not argue for. But it can easily be verified, especially after reading what follows below.) lndeed the most promising syntactic analysis is the one which meets all the objections which we mentioned at the end of the preceding section and which is given by:

(30) You may take an apple or take a pear



If we compute the truth conditions of (3) in accordance with (30), using the principles already mentioned—and in particular clause (29)—then we find that $\langle w, t\rangle$ belongs to $[\![(3)]\!]_c$ iff there is some permissible continuation of w after t in which the addressee of c takes either an apple or a pear; it is not required that his options include continuations in which he may take an apple as well as continuations in which he may take a pear.

How shall we explain that (3) has nevertheless, so often, the stronger reading, according to which continuations of either type belong to the addressee's options? Again our only hope of finding an explanation lies in the theory of conversation:

Suppose it is common knowledge between *A* and *B* that *A* has the relevant authority over *B*. Suppose further that *A* addresses *B* with (3). Then *B* may be expected to infer not only that he has the permission to satisfy the disjunctive verb phrase

(31) take an apple or take a pear;

he is likely to reason moreover: 'There must be a reason why *A* used (3) rather than e.g. (1) or (2). His use of (3) cannot but signify that while he allows me

to satisfy (31) he has left it undecided which of its disjuncts I shall satisfy; for if he had decided this matter he could have used—and conversational propriety would in that case have demanded that he use—the simpler (1) or (2). That he has left the question regarding which disjunct I shall satisfy undecided could have one of two reasons; it is either that he isn't yet in a position to make this decision (e.g. because he doesn't yet possess all the relevant information); or else it is because he doesn't really care.'

In many situations *B* will be in a position to dismiss the first of these two possibilities, and thus to conclude that *A* is indifferent to the question; in other words that he, *B*, may realize either the first disjunct or the second.

Again the argument is unsatisfactory as it stands, and to incorporate it into our theory requires that we articulate general principles which, together with specific assumptions about the utterance in question which can reasonably be imputed to the addressee, yield the desired conclusion by pure logic. This is a task that I shall not undertake here. But even without carrying it out explicitly we can make the following observation: Among the principles that warrant the inference that *A* is *indifferent* as to which disjunct will be realized (and therefore that either disjunct is permitted) there must be at least one which applies only to a small subclass of the set of all indicative sentences which contain disjunctive constructions. For in general disjunctive indicatives do not have the 'strong reading' which the principle must allow us to assign to disjunctive permission sentences.

I mention this fact as it enters into the question whether these principles introduce a pragmatic element into the recursive component of the theory. At this very point the form of the new principles must appear irrelevant to this issue, in as much as we have only contemplated the possibility of adding them to the pragmatic component of the theory, which we assume to provide no feed-back to the truth definition. But the matter with which we are trying to deal is more complicated than the examples at which we have looked so far suggest. Consider, to begin with,

(32) You may take an apple or take a pear, but if you do not peel it you will be sick.

Just as (3) can have its 'strong reading' when used by itself, so it can also have this reading when part of the larger (32). In order for the pragmatic component of the theory to assign this strong reading to certain utterances of (32) it must evidently operate not on the truth conditions of (32) as a whole, but rather on the truth conditions of its conjunct (3).

Cases like (32) are still relatively easy to accommodate. We could for example add a new principle to the effect that an assertoric utterance of a

conjunction may be reinterpreted as a succession of assertions of the respective conjuncts. In fact the need for such a principle has been implicit in the theory of conversation from the very start. For in as much as implicatures are inferred from individual sentences they are usually inferred from *parts* of continuing pieces of discourse. Often such a piece of discourse is equivalent to a single sentence which is the conjunction of the various sentences which constitute the piece; and an assertoric utterance of this conjunction must therefore carry the same implicatures as the discourse fragment. The only way in which the theory can do this is by using as inputs the truth conditions of the individual conjuncts.

There are however other ways in which (3) can occur embedded in larger sentences while yet allowing the strong reading. Take e.g. the two-sentence utterance.

(33) Usually you may only take an apple. So if you may take an apple or take a pear, you should bloody well be pleased.

Here (3) occurs as antecedent of (33.ii). Yet it is most naturally interpreted there as having the strong reading. But how could an essentially conversational argument establish this on the basis of the truth conditions of (33.ii)? The difficulty we are facing is this. The truth conditions our theory assigns to (33.ii) are the same as those which it assigns to

(34) You should be pleased if you may take an apple and you should be pleased if you may take a pear.

The reading which the pragmatic component must produce on the basis of these truth conditions does not *entail* these truth conditions. The effect of applying the pragmatic component must therefore be not just to *add* an implicature to given truth conditions, but rather to *undo* what the recursive component of the theory has already established. For this the theory of implicature, as traditionally conceived, has no provision; and indeed one would be hard-put to introduce such powers into it.

One who takes the conversational arguments we have presented so far at all seriously cannot escape the impression that in certain cases, such as in particular that of (3), an implicature which the sentence typically carries when used by itself, becomes conventionally attached to that sentence in such a way that it may contribute to the interpretation of compounds in which the first sentence is so embedded that it becomes inaccessible as input to the conversational component of the theory in the usual way. We might contemplate adding to the pragmatic component of our theory a principle that makes this intuition precise—although how such a principle should be stated I am unable to say. But in any case this principle would

introduce the feed-back of the pragmatic component into the truth definition which we precisely wanted to exclude.[21]

At this point it appears that the present theory must, just as the theory of Section II, run into conflict with SSP. We should not be too hasty with such a conclusion, however. For rather than introducing a principle into the theory that will somehow make it possible to use the implicatures assigned by the pragmatic component to certain sentences in the computation of the truth conditions of compounds that contain them as parts, we could add directly to the truth definition a clause which gives the strong reading of permission sentences by spelling out the effect that the interaction between 'or' and 'may' produces in them. The clause should be to the effect that:

$$(37)\quad [\![\text{may } \varphi_1 \text{ or } \varphi_2 \text{ or } \ldots \text{ or } \varphi_n]\!]_c(\langle w,t\rangle) = \left\{ u : \bigcup_{i=1,\ldots,n} [\![\varphi_i]\!]^{Per}_{w,t,u} \subseteq Per(w,t,u) \right\}$$

where $[\![\varphi_i]\!]^{Per}_{w,t,u}$ is short for: $\{w' : w$ is a world in which u satisfies φ_1 at some time after t and which belongs to the smallest member of the nest $Pos(w,t,u)$ which contains such worlds$\}$.

(37) allows us to compute for (3) truth conditions that accord with its strong reading. Yet the truth definition still agrees with SSP, at least according to the formal criteria developed in this paper.[22]

---

[21] It may be noted at this point that the complications that arise in connection with examples such as (34) closely parallel a familiar problem in the theory of conditionals. In the very first presentation of the theory of conversational implicature it was proposed that the truth conditions of English conditionals are correctly represented by the truth table for material implication; and that the intensional interpretations which hearers commonly associate with conditional sentences when they are actually used can be explained on the basis of these truth conditions by an appeal to the rules of conversation. This proposal looks promising enough as long as we restrict our attention to conditionals which are not embedded in larger linguistic units. But problems arise in connection with such embedded conditionals as

(35) If aunt Emilia will disinherit us if we do not invite her then we will.

According to the truth functional account of 'if' the truth conditions (35) come out equivalent to those of

(36) If aunt Emilia will disinherit us we will invite her;

and this is not what somebody who utters (35) would naturally be understood as saying. Here too the problem is that the intended reading of (35) cannot be obtained simply by adding an implicature to the reading which is provided by the truth definition. This indeed is the main reason why such accounts of the conditional had to be abandoned.

[22] As disjunctive permission sentences have two different readings (37) must operate alongside (9) and (29), which we already saw were needed to obtain the truth conditions corresponding to the weak readings of these sentences. This means that strictly speaking we

But this is at a price. For the theory now fails to give any account of why there should be besides the separate clauses for 'or' and 'may' the further clause (37), which formulates a joint effect of the two lexical items that no combination of applications of the separate clauses can produce. (The theory of Kamp (1973) was designed precisely to explain why the effect that (37) describes is produced on the basis of the functions which 'or' and 'may' fulfill in general.)

Before we can assess how serious a defect this is I must draw attention to the fact that what we have called the strong reading is possible not only for permission sentences, but for certain other types of disjunctive sentence as well. Examples are:

(38) We may go to France or stay put next summer (in the 'possibility sense' of 'may')
(39) I can drop you at the next corner or drive you to the bus stop.
(40) If John had come to the party with Alice or with Joan we might have had some fun.

If, as it seems reasonable to assume, there are only a finite number of different sentence types which allow this reading then we might accommodate them by adding for each a separate clause similar to (37), and thus preserve compliance with SSP. But our suspicion that the theory misses a significant generalization should now be even stronger. For it is hard to ignore the hypothesis that behind this multiplicity of clauses lies some unifying principle. As I have here presented the theory of Kamp (1973) it does not provide such a unifying principle either. In Kamp (1973) however I did make a suggestion as to what it might be. To develop that suggestion into a respectable doctrine is a major task on which I shall again default. But I will outline once more the general idea.

The sentences (38), (39) and (40) share with (3) that they are all typically used in speech acts whose function it is to bring a certain number of possibilities to the attention of the audience. What distinguishes these speech acts from each other lies only in the different purposes for which they bring these possibilities into focus. In the case of permission sentences the purpose is to convey to the hearer that these possibilities are all acceptable; in the case of sentences such as (38) or (39) it is to point out that these possibilities are to be reckoned with (e.g. in making plans for the future); an utterance of (40)

---

must acknowledge two separate analyses for sentences such as (3), which correspond to the two different computations. Neither of these is a target however for the objections listed at the end of Section II.

offers a set of possible situations in order to claim of each of these that a certain condition—expressed by the consequent of (40)—holds in it. In all these cases the set of possibilities to which the hearer's attention is drawn consists of those in which the truth conditions of the relevant subsentence are fulfilled and which are among the less outlandish possible situations in which this is so.

In Kamp (1973) I implied that the strong readings of disjunctions that are used in any such speech acts arise because there is the option to apply the union operation which 'or' signals to the sets which the two separate disjuncts would have placed in focus had they been used by themselves. However, we can also justify the strong readings by means of the conversational argument which we already considered in connection with (3): In each of these cases the speaker's use of the disjunction can be interpreted as testifying to his indifference whether the possibilities he wants to bring to the hearer's attention satisfy the first or the second disjunct; and so it may be inferred that he wants to bring both kinds of possibilities to his attention.

This latter argument strikes me as particularly persuasive within the present more general setting, in which we try to deal not only with permission sentences. This, together with the implausibility of the syntactic analyses which the theory of Kamp (1973) requires inclines me now to the opinion that that theory should probably be rejected. There remains however the larger question to which it led us, viz. whether it is possible for a comprehensive theory of language to conform to SSP.

The preceding discussion has revealed some of the difficulties we must face if we wish to answer this question. We saw that although it is possible to draft a theory which treats the problems on which we have focussed here correctly, this theory will fail to capture certain systematic correlations between its different parts, and in particular between different parts of its recursive component. On the other hand it appears that we could also formulate the theory in such a way that it does account for these correlations; but so formulated the theory will allow its truth definition to call on the pragmatic component for inputs and therefore would have to violate SSP.

In fact it seems to me that there is still a third alternative. This is suggested by our earlier consideration of the theory of indirect speech acts. We noted that the arguments which are meant to explain the practice of using permission sentences performatively from the fact that they are used assertorically and the truth conditions which govern their assertoric uses cannot be formalized to the point where they could be incorporated into the formal

theory itself. Yet it might be said that these arguments provide a modicum of insight why that theory should be the way it is. This suggests that a combination of (i) a strict formal theory, and (ii) a conglomerate of less formal explanations of various aspects of the formal theory might be the best we can do if we want to give an account of language which is precise and yet optimally illuminating. Within the scientific methodology bequested by the great philosophers of science of the end of the last and the first half of the present century (in particular E. Mach and the Vienna Circle in its various ramifications) there appears to be no room for scientific accounts of this format. It seems to me to be a worthwhile task to see how the methodology of theories of such forms might be worked out; but it is not a task which I shall undertake now.

Let us return to the more specific question which of the proposals we have considered provides the best account of permission sentences. I want to ask this question not because I have an answer to it, but rather because it leads to the further query whether questions of this sort can be answered at all. I said earlier that I could not see why we should prefer one theory over another just because the first satisfies SSP while the second does not, as there appears to be no reason why the computation processes which SSP rules out are any less plausible from the perspective of actual language processing than are those which it licenses. In this consideration there is an implicit acknowledgement that the ways in which speakers process the sentences of their language is relevant to the comparison of linguistic theories. It implies more specifically for the theories we have sketched in this paper that they too must ultimately compare in regard of their faithfulness to the ways in which speakers determine the truth conditions of those sentences that the theories deal with.

As soon as one makes this claim, however, he realizes how hard it is to attach a clear sense to it. What precisely should be understood by the 'ways in which speakers determine the truth conditions' of certain sentences? Remarks throughout this paper have suggested that speakers do this by going through certain computations the successive steps of which are indicated by the syntactic structure of the sentence in question.

But how are these computations to be identified? Questions of this sort are deep and immensely difficult. And at the present time it isn't even possible to make an educated guess at how long it will be before we acquire the degree of understanding of the workings of the human mind necessary to give these questions a coherent interpretation that will render them at least in principle amenable to definite answers. Yet, until then nothing much can be said about the psychological adequacy of linguistic theories.

It is my conjecture that if we ever reach this level of sophistication we will find there to be not just one computation procedure for each (unambiguous) expression, but in general a multiplicity of such procedures—different computations being used by different speakers of the same language and perhaps even by the same speaker on different occasions. If this is indeed so then it might turn out that each of the theories we have considered in this paper correctly mimics a procedure that is actually in use.

This might also be taken to imply that in as much as linguistic theory has the task to account for those aspects of language the knowledge of which is common to all competent speakers, some of the distinctions between alternative theories that came up in the course of this paper are not really relevant to linguistics. But I am not so sure that it would be right to draw this conclusion. For it could well be that once we have reached the understanding of mental processes that will allow us to speak meaningfully about individual computation procedures we shall no longer be content with theories that deal only with the output of those procedures, but demand that a comprehensive account of language deal with the diversity of computation strategies realized, or realizable, in human language users, as well as with the uniformity of their outputs; and that it explain how the shared linguistic experience of those users ensures that uniformity of output without which communication would be impossible, and semantics, as it is generally conceived today, would not exist.

Bibliography

Davidson, D., 1967. Truth and meaning. *Synthese* 17, 304–323.

Frege, G., 1960. On Sense and Reference. In: P. Geach, M. Black (eds.), *Translation from the Philosophical Writings of Gottlob Frege*, 2nd Edition. Basil Blackwell, Oxford, 56–87. English translation of G. Frege, 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik*, NF 100, 1892, 25–50.

Frege, G., 1967. The thought: A logical inquiry. In: P. Strawson (ed.), *Philosophical Logic*. Oxford University Press, Oxford, 17–38. English translation of G. Frege, 'Der Gedanke, eine logische Untersuchung', *Beiträge zur Philosophie des deutschen Idealismus*, 1, 1918, 58–77.

Gabbay, D., 1976. *Investigations into Tense Logic and Modal Logic with Applications to Linguistics and Philosophy*. Reidel, Dordrecht.

Kamp, H., 1971. Formal properties of 'now'. *Theoria* 37, 227–273. Reprinted in: K. von Heusinger, A. ter Meulen (eds.), *Meaning and the Dynamics of Interpretation. Selected Papers of Hans Kamp*. Brill, Leiden, 2013, 11–51.

Kamp, H., 1973. Free choice permission. *Proceedings of the Aristotelian Society* 74, 57–74. Reprinted in: K. von Heusinger, A. ter Meulen (eds.), *Meaning and the Dynamics of Interpretation. Selected Papers of Hans Kamp*. Brill, Leiden, 2013, 169–184.

Kamp, H., 1978. *On Historical Necessity*, unpublished manuscript, Bedford College, London/University of Texas, Austin, TX. Available online: http://www.ims.uni-stuttgart.de/archiv/kamp/.

Kaplan, D., 1978. Dthat. In: P. Cole (ed.), *Syntax and Semantics 9: Pragmatics*. Academic Press, New York, 221– 243.

Kaplan, D., 1989. Demonstratives. An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. In: J. Almog, J. Perry, H. Wettstein (eds.), *Themes from Kaplan*. Oxford University Press, Oxford, 481–563.

Lewis, D.K., 1970. *A problem about permission*, unpublished manuscript, University of California, Los Angeles, CA. Published in: E. Saarinen et al. (eds.), *Essays in Honour of Jaakko Hintikka*. Reidel, Dordrecht, 1979, 163–175.

Morris, C., 1955. *Signs, Language and Behavior*. George Braziller, New York.

Prior, A., 1968. "Now". *Noûs* 2, 101–119.

Searle, J., 1975. Indirect speech acts. In: P. Cole, J. Morgan (eds.), *Syntax and Semantics 3: Speech Acts*. Academic Press, New York, 59–82.

Stalnaker, R., [no date]. *Comments of Lewis' Problem about Permission*, unpublished manuscript, Cornell University, Ithaca, NY.

PART THREE

VAGUENESS

INTRODUCTION TO VAGUENESS

Hans Kamp

On the face of it the two papers in this part may seem to represent diagonally opposite positions. But that isn't really true. Both are concerned with the connections between vagueness and context dependence. But they are concerned with different connections, so should be seen as complementary and certainly not as contradictory.

The motivation for *Two Theories about Adjectives* (TTAA) was a problem that Richard Montague addresses in ***English as a Formal Language***. That paper contains the suggestion that adjectives should be treated as functions from the meanings of simple and complex common nouns—that is, properties—to other such meanings. When the adjective occurs prenominally, the noun following it provides the argument to the function it denotes. When it occurs predicatively within the verb phrase, its argument has to be inferred, for instance by identifying it with the head noun in the subject.

What speaks against such an account is a strong intuition that adjectives—or most of them at any rate—are among the expressions that denote properties, and not functions from properties to properties. On the other hand, there are the apparently incontrovertible data showing that pronominal adjectives cannot be treated as predicates that conjoin with the predicates expressed by their head nouns, with extension intersection as the corresponding semantics—e.g. something can be a small elephant and still be a big animal. How can an account that treats adjectives as predicates account for such data?

The solution pursued in TTAA is to treat adjectives as vague predicates. In that capacity, the thought was, they aren't really different from other property words—virtually all descriptive predicates of our language are vague, when you look at them closely enough. It is just that the vagueness of adjectives is typically of a quite special sort, and subject to particular strategies for resolution or reduction. In particular, when the adjective occurs prenominally, the context, or much of it, is typically supplied by the following noun. Furthermore, when an adjective occurs predicatively—more specifically, when it occurs as complement of some copula, then the head noun of the subject phrase may also play its part in the adjective's interpretation, though in these cases there may be stronger competition from other disambiguating

© HANS KAMP, 2013 | DOI:10.1163/9789004252882_010
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

factors. Such disambiguating factors other than the following noun can play their part in the interpretation of prenominal adjectives as well. In fact, it is not hard to find cases where the head noun itself does play such a role. In particular, application standards fixed earlier in the given discourse may affect the interpretation of an adjective, both when it occurs pronominally and when it occurs predicatively.

TTAA's solution has two components. One is a theory of contextual disambiguation; this I hinted at just now when I spoke of the different disambiguating factors that can enter into the interpretation of a given adjective occurrence. TTAA too does little more than nodding in the direction of such a theory. There has been a good deal of progress on this point in recent years, most prominently through the work of Chris Kennedy. But some of the more difficult problems, which have to do with 'non-linear' adjectives like *clever*, *adroit* and the like, are still in the workshop. The other component is a formal framework for describing vagueness and its potential reductions and resolutions in context. The framework opted for in TTAA was van Fraassen's Supervaluation approach, which in its application to vagueness takes on a rather different character than it had in van Fraassen's own foundations of Free Logic. As far as the use of supervaluation is concerned, TTAA and Kit Fine's ***Vagueness, Truth and Logic*** came up with the same suggestion for dealing with the same general phenomenon—independently and, as far as Fine and I were able to tell at the time, simultaneously. But otherwise the two papers have quite different agendas. Fine speaks of vagueness as a defect of language that should and can be rectified, perhaps only bit by bit in actual practice, leading to increasingly perfect languages. That idea is alien to TTAA, which starts from the observation that adjectives are used in varying contexts affecting their extensions in often conflicting ways. But having been used in one such context disables them in no way from then being used, just as happily, in another, possibly conflicting context. The point here is that, as natural languages function, they do not evolve to more perfect languages. Their adjectives remain, by and large, as vague, no matter how often they are used in varying contextual settings. After each use they return, largely unscathed, to their zero setting, ready to face new and different contextualisations.

When I wrote TTAA, one thing that was not on my mind was the paradox of the heap. To someone familiar with the vagueness literature as it exists today this may sound hardly credible; and there was arguably no excuse for it even then. But those were early days in the logic, philosophy and linguistics of vagueness and it was also a much earlier me. It was Richmond Thomason who pointed out to me how the formal account of vagueness proposed in

TTAA could be used to give an explanation of the paradox, showing why it is a fallacy that we can so easily misunderstand as valid. I should add at this point that an important ingredient to the story that TTAA was able to tell is its reconstruction of intermediate truth values as probabilities over sets of verifying completions. Part of the story is that on intuitively natural ways of fixing the probability function a conditional of the form *If patch n is red, then patch n+1 is red* get values very close to fully true, while long conjunctions of such conditionals, or universal quantifications over them can nevertheless get values as low as 0 (i.e. absolute falsity).

To be told that your theory can deal with a problem that you didn't think of when developing it, is always nice, especially when you are told that others have been looking for an answer to the problem and didn't find it easy to come up with one. But somehow the solution that TTAA provides for the Sorites Paradox just didn't feel right to me. *The Paradox of the Heap* was the result of my attempts to understand better why I felt that way. The main source of the resistance I couldn't overcome, it soon became clear to me, was directly connected with the original motivation for TTAA. Among the factors that help reduce the vagueness of an adjective in use is the stability, or inertia, of standards fixed earlier on in the given discourse, or by the setting in which it takes place. If such contextual effects are strong, and in particular when some of them are hard-wired, as it were, into the grammar of the language, that would do much to strengthen our intuition that sorites reasoning is valid and that would make it into a more real and hardened paradox. The Supervaluation-based explanation that Thomason had suggested seemed too easy—too glib—because it ignored these contextual hysteresis effects. For all I could see, these weren't just pragmatic effects that could be easily overwritten, but mechanisms belonging to the core of how our language works.

The Paradox of the Heap tries to see what happens when one assumes that context fixing of the extensions of vague predicates has a quasi-grammatical aspect, with overwriting as a very last resort, available only in contexts that are deeply compromised by successive, increasingly burdensome commitments. Trying to find a solution to the paradox on those terms was a way of making that task about as hard as it could be made. And that was the central point of the paper. I now feel that the paper probably went too far in pursuing the line of hard-wired contextual effects. I should probably have followed the advice of an early fellow traveler in this territory, Manfred Pinkal, who objected specifically to this kind of hard-wiring already at the time when the paper had just been written. In the meantime there have been a number of partly formal proposals to deal with the paradox within a framework

in which contextual adaptation is treated as a major aspect of the semantics of adjectives and other vague predicates, but which do not go for the hard-wiring I adopted, or which do so in an attenuated form.

The semantics that the paper presents for a formal language containing vague predicates is highly non-standard and it gives rise to formal problems, which, if nothing else, are technically intriguing. Perhaps the technical problems are even worth pursuing for someone who has qualms about some of the assumptions on which the semantics is based, just for the sheer mathematical fun of it. But for all I know the offer has found few takers. A confounding factor is that not everything that is said in the paper on the technical aspects of the semantics it proposes is quite right, something that is made explicit in an appendix that has been added in this reprint. Even if the few things that went wrong in the original paper have been put right, in the way that the appendix tries to, the technical problems to which the approach gives rise may seem too heavily burdened with conceptually problematic assumptions to tempt the philosophically motivated. At least this much can be said for the kind of formal solution that the paper formulates as target: if a reasonable solution could be found of the kind the paper advocates—a logic that reflects the semantics in the sense of the fixed point logics that the paper defines and that also has some reasonable inferential properties (i.e. one that isn't so weak, that there isn't anything of interest you can do in it)—then such a solution might still qualify as illuminating, by virtue of the fact that it has been obtained under what must be about the most unaccommodating preconditions. Even someone who regards some of those premises as conceptually unwarranted, and the whole enterprise as misguided for that reason, should be able to acknowledge that much.

From a current perspective I see these two papers both as precursors, each in its own way, of a growing trend—mostly within linguistics, but increasingly also within philosophy—of seeing context-dependence as a central aspect of vagueness in natural language predicates and to identify and disentangle the different contextual factors that can influence the use and interpretation of such predicates. We now know, largely through the work of Kennedy and his co-authors, how different types of adjectives vary in their sensitivity to contextual factors, and we also have a better grip on what the different contextual factors are. Current work on multi-dimensional adjectives is an indication of how work in this area of research is becoming much more sophisticated, both as regards the methods and tools of analysis and the questions asked. But current work is also in indicator of how much distance still remains to be covered.

TWO THEORIES ABOUT ADJECTIVES[a,*]

Hans Kamp

1.

I will discuss two theories about adjectives. The first theory dates from the late 1960s. It is stated in Montague (1970a) and Parsons (1968). According to this theory the meaning of an adjective is a function which maps the meanings of noun phrases onto other such meanings; e.g. the meaning of *clever* is a function which maps the meaning of *man* onto that of *clever man*, that of *poodle* onto that of *clever poodle*, etc. Predicative uses of adjectives are explained as elliptic attributive uses. Thus, *This dog is clever* is analysed as *This dog is a clever dog*—or as *This dog is a clever animal*, or perhaps as *This dog is a clever being*. Which noun phrase ought to be supplied in this reduction of predicative to attributive use is in general not completely determined by the sentence itself, and to the extent that it is not, the sentence must be regarded as ambiguous.

---

* Since I presented the outline of this paper at the Cambridge conference I—and, I hope, this paper—have profited from discussions with and comments by Michael Bennett, Richard Grandy, Hidé Ishiguro, David Lewis, Richmond Thomason and, in particular, George Lakoff. I was equally fortunate to hear Sally Ginet's paper on comparatives at the summer meeting of the Linguistic Society of America in Ann Arbor, which proposed an approach similar to that taken here. Only after the present paper had already been given its final form did I become acquainted with Kit Fine's article 'Vagueness, truth and logic' which expresses on the topic of vagueness, which is the central theme of the second part of my paper, views very similar to those which can be found here. I know that I would have been able to offer a better contribution to this volume if I had known about Fine's work earlier.

[a] Editors' note: This article was originally published in E. Keenan (ed.), *Formal Semantics of Natural Languages*, Cambridge University Press, Cambridge, 1975, 123–155. We would like to thank Cambridge University Press for its kind permission to reprint the article. We carefully edited the article, corrected some typos, added full bibliographical references and brought some formulas in line with now more common usage. We would like to thank Galit Sassoon for careful proofreading and extremely helpful comments. Hans Kamp provided some additional footnotes, indicated by Roman letters, which clarify some of the arguments in the original text. We consecutively renumbered the original footnotes.

© HANS KAMP, 2013 | DOI:10.1163/9789004252882_011
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

The main virtue of this doctrine is that it enables us to treat, whithin a precise semantic theory for a natural languae—as e.g. that of Montague—adjectives in such a way that certain sentences which are, or might well be, false are not branded by the semantics as logically true. Examples of such sentences are:

(1) Every alleged thief is a thief
(2) Every small elephant is small
(3) If every flea is an animal, then every big flea is a big animal

Each of these sencences would come out logically true in Montague's model theory if it were to treat adjectives as ordinary predicates, so that the logical form of (1), for example, would be $(\forall x)(A(x) \wedge T(x) \rightarrow T(x))$.

Moreover, the theory allows us to express in very simple mathematical terms some important semantic features which some, though not all, adjectives possess. In order to give precise formulations of such features, it is necessary to make some assumptions about the comprehensive semantic theory in which this particular doctrine about adjectives is to be embedded. These assumptions can all be found in Montague (1970a). I regard them as basiclly sound, but would like to point out to those who have strong qualms about possible world semantics that the distinctions drawn by the definitions below do not depend on these assumptions as such.

The assumptions are the following:

(a) Each possible interpretation (for the language in question) is based upon (i) a certain non-empty set W of possible worlds (or possible situations, or possible contexts) and (ii) a set U of individuals.
(b) A property relative to such an interpretation is a function which assigns to each $w \in W$ a subset of U (intuitively the collection of those individuals which satisfy the property in that particular world (or context) w).
(c) The meaning of a noun phrase in such an interpretation is always a property.

Thus, the meanings of adjectives in an interpretation of this kind will be functions from properties to properties.

We may call an adjective *predicative in* a given interpretation if its meaning F in that interpretation satisfies the following condition:

(4) there is a property Q such that for each property P and each $w \in W$, $(F(P))(w) = P(w) \cap Q(w)$.

Once we have singled out a given class $\mathcal{K}$ of admissible interpretations, we can also introduce the notion of being *predicative* simpliciter: an adjective is *predicative* (with respect to the given class $\mathcal{K}$) if and only if it is predicative in each interpretation (belonging to $\mathcal{K}$).

Predicative adjectives behave essentially as if they were independent predicates. If for example *four-legged* is treated as predicative then any sentence *If every N_1 is an N_2 then every four-legged N_1 is a four-legged N_2*, where N_1 and N_2 are arbitrary noun phrases, will be true in each admissible interpretation in all the worlds of that interpretation.

Predicative adjectives are, roughly speaking, those whose extensions are not affected by the nouns with which they are combined. Typical examples are technical and scientific adjectives, such as *endocrine*, *differentiable*, *superconductive*, etc.

We may call an adjective *privative in* a given interpretation if its meaning F in that interpretation satisfies the condition

(5) for each property P and each $w \in W$, $(F(P))(w) \cap P(w) = \emptyset$.

Again, an adjective will be called *privative* if (5) holds on all admissible interpretations.

A privative adjective A is one which, when combined with a noun phrase N produces a complex noun phrase AN that is satisfied only by things which do not satisfy N. If A is a privative adjective then each sentence *No AN is an N* will be a logical truth. Adjectives that behave in this way in most contexts are e.g. *false* and *fake*. I doubt that there is any English adjective which is privative (in the precise sense here defined) in all of its possible uses.

An adjective is *affirmative in* a given interpretation if its meaning satisfies

(6) for each P and w, $(F(P))(w) \subseteq P(w)$.

It is *affirmative* if (6) holds in all admissible interpretations.

Clearly all predicative adjectives are affirmative. But there are many more. In fact the vast majority of adjectives are affirmative. Typical examples of affirmative adjectives which are not predicative are *big*, *round*, *pink*, *bright*, *sharp*, *sweet*, *heavy*, *clever*.

Finally, an adjective is *extensional in* a given interpretation if

(7) there is a function F' from sets of individuals to sets of individuals such that for every P and w $(F(P))(w) = F'(P(w))$

and *extensional* if (7) holds in all admissible interpretations.

Thus an extensional adjective is in essence an operation on extensions of properties: if two properties have the same extension in w then the properties obtained by applying the adjective to them also have the same extension in w.

Clearly all predicative adjectives are extensional. Non-extensional adjectives are, for example, *affectionate* and *skilful*. Even if (in a given world) all and only cobblers are darts players, it may well be that not all and only the

skilful cobblers are skilful darts players;[1] and even if all men were fathers the set of affectionate fathers would not necessarily coincide with the set of affectionate men.[2]

It is an interesting question whether there are any adjectives which are extensional but not predicative. It has been suggested[3] that, in particular, such adjectives as *small*, *tall*, *heavy*, and *hot* belong to this category. Indeed, these adjectives are evidently not predicative, whereas their extensionality follows from a certain proposal according to which they derive from their comparatives in the following way. Let A be an adjective of this kind, and let $\mathcal{R}$ be the binary relation represented by the phrase *is more A than*. The function $\mathcal{A}$ from properties to properties which is associated with A is then characterized by

(8) for any property P and world w $\mathcal{A}(P)\,(w) = \{u \in P(w) : \text{for most } u' \in P(w), \langle u, u' \rangle \in \mathcal{R}(w)\}$

It will soon be evident why I have not much sympathy for analyses of positives in terms of comparatives generally. At this point, however, I only want to express some reservations which concern (8) in particular. That (8) cannot be right is brought out by the fact that it logically excludes the possibility that, for any property P, most Ps are small Ps and only a few are large Ps. Thus, what we usually call a small car in England would according to (8) not be a small car; for we call most English cars small. (One might perhaps reply that this only shows that by *small car* we mean a *car of a small model*. But that does not quite do. After all, it is the *individual* cars we call small.) In this case, the conflict between usage and the consequences of (8) arises from the fact that cars are naturally divided into categories; and it is to these, if anything, that (8) applies.

There is yet another reason why (8) might fail for *small*. We might have a clear concept of what is the normal size of objects satisfying a certain property, even if objects of that size which have the property do not or only rarely occur. It is conceivable to me that we would then call almost all members of a species S small members of S if there was strong biological evidence that only accidental and abnormal circumstances C prevent the majority from growing into a height which most members of the species would reach under conditions we would regard as normal. Yet we might still be unwilling to call them small members of S-under-the-circumstances-C; for, as objects

[1] This example was given at the conference by Professor Lewis.
[2] This example was given to me by Dr. Hidé Ishiguro of University College, London.
[3] Cf. Bartsch & Vennemann (1972), part II.

falling under that second description, they should be expected to have the size they do have. If, moreover, *S* and *S*-under-*C* had precisely the same extensions, the case would tend to show that *small* is not purely extensional. But that it is so difficult to come up with a concrete and convincing example of this sort is perhaps an indication that for all practical purposes *small*, and similar adjectives, are indeed extensional.

2.

This theory of adjectives is of course not new. The observation that *John is a good violinist* cannot be analysed as *John is good and a violinist* is probably too old to be traced back with precision to its origin.[4] What is perhaps new in the doctrine as I have stated it here is the emphasis on the fact that what has for a long time been observed to be a feature of certain adjectives is a common feature of them all. *All* of them are functions from noun phrases to noun phrases. Some adjectives (expressed in (4)), however, possess a certain invariance property that makes them behave as predicates, which when combined with a noun phrase give a complex equivalent in meaning to the conjunction of the predicate represented by the adjective and that represented by the noun phrase.

Even if this theory does accomplish the rather simple-minded tasks for which it was designed, one may feel dissatisfied with it for a variety of reasons. Here I will mention only one (although I believe there are other grounds for dissatisfaction as well). The theory is incapable of providing an adequate treatment for the comparative and superlative. For reasons of convenience, I will concentrate on the comparative and leave the superlative aside; but the theory which will emerge from our considerations will handle the superlative as well.

From a naïve point of view, the comparative is an operation which forms out of an adjective a binary predicate. I believe that this naïve point of view is correct: that when we learn a language such as English, we learn the meanings of individual adjectives, and, moreover, the semantic function which this comparative-forming operation performs *in general*, so that we have no difficulty in understanding, on first hearing, the meaning of the comparative of an adjective of which we had thus far only encountered the

[4] A clear exposition of a view about the adjective *good* which is essentially what is here proposed for adjectives in general can be found in Geach (1956). Notice, however, that not only does *good*, as Geach makes clear, fail to be a predicate; it is not even extensional (cf. *skilful* above).

positive. If this is so, then the meaning of an adjective must be such that the comparative can be understood as a semantic transformation of that meaning into the right binary relation.

It is quite obvious that if adjectives were ordinary predicates no such transformation could exist. How could we possibly define the relation *x is bigger than y* in terms of nothing more than the extension of the alleged predicate *big*?

Could functions from properties to properties serve as the basis for such a transformation? This is a more problematic question. One might for example characterize the transformation as follows:

> For any adjective A with meaning $\mathcal{A}$ in a given interpretation we have for any $u_1, u_2 \in U$ and $w \in W$:

(9) u_1 is more A than u_2 in w iff
  (a) for every property P such that u_1 and u_2 both belong to $P(w)$, if u_2 belongs to $\mathcal{A}(P)(w)$ then so does u_1.
  (b) there is a property P such that $u_1, u_2 \in P(w), u_1 \in \mathcal{A}(P)(w)$, and $u_2 \notin \mathcal{A}(P)(w)$.

This definition is in the right direction. But I doubt that it will do. In particular, I doubt whether (b) is a necessary condition. Take *tall*, for example. According to (9) u_1 is taller than u_2 only if there is a property P that applies to both of them and such that u_1 is a tall P while u_2 is not. But suppose that u_1 is taller than u_2 by a tiny bit. Can we then find a P which satisfies this condition? The question is not easy to answer. Let us suppose, for the sake of argument, that *tall* can be correctly defined by (8) as *taller than most*. Then, the question depends on whether we can find a property P such that u_1 is taller than most Ps while u_2 is not. But this can only be the case if there are enough things in the extension of P which have heights intermediate between those of u_1 and u_2. And perhaps there are no such things at all. Now, in our discussion of extensional adjectives we found that (8) is probably not adequate in any case. So there may be after all a property P which satisfies the condition. In this manner we might succeed in saving (9) by imploring the assistance of some bizarre property whenever we need one. But I find this solution ad hoc and unsatisfactory. What underlies the possibility of making comparative claims is that adjectives can apply to things in various derees. It is my strong conviction that when we learn the meaning of an adjective we learn, as part of it, to distinguish with greater or lesser precision to what degree, or extent, the adjective applies to the various entities to which it applies at all. Once we have learned this, we are able to understand the comparative of the adjective without additional explanation, provided we understand the function of the comparative, in general.

In order to give my view on the primacy of positive over comparative an adequate foundation I will develop a semantic framework in which the idea of a predicate being true of an entity to a certain degree can be made coherent and precise. This specific problem is closely related to such general features of natural languages as vagueness and contextual disambiguation; indeed, I hope that the theory which I will outline will provide an adequate framework for the treatment of these problems as well.

Before stating what, at this point, I believe to be the most promising framework for our purpose, I first want to make some remarks on a theory of formal logic which has been often proposed just for the solution of the problems with which I want to deal, viz. multi-valued, or many-valued logic.

Most systems of multi-valued logic available in the literature are systems of propositional calculus. In view of our purpose, our interest should lie with multi-valued predicate logic. But for reasons of exposition, I will consider the simpler propositional logics.

Multi-valued logics differ from ordinary two-valued logic in the first place by their model theories. Indeed, many such systems are syntactically indistinguishable from standard formulations of ordinary propositional calculus; and I will consider, for the time being, only such systems of multi-valued logic, all of which have the same syntax, based upon an infinite set $q_1, q_2, q_3, \ldots$ of propositional variables. Starting with these we can recursively construct complex formulae, i.e. $\neg(\phi)$, $\wedge(\phi, \psi)$, $\vee(\phi, \psi)$, $\rightarrow(\phi, \psi)$, $\leftrightarrow(\phi, \psi)$ from already constructed formulae ϕ and ψ. (I will write $(\phi \wedge \psi)$ for $\wedge(\phi, \psi)$, etc.) Let us call this language of propositional logic L_0.

A multi-valued semantics for L_0 will provide for this language a model theory based upon some set *TV* (of 'truth values') the cardinality of which is ≥ 2. Two-valued propositional calculus emerges as a case where *TV* contains exactly two elements. A *model* for L_0 according to such a model theory based upon *TV* is a function which assigns to each variable q_i an element of *TV*. Such a function uniquely determines the (truth) values of the complex formulae of *L*, in virtue of another component which specifies for each t_i in *TV* what the value is of $\neg\phi$, given that t_i is the value of ϕ; for each t_i, t_j in *TV*, what the value is of $(\phi \wedge \psi)$, given that ϕ has t_i and ψ has t_j; and similarly for the other connectives.

The definition of logical truth requires a third component of the theory, which singles out a proper non-empty subset TV_t of *TV*, the set of 'designated' truth values. A formula will be regarded as *logically true*, if in each model it has a value belonging to TV_t. *Logical consequence* can be defined in an analogous manner.

Thus, we come to the following formal definition:

A *multi-valued model theory* (in short m.m.t.) *for* L_0 is a triple $\langle TV, TV_t, F\rangle$ where:

(i) TV is a set of cardinality ≥ 2;
(ii) TV_t is a proper, non-empty subset of TV;
(iii) F is a function which maps each n-place connective of L_0 onto an n-place function from TV into itself.

A *model for* L_0 *relative to the m.m.t.* $\langle TV, TV_t, F\rangle$ is a function from $\{q_1, q_2, q_3, \ldots\}$ into TV.

Let $\mathcal{M} = \langle TV, TV_t, F\rangle$. The truth value of a formula ϕ of L_0 in a model M relative to $\mathcal{M}$, $[\phi]_M^{\mathcal{M}}$ is defined by the clauses:

(i) $[q_i]_M^{\mathcal{M}} = M(q_i)$;
(ii) $C(\phi_1, \ldots, \phi_n)[_M^{\mathcal{M}}] = F(C)([\phi_1]_M^{\mathcal{M}}, \ldots, [\phi_n]_M^{\mathcal{M}})$ for any n-place connective C of L_0.

ϕ is *logically true in* $\mathcal{M}$ iff $[\phi]_M^{\mathcal{M}} \in TV_t$ for all models M relative to $\mathcal{M}$.

Clearly, the classical semantics for propositional calculus is the model theory $\langle\{0,1\}, \{1\}, F_c\rangle$, where $F_c(\neg), F_c(\wedge), \ldots\ldots, F_c(\leftrightarrow)$ are the functions defined by the usal two-valued truth tables for these connectives.

It is natural to require of a model theory for L_0 based upon a truth value set of cardinality ≥ 2 that it has the feature:

(10) there are two particular elements of TV—let us call them 0 and 1—such that

(i) $1 \in TV_t$;
(ii) $0 \notin TV_t$; and
(iii) for each connectivc C, the restriction of $F(C)$ to $\{0,1\}$ is the usual two-valued truth table for C (i.e. there are among the truth values two, which we might think of as 'absolute falsehood' and 'absolute truth', with respect to which the connectives behave in the ordinary classical manner).

It appears that those model theories for L_0 which have been seriously proposed in the literature do indeed satisfy (10).[5] The vast majority of these theories assume, moreover, a linear ordering of the members of TV with respect to which 0 is the smallest and 1 the largest element. The formal properties of the theory, as well as its philosophical relevance, then depend on the characterization of the functions $F(C)$. Clearly there are, whenever the cardinality of TV is greater than 2, various such functions which do

[5] For references see Rescher (1969).

not violate (i). The question is which of these 'correctly capture' the function of the connectives *not*, *and*, *or*, ... given a particular interpretation of what the truth values in *TV* really represent. Let us consider the case where $TV = \{0, \frac{1}{n-1}, \frac{2}{n-1}, \ldots, \frac{n-2}{n-1}, 1\}$ and where these numbers represent 'degrees of truth'—the higher the number the higher the degree.[b]

What function would adequately reflect our intuitions about the semantic behaviour of these connectives in this case? I think there are no such functions. The reason is that the connectives *not*, *and*, *or*, ... are not functions of degrees of truth. This becomes evident almost immediately when one reflects upon the definition of $F(\neg)$. The natural suggestion here is that $F(\neg)\left(\frac{k}{n-1}\right) = 1 - \frac{k}{n-1}$, i.e. that the negation of a proposition is true exactly to the degree that that proposition itself fails to be true. And this indeed is a definition of $F(\neg)$ which is commonly accepted.

Now let us assume that n is odd so that one of the truth values is $\frac{1}{2}$. What value would $F(\wedge)$ assign to the pair of arguments $\left(\frac{1}{2}, \frac{1}{2}\right)$? It is plausible that the value should be $\leq \frac{1}{2}$. For how could a conjunction be true to a higher degree than one of its conjuncts? But which value $\leq \frac{1}{2}$? $\frac{1}{2}$ seems out because if $[\phi]_M^{\mathcal{M}} = \frac{1}{2}$ then, if we accept our definition of $F(\neg)$, $[\neg\phi]_M^{\mathcal{M}} = \frac{1}{2}$. So we would have $[\phi \wedge \neg\phi]_M^{\mathcal{M}} = \frac{1}{2}$, which seems absurd. For how could a logical contradiction be true to *any* degree? However, if we stipulate that $(F(\wedge))\left(\frac{1}{2}, \frac{1}{2}\right) = 0$, we are stuck with the even less desirable consequence that if $[\phi]_M^{\mathcal{M}} = \frac{1}{2}$, $[\phi \wedge \phi]_M^{\mathcal{M}} = 0$. And if we choose any number between 0 and $\frac{1}{2}$, we get the wrong values for both $\phi \wedge \neg\phi$ and $\phi \wedge \phi$.[6] This argument indicates why we cannot represent the connectives accurately within the narrow framework of multi-valued semantics based upon linearly ordered truth-value sets. The reason can be expressed thus: the truth value of a complex formula—say $\phi \wedge \psi$—should depend not just on the truth values of the components—i.e. ϕ and ψ—but also on certain aspects of these formulae which contribute to their truth values but cannot be unambiguously recaptured from them.

[6] This argument is certainly not new. It can be found, e.g., in Rescher (1969). Yet it seems to have failed to discourage people from trying to use multi-valued logic in contexts where the argument shows it to be inadequate.

[b] So this is an application of multi-valued logic in which there are n 'truth values' (in other words, $|\ TV\ | = n$) and where the values $\frac{1}{n-1}, \frac{2}{n-1}, \ldots, \frac{n-2}{n-1}$ are the 'intermediate' truth values. The natural choice for TV_t in this case is $\{1\}$.

The possibility of treating the connectives truth-functionally in two-valued model theory (while not in model theories based on larger linearly ordered sets of truth values) is a reflection of the fact that two-element sets are the only linearly ordered truth-value sets which can be regarded as Boolean algebras in the following sense:

If we define the Boolean operations $\cap, \cup, -$ in the usual manner in terms of the ordering relation $\leq$ of a linearly ordered set *TV* (i.e. if we put $C \cap C' = df$ the largest C'' under $\leq$ such that $C'' \leq C$ and $C'' \leq C'$; $C \cup C' = df$ the smallest C'' such that $C \leq C''$ and $C' \leq C''$; and $\bar{C}$ as the largest C' such that $C \cap C' = \emptyset$), then, if *TV* consists of two elements, 0 and 1, we obtain the two-element Boolean algebra $\langle \{0,1\}, \cap, \cup, - \rangle$. But as soon as *TV* contains more than two elements, the resulting algebra is not a Boolean algebra. In particular, the equation $C \cup \bar{C} = 1$ will no longer be satisfied by all elements *C*.

3·

This last observation suggests in which direction a solution to our difficulty might be found: we should choose as truth-value sets not linear orderings, but rather sets which, like the two-valued system, display the structure of the propositional calculus, viz. Boolean algebras. We may then, if we want to, further 'reduce' these Boolean algebras to linearly ordered systems; in this reduction, different Boolean values may be assigned the same element of the linear ordering. But this will no longer affect our semantic characterization of the connectives, as these will now be defined in the Boolean truth-value space, and, thus, not directly on the linearly ordered, 'ultimate' truth values themselves. It may now happen that, even if ϕ and ψ have the same ultimate value, $\phi \cap \chi$ (or $\phi \cup \chi$) has a different ultimate value from that of $\psi \cap \chi$ (or $\psi \cup \chi$); this may happen in cases where the Boolean values of ϕ and ψ which reduce to the same ultimate value are nevertheless distinct.

This idea is by now quite familiar to logicians. Yet it is surprising that its use for the specific problems with which we are here concerned occurred as late as it did; for there is a branch of mathematics, viz. probability theory, in which it has been accepted as the standard solution to what is in many ways the same problem as the one we are facing here. I am referring to the theory of probability in the definitive mathematical form that Kolmogorov (1970) gave it in the 1930s. In this theory, one associates with a proposition in first instance a certain set. With this set is associated in turn a real number in the closed interval $[0 : 1]$. This number gives the probability of

the proposition. Now while the set associated with the conjunction of two propositions is a simple function of the sets associated with the conjuncts, viz. their intersection, there is no way of telling in general the probability of the conjunction on the basis of just the probabilities of the conjuncts; and this is as it should be, for when p and q each have probability $\frac{1}{2}$, the probability of $p \wedge q$ could, intuitively, be anything between 0 and $\frac{1}{2}$.

Perhaps the main philosophical problem which this approach raises is that of giving a plausible interpretation of the sets with which propositions are associated. I will consider here only one doctrine, according to which the elements of the sets are regarded as possible worlds, or possible situations. Thus, the probability of a proposition is measured in terms of the set of those possible worlds in which it is true. Seen in this light, probability theory is closely connected with the possible-world semantics for modal and other types of non-extensional logic. Both theories associate with a given sentence, in any particular interpretation, a set (of possible worlds, points of reference, contexts, etc.).

Of course, probability theory and intensional logic are concerned with different sorts of problems. In intensional logic, we are primarily concerned with the analysis, in terms of the set of all possible 'worlds' (as well as, perhaps, various structural properties of this set), of the semantical function of certain non-truth functional operators, such as *it is necessarily the case that*. In probability theory, one does not consider such intensional operators, but concentrates on the probability function which associates real numbers with the sets, and investigates how the probabilities of certain complex expressions depend on the probabilities of their components—often under certain assumptions about these components, such as independence or disjointness.

There is another theory of formal semantics which fits within the general frame which we are now discussing, viz. the theory of partial interpretations and supervaluations. This theory is, in its simplest form, a generalization of ordinary two-valued model theory, which allows for the possibility that in a given interpretation for a certain formal language (say, ordinary first order logic with a description operator) some sentences of the language are neither true nor false. Yet, in order to avoid the—from a certain standpoint undesirable—consequence that whenever p is without truth value, so are, among others, $p \wedge \neg p$ and $p \vee \neg p$, one considers the collection of all interpretations which extend the given interpretation by filling out its truth gaps in a consistent manner. If a formula comes out true in each of these completions, it will be regarded as true in the interpretation, even if it is not assigned a truth value directly by the (incomplete) recursive definition of

truth. Similarly it will be counted as false in the interpretation if it is false in each of its completions.

One may view this process, again, as one of assigning, in a given interpretation, sets to sentences: each sentence is assigned the set of all completions in which it is true. Sentences already true in the given interpetation, and also such sentences as $p \vee \neg p$ where p itself is not assigned a truth value directly, will be assigned the set of all completions, those already false as well as sentences such as $p \wedge \neg p$ will be assigned the empty set; only sentences which neither have a truth value in virtue of the recursive truth definition, nor have the form of a logical identity or contradiction, may be assigned intermediate sets.

The theory, which was first introduced by van Fraassen (1969) suggests in what way the framework under discussion might be used in an analysis of vagueness. Vagueness is one of the various reasons why certain sentences may be without truth value. Thus, if we regard the world, or any specific speech situation in it, as providing an interpretation for English, what it provides is at best a *partial* interpretation.[7] For such a partial interpretation, we may consider the various completions in which all instances of vagueness are resolved in one way or another. The quantity of such completions in which a certain sentence is true ought, then, to be in some sense a measure for the degree to which the sentence is true in the original interpretation. Such considerations would, of course, apply not only to adjectives, but to other parts of speech as well, in particular to those grammatical categories which, like adjectives, are usually treated as 1-place predicates in simple-minded predicate-logic-symbolizations of English sentences, viz. common nouns and intransitive verbs. I will later try to say something about systematic differences between the semantic behaviour of adjectives and that of these other two categories.

These considerations naturally lead to a modification of the model theory for formal or natural languages which I will exemplify for a rather simple case, viz. first order predicate logic. The example will make it clear enough how one could adapt in a similar fashion more complicated model theories—such as those for intensional logics or for fragments of natural languages.

---

[7] Of course there are other factors whose effect is that a situation of speech will in general determine only a partial interpretation; for example, many predicates are not applicable to individuals of certain kinds—and yet not every statement which attributes a predicate to such a semantically improper object should be regarded as illformed. Such other sources of interpretational incompleteness, however, will not concern us here.

Let us consider the language L for predicate logic, the logical symbols of which are $\neg, \wedge, \exists$, and the variables $\nu_1, \nu_2, \nu_3, \ldots$, and the non-logical symbols of which are the n-place predicate letters $\mathcal{Q}_i^n (n = 1, 2, 3, \ldots; i = 1, 2, 3, \ldots)$.

A *classical model for L* is a pair $\langle U, F \rangle$ where (i) U is a non-empty set and (ii) F assigns to each $\mathcal{Q}_i^n$ an n-place relation on U.

The *satisfaction value of* a formula ϕ of L *in* $M = \langle U, F \rangle$ *by* an assignment a of elements of U to the variables (in symbols $[\phi]_{M,a}$) is defined by the usual recursion:

(i) $[\mathcal{Q}_i^n(\nu_{i_1} \ldots \nu_{i_n})]_{M,a} = 1$ iff $\langle a(\nu_{i_1}), \ldots, a(\nu_{i_n}) \rangle \in F(\mathcal{Q}_i^n)$
(ii) $[\neg\phi]_{M,a} = 1$ iff $[\phi]_{M,a} = 0$
(iii) $[\phi \wedge \psi]_{M,a} = 1$ iff $[\phi]_{M,a} = 1$ and $[\psi]_{M,a} = 1$
(iv) $[(\exists \nu_i)\phi]_{M,a} = 1$ iff for some $u \in U [\phi]_{M,[a]^u/_{\nu_i}} = 1$

A sentence ϕ of L is *true in* M iff $[\phi]_{M,a} = 1$ for some a.

For any model M, Tr_M will be the set of sentences of L which are true in M and Fa_M, the set of sentences of L false in M.

A *partial model* for L is a pair $\langle U, F \rangle$ where (i) U is a non-empty set and (ii) F assigns to each letter $\mathcal{Q}_i^n$ an ordered pair $\langle F^+(\mathcal{Q}_i^n), F^-(\mathcal{Q}_i^n) \rangle$ of disjoint n-place-relations on U.

N.B. I will assume throughout that M is a model and is of the form $\langle U, F \rangle$.

The *satisfaction value* of a formula ϕ in M by an assignment a is now defined by:

(i) (a)$[\mathcal{Q}_i^n(\nu_{j_1}, \ldots, \nu_{j_n})]_{M,a} = 1$ iff $\langle a(\nu_{j_1}), \ldots, a(\nu_{j_n}) \rangle \in F^+(\mathcal{Q}_i^n)$
(b)$[\mathcal{Q}_i^n(\nu_{j_1}, \ldots, \nu_{j_n})]_{M,a} = 0$ iff $\langle a(\nu_{j_1}), \ldots, a(\nu_{j_n}) \rangle \in F^-(\mathcal{Q}_i^n)$
(ii) (a)$[\neg\phi]_{M,a} = 1$ iff $[\phi]_{M,a} = 0$
(b)$[\neg\phi]_{M,a} = 0$ iff $[\phi]_{M,a} = 1$
(iii) (a)$[\phi \wedge \psi]_{M,a} = 1$ iff $[\phi]_{M,a} = 1$ and $[\psi]_{M,a} = 1$
(b)$[\phi \wedge \psi]_{M,a} = 0$ iff $[\phi]_{M,a} = 0$ or $[\psi]_{M,a} = 0$
(iv) (a)$[(\exists \nu_i)\phi]_{M,a} = 1$ iff for some $u \in U$, $[\phi]_{M,a[^u/_{\nu_i}]} = 1$
(b)$[(\exists \nu_i)\phi]_{M,a} = 0$ iff for all $u \in U$, $[\phi]_{M,a[^u/_{\nu_i}]} = 0$

Again a sentence ϕ is said to be *true in* M iff $[\phi]_{M,a} = 1$, for some a; and ϕ is *false in* M if for some a, $[\phi]_{M,a} = 0$; Tr_M is the set of true, and Fa_M the set of false sentences of L in M. But now it is clearly possible that certain sentences are neither true nor false in M, so that $Tr_M \cup Fa_M$ does not coincide with the set of all sentences of L.

The partial model $M = \langle U, F \rangle$ is said to be *at least as vague as* the partial model $M' = \langle U, F' \rangle$ (in symbols: $M \subseteq M'$) iff for each $\mathcal{Q}_i^n$:

$$F^+(\mathcal{Q}_i^n) \subseteq F'^+(\mathcal{Q}_i^n) \text{ and } F^-(\mathcal{Q}_i^n) \subseteq F'^-(\mathcal{Q}_i^n)$$

(Thus $Tr_M \cup Fa_M \subseteq Tr_{M'} \cup Fa_{M'}$).

To each classical model $\langle U, F\rangle$ for L corresponds a unique partial model, viz. the model $\langle U, F'\rangle$ where for each $\mathcal{Q}_i^n$, $F'^+(\mathcal{Q}_i^n) = F(\mathcal{Q}_i^n)$ and $F'^-(\mathcal{Q}_i^n) = U^n \setminus F(\mathcal{Q}_i^n)$. Classical models, as well as the partial models corresponding to them, will be referred to as *complete* models.

A classical model M is called a *completion* of a partial model M' if M' is at least as vague as (the partial model corresponding to) M.

4.

In the theory of supervaluation, one considers partial models in conjunction with *all* their completions. What I want to do here is formally almost the same—but with one crucial exception. Rather than all completions of a given partial model, we consider only a certain subset of them. In addition, we consider a probability function over a field of subsets of this set of completions,[8] which contains, in particular, for each formula and assignment of elements to its free variables the set of all completions in which the formula satisfies that assignment. (This condition warrants that each sentence has a measure.) The complex consisting of the partial model, the set of completions, the field over that set and the probability function over that field I will call a *vague model*. Formally

A *vague model for* L is a quadruple $\langle M, \mathcal{S}, \mathcal{F}, p\rangle$ where

(i) M is a partial model for L;
(ii) $\mathcal{S}$ is a set of classical models for L which are completions of M;
(iii) $\mathcal{F}$ is a field of subsets over $\mathcal{S}$
(iv) for each $\phi \in L$ and assignment a in the universe of M, $\{M' \in \mathcal{S} : [\phi]_{M',a} = 1\} \in \mathcal{F}$; and
(v) p is a probability measure on $\mathcal{F}$.[9]

---

[8] A field of *subsets* of a given set X (or: a *field over* X) is a set $\mathcal{F}$ of subsets of X, such that (i) $X \in \mathcal{F}$; (ii) $\emptyset \in \mathcal{F}$; (iii) if $Y, Y' \in \mathcal{F}$ then $Y \cap Y', X \setminus Y \in \mathcal{F}$.

A *probability function over* a field $\mathcal{F}$ over X is a function p whose domain is $\mathcal{F}$, whose range is included in the real interval $[0,1]$, and which has the properties: (i) $p(X) = 1$; (ii) if $Y \in \mathcal{F}$, then $p(X \setminus Y) = 1 - p(Y)$; and (iii) if $\mathcal{G}$ is a countable subset of $\mathcal{F}$ such that (a) whenever $Y, Y' \in \mathcal{G}$ and $Y \neq Y'$ then $Y \cap Y' = \emptyset$; and (b) if $\cup\mathcal{G} \in \mathcal{F}$, then $p(\cup\mathcal{G}) = \sum_{Y \in \mathcal{G}} p(Y)$.

[9] From the mathematical point of view, this notion is unproblematic only if the universe U is finite. In that case we do not really need to require that p satisfy the condition (iii) of note 8, but only the weaker condition obtained by replacing the word *countable* in (iii) by *finite*. (iii) is necessary when U is denumerable; in that case, however, as well as when U is uncountable, it may happen that no intuitively correct models exist. The only way in which I can see how to cope with these cases involves non-standard analysis. I do not want to go into this here.

Let $\mathcal{M} = \langle\langle U, F\rangle, \mathcal{S}, \mathcal{F}, p\rangle$ be a vague model for L. For any formula ϕ of L and assignment a of elements of U to the variables, the *degee of satisfaction of ϕ by a in $\mathcal{M}$* is defined as $p(\{M' \in \mathcal{S} : [\phi]_{M',a} = 1\})$. Thus, in particular, if $\phi \in Tr_M$ then $[\phi]_{\mathcal{M}} = 1$ and if $\phi \in Fa_M$ then $[\phi]_{\mathcal{M}} = 0$.

The idea behind the notion of a vague model is this. At the present stage of its development—indeed, at any stage—language is vague. The kind of vagueness which interests us here is connected with predicates. The vagueness of a predicate may be resolved by fiat—i.e. by deciding which of the objects that as yet are neither definitely inside nor definitety outside its extension are to be in and which are to be out. However, it may be that not every such decision is acceptable. For there may already be semantic principles which, though they do not determine of any one of a certain group of objects whether it belongs to the extension or not, nevertheless demand that if a certain member of the group is put into the extension, a certain other member must be put into the extension as well. Take for example the adjective *intelligent*. Our present criteria tell us of certain people that they definitely are intelligent, of certain others that they definitely are not, but there will be a large third category of people about whom they do not tell us either way. Now suppose that we make our standard more specific, e.g., by stipulating that to have an I.Q. over a certain minimum is a necessary and sufficient criterion for being intelligent. Further, suppose that, of two persons u_1 and u_2 of the third category, u_1 has a higher I.Q. than u_2. Then, whatever we decide this minimum to be, our decision will put u_1 into the extension if it puts u_2 into it. Finally, let us assume, for the sake of argument, that any way of making the concept of intelligence precise that is compatible with what we already understand that concept to be, is equivalent to the adoption of a certain minimum I.Q. Then there will be no completions in the partial model that reflect the present state of affairs and in which u_2 is put into the extension of the predicate but u_1 is not.

Formally, if Q_1^1 represents the adjective *intelligent* and the model $\langle M, \mathcal{S}, \mathcal{F}, p\rangle$ reflects the situation just described, and M, in particular, the state of affairs which obtains before any of the possible precise definitions has been adopted, then u_1 and u_2 are both members of $U - (F^+(Q_1^1) \cup F^-(Q_1^1))$ and there is no model $M' \in \mathcal{S}$ such that $u_2 \in F(Q_1)$ and $u_1 \notin F(Q_1)$.

My original motivation in setting up this framework was to give a uniform characterization of the operation which transforms adjectives into their comparatives. Let us see if this is now possible.

The relation *x is more A than y* (where A is any adjective) can be defined in terms of the relation *x is at least as A as y* by stating that

(11) x is more A than y if and only if x is at least as A as y and it is not the case that y is at least as A as x

Therefore, a semantic characterization of this second relation will automatically give us one for the first as well. As there are minor, but undeniable advantages in discussing the relation *at least as ... as*, I will concentrate on that concept.

Let us assume that some of the one-place predicates of L represent adjectives, in particular $\mathcal{Q}^1_1$. We add to L the operator symbol $\geq$. $\geq$ forms out of a one-place predicate $\mathcal{Q}^1_i$ a two-place relation $\geq(\mathcal{Q}^1_i)$. $\geq(\mathcal{Q}^1_i)(x,y)$ should be read as *X is at least as $\mathcal{Q}^1_i$ as y*. (What relation $\geq(\mathcal{Q}^1_i)$ might represent when $\mathcal{Q}^1_i$ is not an adjective is of no concern to us now.) Let L' be the language resulting from the addition of $\geq$ to L. Let $\mathcal{M} = \langle M, \mathcal{S}, \mathcal{F}, p\rangle$ be a vague model for L. In order to expand $\mathcal{M}$ to a model for L' we must determine the positive and negative extensions of the relation $\geq(\mathcal{Q}^1_i)$ in M as well as its extensions in all the members of $\mathcal{S}$. To begin, we will consider just the positive extension in M.

Two possible definitions come to mind. According to the first, an element u_1 of U stands (definitely) in the relation to u_2 iff for every member M' of $\mathcal{S}$ in which u_2 belongs to the extension of $\mathcal{Q}^1_i$, u_1 belongs to that extension as well. So we get, representing the positive extension of $\geq(\mathcal{Q}^1_i)$ in M as $F^+(\geq(\mathcal{Q}^1_i))$,

(12) for all $u_1, u_2 \in U$, $\langle u_1, u_2\rangle \in F^+(\geq(\mathcal{Q}^1_i))$ iff $[\mathcal{Q}^1_i(\nu_1)]_{\mathcal{M},a_2} \subseteq [\mathcal{Q}^1_i(\nu_1)]_{\mathcal{M},a_1}$, where a_1 and a_2 are any assignments with $a_1(\nu_1) = u_1$ and $a_2(\nu_1) = u_2$, respectively, and where $[\mathcal{Q}^1_i(\nu_1)]_{\mathcal{M},a}$ is short for $\{M' \in \mathcal{S}: [\mathcal{Q}^1_i(\nu_1)]_{M',a} = 1\}$

According to the second definition u_1 stands in the relation to u_2 if the measure of the set of completions in which u_1 belongs to the extension of $\mathcal{Q}^1_i$ is at least as large as that of the set of completions in which u_2 belongs to the extension. So we obtain

(13) for all $u_1, u_2 \in U$, $\langle u_1, u_2\rangle \in F^+(\geq(\mathcal{Q}^1_i))$ iff $p([\mathcal{Q}^1_i(\nu_1)]_{\mathcal{M},a_2}) \leq p([\mathcal{Q}^1_i(\nu_1)]_{\mathcal{M},a_1})$, where a_1 and a_2 are as above

Before we consider the relative merits of these definitions let us first remove a flaw which they share. Neither (12) nor (13) allows for the possibility that the comparative relation holds between two objects for each of which it is beyond doubt that it satisfies the positive. For if both u_1 and u_2 belong to $F^+(\mathcal{Q}^1_i)$ then $[\mathcal{Q}^1_i(\nu_1)]_{\mathcal{M},a_1} = [\mathcal{Q}^1_i(\nu_1)]_{\mathcal{M},a_2} = 1$, and so both (12) and (13) would exclude $\langle u_1, u_2\rangle$ from $F^+(>(\mathcal{Q}^1_i))$.[c]

[c] The original paper contains the erroneous '$F^+(\geq(\mathcal{Q}^1_i))$'. '$F^+(>(\mathcal{Q}^1_i))$' denotes the strict ordering corresponding to the weak ordering $F^+(\geq(\mathcal{Q}^1_i))$.

It seems that the only way in which we could meet this difficulty without departing too much from our present format is this: Instead of a vague model, consisting of a partial model M, a field $\mathcal{F}$ over a set $\mathcal{S}$ of completions of M and a probability function p over that field, we need to consider models $\mathcal{M}$, in which the set $\mathcal{S}$ comprises besides completions of M also complete models which in certain ways conflict with M. Such models will represent (hypothetical) situations in which the standards for a predicate are set so high that certain objects which already have that predicate in M now fail to have it—or else in which the standards are set so low that objects belonging to the negative extension of the predicate in M now fall within its positive extension. This leads us to the following modification of the notion of a vague model:

A *graded model for* L is a quadruple $\langle M, \mathcal{S}, \mathcal{F}, p\rangle$, where

(i) M is a partial model for L
(ii) $\mathcal{S}$ is a set of classical models for L with universe U
(iii) $\mathcal{F}$ is a field over $\mathcal{S}$
(iv) For each formula ϕ of L and each assignment a to elements of the universe of M, $\{M' \in \mathcal{S} : [\phi]_{M',a} = 1\} \in \mathcal{F}$
(v) $\{M' \in \mathcal{S} : M'$ is a completion of $M\} \in \mathcal{F}$; and
(vi) p is a probability function over $\mathcal{F}$.

We may, then, define the degree of truth of a sentence of L just as before, except that we now consider the conditional probability of a certain set of completions of M on the set of all completions of M in $\mathcal{S}$. On the other hand, the characterization (13) of the comparative of $\mathcal{Q}^1_i$ is now no longer vulnerable to the objection which led us to the introduction of graded models.

Let us consider an example. Suppose that $\mathcal{Q}^1_1$ represents the adjective *heavy*; that all other predicates represent properties of, and relations between, material objects, and (for simplicity) that $\mathcal{Q}^1_1$ is the only vague predicate. Let U be the set of material objects, and let $\mathcal{M} = \langle M, \mathcal{S}, \mathcal{F}, p\rangle$ be a graded representation (restricted to material objects) of the actual world. What should, in this case, $\mathcal{S}$ and p be? As regards $\mathcal{S}$, a simple answer seems possible in this special case: For each particular real number r, there will be a member M of $\mathcal{S}$, in which the extension of $\mathcal{Q}^1_1$ consists of those objects whose weight (in grams) exceeds r.

It is not possible to say precisely what the function p should be. But this much seems beyond doubt: there should be a strictly monotonic function f from the set of all positive real numbers into the interval $[0, 1]$ so that for any object u with weight r, $p([\mathcal{Q}^1_1(v_1)]_{\mathcal{M},a}) = f(r)$ (for some a with $a(v_1) = u$). Thus, the greater u's weight, the larger the class of members of $\mathcal{S}$ in which $\mathcal{Q}^1_1$ is true of u, and the greater the measure (or 'intermediate truth value') of the formula $\mathcal{Q}^1_1(v)$ under a.

5.

We should now compare (12) and (13). According to (12), u_1 is at least as heavy as u_2 just in case the set of models in which u_1 *is heavy* is true includes the class of those which render u_2 *is heavy* true; and this will be the case if and only if u_1 has greater or equal weight. Indeed, within the context of the present example, (12) is precisely the proposal that can be found in Lewis (1979), where it is attributed to David Kaplan.

According to (13), u_1 will be at least as heavy as u_2, provided u_2 *is heavy* is true in a set of models with measure greater than or equal to that of the set of models in which u_2 *is heavy* is true. Again, this will be true if and only if u_1 has greater or equal weight. Thus, for this special case, the two definitions are equivalent.

But this need not always be so. Suppose, for example, that Smith, though less quick-witted than Jones, is much better at solving mathematical problems. Is Smith cleverer than Jones? This is, perhaps, not clear, for we usually regard quick-wittedness and problem solving-facilities as indications of cleverness, without a canon for weighting these criteria against each other when they suggest different answers. When faced with the need to decide the issue, various options may be open to us. We might decide that really only problem-solving counts, so that after all, Smith is cleverer than Jones; or we might decide on a particular method for weighting the two criteria—so that Smith's vast superiority at solving problems will warrant that, in spite of Jones's slight edge in quick-wittedness, Smith is cleverer than Jones; or we might decide that only quick-wittedness counts; and this time Jones will come out as the cleverer of the two.

It is not clear how the probability function of a graded model $\mathcal{M}$ representing this situation should be defined.[d] Yet, if we assume that the third

[d] The next three sentences of this paragraph (up to, but not including 'Given all that …') are, it now seems to me, garbled beyond straightforward reconstruction. What I meant to say is (I hope) more clearly expressed thus:

'But this much can be said: The function should reflect the following intuitions. There are different options for how to combine the criteria of quick-wittedness and problem solving ability in evaluating *clever*. Let us consider just the three options already mentioned: (i) only problem solving counts (O_1); (ii) only quick-wittedness counts (O_2); (iii) both problem solving and quick-wittedness count and they count with equal weight (O_3). Let us assume also that these options differ in how plausible they are, and that O_1 and O_2 are significantly more plausible than O_3. Then those models in $\mathcal{S}$ that are compatible with O_3, but not with O_1 or O_2 should not count heavily (i.e. their presence in a given satisfaction set should only make a small contribution to the probability value of that set). In other words, models in which Jones is in the positive extension of *clever* but Smith is not shouldn't count

decision is less plausible than either the first or the second, then we should expect members of $\mathcal{S}$ which are compatible with that decision to have no more weight than those which are compatible with the other decisions. Further, relatively few models of the first sort will be such that Jones belongs to the extension of *clever* and Smith not; for Jones is not that much quicker in conversation. But because of the disparity in problem-solving ability, many models compatible with the first decision, as well as a good many that are compatible with the second, will have Smith in the extension of *clever* but not Jones. Given all this, we would expect the measure of the set of members of $\mathcal{S}$ in which Smith belongs to the extension to be greater than that of those members where Jones belongs to the extension. So by (13) and (11), we would have to conclude that Smith is cleverer than Jones.

But do we want to say this? I think not. Before any decision has been made, it is true neither that Smith is cleverer than Jones, nor that Jones is cleverer than Smith. This intuitive judgement is in agreement with (12), according to which Jones and Smith are incomparable in respect of cleverness. Indeed, it is (12) which, in my opinion, captures the comparative correctly—at least to the extent that it gives a necessary and sufficient condition for *definite* membership in the positive extension of $\geq$ (Q_1^1). That (13) cannot be right becomes even more evident when we realize that it implies that for any objects u_1 and u_2 and adjective A, either u_1 is at least as A as u_2 or u_2 is at least as A as u_1; and this should fail to be true, in general, whenever we have two, largely independent, criteria for applicability of the adjective, but no clear procedure for weighting them.

We saw that for *heavy* (12) and (13) are equivalent (provided p has been correctly specified). The same is true for a number of other adjectives which, like *heavy*, may be called 'one-dimensional'. With each such adjective is associated a unique measurable aspect. The (numerical) value of that aspect for a given object determines whether or not the adjective applies. For *heavy*,

---

for much. On the other hand, models that are compatible with the plausible options O_1 and/or O_2 should make more substantial contributions to the satisfaction sets that contain them, even when they are incompatible with O_3. Thus models in which Smith is in the positive extension of *clever* and Jones isn't will count for more than those in which Smith isn't and Jones is.'

This is what these three sentences were meant to convey. It isn't much—no more than a vague hint at what constraints the probability function of such a vague model for the general meaning of a multi-criterion predicate like *clever* should satisfy. My current inclination is to see the probability function of such a general model for multi-criterion predicates as underdetermined to such an extent that the notion is well-nigh meaningless. It is only in the context of some specific decision to weight the different criteria that the idea of capturing degrees of satisfaction via probability functions can hope to gain intuitive plausibility.

the aspect is weight. Other examples are *tall* (associated with height) and *hot* (associated with temperature).

But such adjectives are rare. Even *large* is not one of them. For what precisely makes an object large? Its height? or its volume? or its surface? or a combination of some of these? Here we encounter the same phenomenon that has already been revealed by our discussion of *clever*. There is no fixed procedure for integrating the various criteria. Often, it is the context of use which indicates how the critcria should be integrated or, alternatively, which of them should be taken as uniquely relevant.

This is one of the various ways in which contexts disambiguate. Formally, contextual disambiguation can be represented as a function from contexts to models less vague than the ground model. While incorporating this idea into the framework already adopted, I will, at the same time, eliminate a feature of vague and graded models which is unrealistic in any case, but would be particularly out of place in the context-dependent models defined below: Thus far I have assumed that all the members of $\mathcal{S}$ are complete models. But this is unnatural if we want to think of these models as the results of semantic decisions that could actually be made. For most decisions will fail to render the relevant predicates completely sharp. They will only make them sharper. (Indeed, we may with Wittgenstein, doubt that we could ever make any concept completely sharp.) It, therefore, appears more natural to posit that the members of $\mathcal{S}$ are partial models. It is possible, moreover, that one of these contextually determined models is less vague than another, viz. when, intuitively speaking, the semantic decision reflected by the second goes in the same direction, but not as far as, that reflected by the first. Thus, $\mathcal{S}$ will be partially ordered by the relation *as vague as*.

I just suggested that a context picks from this set a particular model—which functions, so to speak, as the ground model of the graded model which represents the speech situation determined by that context. The various sharpenings acceptable from the viewpoint of that context would then be represented by those members of $\mathcal{S}$ which are at most as vague as the new ground model. But I am not convinced that this is absolutely correct. For it could, conceivably, be the case that two different contexts specify for a given predicate two different citeria from the set of those which are prima facie plausible, and which, though they happen to determine the same new ground model, will not permit exactly the same further sharpenings. So the context should select a certain subset of $\mathcal{S}$ of contextually admissible further sharpenings. In addition, the context must select a subset of admissible modifications. This set we could not even hope to reconstruct from the new ground model alone.

Thus far, the probability function p was defined over a class of complete models. This would now seem to be impossible, as we no longer require that $\mathcal{S}$ consists of—or even that it contains any—complete models. Yet the intuition behind the function p—which I tried to convey in the example concerning *heavy*—makes it appear unnatural to define p as a function over sets of *partial* models, especially as these sets may now be expected to contain models one of which is vaguer than the other; it is, so to speak, the number of possible *ultimate* results of repeated sharpening that p should measure, and not, the number of intermediate steps that one may take on the way to these ultimate complete models.

A solution to this dilemma can be found if we assume that all individual cases of vagueness can be resolved, though not all at once; and this assumption does appear to be unexceptionable. Thus, we will impose on the set $\mathcal{S}$ the following condition:

(14) if $\langle U, F_1\rangle \in \mathcal{S}$ and $\langle u_1, \ldots, u_n\rangle \in U^n \setminus (F_1^+(\mathcal{Q}_i^n) \cup F_1^-(\mathcal{Q}_i^n))$, then there is a member $\langle U, F_2\rangle$ of $\mathcal{S}$ in which $\mathcal{Q}_i^n$ is less vague than $\langle U, F_1\rangle$ and such that $\langle u_1, \ldots, u_n\rangle \in F_2^+(\mathcal{Q}_i^n) \cup F_2^-(\mathcal{Q}_i^n)$.

Under this assumption we may construct complete models as the unions of maximal chains in $\mathcal{S}$: Let $\mathcal{S}$ be a set of partial models for L, which all have the same universe U. Then, $\mathcal{S}$ is a *chain under the relation 'vaguer than'* if for any two of its members $\langle U, F_1\rangle$, $\langle U, F_2\rangle$ either

(i) for each predicate $\mathcal{Q}_j^n$ of L, $F_1^+(\mathcal{Q}_j^n) \subseteq F_2^+(\mathcal{Q}_j^n)$ and $F_1^-(\mathcal{Q}_j^n) \subseteq F_2^-(\mathcal{Q}_j^n)$ or
(ii) for each pedicare $\mathcal{Q}_j^n$ of L, $F_2^+(\mathcal{Q}_j^n) \subseteq F_1^+(\mathcal{Q}_j^n)$ and $F_2^-(\mathcal{Q}_j^n) \subseteq F_1^-(\mathcal{Q}_j^n)$.

A subset $\mathcal{S}'$ of a set $\mathcal{S}$ of models with universe U is a *maximal chain in* $\mathcal{S}$ if (i) $\mathcal{S}'$ is a chain (under the relation *vaguer than*) and (ii) for any $M' \in \mathcal{S} \setminus \mathcal{S}'$, $\mathcal{S}' \cup \{M\}$ is not a chain. The *union* of chain $\mathcal{S}$ of models with universe U is the model $\langle U, F_\infty\rangle$ where for each $\mathcal{Q}_j^n$, $F_\infty^+(\mathcal{Q}_j^n) = \bigcup_{\langle U,F\rangle \in \mathcal{S}} F^+(\mathcal{Q}_j^n)$ and $F_\infty^-(\mathcal{Q}_j^n) = \bigcup_{\langle U,F\rangle \in \mathcal{S}} F^-(\mathcal{Q}_j^n)$

If U is countable, then (14) entails that

(15) The union of each maximal chain of $\mathcal{S}$ is complete

However, (15) does not follow automatically from (14) when U is uncountable.[e] Since it is property (15) in which we are primarily interested in

[e] While literally true, this sentence is clearly incorrect in spirit. It is incorrect in that the non-denumerable case isn't different from the denumerable one: there is no way of proving (15) from (14) in either case.

connection with the function p, I will make it, rather than (14), one of the defining conditions of graded context-dependent models.

Here is a counterexample. Let $\mathcal{M}$ be a graded context-dependent model $\langle M, \mathcal{S}, \mathcal{C}, \mathcal{F}, p\rangle$ defined below in the main text with $M = \langle U, F\rangle$ and with the following properties: U is the set of natural numbers: the positive and negative extension of $\mathcal{Q}$ in M are both empty; every model in $\mathcal{S}$ is the ground model M_c of some c in Dom($\mathcal{C}$); and in the extensions of the models M_c 0 plays a special role: once 0 has been put into the positive or negative extension of $\mathcal{Q}$, all thus far unsettled members n of U must be put into the same extension (positive or negative) as 0, if and when they get settled subsequently (and as soon as 0 gets settled, there will be positive natural numbers that testify to this through the extension to which they have been added). But so long as 0 hasn't been settled, the other numbers that get settled get settled in a different and also fully specific way—every even number is added to the positive extension of $\mathcal{Q}$ and every odd number to the negative extension. Furthermore, the set of numbers settled by any model M_c is always a segment of the natural number sequence beginning with either 0 or 1.

To repeat these stipulations:

(i) If M is the ground model of $\mathcal{M}$, then $F^+_M(\mathcal{Q}) = F^-_M(\mathcal{Q}) = \emptyset$.
(ii) For any model M' in $\mathcal{S}$, $F^+_{M'}(\mathcal{Q}) \cup F^-_{M'}(\mathcal{Q})$ is a segment of the natural numbers which either begins with 0 or with 1.
(iii) When the segment is of the form $(1, \ldots, n)$, then the even numbers in it will be in $F^+_{M'}(\mathcal{Q})$ and the odd numbers in it will be in $F^-_{M'}(\mathcal{Q})$.
(iv) When the segment begins with 0, then it is of the form $(0, \ldots, m, m+1, \ldots, n)$, where $m \geq 0$, the numbers between 1 and m are assigned to the positive and negative extension as above (even numbers to the positive and odd numbers to the negative extension) and the numbers from $m+1$ to n are all in the same extension as 0. (So for each segment starting with 1 there is just one model M' that settles all and only the numbers in that segment, whereas for the segments starting with 0 there are in general many.)

Consider the chain of models M_n (for $n = 1, 2, \ldots$) which settle the segments $(1, \ldots, n)$. This chain is maximal and its union fails to settle 0, and so isn't complete. The second claim is self-evident, and the reason why the chain is maximal should be clear as well. There is no way of adding further models from $\mathcal{S}$ to this chain without its ceasing to be a chain. This is so because any M' in $\mathcal{S}$ that is not yet in the chain settles a segment beginning with 0 and for each such M' there is at least one pair consisting of an even and an odd number that are either both in the positive extension of $\mathcal{Q}$ or else both in the negative extension; and that means that M' neither includes nor is included in any model from the chain in which this pair of numbers is settled.

The notion of maximality defined in the paper is quite weak in the sense that it is not too hard for a chain to be maximal in this sense, and so it shouldn't be all that surprising that there are such chains whose unions are not complete. A different, stronger notion of maximality, maximality$'$, which one might hope would give the entailment of (15) by (14) a better run for its money, is one that focuses on the chains' unions: *Ch* is *maximal*$'$ iff there is no other chain *Ch*$'$ of models in $\mathcal{S}$ such that the union M_{Ch} of *Ch* is properly included in the union $M_{Ch'}$ of *Ch*$'$. It is not difficult to see that it is possible for a chain to be maximal in the first sense without being maximal in the second and to construct examples to show this; but I will not do this here.

As far as the relation between (14) and (15) is concerned, the new maximality notion

A *graded context-dependent model for L* is a quintuple $\langle M, \mathcal{S}, \mathcal{C}, \mathcal{F}, p\rangle$ where:

(i) M is a partial model;
(ii) $\mathcal{S}$ is a set of partial models with the same universe as M;
(iii) The union of each maximal chain of $\mathcal{S}$ is complete;
(iv) $\mathcal{C}$ is a function the range of which consists of pairs $\langle M', \mathcal{S}'\rangle$ where (a) $M' \in \mathcal{S}$; (b) $\mathcal{S}' \subseteq \mathcal{S}$ and (c) the union of each maximal chain of $\mathcal{S}'$ is complete;
(v) $\mathcal{F}$ is a field over the set $\bar{\mathcal{S}}$ of unions of maximal chains of $\mathcal{S}$;
(vi) (a) for each formula ϕ and assignment a, the set of members M' of $\bar{\mathcal{S}}$ such that $[\phi]_{M',a} = 1$ belongs to $\mathcal{F}$; (b) $\{M' \in \bar{\mathcal{S}} : M \subseteq M'\} \in \mathcal{F}$; (c) for each $\langle M', \mathcal{S}'\rangle$ in the range of $\mathcal{C}$, if $\bar{\mathcal{S}}'$ is the set of unions of maximal chains of $\mathcal{S}'$, then $\{M'' \in \bar{\mathcal{S}}' : M' \subseteq M''\} \in \mathcal{F}$;
(vii) p is a probability function over $\mathcal{F}$.

We will refer to context-dependent graded models by means of the abbreviation cgm. Henceforth, $\mathcal{M}$ will always be a cgm and will always be equal to $\langle M, \mathcal{S}, \mathcal{C}, \mathcal{F}, p\rangle$; M will be called the *ground model* of $\mathcal{M}$; similarly if $\mathcal{C}(c) = \langle M'_c, \mathcal{S}'_c\rangle$ then M'_c, is called the *ground model* (in $\mathcal{M}$) *with respect to c*.

---

doesn't fair any better than the first one. Our example shows this too, for the chain described there is maximal' as well as maximal. Moreover, the notion of 'maximal chain' that I had in mind when I wrote the paper is captured better by maximality as defined in the paper than by maximality'. The conceptual motivation for this entire section of the paper was that the contextual changes that conversation participants are sometimes able to impose on the context, or that the conversational context imposes on them, can lead stepwise and gradually to the closing of the truth value gaps of vague predicates, even if no single contextual move or change can bring this about in one go. This suggests a notion of chain that is in the first instance a chain of *contexts* c_1, c_2,.. where each new context settles cases that had not been settled so far. But it should be an 'increasing' chain in that the ground models M_{c_i} of the pairs $\mathcal{C}(c_i) = \langle M_{c_i}, \mathcal{S}_{c_i}\rangle$ are increasing in the sense of set-theoretic inclusion.

Whenever 'the Range of $\mathcal{C}$ is all of $\mathcal{S}$' (more precisely: when for each $M' \in \mathcal{S}$ there is a $c \in$ Dom($\mathcal{C}$) and an $\mathcal{S}' \subseteq \mathcal{S}$ such that $\mathcal{C}(c) = \langle M', \mathcal{S}'\rangle$), there will be for every maximal chain M_1, M_2,.. in the sense of the paper one or more chains c_1, c_2,.. of contexts that correspond to it in the sense for some $\mathcal{S}_i\langle M_i, \mathcal{S}_i\rangle = \mathcal{C}(c_i)$. In that case maximality can also be defined as maximality of context chains. However, the counterexample above has been set up in such a way as to show that (15) doesn't hold for this notion of maximality either.

All this supports the suggestion of the paper that (15) should be adopted as an independent principle, rather than trying to derive it from something like (14). But do we really want (15), whether derived or not? As becomes more directly visible in *The Paradox of the Heap* (the successor to this paper, in spirit as well as in actual chronology), insisting on (15) is tantamount to adopting models that describe complete sorites sequences, which run from an unequivocally true proposition 'u_1 is a $\mathcal{Q}$' to an unequivocally false proposition 'u_N is a $\mathcal{Q}$' via incrementations from u_i to u_{i+1} that are too small to make a difference to whether the element belongs to the extension of $\mathcal{Q}$. In certain recent treatments of the context-dependence of vagueness such models are excluded as part of developing a model-theoretically grounded logic of vagueness in which it is not possible to get from 'u_1 is a $\mathcal{Q}$' to 'u_N is a $\mathcal{Q}$' via steps that are licensed by the logic. (See in particular Shapiro 2006).

Again we denote the set of members of $\bar{\mathcal{S}}$ in which ϕ is true under a as $[\phi]_{\mathcal{M},a}$. Similarly, if $\mathcal{C}(c) = \langle M_c, \mathcal{S}_c\rangle$, $[\phi]_{\mathcal{M},c,a}$ is the set of members of $\bar{\mathcal{S}}_c$ in which ϕ is true under a.

The domain of $\mathcal{C}$ should be thought of as the set of contexts. Contexts may be more or less specific; that is, the domain of $\mathcal{C}$ may contain elements c and c' such that $M_c \subseteq M_{c'}$ and $\mathcal{S}_c \subseteq \mathcal{S}_{c'}$; in this case, c' will be at least as specific as c. Thus the members of the domain of $\mathcal{C}$ are partially ordered by the relation $\leq$, defined by: $c \leq c'$ iff $M_c \subseteq M_{c'}$ and $\mathcal{S}_c \subseteq \mathcal{S}_{c'}$. One may wonder if for every member M' of $\mathcal{S}$ there should be a c such that M' is the ground model with respect to c. This would mean that for any possible sharpening of a predicate there is a context which indicates that the predicate should be understood in precisely *that* sharper way. I have no argument to show that this assumption is false; yet, I see no gain from it; thus, I prefer not to make it.

In a cgm, it is possible that while the relation $\geq (\mathcal{Q}^1)$ does not hold in the ground model, it does hold in the ground models of certain contexts. Thus, assume $\mathcal{Q}^1_1$ represents the adjective *clever*; further assume that c_1 represents a context in which *clever* must be understood as '*good at solving problems*'; that c_2 represents a context in which *clever* must be understood as '*quick-witted*'; and that c_3 represents a context in which both quick-wittedness and the ability to solve problems are to be regarded as constitutive of cleverness. Then we may expect that if $a_1(\nu_1) =$ Smith and $a_2(\nu_1) =$ Jones,

(a) $[\mathcal{Q}^1_1(\nu_1)]_{\mathcal{M},c_1,a_2} \subseteq [\mathcal{Q}^1_1(\nu_1)]_{\mathcal{M},c_1,a_1}$; and[f]
(b) $[\mathcal{Q}^1_1(\nu_1)]_{\mathcal{M},c_2,a_1} \subseteq [\mathcal{Q}^1_1(\nu_2)]_{\mathcal{M},c_2,a_2}$

while nothing definite can be said about the relation between $[\mathcal{Q}_1(\nu_1)]_{\mathcal{M},c_3,a}$ and $[\mathcal{Q}^1_1(\nu_1)]_{\mathcal{M},c_3,a_2}$ until more is known about whether, and in what way, c_3 determines how the two criteria for *clever* are to be weighted. In order that (a) and (b) formally guarantee that in c_1 Smith is cleverer than Jones, while in c_2 Jones is cleverer than Smith, we must specify, parallel to (13), that:

(16) if $\mathcal{C}(c) = \langle\langle U, F_c\rangle, \mathcal{S}_c\rangle$ and $u_1, u_2 \in U$, then $\langle u_1, u_2\rangle \in F^+_c(\geq (\mathcal{Q}^1_i))$ if and only if $[\mathcal{Q}^1_i(\nu_1)]_{\mathcal{M},c,a_2} \subseteq [\mathcal{Q}^1_i(\nu_1)]_{\mathcal{M},c,a_1}$

Since not every member of $\mathcal{S}$ is necessarily the ground model with respect to some context, (16) may not define the positive extension of $\geq (\mathcal{Q}^1_i)$ for some of these models. This is of little practical importance. If we insist

[f] For the meaning of $[\mathcal{Q}^1_1(\nu_1)]_{\mathcal{M},c,a}$ see the explanation in (12).

on defining the extensions in these models as well, we may stipulate that for any such model $\langle U, F\rangle$, $\langle u_1, u_2\rangle \in F_1^+(\geq (Q_i^1))$ if and only if for some c, $\mathcal{C}(c) = \langle M_c, \mathcal{S}_c\rangle$, $\langle U, F_1\rangle \in \mathcal{S}_c, M_c \subseteq \langle U, F_1\rangle$ and $\langle u_1, u_2\rangle \in F_c^+(\geq (Q_i^1))$

What is the negative extension of $\geq (Q_i^1)$? It should consist in the first place of those pairs $\langle u_1, u_2\rangle$ of which it is definitely true that u_2 is more Q_i^1 than u_1, i.e. in view of (12), those pairs for which

(17) $[Q_i^1(\nu_1)]_{\mathcal{M},a_1} \not\subseteq [Q_i^1(\nu_1)]_{\mathcal{M},a_2}$.

One might question this condition on the ground that it makes u_2 *is more* Q_i^1 *than* u_1 definitely true also when the difference between u_1 and u_2 is only marginal. But I do not believe that the objection is well-founded. However marginal the difference, if it is a difference in an aspect which is irrevocably bound to the predicate, so that no context can break this tie, then the relation definitely obtains irrespective of whether it is difficult, or even physically impossible, to observe this.

This leaves us with those pairs $\langle u_1, u_2\rangle$ such that neither $[Q_i^1(\nu_1)]_{\mathcal{M},a_1} \subseteq [Q_i^1(\nu_1)]_{\mathcal{M},a_2}$ nor $[Q_i^1(\nu_1)]_{\mathcal{M},a_2} \subseteq [Q_i^1(\nu_1)]_{\mathcal{M},a_1}$. Which of these should go into $F^-(\geq (Q_i^1))$? I think none. As long as there are some acceptable ways of sharpening Q_i^1 which render u_1 at least as Q_i^1 as u_2, the falsehood of u_1 *is at least as* Q_i^1 *as* u_2 cannot be definite.

I introduced the probability function to show how the notion of 'degrees of truth' can be made coherent. But, so far, the function has served no good purpose. In particular, it has proved useless for the characterization of the comparative: once more it turned out to be necessary to define the operation on the sets themselves, rather than on the numerical values to which p reduces them.

However, there are expressions the analysis of which does seem to require the function p. Consider *rather*. *Rather* forms adjectives out of adjectives, e.g. *rather tall* out of *tall*, *rather clever* out of *clever*, etc. When is a person rather clever? Before I can discuss the really important aspects of this question I should first settle a minor point. *x is rather clever* sometimes seems to deny that *x* is clever, while on other occasions it appears to be entailed by the fact that *x* is clever—just as e.g. *most x are F* sometimes seems to entail *not all x are F*, while on other occasions it seems to be entailed by *all x are F*. I think that both cases, as well as a great many similar ones, ask for an explanation involving Grice's theory of implicature: *most x are F* is a consequence of *all x are F*; but when uttered by a speaker whom the hearer assumes to know whether all *x* are *F*, it will convey that not all *x* are *F*—for if all *x* were *F*, why would not the speaker have said so? Similarly, *rather clever* is weaker than *clever*. But one would use the longer phrase only if one had doubts that

the shorter one applies. Thus, *x is rather clever* is weaker than *x is clever*. *x* is rather clever if a certain lowering of the standards for cleverness would make *x* clever, i.e. if the proportion of members of $\bar{\mathcal{S}}$ in which *x* belongs to the extension of *clever* is large enough. Indeed, the closer *x* is to being truly clever, the smaller is the modification of the standards that is required, and thus the larger will be the class of those models where *x* is in the extension.

It should be noted that just as *x* may pass the test of cleverness for different reasons, so he may also pass that of being rather clever in a variety of ways. Thus, it is possible that x, y and z are all rather clever (though not unambiguously clever); *x*, because he is remarkably quick-witted, while hopeless at mathematical problems; *y*, because he is good at such problems, though slow in conversation; and *z*, because he has both capacities to a moderate degree. For any two of x, y, z, there will be certain modifications of standards which will warrant membership in the extension of *clever* for one, but not for the other. Thus, it will be true of the set of members of $\bar{\mathcal{S}}$ in which, say, *x* is in the extension of *clever* and the set of those members of $\bar{\mathcal{S}}$ where the extension contains, say, *y*, that neither will include the other. Yet, they both guarantee membership in the extension of *rather clever*, essentially because they are both large enough. It is this intuition concerning the largeness of sets which *p* tries to capture.

Thus, if *clever* is again represented by $\mathcal{Q}^1_1$, then we may put:

> u_1 is rather clever if and only if $p([\mathcal{Q}^1_1(\nu_1)]_{\mathcal{M},a_1}) \geq p_0$ (where $a_1(\nu_1) = u_1$ and p_0 is some number in $(0,1)$).

Obviously, p_0 should be less than $p(\{M' \in \bar{\mathcal{S}} : M \subseteq M'\})$; but not much more can be said about it.[g] For, of course, p_0 is not fixed. If that were so, *rather clever* would be a sharp predicate, which, evidently, it is not.

---

[g] Lest this observation be puzzling, recall that in a cgm the set $\mathcal{S}$ may contain models that are not extensions of the ground model. Therefore, the set of those members of $\bar{\mathcal{S}}$ that do extend the ground model is in general a proper subset of $\bar{\mathcal{S}}$ and its *p*-value $p_M = p(\{M' \in \bar{\mathcal{S}} : M \subseteq M'\})$ will be in general less than 1. In fact, this *p*-value can now be seen as a measure of what it is to be true in the ground model *M*: To be true in *M* entails only that your *p*-value be $\geq p_M$. This applies in particular to propositions to the effect that an element *u* of *U* belongs to the positive or negative extension of $\mathcal{Q}$. For instance, the requirement that the proposition $\mathcal{Q}(\underline{u})$ be true (where $\underline{u}$ is a name for *u*) is that $p(\{M' \in \bar{\mathcal{S}} : [\mathcal{Q}(\underline{u})]_{M'} = 1\}) \geq p_M$. Speaking intuitively, when *u* only just meets the actually obtaining standards for being a $\mathcal{Q}$, then the *p*-value of the proposition that it is a $\mathcal{Q}$ will be equal to p_M; if *u* satisfies higher standards, then it will also be in the positive extension of $\mathcal{Q}$ in models from $\bar{\mathcal{S}}$ that result from imposing these higher standards and the proposition's *p*-value can be expected to be greater than p_M; and if *u* only meets weaker standards, then the *p*-value can be expected to be less than p_M so that *u* will not be in the positive extension of $\mathcal{Q}$ in *M*. But in this last case it is still possible for *u* to be in the positive extension of *rather*$(\mathcal{Q})$ in *M*, so long as the *p*-value is not too low.

The vagueness of *rather* could also be represented in the following way. We associate with each c in the domain of $\mathcal{C}$ a pair of real numbers r_c^-, r_c^+ between 0 and 1, such that whenever $c \leq c'$, then $r_c^- \leq r_{c'}^- < r_{c'}^+ \leq r_c^+$. The positive and negative extensions of *rather* $\mathcal{Q}_1^1$ in the ground model M_c w.r.t. c are then defined as the sets:

$$\{u \in U : p([\mathcal{Q}_i^1(\nu_1)]_{\mathcal{M},c,a}) > r_c^+ . p(\{M' \in \bar{\mathcal{S}} : M_c \subseteq M'\})\} \text{ and}$$
$$\{u \in U : p([\mathcal{Q}_i^1(\nu_1)]_{\mathcal{M},c,a}) < r_c^- . p(\{M' \in \bar{\mathcal{S}} : M_c \subseteq M'\})\},$$

respectively; finally the intermediate value of *u is rather* $\mathcal{Q}_i^1$ in the ground model is given by

$$p(\{M' \in \bar{\mathcal{S}} : u \text{ belongs to the positive extension of } \textit{rather}\ \mathcal{Q}_i^1 \text{ in } M'\}).$$[h]

---

[h] Galit Sassoon, who was asked for a critical appraisal of the latex version of the paper that had been prepared for its inclusion in the present volume, wondered whether this specification of the 'intermediate truth value' of 'u is rather $\mathcal{Q}_1^1$' could really be what I intended. Not only is this query fully justified, it also reveals an incoherence in the formulation of the clauses immediately above this footnote for the positive and negative extension of *rather*$(\mathcal{Q})$ in the ground models M_c for the different contexts c in Dom$(\mathcal{Q})$. In the 1975 version of the paper these clauses have '$\mathcal{S}_c$' where the clauses in the present version have '$\bar{\mathcal{S}}$'. The older version makes no sense, as p is only defined for subsets of $\bar{\mathcal{S}}$ and not for subsets of $\mathcal{S}$. But I seriously doubt that this was just a typing or type setting error.

Even after the just mentioned correction, the paper's proposal for assessing the intermediate truth values of propositions of the form 'u is rather $\mathcal{Q}_1^1$' is not easy to make sense of. For what is it for a proposition of this form to be true in a member of $\bar{\mathcal{S}}$? The paper says nothing about this question and the answer is not self-evident. Here is a possible answer: Suppose that each M' in $\bar{\mathcal{S}}$ can be reached as limit (that is: as union model) of a chain starting with some model of the form M_c. Let us further suppose that M' inherits from each such M_c the positive and negative extension of *rather*$(\mathcal{Q}_1^1)$ in the sense that the extensions in M_c are included within those in M'. And, finally, let us assume that when M' can be reached as limit from two different ground models M_c and $M_{c'}$, then the extensions of *rather*$(\mathcal{Q}_1^1)$ in M_c and $M_{c'}$ are compatible in that there is no overlap between the positive (negative) extension in M_c and the negative (positive) extension of $M_{c'}$. If these conditions are satisfied, then we can define the positive extension of *rather*$(\mathcal{Q}_1^1)$ in M' as the union of the positive extensions of *rather*$(\mathcal{Q}_1^1)$ in all models M_c from which M' can be reached as limit and likewise for the negative extension of *rather*$(\mathcal{Q}_1^1)$. If we assume in addition that for each u the set of models M' in $\bar{\mathcal{S}}$ in which the proposition 'u is rather $\mathcal{Q}_1^1$' is true is a member of $\mathcal{F}$, then the proposal for assessing the intermediate value of 'u is rather $\mathcal{Q}_1^1$' in the ground model M has been given a well-defined meaning.

But now that we have made sure that the proposal is well-defined, should we adopt it? Sassoon, left helpless by the text with which she had been asked to deal, suggested that one might assess intermediate truth values for propositions of the form 'u is rather $\mathcal{Q}_1^1$' in much the same way that the two clauses which appear in the text immediately above this footnote assess their truth and falsity: if $p(\{M' \in \bar{\mathcal{S}} : u \text{ belongs to the positive extension of } \textit{rather}\ \mathcal{Q}_i^1 \text{ in } M'\})$ is somewhere between $r_c^- . p_{M_c}$ and $r_c^+ . p_{M_c}$, then that p-value should be translatable into an intermediate truth value in M_c for the proposition 'u is rather $\mathcal{Q}_1^1$'. (For instance, a simple linear translation for the cases in question could be that which gives as the intermediate truth

There are a number of words which, like *rather*, form adjectives out of adjectives and which can be analysed along similar lines. Another prominent example is *very*. The extension of *very* $\mathcal{Q}_i^1$ is again a function of $[\mathcal{Q}_i^1(\nu_1)]_{\mathcal{M},a}$. The limit which $[\mathcal{Q}_i^1(\nu_1)]_{\mathcal{M},a}$ must exceed in order that $a(\nu_1)$ belong to the extension of *very* $\mathcal{Q}_i^1$, must be larger, and not smaller, than $p(\{M' \in \mathcal{S}; M \subseteq M'\})$.

6.

For traditional logic, adjectives, nouns and intransitive verbs are all of a kind—viz. one-place predicates. My second theory of adjectives tries to vindicate this view against the one expressed earlier, which puts adjectives into a different category than verbs and nouns. Yet, it is an undeniable fact about ordinary English that while the comparative is in general a natural operation on adjectives, similar operations on nouns are of relatively little importance, and on verbs they are virtually non-existent. This suggests a difference between adjectives on the one hand, and verbs and nouns on the other hand. I will leave verbs out of consideration in the following discussion, as they present problems quite different from those with which this paper is concerned. But I will try to say something about the difference between adjectives and nouns. Why is it that comparisons involving nouns are, in general, so much more dubious than those which involve adjectives? *This is more a table than that* sounds awkward and is perhaps never unequivocally true, except in the cases in which it is evident that this is a table and that is not (but then we can say precisely this, and thus do not need the first phrase). Yet, it appears that nouns too are vague, some of them just as

---

value $\frac{p_u - r_c^-}{(r_c^+ - r_c^-)\cdot p_{\mathcal{S}_c}}$, where '$p_u$' is short for '$p(\{M' \in \bar{\mathcal{S}} : u$ belongs to the positive extension of *rather* $\mathcal{Q}_i^1$ in $M'\})$'.)

The proposal of the paper for determining the intermediate truth values of propositions involving 'rather $\mathcal{Q}_i^1$' tries to maintain the general strategy of defining intermediate truth values of propositions in ground models in terms of p-values of the sets of maximal extensions of those models in which those propositions are definitely true. But once p-values are brought into the truth definition, as the clauses immediately above this footnote do, then it is no longer clear what could be the conceptual benefits of this strategy. So arguably Sassoon's suggestion is just as good, or better.

To conclude: This part of the paper, which is concerned with the semantics of hedges like *rather*, was seriously under-researched and careless. Even at this point in time (March 2013) I am not sure if modelling the semantics of hedges (or any other expressions) along the lines this part of the paper is attempting has any real merit. But if it does, then much more work will be needed.

vague as certain adjectives. Why doesn't their vagueness allow for equally meaningful comparatives? To discover the reasons, it is advantageous to reconsider 'one-dimensional' adjectives.

For any such adjective Q_i^1 it will be the case that, for arbitrary a_1, a_2,

(18) either $[Q_i^1(v_1)]_{\mathcal{M},a_1} \subseteq [Q_i^1(v_1)]_{\mathcal{M},a_2}$ or $[Q_i^1(v_1)]_{\mathcal{M},a_2} \subseteq [Q_i^1(v_1)]_{\mathcal{M},a_1}$;

and this ensures that u_1 *is more* Q_i^1 *than* u_2 always has a definite truth value.

We have already seen that most adjectives do not satisfy (18) unambiguously. u_1 *is cleverer than* u_2 could remain without truth value in the ground model. Yet there should still be a fair proportion of pairs $\langle u_1, u_2 \rangle$ where u_1 and u_2 both lie in the extension gap of *clever*, but for which (18) holds (with $a_1(v_1) = u_1$ and $a_2(v_1) = u_2$). And, for the same reason, there are many contexts c in which (18) holds (with $[Q_i^1(v_1)]_{\mathcal{M},c,a_1}$ for $[Q_i^1(v_1)]_{\mathcal{M},a_1}$, etc.), so that, in c, each comparative sentence involving Q_i^1 is either definitely true or definitely false. On the other hand, if Q_i^1 is a noun, then (18) will, in general, be satisfied for very few pairs of objects which both fall in the extension gap of Q_i^1.

A very rough explanation for this formal distinction is the following: In order for an object to satisfy a noun it must, in general, satisfy all, or a large portion, of a cluster of criteria. None of these we can promote to be the sole criterion without distorting the noun's meaning beyond recognition. We cannot, therefore, compare the degrees to which two different objects satisfy the noun in terms of their degrees of satisfaction of just one of its many criteria; and, in order to compare them by comparing their ratings with respect to a variety of these criteria, we need a method for integrating these various ratings. And such a method is in general not part of the meaning of the noun.

There is another aspect to the difference between nouns and adjectives which is related to the one discussed above, but perhaps even more important in connection with the former's resistance to comparatives. Nouns, though potentially just as vague as adjectives, tend in actual practice to behave much more like sharp predicates. Take *cat*. In principle, there could be all sorts of borderline cases for this predicate—but, in actual fact, there are very few at best. The same is true, be it in slightly varying degrees, of *table*, *rock* or *word*. Thus, nouns often have very small, or no, extension gaps in the actual world—even if it is easy to think of possible worlds in which these gaps would be enormous.[i] This gives an additional explanation of why

[i] There is no display of any awareness here that natural kind terms like *cat* behave very differently from nouns for artefacts like *table* or *word*, and this in spite of the fact that the

comparatives involving nouns should be of relatively little use. For they are particularly important in those cases where neither of the two objects compared belongs unambiguously to the positive or to the negative extension of the predicate in question. And these cases will seldom arise when the predicate is a noun.

It is an interesting question how nouns 'manage' to be as sharp as they are. The explanation must be more or less along the following lines: Even if each of the several criteria for the noun may apply to actual objects in varying degrees, these criteria tend to be, with respect to the actual world, *parallel*: an object which fails to satisfy a few of them to a reasonable degree will generally fail to score well with regard to almost all of them. Consequently, it will either be recognized as definitely inside the extension of the noun, or else as definitely outside. The nature of this parallelism is very much that of a physical law—it is a feature of our world, and thus in essence empirical. This is one of the ways in which the actual structure of the world shapes the contextual frame with which we operate, and one of the reasons why it is difficult to separate the empirical from the purely conceptual.

Where the simple comparative of a predicate is non-sensical, addition of certain special expressions can restore its meaningfulness. Examples, of such expressions are: *in a sense, as far as function is concerned, with regard to shape*.[10]

Let us consider this last phrase. How should we analyse (19)?

(19) with regard to shape, u_1 is more a table than u_2.

First, we should determine the logical type of the expression *with regard to shape*. This is really a problem which does not belong in this paper. I will therefore give an answer which is convenient in connection with the issues which concern us here and does not distort them. I will treat *is more ... than ... with regard to shape* as an atomic, i.e. not further analysable, expression which stands for a new comparative operation—one which again forms

[10] An extensive discussion of such expressions can be found in Lakoff (1972). Lakoff calls such expressions 'hedges', a term I will adopt here too.

discussions of this distinction by Putnam and Kripke must have been known to me at the time when I first presented the paper, and certainly at the time when the version that appeared in print was sent of to the editor. In *The Paradox of the Heap* and even more emphatically in the joint paper *Prototype Theory and Compositionality* with Barbara Partee, there are observations about the relevance of the distinction for the vagueness profiles of natural kind terms and other nouns.

binary relations out of predicates. This new comparative differs from the one considered thus far in the following way: The phrase *with regard to shape* places us, so to speak, in a context where shape is singled out as the only criterion for whatever the property is in respect of which the comparison is made. Let us suppose that there are such contexts—contexts in which those predicates to which shape is at all relevant are evaluated with respect to shape alone. Then (19), should be true in the ground model if and only if u_1 *is more a table than* u_2 is true in each of these contexts.

I should like to make a brief comment, at this point, on the nature of contexts, and the role which, in my opinion, they ought to play in semantic analysis of the sort of which I have tried to give instances in this paper. We could give an alternative, but evidently equivalent, account of (19) by stipulating that the phrase *with regard to shape transforms* the context in which (19) is used *into* one where shape is the only relevant issue. For our account of (19) it does not make much difference which line of explanation we choose. However, I believe that the solution to certain other semantic problems can be found only if we investigate not only the effect of the various aspects of context on the meanings of expressions used in those contexts, but also the mechanisms which *create*, or *modify*, contextual aspects. A proper understanding of these mechanisms seems essential to the analysis of more extended pieces of discourse—such as told or written stories.[11] Given that such understanding must, eventually, be reached in any case, an account of (19) along the lines of the second proposal may well ultimately be the more desirable. It seems, however, too early to pass judgement on this matter.

At any rate it is important to realize that contexts are made up of verbal and nonverbal elements alike. The same contextual aspect may on one occasion be manifested through the setting in which the utterance is made, while on another occasion its presence is signalled by a particular verbal expression. Exclusive preoccupation with shape, for example, can be evident to both speaker and audience, either because they have been discussing shape and nothing but shape all along (think of a session about shape during a conference on industrial design); or because the previous sentence was *But let us now concentrate exclusively on shape*; or because the sentence itself contains the qualifying phrase *with regard to shape*. The three cases differ as

[11] Cf. Isard (1975). Others whom I know to have developed similar ideas are Thomas Ballmer of the Technische Universität, Berlin, and David Lumsden of University Colledge, London.

regards the degree of permanence with which the feature in question is part of the context. In the first case, preoccupation with shape will last throughout the session; and a special verbal effort would be necessary to remove it; in the last, the modification will be in force only during the evaluation of the particular phrase to which *with regard to shape* is attached; the second case is somewhere between the two. Indeed, without further information, it is not possible to say whether the modification is valid just for the present sentence, for everything this particular speaker is going to say right now, or for the remainder of the entire discussion.

Another expression of the sort we have just been discussing is *in a sense*.

What is it to be clever in a sense? That depends on what are the various possible senses of the word *clever*. It will help to consider such related sentences as *Smith is clever in the sense that he is good at solving problems* or *Jones is clever in the sense of being quick-witted*. The expressions following *clever* in these two sentences have, again, the effect of transforming the context, viz. into one where *clever* is given a more specific sense. The truth value of the sentence should, therefore, be the same as it is in any of these contexts. The contexts in question are the same as those created by antecedent specifications like *Let us understand by 'clever': 'good at solving problems'*. Each such specification will single out a set of contexts in which *clever* is understood correspondingly. *x is clever in a sense* is then true if there is such a set of contexts, such that *x is clever* is true in each of its members.

But which are the acceptable specifications of a given noun or adjective? This is a question to which no definite answer can be given; for the notion of an acceptable specification of a given concept is itself subject to just that sort of vagueness with which this paper is concerned. Clearly not every logically possible definition is acceptable; for if this were so, then all statements of the form

(20) *x* is a ... in a sense

would be true. But what is an acceptable specification can if necessary be stretched very far indeed. That is why it is so hard to establish that a particular sentence of the form (20) is false.

I want to conclude this discussion of hedges with a few remarks on the expression *to the extent that*. Let us consider Lakoff's example:

(21) To the extent that Austin is a linguist, he is a good one

Once more, I will leave questions concerning the ultimate logical form of the expression aside. It will be adequate, for our present interests, if we regard

to the extent that as a two-place sentential operator which forms out of two formulas ϕ and ψ the compound formula

(22) to the extent that ϕ, ψ.

The semantic analysis of this connective brings into focus a problem connected with contextual disambiguation which I have so far failed to mention: to what extent does the sharpening of one predicate affect other predicates? Clearly, the decisions concerning two different words cannot in all cases be independent. Sharpening of the noun *leg* will yield sharpening of the adjective *four-legged* as well. Yet there are many pairs of adjectives such that a sharpening or modification of one does not carry with it any perceptible semantic change in the other. This is true, in particular, of *linguist* and *good*. This is important for the following account of (21).

The truth conditions of (22) are essentially these: (22) is true (in its actual context of use c) if ψ is true in all contexts in which ϕ is true and which are as similar to c as is possible, given that they make ϕ true. In the case of (21), these contexts will be contexts in which we have modified the semantics for *linguist* in such a way that Austin is now definitely inside its extension, and have left the semantics otherwise as much the same as the modification of *linguist* permits. In particular *good* would, it seems to me, not be affected seriously by the modification. The truth of the main clause of (21) in such a context is to be understood in the usual manner.

It is interesting to compare (21) with the slightly more complicated

(23) To the extent that Austin and Russell are linguists, Austin is at least as good a linguist as Russell

This sentence will be true in c if in every maximally similar context c' in which *linguist* has been modified in such a way that both Austin and Russell are in its extension, it is true that Austin is at least as good a linguist as Russell. When is it true in c' that Austin is at least as good a linguist as Russell? This will be the case if the pair ⟨Austin, Russell⟩ belongs to the positive extension of $\geq$ (good linguist) with respect to c', i.e. if the set of members of $\mathcal{S}_{c'}$ in which *Austin is a good linguist* is true includes the set of those members in which *Russell is a good linguist* is true. It is important that this account will give us the intuitively correct truth conditions for (23) only if the members of $\mathcal{S}_{c'}$, involve modifications of *linguist*, but not of *good*.

It is clear from this brief discussion that a formal elaboration of such analyses within the framework provided by cgm's, requires a great deal more structure on the set of contexts than I have given.

7·

I have claimed that vagueness is often reduced by context. This doctrine is void, however, unless it is accompanied by a concrete analysis of those contextual factors which contribute to such reduction of vagueness, and of how they succeed in doing so. To provide such an analysis is a difficult task, the completion of which will perhaps forever elude us. Yet, I feel I ought to say something on this topic, more, in fact, than I actually have to offer. But let me mention at least one contextual aspect which plays a central role in almost all cases where adjectives occur in attributive position.

That aspect is the noun to which the adjective is attached. In a great many cases, the noun alone determines, largely or wholly, how the adjective should, in the given context, be understood. Indeed, if we assume that the noun is the *only* factor, we are back with the first theory according to which adjective meanings are functions from noun-phrase meanings to noun-phrase meanings.

But, of course, the noun is not always the only determining factor. *Smith is a remarkable violinist* may be true when said in comment on his afterdinner performance with the hostess at the piano, and false when exclaimed at the end of Smith's recital in the Festival Hall—even if on the second occasion Smith played a bit better than on the first.

It would be desirable to give a general account of how the meaning of the noun determines that of the adjective that combines with it. Here, I will mention just one aspect of this problem. One of the main purposes of the use of an adjective in attributive position is to contribute to the delineation of the class of objects that the complex noun-phrase of which it is part is designed to pick out—or, alternatively, to help determine the particular individual which is the intended referent of the description in which the adjective occurs. In order that the adjective can be of any use at all for these purposes, it should, in the presence of the noun in question, have an extension which, so to speak, cuts the extension of the noun in half—i.e. if we assume for the sake of this argument that both noun and adjective (in the presence of the noun) are sharp, and that $\bar{A}$ is the extension of the adjective, and $\bar{N}$ that of the noun, then both $\bar{N} \setminus \bar{A}$ and $\bar{N} \cap \bar{A}$ should be substantial proportions of $\bar{N}$. Thus, in order to be able to use the adjective profitably in combination with an unlimited number of nouns, we should let the noun determine the criteria and/or standards for the adjective in its presence in such a way that the above condition is in general fulfilled. (The proposal (8), p. 228, obviously meets this requirement.)

The distinctions between nouns and adjectives adumbrated in the previous section are of course far from absolute. *Four-legged*, for example, has virtually no extension gap—which is hardly surprising given the manner in which it is derived from the noun '*leg*'. And, indeed, it yields comparatives as infelicitous as those derived from most nouns. *This is more four-legged than that* would on most occasions sound positively non-sensical. *Blue*, though apparently not derived from a noun, also gives rise to rather strained comparatives. *This is bluer than that* is sometimes a meaningful statement, but would fail to be more often than not. So it seems that *heavy* and *four-legged* are really very far apart and that they will ultimately require analyses that are fundamentally different.

This brings us to a likely objection against the theory I have outlined. Does it not blur fundamental distinctions between different kinds of adjectives? Yes, undoubtedly it does. Still, I feel that what it reveals about adjectives, in general, is important. But this conviction should not bar the way to accounts that deal in detail with small provinces of the wide realm of all those concepts to which it claims to apply. It should be pointed out, in this connection, that the second theory itself can hardly be regarded as comprehending all adjectives. Is *alleged* a predicate, even in the most diluted sense? It seems not. Of course, we can still maintain that in each particular context of use it behaves as a predicate, in so far as the accompanying (or tacitly understood) noun phrase determines to which objects in that context the adjective applies. But this is just a restatement of the first theory in slightly different terms. The original intuition which led to the second theory seems to be inapplicable to *alleged*. The same can be said to be true, to an almost equal degree, of adjectives such as *fake*, *skilful*, or *good*. Where precisely we should draw the boundaries of the class of adjectives to which the second theory applies I do not know. For example, does *skilful* belong to this class? Surely we must always ask 'skilful what?' before we can answer the question whether a certain thing or person is indeed *skilful*; this suggests that the theory is not applicable to the word *skilful*. Yet, there appears to be some plausibility in the view that *having a good deal of skill* does function as a predicate—be it a highly ambiguous one as there are so many different skills. Here, the question whether we face an expression that stands for a function from properties to properties or rather an ambiguous predicate which is disambiguated by accompanying expressions for properties, has perhaps no definite answer. Both views appear to be equally plausible accounts of the same phenomenon. So it may be impossible to determine in a non-arbitrary manner how far the domain of our theory extends. But, then, it probably does not matter whether we can or not. This will certainly

be unimportant once we have a complementary theory which deals specifically with such adjectives as *alleged, fake, skilful* and *good*. It is bad to be left with a semantic phenomenon that is explained by no theory; but it does no harm to have two distinct theories which give equally adequate, albeit different, accounts of those phenomena that fall within the province of both.

8.

To conclude, let me mention some of the questions which I should have liked to discuss, and which I believe can be treated within the framework I have set up.

In the first place, there are intransitive verbs. I have avoided them throughout, even though they too appear to be one-place predicates and to display a good deal of vagueness. In particular, I have failed to give any account of what semantically differentiates verbs from adjectives, or, for that matter, from nouns. My excuse for this is that the proper understanding of these differences involves the consideration of tense, of the time spans during which a predicate is true of an object, and of similar issues, which seem to require for their formal elaboration a framework which incorporates a good deal of tense logic.

Secondly, I have given only the scantest attention to hedges. I think that my framework is basically suitable for their analysis, although more structure on the set of contexts will be needed than I have provided.

Thirdly, I have considered only the simplest kind of comparatives. Examples of comparatives which are considerably more difficult to treat, are

Jones is more intelligent than he is kind
This building is higher than it is long
Smith is much cleverer than Jones and
Smith is more cleverer than Jones than Jones is than Bill

(accepting this as English).

The last two sentences, in particular, present problems of a rather different kind than those I have tackled in this article. Their analysis requires more mathematical structure than has been built into the models here considered. The difference between the formal framework needed there and the one I have presented is essentially that between metric and arbitrary topological spaces. These and other problems I hope to consider in some other paper.

References

Bartsch, R., Vennemann, T., 1972. *Semantic Structures. A Study in the Relation between Semantics and Syntax*. Athenäum, Frankfurt.

van Fraassen, B., 1969. Presupposition, supervaluation and free logic. In: K. Lambert (ed.), *The Logical Way of Doing Things*. Yale University Press, New Haven, CT, 69–91.

Geach, P., 1956. Good and Evil. *Analysis* 17, 32–42.

Isard, S., 1975. Changing the context. In: E. Keenan (ed.), *Formal Semantics for Natural Language*. Cambridge University Press, Cambridge, 287–296.

Kolmogorov, L., 1970. *Foundations of the Theory of Measurement*. Chelsea, New York.

Lakoff, G., 1972. Hedges: A study in meaning criteria and the logic of fuzzy concepts. In: P. Perantean, J. Levi, G. Phares (eds.), *Papers of the 8th Regional Meeting of the Chicago Linguistic Society*. Chicago Linguistic Society, Chicago, IL, 182–228. Reprinted in: *Journal of Philosophical Logic* 2, 1973, 458–508.

Lewis, D.K., 1979. General semantics. In: D. Davidson, G. Harman (eds.), *Semantics of Natural Language*. Reidel, Dordrecht, 169–218.

Montague, R., 1970. English as a formal language. In: B. Visentini, et al. (eds.), *Linguaggi nella Società e nella Tecnica*. Edizioni di Comunitá, Milan, 189–224. Reprinted in: R. Montague, *Formal Philosophy: Selected Papers by Richard Montague*, ed. by R. Thomason. Yale University Press, New Haven, CT, 1974, 188–221.

Parsons, T., 1968. *A Semantics for English*, unpublished manuscript, University of California, Los Angeles, CA.

Rescher, N., 1969. *Many-Valued Logic*. McGraw-Hill Book Company, New York.

Shapiro, S., 2006. *Vagueness in Context*. Oxford University Press, Oxford.

THE PARADOX OF THE HEAP[a,*]

Hans Kamp

1.

(I) (1) One grain of sand cannot make a heap.
(2) If n grains of sand cannot make a heap, then $n + 1$ grains cannot make a heap either; and this is so irrespective of the choice of n.

So:

(3) No matter what n, n grains of sand cannot make a heap.

This argument is paradoxical. For from what appear to be true premises we derive an evidently false conclusion by means of what would seem to be a valid form of inference.

The principle of inference employed, the so-called *principle of mathematical induction*, is an essential tool in mathematics. While important in just about any branch of mathematics it is strictly speaking a principle of number theory, as is clear from the following description of it:

In order to prove that all positive integers possess a given property P it suffices to show that

* As this paper has taken rather a long time to write there are many people who have over a period of several years, helped me with their comments. In fact it has now been so long since a first version of this paper was circulated that I am sure I don't remember everyone of those by whose comments I have profited, or should have, had I only better understood, or heeded their advice. But let this not deter me from mentioning here, with an apology to all those who have erroneously been left out, the colleagues whose assistance I do remember, and gratefully acknowledge. They are Ed Allaire, Bill Hart, Al Martinich, Uwe Mönnich, Mark Platts, Richard Sorabji, Richard Spencer-Smith, Richmond Thomason and David Wiggins.

[a] Editors' note: This article was originally published in U. Mönnich (ed.), *Aspects of Philosophical Logic*, Reidel, Dordrecht, 1981, 225–277. We would like to thank Reidel (now Springer) for its kind permission to reprint the article. We carefully edited the article, corrected some typos, added full bibliographical references and brought some formulas in line with now more common usage. We would like to thank Ewan Klein and David Nicolas for careful proofreading and constructive comments. Hans Kamp provided a second postscript (dated 2013) and some additional footnotes, indicated by Roman letters, which clarify some of the arguments in the original text.

© HANS KAMP, 2013 | DOI:10.1163/9789004252882_012
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

(II) (1′) the number 1 has *P*, and
(2′) for any number *n*, if *n* has *P* then $n + 1$ has *P* too.

Here is a proof—the standard proof!—that the principle of mathematical induction is valid:

Assume that (1′)$P(1)$ and (2′)$(\forall n)(P(n) \rightarrow P(n+1))$ are both true. Now suppose that $(\forall n)P(n)$ is false. Then there is an *n* for which *P* fails. But then there must be a smallest such *n*, n_0 say; since $P(1)$ and not $P(n_0)$, $n_0 \neq 1$. So $n_0 = k+1$ for some positive integer *k*. Since n_0 is the smallest such *n* that not $P(n)$, we have $P(k)$. But then by (2′)$P(k+1)$, i.e. $P(n_0)$, which contradicts the assumption that not $P(n_0)$.

If this argument demonstrates to our satisfaction that the principle of mathematical induction embodies a sound method for showing the universality of the number-theoretic properties considered in mathematics, then why should it fail to establish its validity in connection with such a property of *n* as is expressed by the phrase '*n* grains cannot make a heap'?

An answer which comes to mind immediately is that this last property is *vague* whereas mathematical properties are sharp. Indeed, if *P* is a vague predicate then the justification of induction we just rehearsed seems to break down. For suppose we argue as above. Again we are led to consider the smallest number n_0 such that $\neg P(n_0)$. Again we can conclude that $n_0 = k+1$ for some *k*. But now, since *P* is vague, it is no longer obvious that we can infer from the fact that n_0 is the smallest number for which *P* fails that *P* must be true of *k*. E.g. if by n_0 we were to understand the smallest number for which *P* *definitely* fails, then there would be no reason to expect *k* to unambiguously possess *P*. For the cases in which we are here interested, the vagueness of *P* will show up in the form of a certain 'no-man's-land' situated between the opposing territories of, respectively, those numbers which definitely have *P* on the one side and of those which definitely lack the property on the other. In fact it may not even be meaningful to speak of the 'smallest *n* such that *n* definitely lacks *P*' in the first place. So, for all we know the principle of mathematical induction may be valid for sharply defined properties, and yet fail when applied to properties whose extensions are blurred.

2.

To see whether this is indeed so would involve us in a lengthy analysis. It seems however that we can spare ourselves that effort. For there is a quicker route towards the conclusion which that analysis would establish, which is

that the invalidity of (I) cannot be blamed on the principle of mathematical induction alone.

Besides the defense of the principle which we rehearsed above there is another, almost equally familiar, justification for it, favoured in particular by those mathematicians who are of a constructivist persuasion. It goes as follows.

From the premises $(1')$ and $(2')$ I can for any number k whatever construct a valid proof that $P(\bar{k})$. For (a) I obviously have a 'proof' for the case where $k = 1$: all I have to do is write down premise $(1')$; (b) let k be any number and suppose I can construct a proof of $P(\bar{k})$ from $(1')$ and $(2')$; then I can also construct a proof of $P(\overline{k+1})$. All I have to do is to take the proof of $P(\bar{k})$, then obtain from $(2')$ the special instance

$$P(\bar{k}) \rightarrow P(\overline{k+1})$$

and finally infer $P(\overline{k+1})$ by means of Modus Ponens.

"But wait" you will say, "This is begging the question. You have given me a proof of $P(1)$ and you have shown me how to convert a proof of $P(\bar{k})$ into a proof of $P(\overline{k+1})$. But in order to infer from this that there is a proof of $P(\bar{n})$ for all n, it is precisely the principle of induction one has to appeal to".

This objection is correct: as I have stated it, the justification *is* circular. And yet, when we apply this line of argument to the cases that involve vague predicates we find that something has been gained. For whether or not P itself is vague, the predicate $Q(n)$ expressed by the phrase 'there is a proof of $P(\bar{n})$ from $(1')$ and $(2')$' is *not* vague. Given accepted rules of inference the question whether a given sequence of lines constitutes a proof is a perfectly well-defined—and indeed a decidable—question. So the earlier justification of mathematical induction will warrant the inference from $Q(1)$ and

$$(\forall n)(Q(n) \rightarrow Q(n+1))$$

to $(\forall n)Q(n)$. Moreover, from $(1')$, $(2')$ and the statement:

> 'for any n there is a valid proof of $P(\bar{n})$ from $(1')$ and $(2')$'

we can infer

$(3')$ $(\forall n)P(n)$

To be more precise let Q be defined by:

> Q is true of the number n iff the sentence $P(\bar{n})$ is provable from $(1')$ and $(2')$ by means of the inference principles of classical logic.

Then it follows from the obvious properties of formal proofs that

(4) $Q(1)$ and
(5) $(\forall n)(Q(n) \rightarrow Q(n+1))$

As Q is sharp the induction principle allows us to infer

(6) $(\forall n)Q(n)$

From $(1')$, $(2')$ and (6) we can obtain the conclusion

$(3')$ $(\forall n)P(n)$

if we adopt in addition the principle

(7) $(\forall n)Q(n) \rightarrow ((1') \,\&\, (2') \rightarrow P(n))$.

I claim that (7) is valid; or rather, that it must be accepted as valid by anyone who assumes that the rules of classical logic are valid in the context in which they are here employed. For to accept these as valid is precisely to accept a derivation of $P(\bar{n})$ from $(1')$ and $(2')$ which only makes use of these rules as a guarantee that either one of $(1')$ and $(2')$ is false or else that n satisfies P.

This shows that we cannot hope to explain the paradox by putting the blame exclusively on the principle of induction, while maintaining that the rules of the predicate calculus—specifically those to which the above argument appeals—apply to vague predicates no less than to predicates which are exact.

3.

The existing literature contains two proposals for a resolution of the paradox which I shall briefly review before presenting the account which I have come to prefer.

All three theories which we shall consider agree in denying that the second premise of the argument—i.e. (2), or $(2')$—is true. According to the theory we shall consider first the premise is not false but only 'nearly true'. Here 'nearly true' is to be understood in the terms offered by many-valued logic. In fact the entire account constitutes an application of that branch of many-valued logic which has come to be known as 'fuzzy logic'. It would carry us too far to give a detailed exposition of all those elements of fuzzy logic which are relevant to this application. But the central ideas are easily conveyed. Suppose P is a vague predicate, say 'bald', and that there are lots of individuals to whom we hesitate to attribute this predicate. Among those individuals there will nonetheless be some that are nearer to deserving the predicate than others. Suppose a is nearer to being bald than b. Then you might consider that in a sense the proposition $P(\bar{a})$ is nearer to unqualified truth than $P(\bar{b})$. Indeed, if one assumes that between the completely true and the completely false there are many 'intermediate truth values', linearly

ordered by their respective distances from perfect truth, then what I just said about the 'truth value' of $P(\bar{a})$ can be recast as the claim that $P(\bar{a})$ will differ less from the value of a perfect truth than does the 'truth value' of $P(\bar{b})$. It is common practice to give perfect truths the value 1 and perfect falsehoods the value 0. The intermediate truth values will then correspond to numbers in the open interval $(0, 1)$, closeness to 1 signifying closeness to perfect truth. In particular, for the a and b of our example $P(\bar{a})$ would have a higher value than $P(\bar{b})$.

Even if we accept the contention that (I.2) is not perfectly, but only approximately true, it still isn't obvious how from this near-truth and the unqualified truth (I.1) we can pass by means of the familiar inference rules to a totally false conclusion. According to fuzzy logic this is possible because the rule of M(odus) P(onens), which we need to use many times to arrive at a definitely false conclusion $P(\bar{n})$, no longer preserves truth the way it does in 2-valued logic. In 2-valued logic the truth value of the concusion one draws by means of MP is never smaller than the minimum of the truth values of the premises from which it is drawn. In many-valued logic this is no longer so. If the values of φ and $\varphi \rightarrow \psi$ are both, say, $1 - \varepsilon$ (for some small ε) then the value of ψ can be as low as $1 - 2\varepsilon$. So the result of applying MP may have a value lower than that of either of the premises; and even where each individual application of MP leads only to a small decrease, a large number of applications may lead from the almost true to the completely false. Thus we may conclude that the 'proofs' of $P(\bar{k})$ from $(1')$ and $(2')$ aren't correct proofs; indeed they get 'worse' as k increases: the larger k, the less we can say, on the basis of the proof and the given truth values of the premises, about lower bounds for the truth value of the conclusion we have reached.[1]

How plausible is this account? That depends on the acceptability of what we have just said about the conditional: that where φ and $\varphi \rightarrow \psi$ both have the value $1 - \varepsilon$ the value of ψ can be as low as $1 - 2\varepsilon$. In the most common many-valued treatment of the conditional this claim is a consequence of the stipulation that if the truth values of φ and ψ are respectively α and β then the value of $\alpha \rightarrow \beta$ is 1 iff $\alpha \leq \beta$, and $1 - (\alpha - \beta)$ otherwise.

In my opinion this characterization has little to recommend itself. This is not so much because of inferiority to some alternative specification which

[1] Cf. Goguen (1969). Goguen (1969) proposed a somewhat different definition for the truth values of conditionals. But in substance the account presented here applies to that proposal as well.

could be stated within the same mould, but because the many-valued approach towards the analysis of the familiar logical operators is in my view misconceived from the very start. I argued this point at some length in an earlier paper, (Kamp 1975), and will not repeat the argument in the very same form here. Instead I shall first give a brief sketch of the semantics for vague predicates which Kamp (1975) develops in detail. We need a statement of this type of semantics anyway for our discussion of the second version of the paradox. I shall return to the question of multi-valued truthtables after that.

4.

Let *P* be any vague predicate. The vagueness of *P* manifests itself in there being objects of which it cannot be decided whether or not they belong to the extension of *P*. When I use the phrase 'cannot be decided' I do not just mean that *we* lack the practical means of finding the answer to the question 'is *P* true of this object?' I mean that there simply *isn't* any answer. The predicate *P* does not partition the domain of objects into two sets—the set of objects of which it is true and the set of those of which it is false—but into three sets, the set of objects of which *P* is true (we will refer to this set as the *positive extension* of *P*), the set of objects of which *P* is false (the *negative extension* of *P*), and the set of objects *a* for which the question 'is *P* true of *a*?' has no answer (the *truth value gap* of *P*).

Suppose now we wish to study the logic of vague predicates within the context of a simple formalized language—say a language of first order logic whose only non-logical constants are individual constants $c_1, c_2, c_3, \ldots$ and the 1-place predicate *P*, which is to stand for some vague property—and suppose that we investigate the logic of this language along model theoretic lines.[2] In order that our investigation shall reveal the logical features which are connected with the vagueness of *P* this vagueness must be in some way manifest in the semantic representation *P* receives in the models for the language on which our analysis will be based. These models must therefore at the very least draw the tripartite division which I just mentioned. For the moment this is the only aspect of vagueness that I will explicitly represent.

---

[2] I am persuaded, but I will not try to argue for this conviction here, that only a model theoretic analysis can clarify the logical properties of languages that incorporate aspects which are not yet represented in systems whose logical structure we already understand.

Precisely how we do this is not important; but the exposition of what follows requires that we settle for some particular way of doing it. This is the way I have chosen:

Instead of specifying just the extension of *P* (as is done in the models familiar from the standard model theory for the first order predicate calculus) the model provides two sets, the *positive extension* of *P* and the *negative extension* of *P*. These extensions are always mutually exclusive; but they need not jointly exhaust the universe of discourse. What remains if we subtract from the universe of discourse both the positive and the negative extension of *P* is called the *truth value gap of P*.

This implies that for certain sentences $P(c_i)$ the model may leave the truth value undecided; this indeterminateness will in turn affect the complex sentences which contain such sentences $P(c_i)$, or corresponding open formulae, as parts. Precisely what the truth values of these complex sentences are is no easy question; it is from an attempt to give a satisfactory answer to this question that the theory I am about to explain was born.[3]

Vagueness is typically experienced as a flaw. Often when we come upon an object *a* and our criteria for the predicate *P* do not give us a clear verdict as to whether *a* satisfies *P* or not, and when, moreover, it appears that this is not because of a lack of empirical information about *a*, or some other incidental ineptitude on our part at applying the criteria, then we feel a certain pressure to modify the criteria so that they will decide the case which confronts us as well as similar cases we might come upon subsequently. Generally we shall wish to *extend* or *sharpen* the criteria, i.e. to modify them in such a way that the cases decided by the old criteria are still decided in the same way. In this way we make the associated predicate more precise; its truth value gap has been narrowed.

Just as what is determined by the criteria *before* we have modified them is conveyed by a certain model *M*, so we may assume that what is determined by the criteria after modification is represented by some new model M'. It seems unreasonable to suppose that truth value gaps generally are, or even that they could be, completely eliminated through a single modification. However, it will simplify the exposition at this stage if we assume that such modifications are possible and restrict attention to them. Where

[3] See Fine (1975) and Kamp (1975) for details. The essential idea is, however, already to be found in Dummett (1975), which was written much earlier than the date of its appearance suggests.

moreover we consider only a single predicate P the effect of such a 'definitive' modification will be a model which is essentially like a two-valued model for classical predicate logic.

Suppose once again that we are confronted with the object a and that the criteria for P do not decide whether a satisfies P. What are we to say in such a situation about the truth value of the sentence

$$P(\bar{a}) \vee \neg P(\bar{a})?$$

This is not quite such a simple question as it may seem. One answer that might come to mind is "Well, $P(\bar{a})$ is not fully true; nor is it fully false, so $\neg P(\bar{a})$ isn't fully true either. A disjunction is true only if at least one of its disjuncts is true. But in the present case neither disjunct is true. So the disjunction cannot be true either. In fact, the value of the disjunction could not simultaneously exceed those of each of its disjuncts."

This is essentially what the fuzzy logician would say. It expresses a view which is, so far at any rate, coherent and beyond direct refutation. It is not the only possible answer, however. To see this, observe first that in a case where the criteria *do* decide whether a satisfies P, but where *we* are not able to see this, the truth values of $P(\bar{a}) \vee \neg P(\bar{a})$ must be 1. Now it is simply a fact about language use that the situation where there really is no answer almost always looks, from the user's point of view, just like one where there is an answer which the user cannot find. One reason for this is implicit in an important observation of Putnam. Putnam (1973, 1975) has pointed out that we rely in our use of many technical and semi-technical terms on the knowledge of those more expert than ourselves to tell us, whenever a real need arises, what precisely falls under those terms and what does not. Indeed such reliance on the expertise of others is a practical necessity which none of the speakers of a reasonably expressive language such as our own can escape; and it would be setting standards squarely at variance with what we normally consider linguistic competence to insist that only the entirely self-reliant are genuinely competent speakers of their language.

Putnam makes this observation in the context of an argument that our own application criteria—those which he refers to as our 'stereotypes'—do not in general suffice to determine the extensions of the corresponding terms. This is so, according to the realistic doctrines advocated by him and by Kripke, for those terms which purport to refer to natural kinds. The extensions of such terms are, on this theory, determined by the essential characteristics which demarcate the kind that comprises those individuals with reference to which the term has come into its use, from other kinds.

These characteristics, however, may be hidden from the view of most, or even all of us, in which case it may require significant scientific progress to discover to which individuals the term in fact applies.

Even if we accept this form of realism (as in fact I think we should) there remains the possibility that some terms which we take to refer to natural kind terms have truth value gaps. This could arise either because the field of entities to which the term has been used to refer turns out to be more or less evenly distributed over several natural kinds, or because the demarcation criteria do not yield the neat partition into equivalence classes which the theory must presuppose (some such complication seems to arise for instance, in connection with the classification of certain types of finches). It must be concluded from the possibility of such cases that we can never be quite sure that the extensions of natural kind terms are fully determined.

It is nevertheless reasonable, given the enormous success enjoyed by the scientific enterprise, that we assume natural kind terms to be well-defined irrespective of our own ability to decide what falls within their range. The situation is different with regard to terms which do not stand for natural kinds but for, say, artifacts, professions, character traits, activities or any other universals which have no significant place in the fabric of lawlike regularities. Such predicates cannot transcend, through the power of natural law, the limits of what usage can conventionally impose. Of any such predicate it would therefore be quite unwarranted to assume that it has no truth value gap at all. Nonetheless, when we are faced with a question concerning the extension of the predicate which we are unable to settle, we can rarely if ever be sure that someone more knowledgeable would not know the answer, and so we are disposed towards treating even such predicates as if they were sharp. Consequently we are prepared to acknowledge as true, or false, all those sentences which will come out true, or false, no matter where the dividing line between positive and negative extension will exactly fall. $P(\bar{a}) \vee \neg P(\bar{a})$ and $P(\bar{a}) \;\&\; \neg P(\bar{a})$ are two such sentences; the former will be true no matter where the division runs, and the latter always false. Let us for subsequent use state this principle for the determination of the truth values of compound sentences containing vague predicates explicitly. We shall refer to it as 'SV' (for 'Super Valuation'):

> (SV) speakers tend to feel committed to the assumption that the concepts they use are well defined; this implies in particular that they will accept as true any statement which can be shown to be true on the hypothesis that the predicates it contains do not have truth value gaps; and similarly that they accept a statement as false if on this hypothesis it can be shown to be false;

> they will accept these things irrespective of whether the relevant predicates really have gaps, or whether they only appear to have gaps to them.

Formally the commitment postulated in SV comes to this. Besides the model M, which specifies the extension of P determined by the criteria as they are already accepted, there are other models M' which represent the various ways in which the truth value gap could be eliminated and in which therefore P behaves as a sharply defined predicate. A sentence will count as true if it is true in all these models M' and false if it is false in all of them.[4]

5.

We saw that the many-valued account of the paradox depended on the possibility of distinguishing between different degrees to which objects may fall short of belonging to the positive extension of a predicate. The theory we are in the process of presenting now involves this idea too, albeit in a slightly different form. Let us consider another, but equally notorious, version of the paradox, according to which every human male is bald, a conclusion we are invited to draw from the premises that ($1''$) any man with one hair on his head is bald: and ($2''$) for any n if any man with n hairs on his head is bald then so is any man with $n + 1$ hairs on his head. Suppose you are presented with two heads, sparsely crowned with hair, and that you are unable to decide of either whether or not it is bald; even so you may be positive that

---

[4] It would be an overstatement to claim that speakers always behave in accordance with this principle. People do for example say on occasions such things as:

(1) It is neither red nor is it (plainly) not red.

An utterance of this kind does seem to amount to a rejection of the law of the excluded middle. It would seem possible to accommodate the intention behind such utterances as (1) without abandoning the supervaluation approach, e.g. by analyzing (1) as

(2) It is neither definitely red nor definitely not red.

where 'definitely' could be treated as Fine's operator D (see Fine 1975). There is a real problem with the analysis of D however, that is, where it occurs in negative position (e.g. in the scope of a single negation operator while not in the scope of any other logical operators). A similar difficulty arises in connection with the comparative operator introduced in Kamp (1975). This is a difficulty to which I shall address myself in another paper. It is a definite shortcoming of Kamp (1975) that I failed there, not only to deal with the difficulty but also to point out its existence. Another shortcoming of Kamp (1975) is its apparently implicit suggestion that our attitude towards vague predicates—and, specifically, adjectives—is always as described by SV. I did not intend this very strong thesis when I wrote the paper; but looking at it now I can see that it could have produced misunderstanding on this issue.

the first head is balder than the second. So that, *if* somehow you were to reach the conclusion that the second head is bald, then you would take this to imply that the first must be classified as bald too. On the other hand you might be led to conclude that the first head is bald without feeling compelled thereby to regard the second head as bald. This means that a modification of the criteria for 'bald' would simply not be acceptable to you if it rendered the second head bald but not the first. Such a modification would be incompatible with your present understanding of the concept.

Your understanding of the predicate P thus imposes certain restrictions on the ways in which the modified criteria may partition the truth value gaps. In formal terms, it limits the set S of models M' which represent the semantically admissible partitions. This set S therefore embodies, by virtue of excluding certain models which are possible from a strictly formal point of view, important semantic information about the use of P. Let $P(n)$ mean e.g. 'n grains of sand cannot make a heap'. Let us assume the positive extension of P to consist of the numbers $\{1, \ldots, n_0\}$ and the negative extension of the numbers $\{n_1, n_1+1, \ldots\}$. Mathematically speaking, there are all sorts of ways in which the truth value gap $\{n_0+1, \ldots, n_1-1\}$ can be distributed over the positive and negative extension. But only some of these distributions are semantically admissible: only those, to be explicit, which put, for some number n_2 in the gap, all numbers $\{n_0+1, \ldots, n_2\}$ into the positive, and the remainder into the negative extension. S will thus contain only those models M' in which the positive extension of P is some initial segment $\{1, \ldots, n_2\}$ of the positive integers and the negative extension contains all the others.

It is easily seen that if a satisfies the predicate P 'to a higher degree' than b then the subset S_a of S, consisting of those models M' in which a belongs to the positive extension of P, is a proper superset of the set S_b of models M' in which b belongs to the positive extension of P. So we might hope to find the 'degree of satisfying P' in some fashion reflected by the 'size' of the corresponding subset of S. In Kamp (1975) I tried to make this idea explicit by introducing a measure function p on the appropriate field of subsets of S. With the help of such a function p we can, if we wish, define the 'intermediate truth value' of any sentence, simple or complex, as the p-measure of the set of all M' in S in which the sentence is true. It is well worth remarking that if we think of these p-measures as truth values the connections $\&, \vee, \rightarrow$ are *not* truth-functional. For example, suppose $P(\bar{n})$ has the value $1/2$. Then so does $\neg P(\bar{n})$. It follows that $P(\bar{n}) \vee \neg P(\bar{n})$ has the value 1 while $P(\bar{n}) \vee P(\bar{n})$ has the value $1/2$; that $P(\bar{n}) \,\&\, \neg P(\bar{n})$ has the value 0 while $P(\bar{n}) \,\&\, P(\bar{n})$ has the value $1/2$; and that $P(\bar{n}) \rightarrow \neg P(\bar{n})$ has the value $1/2$, while $P(\bar{n}) \rightarrow P(\bar{n})$ has the value 1. These examples show

that the truth values of the components of a compound sentence are not by themselves sufficient to determine the truth value of the compound. Structural relationships between the components are relevant too—and this seems intuitively correct.

It is difficult to give a fully satisfactory motivation for the function p. In general there is very little that favours the choice of one such function over another.[5] Concerning the predicate P, however, our intuitions about degrees of satisfaction are somewhat clearer than they are in most cases. Indeed it seems quite natural to assume that the degree to which a number n satisfies P is 1 when n is one of the numbers $1, \ldots, n_0$; that it is 0 when n is $\geq n_1$; and that for $n = n_0 + 1, \ldots, n_1$, it diminishes stepwise.

For the purpose of discussion I shall also assume, although this does not seem to follow clearly from our intuitions, that these stepwise reductions are all of the same magnitude e. This implies that $e = \frac{1}{n_1 - n_0}$. Let us put in addition, for each n in $\{n_0, \ldots, n_1 - 1\}$,

> $S_n =_{df}$ the set of all those M' in S in which the positive extension of P includes $\{1, \ldots, n\}$.

If it is agreed that the truth value of $P_0(\bar{n})$ should coincide with the degree to which n satisfies P; and if moreover the function p is to yield these truth values in the way indicated, then we must clearly have, for each n in $\{n_0, \ldots, n_1 - 1\}$,

$$p(S_n) = 1 - (e(n - n_0))$$

This implies that for each such n

$$p(S_n - S_{n+1}) = e$$

If we assume in addition that the domain of p is simply the field generated by these sets $(S_n - S_{n+1})$, then p is completely determined by the stipulations we have made.

Let us assume for the sake of argument that the semantics of P is correctly captured by the model M, the set S and the function p as we have just characterized them. Consider any sentence

---

[5] This is a difficulty that the fuzzy logician must of course also face. It is one thing to observe that a satisfies the criteria for P to a higher degree than b, and quite another to assign numerical values to these degrees. Even the assumption that for any objects a and b in the truth value gap either the degree of satisfaction on the part of a is greater than that of b, or vice versa, or else that the degrees must be the same, does not always seem warranted. That is one of the reasons why IQ tests are so often misleading even in regard of the *order* which they assign to those tested.

$P(\bar{k}) \rightarrow P(\overline{k+1})$

If $k < n_0$ then $P(\overline{k+1})$ is true in all members of S; so $P(\bar{k}) \rightarrow P(\overline{k+1})$ is also true in all members of S and so has p-value 1. If $k \geq n_1$ then $P(\bar{k})$ is false in all members of S and so $P(\bar{k}) \rightarrow P(\overline{k+1})$ is true in all members of S, and so again has p-value 1. In the remaining, and most important, case, where $n_0 \leq k < n_1$, $P(\bar{k})$ is true in the members of S_k while $P(\overline{k+1})$ is true in the members of S_{k+1}. So the only M' in which $P(\bar{k})$ is true while $P(\overline{k+1})$ is false are the members of $(S_k - S_{k+1})$. In all other models $P(\bar{k}) \rightarrow P(\overline{k+1})$ is true. So the p-value of the sentence is $1 - e$.

Now consider the sentence

$(\forall n)(P(n) \rightarrow P(n+1))$

Let M' be any member of S. Suppose the positive extension of P in M' is the set $\{1, \ldots, n_2\}$, for some n_2 in $\{n_0, \ldots, n_1 - 1\}$. Then $P(\bar{n}_2) \rightarrow P(\overline{n_2+1})$ is false in M'. So $(\forall n)(P(n) \rightarrow P(n+1))$ is false in M'. As this holds for arbitrary M' in S, the sentence is false in all members of S, and so has p-value 0.

This solves the paradox of the heap: the second premise of the argument is perfectly false. No wonder then that we can derive from it the plainly false conclusion $(\forall n)P(n)$. Moreover, the solution explains why we could have been misled into supposing that $(\forall n)(P(n) \rightarrow P(n+1))$ was true, or nearly true: although it is correct to say that for any k the sentence $P(\bar{k}) \rightarrow P(\overline{k+1})$ is either fully true or nearly true, it is something else, and indeed wrong, to say that the sentence $(\forall n)(P(n) \rightarrow P(n+1))$ is nearly true.

This solution to the paradox, found independently by Fine (1975) and by Richmond Thomason (1973), struck me when I first learned of it as quite satisfactory. In particular it appears superior to the many-valued account to which I must now devote a few more words.

The many-valued theory would attribute the same intermediate values to the sentences $P(\bar{k}) \rightarrow P(\overline{k+1})$. But it diverges from the above account in attributing to $(\forall n)(P(n) \rightarrow P(n+1))$ the *minimum* of the values of the individual sentences $P(\bar{k}) \rightarrow P(\overline{k+1})$. Why does the many valued theory do this? Well, in giving a semantic account of the quantifiers, many-valued logic finds itself in the same predicament that arises from the attempt to find many-valued truthtables for the connectives. Indeed if one wants to preserve the equivalence between, say, universal quantification and conjunction in the case of finite domains, and one has already accepted the definition of & according to which the value of $(\varphi \,\&\, \psi)$ is always the minimum of the values of φ and of ψ, then there is nothing for it but to accept the definition of the previous paragraph.

Once we regard the second premise of the paradox as approximately true it becomes, as we have already seen, inevitable to block the inference by faulting the inference rules which are used in the proofs of the sentences $P(\overline{k})$. The Fine-Thomason account shows, however, that once we admit that the conditionals $P(\overline{k}) \rightarrow P(\overline{k+1})$ are not absolutely true, there is no evident need to put the blame for the paradox on classical logic and to replace that logic by one which rests on very problematic foundations.

6.

But does this really solve the paradox? Gradually I have become persuaded that it does not. This change of heart has undoubtedly been caused at least in part by the scepticism with which my initially enthusiastic expositions of the supervaluation account often met. I do not know whether the analysis which I want to discuss in the remaining sections of this paper will satisfy the sceptics I have encountered. But at any rate it does dispose of what is perhaps the most telling criticism of the supervaluation account. The point is made by Sanford, who notes in Sanford (1976) that the very same truth definition which confers upon $(\forall n)(P(n) \rightarrow P(n+1))$ the value 0 gives the value 1 to its contradictory $(\exists n)(P(n) \ \& \neg P(n+1))$. But if this latter formula is indeed to be seen as the symbolization of the statement that there is a number n such that n grains of sand can make a heap whereas $n+1$ grains cannot, then this is a dubious result at best. For what number n could possibly satisfy this condition? The defense of the supervaluation theorist must be that it is immaterial that we cannot supply such a number, although we could never say which one it is, inasmuch as any resolution of the truth value gap of P will yield one.

But this is quite unconvincing. The difficulty is not just that we cannot point to some particular number n for which the condition is satisfied. It is rather that for any n we might care to consider, the idea of it satisfying the conditions seems contradictory or incoherent.

Dummett (1975) makes an important point about what he calls 'observational predicates'.[6] By an *observational predicate* we understand a predicate P with the following property, to which I shall refer as E(quivalence of)

[6] The point is discussed at greater length in a paper by Crispin Wright (1975). Although Dummett (1975) and Wright (1975) appeared simultaneously, it seems that the insight originally came from Dummett.

O(bservationally) I(ndistinguishable entities): (EOI). Suppose the objects a and b are observationally indistinguishable in the respects relevant to P; then either a and b both satisfy P or else neither of them does.

Dummett concluded that many observational predicates must be (i) vague; and (ii) incoherent. I will not retrace the argument in its most general context but consider the properties of observational predicates given some further assumptions, which are pertinent to the paradox which occupies us here.

Consider an observational predicate P such that:

(i) the object a definitely belongs to the extension of P;
(ii) the object b definitely does not belong to the extension of P;
(iii) There is a finite sequence of objects $a = a_1, a_2, \ldots, a_{n-1}, a_n = b$ such that any two successive objects in this sequence are observationally indistinguishable in the respects relevant to P.

Evidently there cannot be such a predicate. For (EOI) implies that if a belongs to the positive extension of P then the same must be true of a_2, and of a_3, and … and finally of b. But by assumption b does not belong to the positive extension of P.

This conclusion may seem disturbing. For it implies that many predicates which we might have thought, or wanted, to be observational cannot be that. It entails in particular the impossibility of a certain form of verificationism which proposes to analyse the meaning of every statement in terms of its relations to 'Protokollsätze', statements which can be regarded as directly verified by the observations which they report. The Protokollsätze I have in mind here are not statements about immediate objects of experience such as sense data, but are couched in a language whose terms refer directly to things outside the observer. A typical example would be:

(8) The tip of pointer x coincides with mark y on the dial z, while the hands of clock u coincide with the dial marks v and w.

If this is to be a *direct* report of an observation the predicate 'coincides' (to consider just one of the several general terms of the sentence) must be an observational predicate. However, Dummett's consideration seems to apply. We can easily imagine a series of positions of the tip of x the first one of which coincides with y, the last one of which does not, whereas no difference can be observed between any two successive positions. (We could obtain such a series for instance by photographing the dial repeatedly in rapid succession while the pointer is slowly moving across it.) It follows that if coincidence is a coherent concept it cannot relate to observation in the manner stated by EOI.

How could it be related? Here is one alternative account, inspired by similar constructions in Carnap's *Aufbau* and Goodman's *Structure of Appearance*:

'Coincides', as it occurs in (8), is a relation between *positions*, where a *position* is defined as an equivalence class under the equivalence relation which holds between two material objects *a* and *b* just in case for each object *c*, *a* and *b* stand to *c* in precisely the same spatial relations (such as overlap, complete covering, etc). Two such positions may then be said to *coincide* iff they contain members which overlap. This definition of coincidence will make sense only if the predicates in the definiens, and in particular the spatial relations to which the definiens refers, are coherent; we have just seen however that these spatial relations could not be coherent if they were observational; if they are not observational predicates, however, what are they? Clearly it would get us nowhere to try and interpret them in turn along the lines of our construal of coincidence. For such an interpretation would involve once again new predicates which must either be observational—which they cannot be—or be related to experience in a manner still awaiting elucidation.

What we have considered is only one possible attempt to define concidence as a non-observational predicate. And I am not offering a proof that every other definition must lead to a similar regress. Nevertheless the failure of the particular definition we have looked at would seem to justify the suspicion that no such definition will work and thus that a verificationist account of the sort considered must fail at the very foundation of the elaborate edifice it means to erect.

7.

In view of the multifarious other problems with verificationism the outcome of these reflections is perhaps not all that dramatic. However, this is not an essay about confirmation theory, but one concerned with a paradox whose implications reach beyond the questions of meaning and verifiability in the language of science. The predicates 'bald' and 'heap' for instance, which were central to the two illustrations of the paradox which we have thus far considered, play no part in scientific theory. Indeed the vagueness of *these* predicates appears to be precisely what renders them unsuitable for scientific purposes. Yet, useless as such terms may be to the scientist, they are indispensable tools for communication at a more mundane level.

An important difference between these predicates and a predicate such as 'coincidence' seems to be this. Coincidence, we saw, cannot be coherent

if it is an observational predicate. The conclusion we drew from this—and apparently the only reasonable conclusion that can be drawn from it—was that 'coincidence' is not an observational predicate—a conclusion which only the missionaries of a certain verificationist programme would seriously regret. But with predicates such as 'bald' the predicament is more serious. Denying that 'bald' is an observational predicate won't do, for it seems to be "part of the logic of 'bald' " that where two heads look indistinguishable, one should qualify as bald if and only if the other does. In fact it could be argued that this equivalence retains its validity even for certain cases where the difference between the two *is* detectable by direct observation, but deemed so small as to be irrelevant. A clear case of this is the predicate 'small' as applied to natural numbers. The number predicate 'small' gives rise to another illustration of the paradox, which has come to be known as Wang's Paradox: since 1 is a small number, and since whenever n is a small number $n + 1$ is a small number too, every number is small. Evidently there is a detectable difference between n and $n + 1$; Wang's Paradox derives its force from the circumstance that a difference of 1 does not seem big enough to separate what is small from what is not.

To see more clearly what the *logic* of predicates such as 'bald' and 'heap' might be, let us contemplate yet another instance of the paradox.

In front of us is a large screen. Its extreme left is green, its extreme right yellow, and there is a gradual transition from the one colour to the other. The screen is subdivided into many small squares, so small that each square appears to have a uniform hue and moreover the colours of no two adjacent squares can be distinguished by sight. Compare the following two experiments.

(1) We are both facing the screen which is entirely visible to you. I begin by pointing at a little square on the extreme left and ask you what its colour is. Assuming you are not colourblind you will surely answer: "green". I then point to the adjacent little square on its right and ask the same question. Probably you will again say "green". Then I point at the square to the right of this one, and so on. After a while your answers "green" will become hesitant, increasingly so, until the point is reached where you either say: "Now I really don't know what to say any more," or else some such thing as "this one really looks more like yellow."

(2) This time the big screen is completely covered. I ask the same question about the same squares in the same order. But I now proceed as follows. When I ask my first question I *only* uncover the first little square. After you have answered I reveal the square next to it. Then, after your second answer, I cover the first square and uncover the third; after you have made your third reply, I cover the second and uncover the fourth, etc.

How will the outcomes of these two experiments compare? In all likelihood you will carry on answering "green" for a longer time in the second trial than in the first.[7] The reason why I think we should expect this discrepancy is the following. During the first trial you are subject to two forces which eventually work against each other. On the one hand there is a commitment to EOI which requires you to keep answering "green", as it never seems quite right to suddenly stop saying "green", after you have already said "green" to the previous square—after all, the two squares look just the same. On the other hand there are the indubitably green squares on the extreme left and the indubitably yellow squares on the extreme right. These serve as it were as *anchor points*. And there comes a time when the square about which you are asked looks as much like the yellow squares on the right as it looks like the green squares on the left. At that point, or shortly afterwards, the second force comes to dominate the first, and you either suspend or switch judgment.

During the second trial the visible anchor points are absent; and so it might be expected that your commitment to EOI, unopposed this time by competing visual impressions, will cause you to persist with your verdict "green" beyond the point where, in the course of the first trial, you suspended it. In fact, if EOI were entirely unopposed you would presumably carry on with your verdict until the bitter, yellow, end. But of course even in the second trial EOI does not operate unchallenged. Even if you know nothing about the big screen in advance, you will nevertheless have an understanding of what it is for something to be yellow, or green, in absolute terms. And this understanding, the ability to judge the squares, to some extent at least, on their own merit will eventually defeat your commitment to EOI. Thus in spite of the differences the two trials have something important in common: in each case you are subject to two increasingly conflicting semantic principles. The second principle eventually wins—but not without leaving you with the feeling that there is some sort of inconsistency in the sequence of your responses.

[7] In fact I have not carried out these experiments. And so it is not quite legitimate to build on their conjured outcomes. However, all I wish to do for the moment is to give a graphical illustration of the general theoretical account that will follow shortly. Of course it would be somewhat embarassing if the outcome of the experiment went against what I am here conjecturing. In any case I would be much interested to know what actually happens in these trials. I have been unable to verify whether the experiments have been carried out by others.

The interest of such Gedanken experiments lies, I feel, in the force with which they expose the nature of the paradox as well as its depth. They appear to show that a substantial class of terms which our language contains, and which it could hardly do without, are ruled by semantic principles which are strictly incompatible. This is a fact we must take very seriously indeed. It is equally remarkable that in actual practice we manage so well with such terms in spite of the lurking contradictions which beset them. One almost feels that the man who conducts such an experiment and leads the guileless subject to the embarrassing contradictions in which he will inevitably get entangled, may be accused of malice.

And yet, is there any sense in which these experiments can be considered illegitimate?

To answer this last question we must begin by drawing attention to a phenomenon that goes hand in hand with vagueness and to which I shall refer as *contextual disambiguation*. It is typical of a vague predicate that what objects it is true of depends on the context in which it is used. There are certain adjectives in particular—such as e.g. 'large', or 'soft' or 'clever'—about whose extensions we can say hardly anything in abstraction from any contextual setting; it is only with respect to a given context of use that we can meaningfully ask whether a certain object is large, and there are very few, if any, objects of which it is clear absolutely whether or not they are clever or soft. Different contexts resolve these questions in different ways; the same object may count as definitely clever in one context and yet as definitely not in another. Such reduction of ambiguity is, moreover, not just a privilege of those adjectives whose extensions are completely, or virtually, indeterminate in the absence of particular contexts. It also applies to words such as 'liquid' or 'green' which are definitely true (or false) of at least some objects irrespective of the context of use.

The contextual factors that can help to reduce vagueness are so varied as to render the task of an exhaustive analysis of contextual disambiguation a virtual impossibility. Yet some factors operate in a sufficiently general and systematic manner to allow (and deserve) articulation. One such factor, to which I drew attention in Kamp (1975), concerns prenominal occurrences of adjectives. Adjectives in such positions depend as a rule heavily for their extensions on the nominals which they precede. Here we shall be concerned with another factor, which influences the extensions of those predicates which are ruled by the principle EOI. One kind of contribution which the context can make to the evaluation of an utterance is mediated by the *background* that context provides, i.e. by the information that is taken for granted by the participants in the discourse of which the utterance is part.

It has by now become commonplace that the contextual background has to be taken into account in the analysis of such phenomena as presupposition, implicature and anaphora.[8] In connection with the predicates that concern us here the background is important in view of the fact that, as EOI asserts, we must recognise an object as falling within the positive or negative extension of such a predicate if it is indistinguishable (in the relevant respect) from an object which has already been *accepted* as belonging there. The backround of a context of an utterance is in general determined, at least in part, by that section of the ongoing discourse which precedes the utterance; and therefore evolves as the discourse unfolds and more information becomes accepted.[9] Of particular significance here is the effect exerted upon the context by the acceptance of a new sentence of the form $P(\bar{a})$ or $\neg P(\bar{a})$. This modifies the context into one in which one must now also acknowledge as P (or not P) all those objects which are indistinguishable from a (including those which *can* be distinguished from all objects which had already been accepted as P, or not P, in the original context.)

Consider again our experiment with the screen. We can now describe the sequence of exchanges between experimenter and subject in the following way. When I ask my question about the first square I do so in a context c_0 in which relatively little has been accepted concerning the extensions of green and yellow. We may assume that the first square to which I refer in my question is to be accepted as green not because of a commitment made earlier in the course of discussion but because it is context-independent: we can regard *any* normal context as carrying the commitment that this square is green just as we may assume that any normal context carries the commitment that the rightmost square be accepted as yellow, and thus as not green. As regards intermediate squares, however, there need be no such commitments at the outset. You answer that the square is green. In this case there is no reason to suppose that the context changes as a consequence of your verdict. After all we assumed that c_0 already carries the acceptance of this judgement.

I ask you about the second square. Let us suppose that in the context c_0 there is no *explicit* commitment to accepting this square also as green. Even so EOI demands that you call it green since it is indistinguishable from a square concerning which the context in which the question was put

[8] See e.g. Gazdar (1979).

[9] For example of this two-way interaction between context and utterance see Lewis (1979b).

to you does carry such a commitment. Suppose you qualify this square as green. (Indeed this seems the right response, if only because the square you are watching looks just like the square you already said was green, while we may assume there are as yet no other factors which could override this consideration.) Once again I do not challenge your reply and so its content is incorporated into the contextual background. And in this way we continue. As we continue the context gets ever more heavily burdened with assumptions—until, at last, the assumptions produce a direct clash: the context carries the commitment that a certain square is green and at the same time the commitment that it is not green.

Such a context is clearly incoherent. In what follows contextual incoherence will play an important part. I shall refrain however from trying to define exactly what incoherence consists in, from defining, in other words, precisely which contexts are coherent and which are incoherent. But I will give at least one sufficient condition for incoherence, which generalizes the observation we just made:

(9) If the background assumptions of the context entail for some sentence φ both φ and $\neg\varphi$, then the context is incoherent.

It would not seem warranted to treat (9) also a necessary condition for incoherence. For suppose that c is a context which does not satisfy (9) but whose background entails both that a given object a is green and that some other object d is not green, whereas there is an object b which is indistinguishable in colour both from a and from d. Could such a context be regarded as anything but incoherent?

The basis for insisting that c is not incoherent appears decidedly slim. Note however that once we admit that c is incoherent we seem to have little justification for denying the incoherence of a context c', where c' is like c except that instead of object b there are now two objects b_1 and b_2 such that b_1 is indistinguishable from a and from b_2 while b_2 is also indistinguishable from d. Evidently, as these considerations reveal, incoherence is itself a vague predicate; and so is its positive counterpart, coherence. The characteristics for which these predicates stand 'come', one is inclined to say, 'in degrees'. When we start out e.g. with the screen experiment we do so, it may be assumed, in a context which is perfectly coherent. But as we proceed the context becomes progressively less so, until, if only we carry on for long enough, all semblance of coherence will have been lost.

It might be thought in the light of these reflections that coherence and incoherence are themselves answerable to the principle EOI, in just the way in which that principle applies to such predicates as 'green' which led us to

consider these contextual attributes. As a matter of fact I do not believe that there are compelling reasons for thinking that they must be so answerable. But this is a question which we must leave for the time being. I shall return to it in Section 9.

9.[b]

The ideas of contextual influence on truth, mediated by the principle EOI; of contextual incoherence; and of the possible deterioration of an initially coherent context through repeated incorporation of truths constitute the basic elements of the formal account which we shall now proceed to expound. The main formal feature of this account, which sets it apart both from the supervaluation theory and from that supplied by fuzzy logic, is that a universal sentence may be false even though each of its instances is true. This allows us to maintain that the second premise of II, $(\forall n)(P(n) \rightarrow P(n+1))$, is false, while yet for each number k the conditional $P(\overline{k}) \rightarrow P(\overline{k+1})$ is (perfectly) true. And this is, I wish to claim, precisely what we should say about $(2')$ and its instances in relation to any of those situations in which the repeated application of EOI leads to incoherence. Evidently this requires a non-standard analysis of the universal quantifier. In fact we shall also adopt a non-standard account of the conditional, and I shall begin by explaining how I propose to analyze that construction.

Crucial to that analysis is the two-way interaction between context and utterance which we encountered in our last description of the screen experiment. The context affects the truth value of the sentence uttered in it, and the utterance modifies the context when it is accepted. The rationale for the proposal is perhaps most clearly explained through a quick look at another linguistic phenomenon which depends on a similar interplay between language and context. I mean the phenomenon of pronominal anaphora. A detailed account of pronominal anaphora, from the viewpoint that is relevant for what I want to say here, can be found in Kamp (1981). There is no need, however, given our present purpose, to reiterate the exact contents of that paper. It will suffice to give the merest outline of some of the underlying ideas, ideas which, incidentally, can already be found in Chastain (1975). Chastain considers, among others, the following three bits of discourse:

[b] Editors' note: Please note that there is no section 8 in the original text.

(10) Three years ago a man deposited $5000 in the bank at 5% interest compounded semi-annually. Now he has over $5790 in his account.
(11) Suppose a man deposits $5000 in the bank at 5% interest compounded semi-annually. After three years he has over $5790 in his account.
(12) If a man deposits $5000 in the bank at 5% interest compounded semi-annually then after three years he has over $5790 in his account.

How are pronouns such as the 'he' of (10) linked with their anaphoric antecedents? It would seem natural to think that the anaphoric function of pronouns must be intimately related to their deictic function (which is exemplified in my saying: "He is a fool", while pointing at some particular person); and that in both types of use the pronoun refers to some independently available entity. Where the pronoun is used deictically this entity is an object in the real world; in case it is used anaphorically the 'referent' is established through the preceding bit of discourse in which the anaphoric antecedent is contained. In this second case the referent is, we might say, contextually supplied. But if we say this we must acknowledge that the preceding discourse has modified the context in which it was evaluated, endowing it with the information which the anaphoric pronoun may then take advantage of. Thus if (10) is uttered in a context c then the first of its sentences will transform c into another context c' in which the second sentence can then be evaluated appropriately.

The same mechanism must be responsible for the possibility of anaphoric reference in (11). There again the second sentence is to be evaluated with respect to a context which has been modified through an antecedent evaluation of the first sentence. The difference between (10) and (11) lies in the ways in which the two evaluations are connected. (10) is true in a context c provided its first sentence is true in c and the second is true in the context which results from evaluating the first. (11) on the other hand is true in c, roughly, iff, provided the evaluation in c of its first sentence is positive (i.e. shows the sentence to be true), the second sentence is true in the context resulting from that evaluation.

What we have just said about (11) applies, without substantial changes, also to (12). The only distinction between (11) and (12) lies in that (12) consists of a single sentence instead of a couple of them. If we are right in assuming such a close kinship between the two discourses it follows from what we have said about (11) that the evaluation of (12) in c must involve evaluating one *part* of that sentence (viz. its consequent) not in c itself but in the context into which c was modified on account of evaluating another part of the same sentence. This leads us to the following assessment of the truth conditions of a conditional:

(13) A conditional 'if φ then ψ' is true in a context c iff, provided the evaluation of φ in c is positive, the evaluation of ψ in the context modified by this evaluation is positive too.

(13) is applicable, I maintain, not only to conditionals whose consequents contain context-dependent elements of the sort we just considered, but whenever the consequent involves context-sensitive material. It applies, in particular, to those conditionals whose consequent contains context-sensitive predicates such as P; in that case the modification effected by the evaluation of the antecedent should be understood in the sense in which we talked about modification in the preceding section, viz. through acceptance of the evaluated sentence as true. Thus for instance if k and $k+1$ count as not discriminable in any way that is relevant to P, the conditional $P(\overline{k}) \rightarrow P(\overline{k+1})$ will be true in any normal context c. For acceptance of $P(\overline{k})$ creates, as we argued already, a context in which, $k+1$ being not relevantly distinguishable from the number k which has already been committed to the positive extension of P, $P(\overline{k+1})$ must be true as well.

Note by the way that (13) reduces to the familiar truth condition for the conditional in all those cases where the evaluation of φ has no effect on that of ψ, i.e. where it makes no difference whether we evaluate ψ in the modified context or in c itself.

There is not a very great deal that I can say in support of the principle that allows $(\forall n)(P(n) \rightarrow P(n+1))$ to be false even though all conditionals $P(\overline{k}) \rightarrow P(\overline{k+1})$ are true. The principle embodies the idea that a sentence such as $(2')$ must be reckoned false whenever its acceptance would lead to inconsistency. Although I personally find this idea intuitively plausible I have no real argument to show that the principle is correct; and I am aware that this, more than any other, part of my proposal is likely to meet with scepticism. However, leaving the intrinsic merits of the principle momentarily aside, we can note at least that it will be operative only where acceptance may affect subsequent evaluation. To be explicit, suppose we stipulate that a sentence of the form $(\forall x)\varphi$ is true in a coherent context c iff (i) each of its instances is true, and (ii) acceptance of the sentence preserves coherence. Then this condition will reduce to the standard truth condition for universal sentences if acceptance cannot produce inconsistent subsequent evaluations and thus cannot transform a context which is coherent into one that is not.

10.

Contexts play so central a role in our theory that it seems right to devote a little space to the question of how they should be represented. The existing literature contains in particular two answers to this question which demand attention, that of Lewis (1979b) and that of Cresswell (1973).

According to Lewis a context is a sequence of features which are relevant to the evaluation of those context-sensitive elements which are part of the language that is under study. Among these features there may be e.g: the speaker, the hearer, the time of the utterance, the place of the utterance, etc. Cresswell objects to this identification on account of its lack of flexibility. When we move from a given language L to a more comprehensive language L' which contains context dependent elements of a type not present in L we may have to add new features to the contexts in terms of which these new elements must be analysed. One is thereby forced to revise the notion of context itself, and this seems unnatural. Rather the additional contextual information to which the new elements of L' are sensitive should be regarded as already present in the contexts of utterance of expressions of L—it is just that the evaluation of such an utterance never exploits this information. Cresswell himself proposes to identify contexts with certain *properties* of utterances. This gives the flexibility which he misses in Lewis's approach. For besides the context which an utterance u has iff it is uttered at time t and by speaker s, there is also the context which u has iff u is uttered at t by s to hearer h in place p. The contexts of Lewis's theory corresponding to these two properties would be sequences of different length. Of course one could introduce a similar flexibility into Lewis's theory by admitting as contexts sequences of varying length and allowing for the possibility that in any one such context c a sentence φ may lack a truth value because c does not specify some information upon which that truth value critically depends. But this remains unsatisfactory in so far as the actual speech situations of which these contexts are the formal representations normally contain that information, e.g. the information regarding audience or place. In these cases the lack of truth value is evidently an artifact of the theory. Thus there is no simple way it seems in which the theory can distinguish between such spurious truth value gaps and the genuine gaps which arise when the speech situation itself does not supply the necessary information—such genuine indetermination of truth value may arise, e.g., when a second-person sentence is uttered in the presence of several people while it is unclear who is being addressed. (The speaker may be so confused about the identity of his intended audience

that even his intentions would provide no sufficient basis for identifying the addressee.)

There is nevertheless something unnatural about Cresswell's proposal too. Intuitively it seems perfectly meaningful to speak of *the* context of a given utterance. Each utterance is placed in a specific set of circumstances and these provide all the contextual information concerning the utterance that could be provided at all. Much of that information to be sure, and often all of it, is left unused as it is not needed to evaluate the sentence uttered. But it is nonetheless there to be had; and it is part of what we think of as the utterance's context. Cresswell's proposal however seems to offer no means of identifying what the context of an utterance u might be. u will have a multitude of properties; which of these should qualify as 'the' context of u, to the exclusion of all others?

The simplest way out of this minor predicament is, I believe, this. We assume contexts as separate entities, distinct both from the features they determine and from the utterances which take place in them. Each type of feature—such as speaker, hearer, time, place, background assumptions, etc.—is a function from contexts to items of the appropriate sort. This leaves it admittedly quite indeterminate what contexts 'really' are. But as long as the ranges of the feature functions consist of well-defined entities this will put us in a position which is no worse than that we would have been in had we adopted either of the proposals we just discarded. For suppose that L is a language whose expressions are sensitive to no other contextual features than those given by the functions $f_1, \ldots, f_n$. Then we can translate any semantic theory for L articulated within the framework proposed here into a theory which refers to n-tuples of features rather than to our contexts, viz. by reinterpreting every reference to a context c in our sense as a reference to the tuple $\langle f_1(c), \ldots, f_n(c)\rangle$. Similarly we can reduce our theory to one that fits Cresswell's characterization of contexts. For each of our contexts determines, via $f_1, \ldots, f_n$, a unique property of utterances, viz. that property which an utterance u has iff it is made in a context c to which each of the functions $f_1, \ldots, f_n$ assigns the same value it assigns to the given context.

11.

In the formal theory we shall state below, only one contextual feature will figure explicitly, that of the background assumptions of the context. Some existing studies in which this notion is formally exploited take the back-

ground to be a set of propositions. Such an identification is possible only if one supplies a separate account of what a proposition is. Unfortunately we are not in a position to do this, inasmuch as any characterization of the realm of propositions must presuppose, or else simultaneously provide, the logic by which these propositions are governed; and it is precisely one of the purposes of our study to determine what that logic could be like. We are reduced therefore to identifying the background of c as a set of sentences, belonging to the very object language for which we shall give a definition of truth.

Our analysis of the paradox essentially depends, as we saw, on the possibility of modifying a context by accepting one or more new statements as part of its background. And, as may have been expected, contextual modification plays an important part in our formalization too. We shall represent modification as a function m which associates with a context c and a set of sentences Γ a new context $m(c, \Gamma)$. As $m(c, \Gamma)$ results from incorporating into the background $B(c)$ of c all the sentences in Γ we must have:

$$\Gamma \subseteq B(m(c, \Gamma)).$$

Acceptance, moreover, can, or so at any rate we shall assume, only increase contextual commitment; which means that we should also impose the general condition

$$B(c) \subseteq B(m(c, \Gamma)).$$

Other conditions relating B and m could be contemplated as well, but they won't be needed in the formal developments to which I have restricted myself in the present essay.

The particular effect of context on truth with which we are concerned here involves the notion of relevant difference: where a does not differ from b in a way relevant to P and the context carries the commitment that b belongs to the positive, or negative, extension of P then $P(a)$ is true, or false, in that context. The relation which holds between a and b if they are not relevantly different is evidently reflexive and symmetric (though, in the cases that interest us, it is of course not transitive). We shall symbolize this relation as $\sim$ and assume reflexivity and symmetry for it. (As we will consider only one predicate P there will be no necessity to introduce more than one such relation.)

We already indicated the role that $\sim$ is to play: an atomic sentence $P(\bar{a})$ will be true, or false, in a context c if there is a $b \sim a$ such that c carries the commitment to the truth, or falsity, of $P(\bar{b})$. But how should we understand 'commitment'? One possibility, it might be thought, would be

this: c commits us to the truth of $P(\bar{b})$ iff $P(\bar{b})$ belongs to the background of c (and it commits us to $P(\bar{b})$'s falsehood iff the background contains $\neg P(\bar{b})$). But this surely is too restrictive a notion of commitment. Suppose for instance that the background contains the sentence $P(\bar{b})$ & $P(\bar{a})$, but that it does not contain $P(\bar{b})$ as a separate sentence. Would it not be reasonable nevertheless to regard this background as committing us to the truth of $P(\bar{b})$? Or suppose the background contains the sentence $(\forall x)(\varphi(x) \rightarrow P(x))$ and $\varphi(\bar{b})$, but that $P(\bar{b})$ is not one of its actual members. Again we might see ourselves as committed by this background to the truth of $P(\bar{b})$. For once more $P(\bar{b})$ follows from sentences the background explicitly contains.

But does $P(\bar{b})$ really follow from those sentences? The inferences are clearly valid in classical logic; but we have as yet no good reason to expect that the logic which our semantic analysis will generate will turn out to be that logic. In fact we shall soon see than on many, and possibly all, of the semantic proposals we are now in the process of formulating, it is not. Even so, our analysis must generate some logic or other, and whatever can be inferred according to that logic from sentences which explicitly belong to the contextual background must count among the statements to which the context commits us.

As we cannot know this logic in advance there is only one way in which we can proceed, and that is simply to assume that we have the inference relation, $\vdash$, already. This relation, concerning whose extension we are at present in the dark, will determine, albeit in ways we cannot as yet fully assess, what commitments are carried by sets of sentences which constitute contextual backgrounds and, thereby, what inferences should count as semantically valid.

There are two grounds why a sentence $P(\bar{a})$ may be true in a given context: either a belongs to the positive extension of P for contextually independent reasons, or there is a contextual commitment that carries with it the truth of $P(\bar{a})$. In general the truth of $P(\bar{a})$ in c will result from an interaction between these two factors; e.g. $\bar{b}$ may belong to the positive extension of P absolutely while the background of c contains the sentence $P(\bar{b}) \rightarrow P(\bar{a})$. Then c should presumably count as carrying a commitment to the truth of $P(\bar{a})$. For this sentence would seem to follow from the background sentence $P(\bar{b}) \rightarrow P(\bar{a})$ together with the sentence $P(\bar{b})$, to which there is a commitment that is not even contextually dependent. (I shall assume for the sake of this discussion that the rule of Modus Ponens, which the inference of $P(\bar{b})$ from $P(\bar{b}) \rightarrow P(\bar{a})$ and $P(\bar{b})$ instantiates, is validated by the inference relation $\vdash$ in terms of which we have decided to character-

ize commitment; but see Sections 15–17 for further discussion of this and related issues.) It should be clear that if the set of sentences to which c commits us is to be determined as consisting of all those sentences which follow, in the sense of $\vdash$, from a specified set of premises, and if the sentence $P(\bar{a})$ is to be one of those to which c commits us, then the premise set will have to include those sentences which, like $P(\bar{b})$, express commitments that are absolute—rather than resulting from the adoption of a sentence somewhere along the path which led to the context c in question.

Just as the sentence $P(\bar{b})$ conveys that b belongs to the positive extension of P, so does $\neg P(\bar{b})$ express that b is a member of P's negative extension. Where b belongs to P's negative extension on context-independent grounds the sentence $\neg P(\bar{b})$ will be as indispensable to the assessment of which atomic sentences are false in c as sentences of the form $P(\bar{d})$ are needed to determine which atomic sentences are true. Thus, if the commitments carried by a context c are to be characterized as those sentences which are implied, in the sense of $\vdash$, by a given set of premises, we must include among those premises, not only all sentences $P(\bar{b})$ where b is in the positive extension of P for context-independent reasons, but also all sentences $\neg P(\bar{b})$ where b is for such reasons in the negative extension of P. We shall refer to the set of all these sentences as the *diagram* (*of* P).

For the formalization below it will be convenient to assume the background sets $B(c)$ to be closed under deduction, and also to take them to always include the absolute commitments embodied in the diagram. This may seem somewhat unrealistic inasmuch as the *explicit* commitments made in the course of discussion would normally not be so comprehensive. There should, however, be no real difficulty in applying this more abstract notion of background to actual situations of speech.

12.

The last notion we must discuss, once more, before starting on our formalization, is contextual coherence. Some contexts are incoherent. Among them, in particular, are those which assign opposite truth values to one and the same sentence. On the other hand some contexts should be acknowledged as definitely coherent. Most 'ordinary' contexts we encounter in daily life should count as coherent, in particular those contexts in which a discussion begins. It was an implicit assumption of Section 7 that the experiments we discussed there always start off in coherent contexts.

Where lies the boundary between coherent and incoherent contexts? We have already observed that coherence must be a vague concept. A context can become gradually less and less coherent but it is impossible to say for sure precisely when it is coherent no longer.

To remain as close as possible to a classical metatheory I shall assume two context predicates, Coh(erent) and Inc(oherent), which are mutually exclusive but not necessarily jointly exhaustive of the set of all contexts, and of whose extensions we may think as the positive and negative extension of a single vague coherence concept. For my present limited purposes this will be adequate, although from a conceptual standpoint it is not fully satisfactory. For even if we assume that the two predicates are not jointly exhaustive what intuitive justification could there be for the sharp division which is thus introduced between, say, the coherent contexts and those which are neither coherent nor incoherent?

I think it is worthwhile to briefly interrupt our exposition at this point for the sake of a general observation on formal analyses of vagueness. The point to which I want to draw attention is simple enough; but, even so, it is not always kept in mind as firmly as it deserves to be, and as far as I know there is as yet no explicit statement of it in the literature.

The description of vagueness offered in Kamp (1975) suffers from the same deficiency that I just pointed out about my present proposal for the representation of contextual coherence. No intuitive significance can be attached to the sharp boundary between the truth value gap of a vague predicate and its positive or negative extension. One feels there just *could* not be such a sharp demarcation. Fine (1975) is aware of this problem and proposes an ingenious construction to circumvent it. He reasons roughly as follows. It is not fully determined where the boundary between, say, the truth value gap and the positive extension of a vague predicate P falls. It could fall in a number of different places. This suggests that the model in which the boundary falls at some particular place is only one of a collection M of similar models in each of which it runs differently. However, given this class M of models there is still a distinction between on the one hand the objects which belong to the positive extension of P in *each* of the models in M and on the other hand those which lie outside the positive extension in at least one model in the class. This sharp distinction is also unreasonable. The distinction is consequent upon the choice of *this particular class M*. But that choice is artificial, inasmuch as nothing in the semantics of the predicate P could possibly determine precisely which boundaries between positive extension and truth value gap are possible and so precisely which models should belong to M. Thus M is to be seen as only one of a number of

possible such sets of models. Together these various possible sets of models form a super class M^2. But of course with regard to this super class the same embarrassing question arises again. What significance is there to the sharp division between on the one hand those objects *a* which belong to the positive extension of *P* in all models of all sets *M* that belong to M^2, and on the other hand those *a* which fail to belong to the positive extension of *P* in at least one model belonging to a set *M* that is a member of M^2? This suggests that even the choice of M^2 is to some extent arbitrary. So we should regard it as being one of a group M^3, of similar such classes. M^3, however, will give rise to the same question in its turn; so, presumably, it is only one of a number of such possible groups all belonging to M^4. There appears to be no natural end to this regressive sequence of super classes $M^1, M^2, M^3, \ldots$.

Fine does assume that the regress may be infinite, and uses the rather complicated set theoretic structures that arise in this way to state the semantics of a language which contains besides vague predicates and the familiar connectives and quantifiers also a new 1-place sentential operator *D* (meaning 'it is definitely the case that'). The logical structure of this language seems intriguing and very complex. It is distressing however that in spite of these complexities the original question still causes embarrassment. For what justification is there for the perfectly sharp division between on the one hand those objects *a* which belong to the positive extension of *P* in all models belonging to all classes belonging to all classes, belonging to all classes, belonging to ..., and on the other hand those which fail to belong to some model belonging to some class belonging to some class belonging to ...?

This predicament is of course inescapable: any semantic account of a vague predicate *P* according to which at least some objects are definitely *P* and some others are either definitely not *P* or belong to the truth value gap of *P*, will produce such a sharp distinction if the language in which this account is given contains only sharply defined predicates and the apparatus of classical logic and set theory. For whatever the condition may be which separates the objects that are definitely *P* from the others, it is bound to mark a sharp division—such is the classical theory of sets.

The only escape from this predicament lies in the adoption of a non-classical meta-language. I do not know however of any proposal of this sort that has been worked out in full detail. Nor will I attempt to proceed in this direction myself. Indeed I believe that any such attempt would be premature. For the value of any semantic analysis depends on the adequacy of the logic which underlies the meta-language in which the analysis is carried out. But what is the logic of a language that contains vague concepts? This is precisely the question which the paradox of the heap poses with so

much force. The purpose of this essay is to elucidate that question. To adopt a vague meta-language would mean either that we put our cart before the horse or else that by defaulting on the theorist's obligation to make the logic of his meta-language explicit, fail to attach the cart to the horse in any way whatsoever.

13.

We have now reached the point where we can attempt a formal truth definition that incorporates the ideas we have gathered along the way. We shall begin by making things as easy for ourselves as is compatible with the aim of giving a formal account of the paradox along the lines sketched above, by concentrating on a language which contains, apart from negation, just the logical operators which figured explicitly in our earlier observations. Let L_0 be the language whose vocabulary consists of:

(i) variables: $\nu_0, \nu_1, \nu_2, \ldots$;
(ii) individual constants $c_0, c_1, c_2, \ldots$;
(iii) a 1-place predicate P;
(iv) the logical constants $\neg, \rightarrow, \forall$.

The *well-formed formulae* of L_0 are defined as usual. We shall refer to the set of sentences of L_0 as S_0.

The models for L_0 relative to which we shall define truth and falsity must incorporate all those structural elements which we have earlier singled out as relevant to the truth values of sentences containing the vague context-sensitive predicate P. In the light of what has already been said there would seem to be no further need to argue that this leads to the following notion of what we shall call a general context-sensitive model for L_0.

Definition 1 *A general context-sensitive model for* L_0 is a tuple $\langle U, F, C, \mathrm{Coh}, \mathrm{Inc}, B, m, \sim\rangle$ such that

(i) U is a non-empty set (the *universe of discourse*);
(ii) F is a function which assigns to each individual constant c_i of L_0 an element of U (the *referent of* c_i), and to the predicate P a pair of disjoint subsets of U (the *positive* and the *negative* extension of P, which we shall denote as $F^+(P)$ and $F^-(P)$, respectivly);
(iii) C is a non-empty set (of *contexts*);
(iv) Coh and Inc are disjoint subsets of C (the sets of *coherent* and of *incoherent* contexts);
(v) B is a function from C into $\mathfrak{B}(S_0)$; ($B(c)$ is the *background of* c);
(vi) m is a function from $C \otimes \mathfrak{B}(S_0)$ into C; ($m(c, \Gamma)$ is the *modification of* c *resulting from incorporating into* it the set Γ);

(vii) $\sim$ is a binary relation on U which is reflexive and symmetric;
(viii) $B(c) \subseteq B(m(c,\Gamma))$ and $\Gamma \subseteq B(m(c,\Gamma))$;
(ix) if $\Gamma \subseteq B(c)$ then $m(c,\Gamma) = c$.

To facilitate the formulation of various propositions which we shall need to express below I shall make the inessential assumption that each of the elements of the universes of the models we will consider is named by some individual constant. The techniques for generalizing the truth definition with respect to models which satisfy this assumption to those which do not are all too familiar and I shall take it that the reader will be able to carry this generalization through himself should he feel any desire to do so.[10] With regard to those models M for which the condition is satisfied it is moreover convenient to assume the existence of a function which for each $a \in U_M$ selects a particular constant $\bar{a}$ such that $F_M(\bar{a}) = a$. By D_M we shall understand the set

$$\{P(\bar{a}) : a \in F_M^+(P)\} \cup \{\neg P(\bar{a}) : a \in F_M^-(P)\}$$

The clauses (i)–(ix) do not exhaust the constraints suggested by our earlier informal considerations. But when trying to formalize the remaining conditions we face a difficulty. Several times in the course of our informal exploration I had cause to refer to an entailment relation between sets of sentences of L_0 (viz. contextual backgrounds) and single sentences. I did this with the promise that more would be said about that relation later on. It would seem that the moment has now come to discharge this obligation. Alas we are still in a poor position to do so. The problem is this. The truth definition which will be presented below will yield a notion of logical consequence different from the familiar classical notion. In fact there is little that we can say at this point about the formal properties which that notion is going to have. Now it ought to be the case that the notion of entailment for L_0 which enters into the definition of truth from which the notion of logical consequence derives, coincides with that consequence relation—this is desirable for much the same reasons that we have for wanting deductive completeness. But since we cannot define truth—or, for that matter even the intended class of models—without having the entailment relation already at hand, we can do no better than assume some such relation as given and hope for the best. So let us suppose that $\vdash$ is a given relation between sets of sentences of L_0 and sentences of L_0.

---

[10] The matter is dealt with in one way or another in every text on symbolic logic which contains the basics of model theory.

Definition 2 By a *Context-sensitive model for L_0, relative to* $\vdash$ we understand an 8-tuple $M = \langle U, F, C, \mathrm{Coh}, Inc, B, m, \sim \rangle$ such that:

(i) M is a general context-sensitive model for L_0;
(ii) if for some $\varphi \in S_0$ $B(c) \vdash \varphi$ and $B(c) \vdash \neg\varphi$, then $c \in \mathrm{Inc}$;
(iii) for each $c \in \mathrm{Coh}_M$, $B(c)$ is closed under $\vdash$ (i.e. for any $\Gamma \subseteq B(c)$ and φ if $\Gamma \vdash \varphi$, then $\varphi \in B(c)$);
(iv) for each $c \in C$, $D_M \subseteq B(c)$.

As soon as we try to give an explicit truth definition we find that our earlier reflections still leave considerable room for play. Which of the various definitions that are compatible with our informal stipulations is, or are, conceptually or formally optimal is a complicated and ramified question which I shall not try to answer in the present paper but which I hope to address in a subsequent essay. Here I shall present only a very small selection from the various alternatives. The definition is not particularly elegant, and has been partly inspired by the wish to insure in the most obvious way possible, that the resulting truth concept be coherent, in the sense that for no coherent context c and sentence φ the value of φ at c comes out both as 0 and as 1.

The main complication for the definition arises from the principle, formulated earlier in connection with the generalized conditional $(\forall n)(P(n) \rightarrow P(n+1))$, that a sentence can be false in a context c on the grounds only that accepting it would lead to a modified context that is incoherent. When I laid down this principle earlier I did not raise the question whether this principle should be seen as particular to sentences of universal form or whether it should apply across the board. In fact I have no conclusive argument for either option. But a few, partly opportunistic, decisions on other points have led me towards the former. First, to provide a simple guarantee of consistency it is expedient to formulate the conditions for truth and those for falsity in such a way that they can, at least for coherent contexts, be immediately recognized as mutually exclusive.[11] Where the falsity condition for φ is a disjunction, one of the disjuncts of which says that incorporation of φ in c yields an incoherent context, a 'safe' matching positive clause will provide a conjunctive condition for the truth of φ one conjunct of which demands that φ can be coherently incorported into c. It would be undesirable however to have positive clauses of this form generally. An atomic sentence for instance should be allowed as true in certain coherent contexts c even though its incorporation may produce a context which is no longer

[11] See postscript.

coherent. For only in this way will the present semantics be able to capture the central intuition that we may move from a perfectly coherent context to one which is plainly not by successive incorporation of a series of atomic sentences each one of which is true in the context created by its predecessors.

It is equally undesirable to make coherent incorporability a necessary condition for the truth of a conditional. It has been one of our aims to provide an analysis according to which in certain contexts of certain models each conditional $P(\bar{k}) \rightarrow P(\overline{k+1})$ is true. In the framework in which we have chosen to work we shall have to admit coherent contexts c such that for a given k, $B(c) \vdash P(\bar{k})$ while there is no coherent context c' such that $B(c') \vdash P(\overline{k+1})$. This implies that $m(c, P(\overline{k+1}))$ cannot be coherent. Now it seems an essential feature of conditionals that where φ is both true and accepted as true the truth values of $\varphi \rightarrow \psi$ and of ψ should be the same. Translated into our formalism this should mean that if $B(c') \vdash \varphi$ and φ are true in c' then $\varphi \rightarrow \psi$ and ψ should have the same truth values in c'. So if we assume in addition—as we shall have to so as to arrive at the intended account of the paradox—that $k \sim k{+}1$, then $P(\overline{k+1})$ will be true in c' and so $P(\bar{k}) \rightarrow P(\overline{k+1})$ should be true there as well. But just as it would be wrong to allow $\varphi \rightarrow \psi$ to have a different truth value from that of ψ in any context c' such that $B(c') \vdash \varphi$ and φ are true in c' so it would be equally implausible to assume that where φ is true in c' and $B(c')$ entails φ incorporation of $\varphi \rightarrow \psi$ into c' would yield a different new context than incorporation of ψ. We had already assumed that $m(c, P(\overline{k+1}))$ is not coherent. So by the principle just enunciated $m(c, P(\bar{k}) \rightarrow P(\overline{k+1}))$ is not coherent. c therefore is a context in which $P(\bar{k}) \rightarrow P(\overline{k+1})$ is to be true but not coherently incorporable.

The clauses I have chosen for negation reflect the simple intuition that $\neg\varphi$ is true when φ is false and false when φ is true. Here too a special condition concerning the incorporation of $\neg\varphi$ does not seem called for.

At last we are ready for the truth definition itself.

Definition 3 Let an ('entailment'-) relation $\vdash$ and a context-sensitive model M relative to $\vdash$ be given. The *truth value of* a sentence φ *in* M *in* a context $c \in C_M$ *relative to* $\vdash$, $[\varphi]^{\vdash}_{M,c}$, is determined by the following clauses:

(i) $[P(c_i)]^{\vdash}_{M,c} = 1$ iff $(\exists b \in U_M)(b{\sim}F_M(c_i)\ \&\ P(\bar{b}) \in B_M(c))$;
(ii) $[P(c_i)]^{\vdash}_{M,c} = 0$ iff $(\exists b \in U_M)(b{\sim}F_M(c_i)\ \&\ \neg P(\bar{b}) \in B_M(c))$;
(iii) $[\neg\varphi]^{\vdash}_{M,c} = 1$ iff $[\varphi]^{\vdash}_{M,c} = 0$;
(iv) $[\neg\varphi]^{\vdash}_{M,c} = 0$ iff $[\varphi]^{\vdash}_{M,c} = 1$;
(v) $[\varphi \rightarrow \psi]^{\vdash}_{M,c} = 1$ iff $[\varphi]^{\vdash}_{M,c} = 0$ or $[\psi]^{\vdash}_{M,m(c,\varphi)} = 1$;
(vi) $[\varphi \rightarrow \psi]^{\vdash}_{M,c} = 0$ iff $[\varphi]^{\vdash}_{M,c} = 1$ and $[\psi]^{\vdash}_{M,m(c,\varphi)} = 0$;

(vii) $[(\forall \nu_i)\varphi]^{\vdash}_{M,c} = 1$ iff $m_M(c, (\forall \nu_i)\varphi) \in \mathrm{Coh}_M$ & $(\forall a \in U_M)[\varphi^{\bar{a}}_{\nu_i}]^{\vdash}_{M,c} = 1$;
(viii) $[(\forall \nu_i)\varphi]^{\vdash}_{M,c} = 0$ iff $m_M(c, (\forall \nu_i)\varphi) \in \mathrm{Inc}_M$ or $(\exists a \in U_M)[\varphi^{\bar{a}}_{\nu_i}]^{\vdash}_{M,c} = 0$.

Note that in phrasing the clauses for conditional and universal sentences I have stayed as close to the classical definitions of $\rightarrow$ and $\forall$ as is compatible with the specific requirements we imposed in the preceding sections.

14.

To see some of the implications of these definitions for the particular uses of vague predicates which motivated them consider the following example.

Let $\vdash_c$ be the consequence relation of classical logic. Let r, n_0, n_1, n_2, n_3 be positive integers such that for $i = 0, 1, 2, n_{i+1} - n_i > 2r$. Let M_0 be the context-sensitive model $\langle U_0, F_0, C_0, \mathrm{Coh}_0, \mathrm{Inc}_0, B_0, m_0, \sim_0\rangle$ relative to $\vdash_c$, such that

(i) $U_0 = \{1, 2, 3, \ldots\}$;
(ii) $F_0^+(P) = \{1, \ldots, n_0\}$;
(iii) $F_0^-(P) = \{n_3, \ldots\}$;
(iv) $C_0 = \{Cl(S') : D_M \subseteq S' \subseteq S_0\}$, where $Cl(S')$ is the closure of S' under $\vdash_c$;
(v) $\mathrm{Coh}_0 = \{S' \subseteq C_0 : (\forall n)((P(\bar{n}) \in S' \rightarrow n < n_1) \,\&\, (\neg P(\bar{n}) \in S' \rightarrow n_2 < n))\}$;
(vi) $\mathrm{Inc}_0 = \{S' \subseteq C_0 : (\exists n, n')(n' - n \leq r \,\&\, P(\bar{n}) \in S' \,\&\, \neg P(\bar{n}') \in S')\}$;
(vii) for all $c \in C_0, B_0(c) = c$;
(viii) for all $c \in C_0$ and $\Gamma \subseteq S_0, m_0(c, \Gamma) = Cl(c \cup \Gamma)$;
(ix) for all positive integers $n, n', n \sim_0 n'$ iff $|n - n'| \leq r$.

Let c_0 be the set $Cl(D_{M_0})$, (c_0 is what one might call a 'minimal' context of M_0, the context in which the minimum of commitments have been made as regards the extensions of P). For this context we have

(a) $[P(\bar{k}) \rightarrow P(\overline{k+1})]^{\vdash_c}_{M_0,c_0} = 1$ for each positive integer k;
(b) $[(\forall n)(P(n) \rightarrow P(n+1))]^{\vdash_c}_{M_0,c_0} = 0$ (as $Cl(c_0 \cup \{(\forall n)(P(n) \rightarrow P(n+1))\})$ belongs to Inc_0);
(c) $[P(\bar{k})]^{\vdash_c}_{M_0,c_0} = 1$ iff $k \leq n_0 + r$;
(d) $[\neg P(\bar{k})]^{\vdash_c}_{M_0,c_0} = 1$ iff $k \geq n_3 - r$.

That (a)–(d) are true is some indication that our semantics captures at least part of the intuitions which we are attempting to formalize.

15.

As usual we can define validity by quantifying over those parameters on which the truth value of a sentence depends—here, models and contexts.

Before putting down a definition along these lines we should remind ourselves, however, that the notions 'true' and 'false', as we have specified them, can be expected to behave reasonably only with respect to contexts which are coherent. It is advisable therefore to take only such contexts into account. Thus we come to:

Definition 4 A sentence φ of L_0 is a *logical consequence of* a set of sentences Γ, *relative to* $\vdash$, in symbols $\Gamma \models(\vdash)\varphi$, iff for every context-sensitive model M relative to $\vdash$, and every $c \in \mathrm{Coh}_M$, if for all $\psi \in \Gamma$ $[\psi]^{\vdash}_{M,c} = 1$ then $[\varphi]^{\vdash}_{M,c} = 1$.

Within the present framework this is however not the only possible way of characterizing validity. It might be argued that the statement that "φ can be validly inferred from Γ" is not really captured by Definition 4 but should be interpreted, rather, as the claim that φ is true in any context in which Γ *has already been accepted*. When formally implementing this interpretation we should again make sure to involve coherent contexts only. Thus we are led to the following definition of a relation to which I shall refer as "φ can be validly inferred from Γ" and which I shall symbolize as $\models\!\!\models$:

Definition 5 A sentence φ of L_0 *can be validly inferred from* a set of sentences Γ *relative to* $\vdash$, in symbols $\Gamma \models\!\!\models(\vdash) \varphi$, iff for every context-sensitive model M relative to $\vdash$ and every $c \in C_M$, if $m(c, \Gamma) \in \mathrm{Coh}$ then $[\varphi]^{\vdash}_{M,m(c,\Gamma)} = 1$.[12]

Having defined these relations we are bound to ask the familiar completeness questions. Are they axiomatizable; and what would an axiomatization look like? As the matter stands, however, these questions are indefinite. Before we can make any attempt at answering them we must first fix the extension of the relation $\vdash$; the problem is: which extension should we choose?

In a way this is a kind of completeness question too. For ideally $\vdash$ should be chosen in such a way that it coincides with the semantic consequence relation which it generates. Let us suppose that it is $\models\!\!\models$, rather than $\models$, which captures the intuitive concept of valid inference. Then $\vdash$ ought to be such that for all Γ, φ $\Gamma \vdash \varphi$ if and only if $\Gamma \models\!\!\models(\vdash) \varphi$. We can recast this requirement as follows. Both $\vdash$ and $\models\!\!\models$ denote subsets of the set $\mathfrak{B}(S_0) \otimes S_0$. $\models\!\!\models$, moreover, is determined by $\vdash$, i.e. we can think of $\models\!\!\models$ as a function which maps certain subsets of $\mathfrak{B}(S_0) \otimes S_0$ to such subsets. Our problem is to find a fixed point of this function. Let us note the particular ways in which $\vdash$ influences the extension of $\models\!\!\models$. In the first place the extension of $\vdash$ may affect the size of

[12] I shall omit henceforth the subscript 'M' from both 'm_M' and 'B_M'.

the class of context-sensitive models. Inspection of the definition of that class shows that where $\vdash_1$ and $\vdash_2$ are two relations such that $\vdash_1 \subseteq \vdash_2$ the class of context-sensitive models relative to $\vdash_2$ is included in the class of context-sensitive models relative to $\vdash_1$. In the second place, $\vdash$ enters into the truth definition via the clauses which concern atomic formulae. Here it is clear that if $\vdash_1 \subseteq \vdash_2$ and M is any context-sensitive model relative to $\vdash_2$ then for any sentence φ and $c \in \mathrm{Coh}_M$, if $[\varphi]^{\vdash_1}_{M,c} = \{^1_0$ then $[\varphi]^{\vdash_2}_{M,c} = \{^1_0$.

It is easy to infer from these two observations that if $\vdash_1 \subseteq \vdash_2$ then $\models(\vdash_1) \subseteq \models(\vdash_2)$. Thus $\models$ is a monotone function from some power set $\mathfrak{B}(X)$ into $\mathfrak{B}(X)$ and so it has a fixed point. But what might these fixed points be like? This is an issue on which I have little to say at present. The problem appears to be technically non-trivial and to have many ramifications. I expect that there may be a great many fixed points and that the logics represented by most of them will be quite unreasonable.

One way to secure fixed points which satisfy certain predetermined constraints is the following. Suppose C is a condition such that (i) whenever $\vdash$ satisfies it, then so does $\models(\vdash)$; (ii) whenever E is a set of relations satisfying C then $\bigcap E$ satisfies C; (iii) whenever E is a $\subseteq$-chain of relations satisfying C then $\bigcup E$ satisfies C. Then we can first form the intersection $\vdash_0$ of all relations which satisfy C. By (ii) this relation satisfies C. According to (i) $\models(\vdash_0)$ will satisfy C too and so $\vdash_0 \subseteq \models(\vdash_0)$. From this it follows that the union of the relations

$$\models(\vdash_0), \models(\models(\vdash_0)), \models(\models(\models(\vdash_0))), \ldots$$

is a fixed point. By assumption (iii) this union satisfies C.

Among the conditions for which (i), (ii) and (iii) hold there are in particular those of the form $R \subseteq \vdash$, where R is some fixed relation.[c] Such conditions

[c] The considerations from here to the bottom of p. 303 (paragraph ending with '... will also be validated by $\models$') seem strangely at odds with the general worry of this and the following sections, viz. that the emerging fixed point logics may turn out to be so weak as to be of little conceptual interest. It is argued below that quite strong closure conditions can be imposed on $\vdash$, and in such a way that they are passed on from $\vdash$ to $\models(\vdash)$. This entails that there will be a fixed point (of the function $\models(.)$) that also has those closure properties. One of the closure properties in question is closure of Modus Ponens, the inference principle that (as emphasised in the 2013 Postscript below) is the principal culprit in the construction of paradoxical sorites arguments and thus a first target for revision (by restricting its applications or abandoning it altogether). So the question is: How can this all fit together?

In a nutshell the answer is this. It *is* true that we can obtain fixed point logics that are quite strong. But do we want them? As noted in the text, the stronger the relation $\vdash$ the more difficult it is for a structure to qualify as a model relative to $\vdash$. (Stronger $\vdash$'s lead to smaller

are of particular interest where R is what we might call an 'inference rule', i.e. the set of all pairs $\langle \Gamma, \varphi \rangle$ which correspond to some inference schema. Rather than make the general notion of 'inference rule' I have in mind precise, let me just present the particular instances which I wish to consider here. The inference schema which we know as Modus Ponens can be given in the form

$$S_1 : \frac{\psi, \psi \rightarrow \varphi}{\varphi}$$

To this schema corresponds the inference rule R_1 consisting of all pairs $\langle \Gamma, \varphi \rangle$ where for some ψ both ψ and $\psi \rightarrow \varphi$ belong to Γ for some ψ. Similarly to the schema

$$S_2 : \frac{\neg(\varphi \rightarrow \psi)}{\varphi}$$

corresponds the relation R_2 whose members are the pairs $\langle \Gamma, \varphi \rangle$ such that for some ψ, $\neg(\varphi \rightarrow \psi)$ is in Γ for some ψ. In this fashion we can also form relations R_3 and R_4 corresponding to the schemata

$$S_3 : \frac{\neg(\psi \rightarrow \varphi)}{\neg\varphi} \text{ and } S_4 : \frac{(\forall \nu_i)\varphi}{\varphi^{c_i}_{\nu_i}}$$

---

classes of compatible models.) In this way we may well lose some of the models we might want to include in our model class. And furthermore, some of the models that do survive the test imposed by strong input inference relations may display a kind of instability that is a direct reflection of the paradoxical sorites arguments. The model M_0 of section 14 is a case in point. Here the input inference relation is assumed to be classical logic (no less!). This has the effect that even the minimal context c_0 for which $B(c_0)$ is the classical closure of the diagram of P is not *sound* in that it isn't possible to incorporate into $B(c_0)$ every sentence that is true in M_0 in c_0 without turning c_0 into an Incoherent context. (The resulting context would obviously be Incoherent, since closure of this extended background under classical logic, and thus in particular under MP, would put $P(\overline{n_3})$ into its background alongside with $\neg P(\overline{n_3})$.)

Classical logic is of course an input inference relation that will never be returned as consequence relation on either of the two main logical consequence definitions presented in the paper (Def. 4 and Def. 5). This follows from the observation in section 17, amplified in the new Postscript, that the model classes that Def. 2 associates with different input inference relations $\vdash$ will always contain models that interpret P as a vague but context-insensitive predicate. The point made about such models in the new Postscript entails that when the input relation $\vdash$ is the inference of classical logic the consequence relation generated from it in the manner the paper describes will 'drop below' the Strong Kleene logic K3. On the other hand, K3 does verify MP. And it is one of the corollaries of the argument below that the input inference relation that consists just of the rule MP will yield a fixed point logic that verifies MP. For the reasons emphasised in the new Postscript that is not going to be an acceptable logic since it won't block the multi-sentence sorites paradox. It will be unacceptable not because it is too weak but because it is too strong.

Let $R_0 = \bigcup_{i=1}^{4} R_i$ and let $C_0(\vdash)$ be the condition: $R_0 \subseteq \vdash$. Then C_0 satisfies all the assumptions we need to make about a condition C to prove the existence of a fixed point satisfying C. Indeed it is obvious that any condition of the form $R \subseteq \vdash$ satisfies (ii) and (iii).

To show that $\models\!\!\models(\vdash)$ satisfies C_0 provided $\vdash$ does is a little more complicated. We first show that if $\vdash$ satisfies C_0, M is a context-sensitive model relative to $\vdash$, and if φ is a sentence of L_0 then for all $c \in \mathrm{Coh}_M$

(1) If $\varphi \in B(c)$ then $[\varphi]^{\vdash}_{M,c} = 1$, and
(2) If $\neg\varphi \in B(c)$ then $[\varphi]^{\vdash}_{M,c} = 0$.

The proof is by induction on the complexity of φ. Evidently (1) and (2) are satisfied for atomic φ (see the truth def.), and if they are satisfied for φ they are also satisfied for $\neg\varphi$. Suppose that φ has the form $(\forall\nu_i)\psi(\nu_i)$, that $c \in \mathrm{Coh}_M$ and that $\varphi \in B(c)$. $B(c) \vdash \varphi$ since $\vdash$ satisfies C_0, and so by Def. 1 (ix) and Def. 2 (iii) $m(c,\varphi) \in \mathrm{Coh}_M$. Also, since $R_4 \subseteq \vdash$, $B(c) \vdash \psi(\bar{a})$ for all $a \in U_M$. So by induction hypothesis $[\psi(\bar{a})]^{\vdash}_{M,c} = 1$ for each $a \in U_M$. It follows that $[(\forall\nu_i)\psi(\nu_i)]^{\vdash}_{M,c} = 1$. Now suppose that $\neg\varphi \in B(c)$. Then $\varphi, \neg\varphi \in B(m(c,\{\varphi\}))$. From this it follows, in virtue of Definition 2 (ii), that $m(c,\{\varphi\}) \in \mathrm{Inc}_M$. So $[\varphi]^{\vdash}_{M,c} = 0$.

Next assume that φ has the form $\psi \to \chi$. First let us suppose that $\varphi \in B(c)$. Then $\psi, \psi \to \chi \in B(m(c,\{\psi\}))$. So, because $R_1 \subseteq \vdash \chi \in B(m(c,\{\psi\}))$ and so on the strength of the induction hypothesis $[\chi]^{\vdash}_{M,m(c,\{\psi\})} = 1$. So $[\varphi]_{M,c} = 1$. Finally suppose $\neg\varphi \in B(c)$. Then $\psi \in B(c)$ since $R_2 \subseteq \vdash$, and so by induction hypothesis $[\psi]^{\vdash}_{M,c} = 1$. Also $B(m(c,\{\psi\})) = B(c)$ by Def. 2 (iii) and Def. 1 (ix). Since $R_3 \subseteq \vdash$, $\neg\chi \in B(c) = B(m(c,\{\psi\}))$. So by induction hypothesis $[\chi]^{\vdash}_{M,m(c,\{\psi\})} = 0$. So $[\varphi]^{\vdash}_{m,c} = 0$.

To establish that $\models\!\!\models(\vdash)$ satisfies C_0 we must show that $\models\!\!\models(\vdash)$ contains each of the relations R_1–R_4. Consider e.g. R_1. To show that $R_1 \subseteq \models\!\!\models$ we must show that for all M and $c \in C_M$, if, putting $m(c,\{\psi,\psi\to\varphi\}) = c'$, $c' \in \mathrm{Coh}_M$ then $[\varphi]^{\vdash}_{M,c'} = 1$. But this now follows from the fact that since $R_1 \subseteq \vdash$ $B(c') \vdash \varphi$, so that, by Def. 2 (iii), $\varphi \in B(c')$; and then, by the result just proved, $[\varphi]^{\vdash}_{M,c'} = 1$. Containment by $\models\!\!\models$ of the remaining relations R_i is proved in the same way.

Note that if we assume that $\vdash$ satisfies certain additional schemata such as, for instance

$$S_5 : \frac{\neg\neg\varphi}{\varphi} \text{ or } S_6 : \frac{\psi, \neg\psi}{\varphi}$$

these schemata will also be validated by $\models\!\!\models$.

While these last considerations suggest that fixed points exist which have at least some reasonable deductive properties, they by no means guarantee

that we might find a fixed point that really deserves to be called a deduction relation. Thus for instance I haven't even managed to determine yet whether there is a fixed point satisfying the conditions[d]

(d_0) if $\varphi \in \Gamma$ then $\Gamma \vdash \varphi$,
(d_1) if $\Gamma \subseteq \Delta$ and $\Gamma \vdash \varphi$ then $\Delta \vdash \varphi$, and
(d_2) if $\Gamma \vdash \varphi$ and $\Delta \cup \{\varphi\} \vdash \psi$ then $\Gamma \cup \Delta \vdash \psi$,

which may well be thought to be essential to any concept of deduction.

We can also pose the fixed point problem in connection with the consequence relation $\models(\vdash)$. Here we encounter the difficulty that there is apparently no simple way of showing that $\models$ is a monotone function. The reason for this is that the valuation function [ ], which *is* monotone in $\vdash$, figures not only in the consequent of the definiens of $\models$ but also in its antecedent. This negative occurrence of [ ] prevents us from concluding that $\models$ is monotone as well. There is however the possibility of slightly modifying the truth definition in a way which restores monotonocity of $\models$, and which is not altogether indefensible from a conceptual point of view. Suppose we rephrase the truth clauses for the atomic sentences in such a way that they provide sufficient but not necessary conditions for the truth or falsity of atomic sentences—i.e. replace the *iff* of (i) and (ii) of Definition 3 by *if*. We thereby associate with any given model not just a single valuation function, but a collection of such functions, each compatible with the clauses of the modified definition of truth.

We might defend this alteration as expressing the intuitive idea that factors other than those we have explicitly captured in our formalization might determine the truth values of atomic sentences. However, if we are to modify the truth definition in this way, we must consider the question whether we should not impose further restrictions on the collection of valuation functions than are provided by the modified truth definition itself. We should certainly want it to remain true that in any coherent context an atomic sentence cannot be both true and false. So, unless there is some other

[d] The author's declaration that he 'hasn't even managed to determine yet whether there is a fixed point satisfying the conditions' (d_0)–(d_2) that are given on the next three lines, if the text should be read in the context of the unfounded worry that adoption of Def. 4 might lead to non-monotonic mappings from input inference relations to output logical consequence relations. Once this worry has been set aside—as the new Postscript says it can and should—then the quoted confession of inability no longer makes sense. For it is obvious that (d_0)–(d_2) must hold for any logical consequence relation that results from applying Def. 4.

means of insuring this we must supplement the new truth 'definition' with the clause

(3. ix) if $[P(\bar{a})]_{M,c} = 1$ and $[P(\bar{a})]_{M,c} = 0$ then $c \in \mathrm{Inc}_M$.

Let us for given $\vdash$ and M denote the set of valuations compatible with the revised definition as $V(\vdash, M)$. Then, clearly, if $\vdash_1 \subseteq \vdash_2$ then $V(\vdash_2, M) \subseteq V(\vdash_1, M)$.

Our modification of the truth definition requires a slight alteration of the definition of $\models$, which should now read

Definition 6 $\Gamma \models'(\vdash)\varphi$ iff for all context-sensitive models M relative to $\vdash$, all $c \in \mathrm{Coh}_M$ and all valuations $[\ ] \in V(\vdash, M)$, if for all $\psi \in \Gamma$ $[\psi]^{\vdash}_{M,c} = 1$ then $[\varphi]^{\vdash}_{M,c} = 1$.

It must be evident that $\models'(\vdash_1) \subseteq \models'(\vdash_2)$ whenever $\vdash_1 \subseteq \vdash_2$. So $\models'$ will have a fixed point. It is easy, moreover, to show that $\models'$ (as well as the earlier relation $\models$) satisfies the conditions (d_0)–(d_2) which, as I said, any deduction relation worthy of the name must honour. It also automatically verifies a third, equally essential, condition, viz. that $\Gamma \vdash \varphi$ whenever $\varphi \in \Gamma$. On the other hand it is worth noting that Modus Ponens for instance cannot be a valid inference principle in the sense of $\models'$. This is a direct consequence of our semantics, which was designed to allow both $P(\overline{k+1})$ and $P(\overline{k+1}) \rightarrow P(\overline{k+2})$ to be true in certain contexts in which $P(\overline{k+2})$ is not.[13]

16.

There is an objection to my characterization of the relation $\models\!\models$ which the reader may have realized as soon as he became acquainted with Definition 5. If it has been a shortcoming of my presentation not to have drawn attention to it earlier let me make up for the omission now. The problem with Definition 5 is that it demands that φ be true in $m(c, \Gamma)$ irrespective of whether the members of Γ are themselves true in that context. Nothing we have said, however, precludes the possibility of a context c such that $\varphi \in B(c)$ while yet $[\varphi]^{\vdash}_{M,c} \neq 1$. Indeed, one might well suppose that such contexts must exist, in so far as the participants in a discourse sometimes accept, and thus incorporate into the stock of shared background assumptions, sentences which are in fact false—and which become no truer just by being adopted as true. Note however that contexts of this sort do not arise in the particular

[13] I have also not verified, but suspect, that some of the other schemata S_i discussed above will also fail for $\models$ and $\models'$.

analysis of the paradox of the heap for which we are trying to provide a suitable formal framework. Indeed, as the example of section 14 illustrates, the contexts which represent the various stages of discussions such as those described in section 7, all develop out of a 'minimal' context (i.e. a context c such that for every other context $c'B(c) \subseteq B(c')$) through successive incorporations of sentences that are true in the contexts into which they are being incorporated. Even of such contexts c it isn't evident that each sentence entailed by $B(c)$ is true in c. It does appear reasonable however to regard only contexts for which this additional condition is fulfilled as sufficiently proper to merit consideration in connection with valid inference. There is a technical difficulty we must face when implementing this condition, which is caused by the circumstance that the condition can be stated only with reference to truth and, therefore, only after the truth definition has already been given.

Let us call contexts which, besides being coherent, satisfy the condition *sound*.

Definition 7 Let $\vdash$ and a context-sensitive model M relative to $\vdash$ be given. A member c of C_M is called *sound* iff (a) $c \in \mathrm{Coh}_M$ and (b) for all $\varphi \in S_0$ if $B(c) \vdash \varphi$ then $[\varphi]^{\vdash}_{M,c} = 1$.

We can now replace Definition 5 by

Definition 8 Let $\vdash$ be as in Definition 7. For any subset Γ of S_0 and $\varphi \in S_0$ $\Gamma \models'(\vdash)\varphi$ if for every context-sensitive model M relative to $\vdash$ and every $c \in C_M$ if $m(c, \Gamma)$ is sound then $[\varphi]^{\vdash}_{M,m(c,\Gamma)} = 1$.

Alas, this new definition does not guarantee that the relation it defines is a monotone function of $\vdash$. For the truth definition now enters, via the concept of soundness, into the characterization also in a negative way. So we cannot conclude that $\models'$ has any fixed points. If, however, we modify the truth definition as described in the previous section, this problem will disappear in just the way we succeeded in removing the analogous difficulty that arose in connection with $\vDash$.

Since by adopting Definition 7 instead of Definition 5 we can only narrow the class of relevant contexts, the earlier argument that if $R_0 \subseteq \vdash$ then $R_0 \subseteq \models(\vdash)$ also shows that $R_0 \subseteq \models'(\vdash)$ if $R_0 \subseteq \vdash$. Note that $\models'$ also satisfies conditions (d_0)–(d_2) provided $\vdash$ does. (d_0) and (d_1) are easy. As for (d_2) suppose that $\Gamma \models'(\vdash)\varphi$ and $\Delta \cup \{\varphi\} \models'(\vdash)\psi$; in other words: for every context-sensitive M relative to $\vdash$

(i) for every $c \in C_M$ such that $m(c, \Gamma)$ is sound $[\varphi]^{\vdash}_{M,m(c,\Gamma)} = 1$; and
(ii) for every $c \in C_M$ such that $m(c, \Delta \cup \{\varphi\})$ is sound $[\psi]^{\vdash}_{M,m(c,\Delta\cup\{\varphi\})} = 1$.

Let c be such that $m(c, \Gamma \cup \Delta)$ is sound. Then, since by assumption $\vdash$ satisfies (d_1), $B(m(c, \Gamma \cup \Delta)) \vdash \varphi$. So, by (iii) of Definition 2, $\varphi \in B(m(c, \Gamma \cup \Delta))$. Also $\Delta \subseteq B(m(c, \Gamma \cup \Delta))$. By (ix) of Definition 1 $m(m(c, \Gamma \cup \Delta), \Delta \cup \{\varphi\}) = m(c, \Gamma \cup \Delta)$. So there is a context c', viz. $m(c, \Gamma \cup \Delta)$, such that $m(c, \Gamma \cup \Delta) = m(c', \Delta \cup \{\varphi\})$. It follows from (ii) above that, since $m(c', \Delta \cup \{\varphi\})$ is sound, $[\psi]^{\vdash}_{M, m(c, \Gamma \cup \Delta)} = 1$.

From what has been said here it would appear that of all the possible inference relations we have considered $\models'$ is likely to prove the most satisfactory, both as regards its conceptual motivation and for the logic it is capable of generating. But this depends, for all that has been established here, on the acceptability of the modified truth definition of section 15.

To conclude this section let us note that various other alternatives to Definition 7 could be considered too. Thus we may require for $\Gamma \models \varphi$ only that $[\varphi]_{M, m(c, \Gamma)} = 1$ just for those contexts c such that (i) $m(c, \Gamma)$ is coherent (or sound) and (ii) c itself is coherent and/or all members of Γ are true in c. It is quite possible that, either on our present definitions of *model* and *true* or after some further sensible modifications of these definitions, two or more of these alternative ways of characterizing valid inference coincide. This too, however, is a matter for further exploration.

17.

Intuitively the peculiar properties of our semantics are consequent upon the particular kind of context-sensitivity of the predicate P which we tried to capture through the clauses (i) and (ii) of Definition 3. This observation kindles the expectation that were P replaced by some sharp predicate Q the truth definition should reduce to the familiar classical definition.

Let L_1 be the language which is like L_0 except for having the 1-place predicate Q instead of P. Let $\vdash_c$ be once more the consequence relation of classical predicate logic and let M be a context-sensitive model relative to $\vdash_c$, such that $F^+_M(Q) \cup F^-_M(Q) = U_M$. Let us replace, moreover, in the truth definition of section 13, the clauses (i) and (ii) by the new clauses

(i′) $[Q(\bar{a})]_{M,c} = 1$ iff $a \in F^+_M(Q)$, and
(ii′) $[Q(\bar{a})]_{M,c} = 0$ iff $a \in F^-_M(Q)$.

These clauses express the intuition that the extension of Q is not dependent on the context in which Q is being evaluated. One implication of this is that no context in a model for L_1 is incoherent. For incoherence, in the sense in which we intend the term in this essay, arises there where a context forces upon a predicate an extensional inconsistency. But with regard to

the predicate Q no such inconsistency can come about through contextual modification. We may assume therefore that $\mathrm{Coh}_M = C_M$ (and consequently that $\mathrm{Inc}_M = \emptyset$).[14] Under these assumptions on M the truth definition, with its new clauses (i′) and (ii′), coincides with the classical characterization.

Various other assumptions will also reduce the truth definition to one of simpler or more familiar form, even though the final product will not be the classical notion of truth.

A partial simplification arises if we assume that Q is vague, but not dependent on context in the manner of (i) and (ii). Formally this means that we should always have that $\mathrm{Coh}_M = C_M$ and that the truth values of atomic formulae involving Q are given by (i′) and (ii′); but we do not assume that $F^+_M(Q) \cup F^-_M(Q) = U_M$. Another simplifying assumption concerns the relation $\sim$. There may well be predicates which while in principle subject to EOI nonetheless happen to have a meaning such that any two distinct objects can be recognized as distinct in a respect that is relevant to Q. Formally this amounts to the assumption that $\sim_M$ is the identity relation on U_M. Models for which this assumption holds differ from those considered in the preceding pragraph in that they do allow for contextual modification of the extension of Q, although they will not produce such modifications merely through contextual incorporation of what is already true. How the possible logics generated under such special assumptions differ from those which are compatible with the general semantics developed here is again an issue which further research would have to clarify.

18.

So far we have paid attention to only some of the logical constants which standard developments of symbolic logic have studied. Of the remaining ones, $\&$, $\vee$, $\exists$ and $\leftrightarrow$, the last one will, on any resonable interpretation, be determined by what we say about the others. For I take it that $\varphi \leftrightarrow \psi$ should, as the phrase 'if and only if' implies, always be analysed as the conjunction of the conditionals $\varphi \rightarrow \psi$ and $\psi \rightarrow \varphi$. $\&$ poses, within our framework, much the same problems as $\forall$: One reason why a conjunction such as

[14] Alternatively we could, in an effort to maintain the letter, rather than the spirit, of Definitions 1 and 2, define Coh_M as the set of contexts c such that $B(c)$ is consistent and Inc_M as consisting of the remaining c's. The truth definition will still coincide with the classical definition for all contexts which are sound.

$$(P(\bar{n}) \to P(\overline{n+1})) \,\&\, (P(\overline{n+1}) \to P(\overline{n+2})) \,\&\, \ldots$$
$$\&(P(\overline{n+k}) \to P(\overline{n+k+1}))$$

will in general not be true in the models which reflect such cases as we considered, e.g. in Section 7, is that we cannot always incorporate all the conjuncts simultaneously without producing a context that is incoherent. This point is more easily appreciated in connection with long conjunctions such as the one displayed; but little reflection should suffice to see that it must apply to binary conjunctions as well. If we wish to maintain the natural connections between conjunction and universal quantification we are naturally led to the following truth clauses for $\&$:

(&i) $[\varphi \,\&\, \psi]^{\vdash}_{M,c} = 1$ iff $m(c, \varphi \,\&\, \psi) \in \mathrm{Coh}_M$ and $[\varphi]^{\vdash}_{M,c} = [\psi]^{\vdash}_{M,c} = 1$.
(&ii) $[\varphi \,\&\, \psi]^{\vdash}_{M,c} = 0$ iff either $m(c, \varphi \,\&\, \psi) \in \mathrm{Inc}_M$ or $[\varphi]^{\vdash}_{M,c} = 0$ or $[\psi]^{\vdash}_{M,c} = 0$.

There is an alternative to this characterization of *and* which is sugested by the serial character of conjunctions. It might be held that it is sufficient for $\varphi\&\psi$ to be true in c that φ be true in c and ψ be true in $m(c, \varphi)$. This conception of conjunction—which in natural language is perhaps less ambiguously expressed by the locution 'φ, and moreover ψ'—is captured by the following two truth clauses:

(&̇i) $[\varphi \,\dot{\&}\, \psi]^{\vdash}_{M,c} = 1$ iff $m(c, \varphi \,\dot{\&}\, \psi) \in \mathrm{Coh}_M$ and $[\varphi]^{\vdash}_{M,c} = 1$ and $[\psi]^{\vdash}_{M,m(c,\{\varphi\})} = 1$;
(&̇ii) $[\varphi \,\dot{\&}\, \psi]^{\vdash}_{M,c} = 0$ iff either $m(c, \varphi \,\dot{\&}\, \psi) \in \mathrm{Inc}_M$ or $[\varphi]^{\vdash}_{M,c} = 0$ or $[\psi]^{\vdash}_{M,m(c,\{\varphi\})} = 0$;

Disjunction can, it seems, be treated along the lines of our account of $\to$:

($\dot{\vee}$i) $[\varphi \,\dot{\vee}\, \psi]^{\vdash}_{M,c} = 1$ iff $[\varphi]^{\vdash}_{M,c} = 1$ or $[\psi]^{\vdash}_{M,m(c,\{\neg\varphi\})} = 1$;
($\dot{\vee}$ii) $[\varphi \,\dot{\vee}\, \psi]^{\vdash}_{M,c} = 0$ iff $[\varphi]^{\vdash}_{M,c} = 0$ and $[\psi]^{\vdash}_{M,m(c,\{\neg\varphi\})} = 0$.

But here, as in the case of *and*, we cannot dismiss out of hand the more symmetric alternative provided by the clauses:

($\vee$i) $[\varphi \vee \psi]^{\vdash}_{M,c} = 1$ iff $[\varphi]^{\vdash}_{M,c} = 1$ or $[\psi]^{\vdash}_{M,c} = 1$;
($\vee$ii) $[\varphi \vee \psi]^{\vdash}_{M,c} = 0$ iff $[\varphi]^{\vdash}_{M,c} = 0$ and $[\psi]^{\vdash}_{M,c} = 0$.

As for the existential quantifier, there is only one reading for it that strikes me as natural within the present setting. It is given by the following two clauses:

($\exists$i) $[(\exists\nu_i)\varphi^{\vdash}(\nu_i)]_{M,c} = 1$ iff for some $a \in U_M$ $[\varphi^{\vdash}(\bar{a})]_{M,c} = 1$;
($\exists$ii) $[(\exists\nu_i)\varphi^{\vdash}(\nu_i)]_{M,c} = 0$ iff for every $a \in U_M$ $[\varphi^{\vdash}(\bar{a})]_{M,c} = 0$.

Note that according to these clauses the formula $(\exists n)(P(n) \,\&\, \neg P(n+1))$ comes out false in the model M_0 of section 14 in the minimal context c_0, even

though $(\forall n)(P(n) \rightarrow P(n+1))$ is false in M_0 in c_0 also. Thus the present theory is proof against Sanford's damning criticism of the supervaluation account.

19.

We have seen how we might give an explanation of the invalidity of the argument I of section 1 and yet preserve the inference schemata S_1–S_6 which I listed in section 15, in particular the schema of Modus Ponens. In this respect our theory agrees with the Fine-Thomason account, rather than with the many-valued treatment—according to which MP may yield conclusions whose values are below those of both of the corresponding premises, and thus is not a valid principle in the strict sense. The supervaluation account, however, manages to hold on to MP only at the cost of branding the conditionals $P(\bar{k}) \rightarrow P(\overline{k+1})$, which play such a central part in the versions of the paradox we have here considered, as being less than perfectly true. The present theory indicates how one could avoid paying that price. But, as so many questions are still waiting to be answered, it is too early to assess whether this can be achieved without incurring even greater expense elsewhere.

I cannot emphasize enough in this connection that virtually all metamathematical work in the area of which this paper has tried to produce a first, sadly incomplete chart, is yet to be done. The scope for such work appears, from the modest efforts I have so far made in this direction, almost unlimited; yet anyone who launches himself into the labyrinth of possibilities easily succumbs to doubts whether the work deserves to be done at all. For it appears less than likely that any reasonable logic could be found there.

Understandable though such defeatism may be, it is, I believe, ultimately without justification. This is beause a 'negative' result—i.e. a proof that a semantic theory of the sort I have sketched *cannot* generate a reasonable logic—is as important as a 'positive' result, which would consist in finding a logic that does not preposterously diverge from the canons of inference on which we rely in most of our scientific and non-scientific reasoning. To see the importance of the negative result, let us, once more, ask what precisely it is about the paradox that is so disturbing. In the preceding sections we found that the semantic principles which seem to lie at the root of the paradox can be combined in a formally consistent manner. However, the success of this formal exercise must be measured, in part at least, in terms of the distance between the concept of validity which it yields and the classical

concept that we know to be adequate in connection with predicates of a less troublesome nature than those which have occupied us here. If we could establish that this distance cannot be reduced beyond a certain substantial minimum we would thereby confirm the depth of the problem which the paradox pinpoints.

I should hasten to add that the predicament cannot be a question only of how small this distance can be made. Even if we found the classical consequence relation itself to be a fixed point, this by itself would do nothing to reduce the apparent contradictions to which a person can be reduced in the concrete instantiations of the paradox which we discussed earlier. Let us consider once again the experiment with the screen, supposing that you, the subject of the experiment, have had the benefit of the present essay. You just qualified the k-th square as green, and now you refuse to apply the predicate to the $k + 1$-st square. When I challenge this seemingly inchoate succession of verbal acts you retort that on acceptance of the statement 'the k-th square is green' the context ceased to be coherent, so that there was no longer a sound basis for making judgements involving the predicate 'green'. I would then ask: "But why did the context cease to be coherent just *now*?" You might reply that contextual coherence is not the kind of concept that is ruled by the principle EOI. After all the notion of a 'context' is an artifact of linguistic theory, and so is, *a fortiori*, that of contextual coherence. And what could prevent us from stipulating that this predicate is not to function in the fashion of those terms whose peculiar features prompted its introduction into the machinery of linguistic description?

I do not see any way of refuting this position; but this is not to say that I am convinced of its being correct. Too much has been left unsaid about the role of context in linguistic explanation to see with any clarity whether the constraints on the concept of contextual coherence are compatible with the stipulation of which we just spoke.

But let us assume for the sake of argument that your last reply is unassailable. Even then you could hardly claim to have *justified* your behaviour. For, although you may have shown that that behaviour is not formally inconsistent you have done nothing to show that it was correct. To show that you must give me a reason why the context stopped being coherent just now, and not a little earlier or a little later. Your only possible argument here is, it would seem, to contend that no such reason need be given—that the truth value gap of the context predicate 'coherent' can, like those of other predicates that are vague but not subject to EOI, be resolved in any one of a number of ways. How it is to resolved is a matter for decision. Sometimes such decisions must be made; and then somebody has to make them. This

is what you just did. The justification of your decision lies, as it must in the case of any genuine decision, in the circumstance that a decision *had* to be made, rather than the superiority of the option you chose over its possible alternatives.

20.

This last discussion suggests that predicates subject to the conditions (i)–(iii) of section 6 need not be incoherent after all. This conclusion is clearly at variance with the pessimism which permeates the earlier mentioned papers of Dummett and Wright. Should we conclude that, at least as matters stand, that pessimism is not quite warranted?

No, I think we should not. Dummett is concerned with two sorts of predicates, viz. (a) the observational predicates which, according to certain doctrines, must figure in the simplest reports of what is observed; and (b) the predicates that would form the substance of strictly finitistic mathematics if such mathematics were possible. And these predicates, even if it is true that they are not strictly incoherent in themselves, nevertheless behave in a manner that is incompatible with a fundamental aspect of mathematical and scientific language. It is the task of mathematics as well as that of science not only to reveal truths that are eternal, but also to develop languages in which these truths can be 'eternally' expressed—expressed, that is, in a manner invariant under all change of context in which these truths might be contemplated or communicated. For those who are committed to that enterprise our semantics does not provide a solution. For it vindicates these predicates by placing them within a linguistic framework of precisely the sort which it has been their acknowledged task to expunge.

Postscript (1981)

On page 297 I revealed my concern to formulate the truth definition so that it would be evident that no sentence could come out as both true and false in one and the same coherent context. But as a matter of fact the truth definition 3 of p. 297 does not guarantee this at all. It doesn't because the clauses (v) and (vi) reduce the question of truth and falsehood of the conditional $\varphi \to \psi$ in c for certain cases to the question of truth or falsehood of ψ in $m(c, \varphi)$, and this latter context may be incoherent even if c itself is coherent.

Thus consider for instance the sentence

(1) $P(\overline{n_0+1}) \to (P(\overline{n_0+2}) \to (\ldots \to (P(\overline{n_3-2}) \to P(\overline{n_3-1})))\ldots)$.

This sentence is true in the minimal context c_0 of the model M_0 described in section 14. For it is sufficient for the truth of (1) in c_0 that

(2) $P(\overline{n_0+2}) \to (\ldots \to (P(\overline{n_3-2}) \to P(\overline{n_3-1}))\ldots)$

is true in $m_0(c_0, P(\overline{n_0+1}))$. This is so if the consequent of (2) is true in $m_0(m_0(c_0, P(\overline{n_0+1})), P(\overline{n_0+2}))$, which is the same as $m_0(c_0, \{P(\overline{n_0+1}), P(\overline{n_0+2})\})$. By continuing this line of reasoning we conclude that (1) is true in c_0 if $P(\overline{n_3-1})$ is true in $m_0(c_0, \{P(\bar{k}) : k = n_0+1, \ldots, n_3-2\})$, which it is, since $P(\overline{n_3-2})$ belongs to the background of this last context and $(n_3-1)-(n_3-2) \leq r$. On the other hand (1) is also false in c_0. For (i) $P(\overline{n_0+1})$ is true in c_0; and (ii) (2) is false in $m_0(c_0, P(\overline{n_0+1}))$. The latter claim is correct since it reduces to the claims (ii.a) that $P(\overline{n_0+2})$ is true in $m_0(c_0, P(n_0+1))$, and (ii.b) that the consequent of (2) is false in this last context. (ii.b) reduces in turn to a conjunction of two claims, the first of which is immediate while the second reduces to a further pair of claims. Ultimately we arrive at the claim that $P(\overline{n_3-1})$ is false in $m_0(c_0, \{P(\bar{k}) : k = n_0+1, \ldots, n_3-2\})$, which is correct since $P(\overline{n_3-1})$ is true in every context of M_0. It should be noted that this conclusion does not depend on the ultimately indefensible assumption I made in defining M_0, viz. that the underlying relation of inference is that of classical logic, for we have only made use here of the information that various sentences are members of the background on account of having been added explicitly.

There are various ways in which the uniqueness of truth value assignments in coherent contexts could be secured. We could e.g. replace (v) of the definition on p. 297 by the clause

(v′) $[\varphi \to \psi]_{M,c} = 1$ iff $[\varphi]_{M,c} = 0$ or $(m(c,\varphi) \notin \text{Inc}\ \&[\psi]_{M,m(c,\varphi)} = 1)$

while making it a condition on models that whenever (according to this modified truth definition) some sentence is both true and false in a context c then c belongs to Inc.

This suffices to insure unequivocal truth value assignments for coherent contexts, but a more natural modification would involve a corresponding alteration in clause (vi), i.e. by adding '$m(c,\varphi) \notin$ Inc' as an extra conjunct to its right hand side. This last alteration would of course also suffice to block multiple truth value assignments in coherent contexts by itself. The same effect could be achieved moreover, if we added '$m(c,\varphi) \in$ Coh' instead of '$m(c,\varphi) \notin$ Inc', to either or both of (v) and (vi). A different type of solution would be to stipulate that the formulae $\ulcorner[\psi]_{M,m(c,\varphi)} = x\urcorner$, which appear as part of the right hand sides of (v) and (vi) are to be understood as saying

that x is the one and *only* truth value of ψ in $m(c, \varphi)$. Once more it would be enough to so modify just one of the two clauses, but again the resulting asymmetry would seem unwarranted.

Each of these possible modifications would seriously affect the metamathematical discussions of sections 15 and 16. Thus the second type of alteration would appear to destroy the monotonicity of the truth value assignment function (as a function of $\vdash$), and thereby any simple argument I can see of showing the existence of fixed points for the inference function, whereas the former type of modification invalidates the arguments of pp. 299–301. As the substantial changes in the essay that any of the alterations discussed here would therefore have required were out of the question at the point I came to contemplate their desirability (a point at which the production of the present volume had already reached an advanced stage) I thought it better to leave the truth definition as it is given on p. 297. I hope to deal with the issues raised in this postscript more extensively in the projected technical paper of which I spoke already in the main body of this essay.

While it certainly was my original aim to set up the truth definition in such a way that no sentence can be both true and false in one and the same coherent context, I should nevertheless wish to raise the question now how important this aim really is. The sentences which are liable to get both truth values in coherent contexts (according to the definition of p. 297) are, as far as I can see, always sentences about the truth values of which we have no very clear intuitions anyway, and the fact that such a sentence may end up with both truth values might even be interpreted as an indication of why our intuitions about such sentences are uncertain or confused. Moreover we could combine the truth definition of p. 297 with a further stipulation that a sentence is to be counted as *genuinely true* (or *false*) in a context c iff the truth definition assigns it the value 1 (0) but *not* the value 0 (1).

Of course, if we then proceed to refer to these last notions of truth and falsehood in the subsequent characterizations of valid inference and logical consequence, we encounter the same metamathematical difficulties that are the penalty for the second type of modification of (v) and/or (vi) discussed above. I am by no means convinced however that it would be wrong to retain the definitions of inference and consequence as given in sections 15 and 16 even in the light of the observations of this Postscript. The only really clear intuition concerning validity is, I believe, that where φ can be validly inferred from Γ it is impossible for φ to be (genuinely) false while all the members of Γ are (genuinely) true. And this intuition is respected by the definitions of sections 15 and 16 precisely when we are right to characterize genuine truth and falsity in the manner I just proposed.

Postscript II to 'The Paradox of the Heap' (2013)

[The following Postscript owes much to feedback from the audience of a seminar on Vagueness taught at the University of Chicago in the fall of 2012. What is identified below as the seminal misconception became fully clear to me only through a penetrating question from Anubav Vasudevan.]

The technical observations of this section and the following ones are flawed in more than one way. The principal mishap took the form of an unfounded worry about the monotonicity of the function that maps inference relations $\vdash$ as inputs to logical consequence relations as outputs. Recall the central problem stated in this part of the paper and the main strategy developed to find solutions for it: The truth definition needs an inference relation as input, but that relation should be the very same as the one that is then generated by the truth definition in combination with one from a number of possible definitions of logical consequence. In other words, the given truth definition together with the consequence definition of one's choice jointly define a function from inference relations to consequence relations. And any such combination presents us with the problem of finding out whether the function it determines has a suitable fixed point and to identify that logic and its mathematical properties.

A minimal precondition for success of that search is that the function has any fixed points at all. So the first concern of the paper at this point is to make sure that at least that condition is satisfied. And the strategy chosen in this connection is that of adopting a consequence definition such that the resulting function from inference relations to consequence relations is monotonic. For in that case Tarski's Fixed Point Lemma will guarantee the existence of fixed points. In fact, the Lemma gives us something stronger, viz. that for any inference relation $\vdash$, if the consequence relation $f(\vdash)$ it generates includes it (i.e. $\vdash \subseteq f(\vdash)$), then f will have a fixed point including this $\vdash$.

This looked like a promising starting point for the quest that becomes the central focus of the paper from section 15 onwards; and so everything was geared towards this way of finding a solution. Making sure that the function f was monotonic, however, seemed to impose constraints on the various components that go into its definition, and it was in that connection that the paper could be said to go off the rails (even on its own terms). The central mistake I made comes out most clearly and directly in relation to the definition of the static logical consequence relation $\models$ (Def. 4 in section 15). As the paper has it, that definition is problematic for the fixed point strategy

because it contains negative occurrences of the truth predicates; the worry expressed in the paper is that when we pass from a weaker input inference relation $\vdash_1$ to a stronger relation $\vdash_2$ (i.e. $\vdash_1 \subset \vdash_2$), then that may have the effect of *strengthening* the truth definition: it may be the case that for certain models M relative to $\vdash_2$, certain contexts c in C_M and certain formulas ϕ, $[\phi^{\vdash_2}_{M,c} = 1\,(0)$, while $[\phi^{\vdash_1}_{M,c}$ is undefined.

If this had been a real possibility, then negative occurrences of the truth predicate in the definition of logical consequence would have been a threat to the monotonicity of *f*. (Or, put in the terms of the paper, where the symbol '$\models$' is used also to denote the function f in this case, $\models(\vdash)$ might not have been a monotonic function of the input relation.)

But this was a needless worry. (As far as I am able to reconstruct the error, it may have slipped in when changes were made in the definition of 'context-sensitive model' and/or the truth definition in the final stages of writing, and without my carefully checking the implications of those changes.) The truth definition given in the paper as it appeared in print nowhere mentions $\vdash$. It does mention the predicates *Coh* and *Inc*. And it is true that what the extensions of these predicates can be does depend on the choice of $\vdash$. But that dependence is fully accounted for in the notion 'model relative to $\vdash$': the stronger $\vdash$ is, the smaller is the class of those models. But that relationship between the class of models relative to $\vdash_1$ and the class of models relative to $\vdash_2$ was already taken into account and it was (correctly) argued that it is no threat to monotonicity.

The upshot of this is that no truth definition in which there is no explicit mention of $\vdash$ can interfere with monotonicity (even if it may contain oblique references to $\vdash$ through mentioning *Coh* or *Inc*). So there is no problem, from this point of view, with Def. 4; nor is there with a variant of the dynamic Def. 5, in which the antecedent says not that the sentences in Γ are accepted but that they are true in *c*. In fact, once this imaginary difficulty has been recognized for what it is, we can see that we have far more room for manoeuvre—and thus also more leeway to cater for other conceptual constraints in pursuing the fixed point strategy—than the paper suggests. So the field of options for this approach to finding a viable logic of vagueness is much larger than the paper suggests.

While the prospects for a decent fixed point logic (more specifically: one that is not exceedingly weak) seem better now that this misconception has been removed, it is important to observe that there are certain absolute limits to how strong the logic can be. There is an observation that is relevant to this in Section 17 of the paper, but unfortunately the paper fails to draw the

relevant conclusion from this. Section 17 briefly addresses the question what happens when we assume that the predicate P is vague but without giving rise to the sorites paradox. (A well-known example of such a predicate is due to Sainsbury, who notes that at least for some speakers the noun 'pearl' is vague insofar as it is not settled for them whether artificial pearls should be counted as part of its extension or not.) The truth values of atomic sentences $P(a)$ where P is a vague predicate of this kind are not sensitive to tolerance effects.

One way to capture this insensitivity formally is to assume that the tolerance relation $\sim$ is just the identity relation *Id*. This has the effect that the truth value of $P(\overline{a})$ in c in M only depends on the positive and negative extension that are assigned to P in c in the model. ($P(\overline{a})$ will be true in any c if and only if a belongs to the positive extension of P in c, and likewise for the falsity conditions of $P(\overline{a})$.) If we assume in addition that the Incoherent contexts of the model are those in which some formula is both (i) false and (ii) either true or on the scoreboard (i.e. for some ϕ, either ϕ is true in c or ϕ is in $B(c)$), and that the Coherent contexts all satisfy the condition that everything on their scoreboard is true in them (i.e. if $\phi \in B(c)$, then ϕ is true in c) then it is not hard to see that for Coherent c the truth clauses for the universal quantifier reduce to the classical ones: $(\forall x)\phi(x)$ is true in c iff for all $a \in U$ $\phi(\overline{a})$ is true in c and $(\forall x)\phi(x)$ is false in c iff for some $a \in U$ $\phi(\overline{a})$ is false in c. So in such a model the truth value of any formula ϕ in any Coherent c will be the same as the one it gets according to the truth definition for partial models in a partial model M' in which the positive and negative extension are given by $B(c)$: $(F^{+}(P)) = \{a \in U\colon P(\overline{a}) \in B(c)\}$ and likewise for the negative extension).

Not only do such models behave like partial models for the purpose of truth and falsity, it is also fairly obvious that for any partial model M we can construct a context-sensitive model M' of the kind just described and a context c in the context set of M such that the truth value of any sentence ϕ in M' in c coincides with its truth value in M. This means that if we define logical consequence by means of Def. 4, then the logical consequence relation we would obtain by restricting the quantification over models to just the models of this special kind would be that of the 'logic of partial models' known as *Strong Kleene* (also sometimes referred to as *K3*). The properties of this logic are now quite well understood; for one thing we have explicit formulations of complete axiomatisations (see Cobreros et al. 2012).

Since there is no good reason why models of this kind should be excluded from the class of models over which logical consequence is defined, no consequence relation thus defined can be stronger than K3.

As stated here, the argument that 'Strong Kleene' is an upper bound for the logic of vagueness, it applies directly only to the static logical consequence definition Def. 4. But similar arguments can be given to show that K3 is also an upper bound for the logical consequence relations generated by other definitions of logical consequence.

The second point I want to raise in this Postscript about the technical part of the paper concerns the truth conditions of universal quantifications and conditionals. The truth conditions proposed in the paper for the universal quantifier were motivated by the thought that there had to be a reason why the second premise of the two-premise version of the sorites argument—the formula $(\forall x)(P(x) \rightarrow P(f(x)))$—can be false in a model describing a sorites situation, even though all its instances $P(\overline{a}_i) \rightarrow P(f(\overline{a}_{i+1}))$ are true. I felt at the time that there had to be a way to maintain the falsity of the universally quantified premise somehow, and the only reason I could think of was that a universally quantified sentence could be false because of its capacity to produce an inconsistency or incoherence when adopted, even in circumstances where the usual requirements for its truth are fulfilled.

The paper is rather apologetic about the dubiousness of this motivation and for its inability to offer more persuasive arguments for truth and falsity conditions for the universal quantifier it proposes. What I somehow failed to appreciate was that in a set-up designed to query the logic of vagueness there really is no pressing reason why the universal sorites premise should be false. In fact, much of what is said early on in the paper points in the direction of classical logic as the most likely culprit: The paradox is to be resolved not by showing that its major premise proves, when looked at closely and properly, to be false, but that for predicates of the kinds in question it is the reasoning that is invalid.

In fact, it is the multi-sentence version of the sorites argument, in which the universally quantified conditional is replaced by a long conjunction of non-quantified conditionals, that ought to be the focus of attention. As indicated in the Postscript from 1981, I had come to realize this by the time the book in which the paper was to appear was already in production, and at that time there was no more that I could do than to add that Postscript.

Conjunction is not part of the primitive vocabulary of the object language of the paper, but there are various ways in which a conjunction operator can be introduced into this language. One of these is to define $\&$ in the manner familiar from classical logic; for instance, $\phi \,\&\, \psi$ is introduced as a shorthand for $\neg(\phi \rightarrow \neg\psi)$. On this definition of $\&$ and the truth clauses that the paper specifies for $\neg$ and $\rightarrow$, $\phi \,\&\, \psi$ has the truth conditions given below in (3).

(3) Def. (truth clauses for &, when defined as $\neg(\phi \rightarrow \neg\psi)$)

$[\phi \,\&\, \psi]^{\vdash}_{M,c} = 1$ iff $[\phi]^{\vdash}_{M,c} = 1$ and $[\psi]^{\vdash}_{M,m(c,\{\phi\})} = 1$;
$[\phi \,\&\, \psi]^{\vdash}_{M,c} = 0$ iff $[\phi]^{\vdash}_{M,c} = 0$ or $[\psi]^{\vdash}_{M,m(c,\{\phi\})} = 0$.

Note that a conjunction operator with these truth conditions is dynamic in the sense that the contribution made by the second conjunct is relative to an update of the given context with the first conjunct. In this respect it resembles the conjunction operator; that is central to the systems of Dynamic Semantics first developed by Groenendijk and Stokhof (1990, 1991). (In those systems the conditional is defined in terms of ; and ¬, with the effect that it is 'locally dynamic'—the truth conditions of $(A \rightarrow B)$ in c involve the truth value of B in $m(c, A)$—while 'globally static'—its evaluation in c does not lead to a modification of c.)

Now consider the formula

(4) $(P(\overline{a_0}) \rightarrow P(\overline{a_1})) \,\&\, (P(\overline{a_1}) \rightarrow P(\overline{a_2})) \,\&\ldots\&\, (P(\overline{a_{N-1}}) \rightarrow P(\overline{a_N}))$

This formula too can serve as second premise in a sorites argument. But in a context-sensitive model M, the domain U_M of which consists of the elements a_1, a_2,.., where for all i = 1, .. ,N-1, $a_i \sim a_{i+1}$ and in a coherent context c of M such that $B(c_0)$ is the closure under $\vdash$ of the set $\{P(\overline{a_0}), \neg P(\overline{a_N})\}$, (4) will, according to the truth definition of the paper, be true. This means that the logic generated by our model theory must be weak enough to disable the inference of $P(\overline{a}_N)$ from (4) and $P(\overline{a}_0)$.

So which parts of the proof theory of classical logic is it that have to go? The clearest window on this question is provided neither by the two-sentence version in which the second premise is a universalised conditional, nor by the two-sentence version in which that premise has been replaced by the conjunction (4), but by the multi-sentence version which has instead of the second premise as many conditionals of the form $P(\overline{a_i}) \rightarrow P(\overline{a_{i+1}})$ as it takes to get from one end of a given sorites sequence to the other. For, as is in fact shown in the first sections of the paper, in these sorites arguments there can be only one culprit, viz. Modus Ponens.

So at the very least MP cannot be upheld in full generality. How much of it can be retained, and what else, if any, of classical logic may have to go, remains, I believe, an unresolved question, in spite of the many striking advances that the study of vagueness has made over the more than three decades that have passed since *The Paradox of the Heap* was published.

References

Chastain, C., 1975. Reference and context. In: K. Gunderson (ed.), *Language, Mind and Knowledge*. University of Minnesota Press, Minneapolis, MN, 149–270.

Cobreros, P., Egre, P., Ripley, D., van Rooij, R., 2012. Tolerant, classical, strict. *Journal of Philosophical Logic* 41, 347–385.

Cresswell, M., 1973. *Logics and Languages*. Methuen, London.

Dummett, M., 1975. Wang's paradox. *Synthese* 30, 301–324.

Fine, K., 1975. Vagueness, truth and logic. *Synthese* 30, 265–300.

Gazdar, G., 1979. *Pragmatics, Implicature, Presupposition and Logical Form*. Academic Press, New York.

Goguen, J., 1969. The logic of inexact concepts. *Synthese* 19, 325–373.

Groenendijk, J., Stokhof, M., 1990. Dynamic Montague Grammar. In: L. Kalman, L. Pólos (eds.), *Papers from the Second Symposion on Logic and Language*, Budapest. Eotvos Lorand University Press, Akademiai Kiadeo, 3–48.

Groenendijk, J., Stokhof, M., 1991. Dynamic predicate logic. *Linguistics and Philosophy* 14, 39–100.

Kamp, H., 1975. Two theories about adjectives. In: E. Keenan (ed.), *Formal Semantics of Natural Language*. Cambridge University Press, Cambridge, 123–155. Reprinted in: K. von Heusinger, A. ter Meulen (eds.), *Meaning and the Dynamics of Interpretation. Selected Papers of Hans Kamp*. Brill, Leiden, 2013, 225–261.

Kamp, H., 1981. A theory of truth and semantic representation. In: J. Groenendijk, T.M.V. Janssen, M. Stokhof (eds.), *Formal Methods in the Study of Language, Proceedings of the Third Amsterdam Colloquium*. Mathematical Center, Amsterdam, 277–322. Reprinted in: J. Groenendijk, T.M.V. Janssen, M. Stokhof (eds.), *Truth, Interpretation and Information, Selected Papers from the Third Amsterdam Colloquium*. Foris, Dordrecht, 1984, 1–41, and reprinted in: K. von Heusinger, A. ter Meulen (eds.), *Meaning and the Dynamics of Interpretation. Selected Papers of Hans Kamp*. Brill, Leiden, 2013, 329–369.

Lewis, D.K., 1979a. General semantics. In: D. Davidson, G. Harman (eds.), *Semantics of Natural Language*. Reidel, Dordrecht, 169–218.

Lewis, D.K., 1979b. Score keeping in a language game. In: R. Bäuerle, U. Egli, A. von Stechow (eds.), *Semantics from Different Points of View*. Springer, Berlin, 172–187.

Putnam, H., 1973. Meaning and reference. *The Journal of Philosophy* 70, 699–711.

Putnam, H., 1975. The meaning of "meaning". In: K. Gunderson (ed.), *Language, Mind and Knowledge*. University of Minnesota Press, Minneapolis, MN, 131–193. Reprinted in: H. Putnam, *Mind, Language and Reality: Philosophical Papers, Vol. 2*. Cambridge University Press, Cambridge, 1975, 215–271.

Sanford, D., 1976. Competing semantics of vagueness: Many values vs. super truth. *Synthese* 33, 195–210.

Thomason, R.H., 1973. *Supervaluations, the Bald Man and the Lottery*, unpublished manuscript, Yale University, New Haven, CT.

Wright, C., 1975. On the coherence of vague predicates. *Synthese* 30, 325–365.

PART FOUR

DISCOURSE REPRESENTATION

INTRODUCTION TO DISCOURSE REPRESENTATION

Hans Kamp

The first paper of this part, *A Theory of Truth and Semantic Representation* (TTSR), is probably the most widely known of those in this part, and perhaps, at this point in time, of all the papers in this collection. It appears to be the document most often cited as one of the two starting points of Dynamic Semantics, as a discipline with an explicit compositional semantics and a well-defined logic, rather than as a mere collection of informal intuitions about how discourse works and a list of promissory notes for making those into a theory. The other document standardly mentioned in this connection is Irene Heim's ***The Semantics of Definite and Indefinite Noun Phrases***. Often, in fact, her dissertation and TTSR have been cited in tandem; rightly so, it might be added, given that where the application domains of the two approaches overlap primarily in what they have to say about donkey anaphora, they make the same predictions about meaning and semantic interpretability. As with the application of Supervaluation Theory to the analysis of vagueness, where Kit Fine and I independently arrived at the conviction that this was an approach worth developing, Heim and I also arrived at our formulations of these mechanisms independently. But her dissertation was the far more substantial work, which, for one thing, covers definite descriptions as well as pronouns. It rightly has become the yardstick for subsequent work on the topics it addressed. It deserves emphasis in this context, however, that TTSR wasn't meant to be just a theory of the behavior of pronouns and certain other noun-phrases. As indicated in the introduction to the first part, the paper was rather the outcome of a process that started with observations about the discourse-linking properties of tenses. I was groping for evidence at that point that this was not an isolated phenomenon. For, as Robert Stalnaker had made plain as early as the early seventies, the contributions made by a new sentence or utterance in an ongoing discourse or text typically take advantage of information that has already been provided by earlier parts of the text or discourse. Since linguistic communication seems to rely on this to such a very large extent, it would have been remarkable if languages had not developed grammaticised mechanisms for exploiting this possibility of providing information cumulatively, or if the tenses

© HANS KAMP, 2013 | DOI:10.1163/9789004252882_013
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

in languages like French and English had been the only instances of such grammaticisation.

In this light the discovery that some of the main facts about donkey pronouns, essentially as they had been laid out by Peter Geach in ***Reference and Generality***, could be accounted for by assuming discourse-related mechanisms closely reminiscent of what I thought I had observed for the tenses of French, seemed just the confirmation I was looking for. At the same time the crisp judgments about the acceptability and the truth conditions of donkey sentences—Geach's actual target, to be distinguished from the donkey discourses that TTSR included as part of the data—made it possible to turn this small part of the discourse-oriented enterprise into a theory that could be used to make explicit predictions about robust linguistic observations. Indeed, only when the formal framework that emerged from this attempt to deal with these particular data was in place, was the road free to deal in similarly formally precise ways with the sentence-transcending semantic effects of tenses and other parts of temporal vocabulary. This is also the reason why difficulties with some of the predictions made by TTSR and ***The Semantics of Definite and Indefinite Noun Phrases*** did not seem as troubling to the general DRT enterprise, as they may have seemed to Heim. In particular, there are the recalcitrant facts about donkey discourses like *A woman came into the lobby. She went up to the reception desk and got herself a double room with a view of the river*. or *A fat man was pushing his bicycle. He seemed at the end of his tether*. There is a strong intuition that such discourses are inappropriate, unless the speaker is talking about some particular woman or fat man she could identify upon demand. Of course, these data, and others that have to do with donkey sentences rather than with donkey discourses, have to be taken seriously by anybody who wants to use DRT as a framework. But once one is convinced that dynamics is an important aspect of linguistic meaning and interpretation and that it would have been a miracle if things had been otherwise, the natural response to these counterexamples is not to go and look for an alternative framework. Instead one will be inclined to see such data as that this part of the theory isn't quite right, so it needs rethinking and adjusting. If there are other parts of the general approach that do not seem directly affected by the difficulties observed, then concluding that the approach was wrong after all *in toto*, that it should scuttled and that we ought to go back to a well-tried, well-established form of sentence semantics such as Montague Grammar, seems overly alarmist and overly defeatist.

It is now more than thirty years after TTSR appeared and in the course of that time the scientific landscape has changed considerably. Dynamic Semantics, in a broad sense of the term, according to which it includes DRT

as one among a number of alternative approaches, now seems quite firmly established at least within linguistics. In a substantial part this is because it has proved helpful—some practitioners in question might well want to go further and say 'indispensable'—in exploring the semantics of many non-European languages, in which the articulation of discourse into distinct, grammatically closed sentential units is often far less straightforward than for the languages of Europe. Mandarin Chinese is a prominent example. In the course of applying the principles of Dynamic Semantics to the variety of problems that have presented themselves and which seemed to demand such a treatment, a range of new frameworks have been developed. As a consequence, the dynamic approach seems as well entrenched as it ever was. On the other hand, it has become increasingly difficult to answer exactly what Dynamic Semantics is in general. It has even become difficult to say exactly what DRT is, for here too there now exist a number of distinct versions, which are currently being developed and/or used. If pressed for succinct characterizations of either, this is what I would offer:

(i) Dynamic Semantics is characterized by the principle of a two-way interaction between linguistic expressions and contexts: the content of an expression E used in a context C depends on C, but conversely, once this content has been determined, it leads to an update of C to a new context C' and this updated context C' helps determine the content of the next expression. DRT, where context update is analyzed as updating of DRSs, makes a point of identifying content representations and context representations.

(ii) My proposal to distinguish versions of DRT from other forms of Dynamic Semantics is that they make use, at some level, of semantic representations. This can mean one or both of two things:
   (a) The compositional characterization of the semantics of syntactically complex expressions makes an essential use of the form of the semantic representations of the constituents of complex expressions.
   (b) The interpretations that the compositional account of meaning associates with complete expressions (complete sentences, but also coherent multi-sentence discourses and texts) are assumed to model the ways in which human interpreters identify the linguistic inputs that come their way, and which they then exploit for various cognitive tasks (such as reasoning and planning), again making use of the form of those representations.

Within this diverse range of different frameworks the DRT of TTSR is just one in an expanding playing field. It is one of the minor players insofar as its explicit application range is quite limited. It may still have its use as a way of introducing a novice to some of the distinctive features that its version of DRT has in common with other proposals from within the DRT family, as, for instance, the theory of ***From Discourse to Logic***, which has a much

broader scope, including applications to plural, as well as singular noun phrases and to a fairly broad range of words and constructions in the area of tense and aspect. But even within the DRT family there are many theories that differ from these two in fundamental respects, not only the more recent versions designed to deal with presuppositions and propositional attitudes that are discussed and applied in the more recent papers of this collection, but also theories that differ from the one presented in ***From Discourse to Logic*** in more sweeping ways, such as Uwe Reyle's UDRT, Nicholas Asher's SDRT or the Layered DRT of Bart Geurts and Emar Maier. Then there are the type-theoretical reductions of DRT along the lines of, in particular, Reinhard Muskens and a variety of systems that have been developed more recently along somewhat similar lines, such as the Infotention Theory of Maria Bittner and her students. Of those I am not quite sure whether they should be classified as versions of DRT according to the criterion (ii.a) given above. One problem is what do we want to say about the new types that such reductions of DRT to type theory resort to. Is the introduction of such new types to be taken as a sign that we have left the domain of content and entered that of semantically unreflected aspects of form, or do they signal a transition to a more refined notion of content?

Context, Thought and Communication is a comparatively slight piece, which tries to explain to a primarily philosophical audience how DRT works and what I thought of then (and still think of now) as its potential for philosophical analysis. The paper was included in this collection because it is sometimes regarded as a useful informal introduction to the basic ideas of DRT, stripped from the technicalities that make parts of TTSR difficult to read. Moreover, the paper gives a flavor of how DRT can be employed in the analysis of double vision problems, like Kripke's puzzle about belief and Geach's puzzle about Hob and Nob, foreshadowing subsequent work by Nicholas Asher and also by myself, as illustrated in the last two papers in this collection.

Disambiguation in Discourse is an example of applying DRT to problems of a quite different sort. It deals with an example of lexical ambiguity—the German word *ausleihen*, which can be used to mean either 'borrow' or 'lend'—and various discourse contexts in which it is naturally resolved in one way or the other. The purpose of the paper is to show how DRSs can serve at the same time as representations of utterances or textual passages and as forms of knowledge representation, of the kind wanted and advocated within AI: as representations, that is, of information that can be used in formal and automated reasoning. In the cases discussed in the paper the reasoning has to do with whether the subject of *ausleihen* was or was not in

possession of the direct object immediately before the event that the given occurrence of *ausleihen* is used to describe. In exploring these possibilities the paper fits within the larger enterprise of using DRT as s foundation for logic-based text and discourse processing. In DRT's early days this field of application looked like it was going to be its principal claim to fame, with large scale implementation projects as IBM's LILOG, which ran for a good part of the eighties.

Disambiguation in Discourse was written at a time when the enchantment with the kind of automated text processing in which it could have found its niche was already on the wane. You could say that the paper provides an (unwitting) explanation of why disenchantment should have set in as soon as it did. For it exposes the fragility of the analyses it presents as an almost inevitable concomitant of their potential. An enormous amount of theoretical analysis of the kind the paper exemplifies would be needed before language processing systems built on such foundations could work effectively. Arguably, the task is not only one of gigantic proportions, but also just too open-ended to ever be completed to the extent that general purpose text processors would require, even if cost and manpower weren't an issue. No wonder that computational linguists, who can survive only when they can deliver something that works, have turned away from these methods. In the approaches that have taken their place, in which statistics plays the dominant part, there seems to be little room for data structures like DRSs. But there are some indications that the tide may be turning again and that we are moving in a direction where logical and statistical methods are made to work hand in hand, by a community of computational and non-computational linguists.

A THEORY OF TRUTH AND SEMANTIC REPRESENTATION[a,*]

Hans Kamp

1. Introduction

Two conceptions of meaning have dominated formal semantics of natural language. The first of these sees meaning principally as that which determines conditions of truth. This notion, whose advocates are found mostly among philosophers and logicians, has inspired the disciplines of truth-theoretic and model-theoretic semantics. According to the second conception meaning is, first and foremost, that which a language user grasps when he understands the words he hears or reads. This second conception is implicit in many studies by computer scientists (especially those involved with artificial intelligence), psychologists and linguists—studies which have been concerned to articulate the structure of the representations which speakers construct in response to verbal inputs.

* This paper was written when I held a Post-Doctoral Fellowship at the Center for Cognitive Science of the University of Texas at Austin. Anybody who has the faintest acquaintance with my personality will realize that it would not have been written had the Directors of the Center not given me this opportunity, and thus understand the depth of my indebtedness to them. I would also like to thank, among the many who helped me during my stay in Austin, Kate Ehrlich, Alan Garnham, Lauri Karttunen and Stanley Peters for their comments and suggestions.

[a] Editors' note: This article was first published in J. Groenendijk, T.M.V. Janssen & M. Stokhof (eds.), *Formal Methods in the Study of Language, Proceedings of the Third Amsterdam Colloquium*, Mathematical Center, Amsterdam, 1981, 277–322. It was then reprinted in: J. Groenendijk, & T.M.V. Janssen & M. Stokhof (eds.), *Truth, Interpretation and Information, Selected Papers from the Third Amsterdam Colloquium*, Foris, Dordrecht, 1984, 1–41. We would like to thank Foris (now de Gruyter) for the kind permission to reprint the article, which in its 1984 version can be found at: http://www.degruyter.com/view/product/12375?rskey=WTa2Gk&result=1&q=Truth,%20Interpretation [8 March 2013]. We carefully edited the article, corrected some typos, updated the bibliographical references and brought some formulas in line with now more common usage. We would like to thank Tillmann Pross and Dominik Schlechtweg for careful proofreading and helpful comments.

© HANS KAMP, 2013 | DOI:10.1163/9789004252882_014
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

It appears that these two conceptions, and with them the theoretical concerns that derive from them, have remained largely separated for a considerable period of time. This separation has become an obstacle to the development of semantic theory, impeding progress on either side of the line of division it has created.

The theory presented here is an attempt to remove this obstacle. It combines a definition of truth with a systematic account of semantic representations. These two components are linked in the following manner. The representations postulated here are (like those proposed by others; cf. e.g. Hendrix (1975) or Karttunen (1976)) similar in structure to the models familiar from model-theoretic semantics. In fact, formally they are nothing other than partial models, typically with small finite domains. Such similarity should not surprise; for the representation of, say, an indicative sentence ought to embody those conditions which the world must satisfy in order that the sentence be true; and a particularly natural representation of those conditions is provided by a partial model with which the (model describing the) real world will be compatible just in case the conditions are fulfilled.

Interpreting the truth-conditional significance of representations in this way we are led to the following characterization of truth: A sentence *S*, or discourse *D*, with representation *m* is true in a model *M* if and only if *M* is compatible with *m*; and compatibility of *M* with *m*, we shall see, can be defined as the existence of a proper embedding of *m* into *M*, where a *proper embedding* is a map from the universe of *m* into that of *M* which, roughly speaking, preserves all the properties and relations which *m* specifies of the elements of its domain.

A theory of this form differs fundamentally from those familiar from the truth-theoretical and model-theoretical literature, and thus a substantial argument will be wanted that such a radical departure from existing frameworks is really necessary. The particular analysis carried out in the main part of this paper should be seen as a first attempt to provide such an argument. The analysis deals with only a small number of linguistic problems, but careful reflection upon just those problems already reveals, I suggest, that a major revision of semantic theory is called for.

The English fragment with which the analysis deals contains sentences built up from these constituents: common nouns, certain transitive and intransitive verbs (all in the third person singular present tense), personal and relative pronouns, proper names, and the particles *a*, *every*, and *if ... (then)*. These can be combined to yield the following sorts of compounds:

1 complex singular terms such as *a man, every woman, a man who loves every woman, every woman whom a man who owns a donkey loves*, etc. (We can embed relative clauses inside others and there is no upper bound to the depth of embedding!);
2 singular terms—i.e. complex terms of the kind just exemplified, proper names and personal pronouns—can be combined with verbs to yield sentences;
3 sentences may be joined with the help of *if* to form larger sentences of conditional form; sentences serve moreover as the sources of relative clauses.

The choice of this fragment is motivated by two central concerns: (a) to study the anaphoric behaviours of personal pronouns; and (b) to formulate a plausible account of the truth conditions of the so-called "donkey-sentences" (which owe their name to the particular examples in Geach (1962), the work that kindles contemporary interest in sentences of this type). As these donkey-sentences will play a prominent role in the theory developed below, let me briefly review the problem that they have been taken to present. We shall concentrate on the following instances:

(1) If Pedro owns a donkey he beats it.
(2) Every farmer who owns a donkey beats it.

For what needs to be said at this point it will suffice to focus on (1). For many speakers, including the author of this paper, the truth conditions of (1) are those determined by the first order formula

(3) $(\forall x)(\text{Donkey}(x) \wedge \text{Owns}(\text{Pedro}, x) \rightarrow \text{Beats}(\text{Pedro}, x))$.

(As a matter of fact not all English speakers seem to agree that (3) correctly states the truth conditions of (1). Unfortunately an adequate discussion of diverging intuitions is not possible within the confines of the present contribution.)

The problem with (1) and (3) is that the indefinite description *a donkey* of (1) reemerges in (3) as a universal quantifier. How does an expression of a type which standardly (or so it always seemed) conveys existence manage to express universality in a sentence such as (1)? One way in which one might hope to explain this is by referring to the familiar equivalence between universal quantifiers with wide and existential quantifiers with narrow scope. Sentence (4), for instance, can be symbolized not only as (5) but also as (6).

(4) If Pedro owns a donkey he is rich.
(5) $(\forall x)(\text{Donkey}(x) \wedge \text{Owns}(\text{Pedro}, x) \rightarrow \text{Rich}(\text{Pedro})$
(6) $(\exists x)(\text{Donkey}(x) \wedge \text{Owns}(\text{Pedro}, x)) \rightarrow \text{Rich}(\text{Pedro})$

Out of these two (6) would appear to be the "natural" symbolization of (4) as it renders the indefinite *a donkey* as an existential quantifier. (5), we might be inclined to say, is adequate only for indirect reasons, viz. in virtue of its logical equivalence to (6). Note, however, that (1) cannot be captured by an analogue of (6). For in such a formula the scope of the existential quantifier would have to be restricted, just as it is in (6), to the antecedent alone; but then the quantifier would be incapable of binding the position corresponding to that occupied by *it* in the main clause of (1).

No one of the solutions to this problem that can be found in the existing literature strikes me as fully satisfactory. As I see the problem a proper solution should provide: (i) a general account of the conditional; (ii) a general account of the meaning of indefinite descriptions; and (iii) a general account of pronominal anaphora; which when jointly applied to (1) assign to it those truth conditions which our intuitions attribute to it. These requirements are met, I wish to claim, by the theory stated in the next two sections.

As earlier remarks implied, there are three main parts to that theory:

1 A generative syntax for the mentioned fragment of English (I have cast the syntax in a form reminiscent of the syntactic descriptions which are used by Montague; the reader may verify, however, that many other syntactic descriptions would be equally compatible with the remaining components of the theory);
2 a set of rules which from the syntactic analysis of a sentence, or sequence of sentences, derives one of a small finite set of possible non-equivalent representations; and
3 a definition of what it is for a map from the universe of a representation into that of a model to be a proper embedding, and, with that definition, a definition of truth.

The analysis thus obtained not only yields an account of the truth conditions of the donkey sentences (as well as of certain other notoriously problematic sentences which the fragment admits, such as e.g. some types of Bach-Peters sentences), it also reveals two more general insights concerning, respectively, personal pronouns and indefinite descriptions.

1 Personal pronouns, it has been pointed out, have a number of apparently distinct functions. Sometimes they seem to behave as genuinely referential terms, as e.g. the *he* in *Pedro owns a donkey. He beats it*. Sometimes, as the *him* of *Every man who loves a woman who loves him is happy*, they appear to do precisely what is done by the bound variables of formal logic. Yet another occurrence, noted in particular by Evans (1977, 1980), who coined the term "E-type pronoun" for it, cannot be understood, or so it has been claimed, either on the model of a simple referential expression or on that of a bound

variable. An example is the occurrence of *it* in *If Pedro owns a donkey he beats it*. The present theory brings out what these three different types have in common in that it offers, at the level of representation-formation a single rule which equally applies to each of them. This rule may interact in various ways with other rules, which are associated with different syntactic constructions, and this gives rise to the seeming multiplicity of functions which the recent philosophical and linguistic literature has noted. (There are several pronoun uses, such as "pronouns of laziness" and deictic pronouns, which have no instances within the fragment of English studied in this paper and which, therefore, cannot be discussed here. Such occurrences, however, can also be accommodated along the lines sketched in this paper.)

2 Indefinite descriptions are, on the account given here, referential terms, not existential quantifiers. When an indefinite has existential force it has that force in virtue of the particular role played by the clause containing it within the sentence or discourse of which it is part. It is true that the clausal roles which impose an existential, rather than a universal, reading upon indefinites are the more prominent; and this, I take it, has been responsible for the familiar identification of the indefinite article as a device of existential quantification. But they are not the only roles. The antecedent of a conditional, for instance, plays a role which is not of this kind; a simple clause which occurs in this role confers a universal interpretation on the indefinite descriptions it contains.

There is much that ought to be said about the conceptual implications of the present theory and about the range of its possible applications. But, as space is limited, I shall confine myself to a couple of brief remarks.

1 It should be stressed that truth as it is defined here applies not only to single sentences but also to multi-sentence discourse. This is of special importance where intersentential relations within the discourse (such as intersentential anaphoric links) contribute to its meaning. As will be seen below the links between anaphoric pronouns and their antecedents invariably have their impact on the discourse representation (irrespective of whether pronoun and antecedent occur in the same, or in different sentences) and thus on the truth conditions of the discourse, which the discourse representation embodies. Other intersentential relations, such as the relation which obtains between the sentences of past tense narratives on account of their sequential order—which is typically understood to convey the temporal relations between the events which the sentences report—can be encoded into the discourse representation with equal ease.

2 The role representations are made to play within the theory developed in this paper places substantial constraints on their internal structure. (Careful reading of the subsequent sections will, I hope, confirm this assessment.) This is of particular significance if, as I have already more or less implied, discourse representations can be regarded as the mental representations which speakers form in response to the verbal inputs they receive. I should point out that the specific theory that is presented below does not render

> such identification essential. Even if the representations it posits are thought of as purely theoretical devices whose raison d'être is to be found solely in the contribution they make to an effective account of certain semantic properties of sentences and sentence complexes, the theory may merit comparison with other schemes of linguistic description which have been applied to the same phenomena. But this is not how I would like to see the proposal of this paper myself. I conjecture that the structures which speakers of a language can be non-trivially described as forming to represent verbal contents are, if not formally identical, then at least very similar to the representations here defined.

If this identification is legitimate then a theory of the sort I have tried to develop brings to bear on the nature of mental representation and the structure of thought, a large and intricate array of data relating to our (comparatively firm and consistent) intuitions about the truth-conditions of the sentences and sentence sequences we employ. I very much hope that along these lines it may prove possible to gain insights into the objects of cognitive operations, as well as into these operations themselves, which are unattainable if these data are ignored, and which have thus far been inaccessible to psychology and the philosophy of mind precisely because those disciplines were in no position to exploit the wealth of linguistic evidence in any systematic fashion.

2. The Theory: Informal Preliminaries

2.1. *Anaphoric Pronouns*

The analysis of pronominal anaphora I shall sketch is informed by the conviction that the mechanisms which govern deictic and anaphoric occurrences of pronouns are basically the same. This is an intuition that has guided many recent theories of pronominal reference; inevitably the account given here will resemble some of these in various respects.[1]

---

[1] Theories that to a greater or lesser degree accord with this intuition have emerged within Artificial Intelligence and Computer Science, as well as within Linguistics. A significant contribution of this kind that comes from the first field is Webber (1978). Examples of such theories that have been proposed by linguists are: the theories outlined in Bartsch (1976, 1979), Cooper (1975, 1979), Hausser (1974, 1979), Karttunen (1976).

By no means every recent account of pronouns is predicated on the assumption that all cases of pronominal reference can be brought under one unifying principle. Cf. for instance Evans (1977, 1980), Lasnik (1976), Partee (1978).

Our point of departure will be the hypothesis that both deictic and anaphoric pronouns select their referents from certain sets of antecedently available entities. The two pronoun uses differ with regard to the nature of these sets. In the case of a deictic pronoun the set contains entities that belong to the real world, whereas the selection set for an anaphoric pronoun is made up of constituents of the representation that has been constructed in response to antecedent discourse.

About deixis I shall have no more to say in this paper. But a little more needs to be said about anaphoric pronouns before we can proceed to the detailed analysis of some particular pieces of discourse.

The strategies used in selecting the referents of anaphoric pronouns are notoriously complex; they usually employ background assumptions about the real world, "grammatical" clues, such as the requirement of number and gender agreement between the anaphor and its antecedent, and the order in which the potential referents were introduced by the preceding discourse.[2]

The integration of these various factors often involves, moreover, what seem to be quite intricate patterns of inference. Efforts to understand these strategies have claimed much thought and hard work, but, in its general form at least, the problem appears to be far too complex to permit solution with the limited analytic tools that are available at the present time.[3]

About the strategies I shall have nothing more to say. Our concern will be, rather, with the sets of referential candidates from which they select. These entities will constitute the universes of the representations of which I spoke in Section 1. I have already said that these discourse representations,

---

[2] There seems to be a rough preference for referents introduced by terms that appear in the discourse *before* the anaphoric pronoun over those that are introduced by subsequent terms, as well as a preference for referents that are introduced by terms that occur *near* the anaphor. (Thus the referent introduced by the last referential term preceding the anaphoric pronoun is, other factors permitting, a strong referential candidate.)

[3] A large part of the research that has been done on anaphora by computer scientists and people working in Artificial Intelligence has been concerned with this problem—understandably enough, as the lack of effective routines for the detection of anaphoric antecedents has for many years been one of the main obstacles to producing satisfactory computer systems for question answering and translation. However useful some of this work may have been, I have the impression that its theoretical significance is rather limited. Indeed I much incline to the opinion expressed, for example in (Partee 1978: p. 80) that all we can reasonably expect to achieve in this area is to articulate orders of preferences among the potential referents of an anaphoric pronoun, without implying that the item that receives the highest rating is in each and every case the referent of the anaphor.

or DR's as I will call them for short, are formed in response to the discourses they represent and that their formation is governed by certain rules. These rules—and this is a new, and crucial, assumption of the theory—operate on the syntactic structures of the sentences of the discourse, and it is via them that syntactic form determines what the resulting DR will be like. This determination is not complete however. The syntactic structure does not, for instance, determine the anaphoric links between pronouns and their antecedents, which the DR makes explicit.

Most of the real work that the present theory will require us to do concerns the exact formulation of the rules of DR-formation. The exact formulation of the rules will be rather compact, and will betray, I suspect, little of either motivation or empirical implications to any but the initiated. I have decided therefore to first present a number of applications of the theory. I hope that if we proceed in this manner its formal features will reveal themselves more naturally and that the subsequent reading of the exact definitions in Section 3 will thus be less disagreeable than it would be without such preparation.

Let us begin by considering the two sentence discourse:

(7) Pedro owns Chiquita. He beats her.

The DR for the first sentence of (7) will contain two elements, call them u and v, which represent, respectively, Pedro and Chiquita, and furthermore the information that the first of these, u, owns the second, v. Schematically we shall represent this information as follows:

$m_1(7)$

u	v
.	.
Pedro owns Chiquita	
u = Pedro	
v = Chiquita	
u owns v	

To incorporate the information contained in the second sentence of (7) we must extend structure $m_1(7)$. But to do that we must make two decisions, regarding the reference of, respectively, *he* and *her*. It is natural to understand *he* as referring back to *Pedro* and *her* as referring back to *chiquita*. Let us agree to interpret the pronouns in this way and to expand $m_1(7)$ accordingly. What we get is:

$m(7)$

u	v
.	.
Pedro owns Chiquita	
u = Pedro	
v = Chiquita	
u owns v	
He beats her	
u beats her	
u beats v	

I said that linking *he* with *Pedro* and *her* with *Chiquita* yields what seems the most natural reading of (7). "But", you might ask, "what *other* readings could (7) have?" The answer to that question depends on the setting, or context, in which (7) is supposed to be used. If (7) were uttered by a speaker who points at some individual other than Pedro while saying *he*, or at some being distinct from Chiquita when he says *her*, the gesture would recruit this demonstrated individual as referent for the pronoun. Similarly, if (7) were part of a larger discourse *he* or *her* could conceivably refer back to some other individual introduced by an earlier part of that discourse; and this could result in a genuine referential ambiguity. However, if (7) is used by itself, i.e., without preceding verbal introduction, and also in the absence of any act of demonstration, then—and this is another important hypothesis of our theory—there are no other potential referents for *he* and *her* than the discourse referents which have been introduced in response to *Pedro* and *Chiquita*. Let us agree that henceforth (except where the contrary is indicated explicitly) all our examples of simple and multi-sentence discourses shall be understood in the last of these three ways, i.e., as used without accompanying deictic gestures and not preceded by any related discourse.

Even when we understand (7) in this third way its anaphoric links are not fully determined by what we have said. For why cannot *he* and *her* both refer to u, say, or *he* to v and *her* to u? The reason is of course obvious: *he* must refer to a male individual, and *her* to a female one. But, obvious as the determining principle may be, it is not quite so easy to state it in a form that is both general and accurate. For what is it that determines an antecedently introduced discourse referent as male, rather than female, or neither male nor female? (7) allows us to infer that u is male because we know that *Pedro*, typically, refers to male individuals. But often the antecedent term which led to the introduction of a discourse item is not quite so explicit about the gender of its referent. Consider for example such terms as: *Robin,*

Hilary, the surgeon, the president, an officer in the Air Force, the professor, the professor's secretary, the first inhabitant of this cave. Often we can do no better than guess whether the referent is male or female, or human or non-human. Some of these guesses are more educated than others. And not infrequently where the anaphoric link between the antecedent and some particular pronoun is clear on independent grounds it is in fact the gender of the pronoun which resolves the uncertainty.[4]

Applying the principle of gender agreement will thus often involve drawing various inferences from the information that is given explicitly; and as in all other processes where inference can be involved, there appears to be no clear upper bound to its potential complexity.

There is a further complication that an exact statement of the principle must take into account. The gender of the pronoun that is used to refer to a certain object is not exclusively determined by the nature of that object, but, to some extent, also by the actual *form* of the anaphoric antecedent which made it available as a referent. Thus let us suppose that the name *Chiquita* in (7) actually refers to a donkey. In most situations we refer, or at any rate may refer, to a donkey by means of *it*. But in a discourse such as (7) this would be inappropriate. The name *Chiquita* highlights, one might wish to say, the fact that its referent is female, and this makes *she* the correct resumptive pronoun. But nonetheless the task of giving even an approximate formulation of the principle appears to be well beyond our present means. In what follows we shall ignore the principle of gender agreement, just as we ignore all other factors that help to disambiguate the reference of anaphoric pronouns. But where, in subsequent examples, the need for gender agreement clearly excludes certain anaphoric links I shall not bother to mention those without referring to the principle explicitly.

Clearly (7) is true, on the reading of it that is given by $m(7)$ if and only if the real Pedro stands to the real Chiquita in a relation of ownership and also in the relation expressed by the verb *beat*. Put differently, if M

[4] We are much assisted in our making of such guesses by the spectrum of our social prejudices. Sometimes, however, these may lead us astray, and embarrassingly so, as in the following riddle which advocates of Women's Lib have on occasion used to expose members of the chauvinistic rearguard: In a head-on collision both father and son are critically wounded. They are rushed into hospital where the chief surgeon performs an emergency operation on the son. But it is too late and the boy dies on the operation table. When an assistant asks the surgeon, "Could you have a look at the other victim?" the surgeon replies "I could not bear it. I have already lost my son." Someone who has the built-in conception that chief surgeons are men will find it substantially more difficult to make sense of this story than those who hold no such view.

is a model, representing the world—consisting of a domain U_M and an interpretation function F_M which assigns to the names *Pedro* and *Chiquita* members of U_M and to the intransitive verbs *own* and *beat* sets of pairs of such members—then (7) is *true in M* iff the pair $\langle F_M(Pedro), F_M(Chiquita)\rangle$ belongs both to $F_M(own)$ and to $F_M(beat)$. Moreover, the right hand side of this last biconditional is fulfilled if there is a map f of the universe of $m(7)$, i.e. the set $\{u, v\}$, into U_M so that all specifications of $m(7)$ are satisfied in M—i.e., $f(u)$ is the individual denoted in M by *Pedro*, $f(v)$ is the individual $F_M(Chiquita)$, and it is true in M that $f(u)$ both owns and beats $f(v)$, in other words, that $\langle f(u), f(v)\rangle$ belongs to both $F_M(own)$ and $F_M(beat)$.

Let us now consider

(8) Pedro owns a donkey. He beats it.

The first sentence of (8) induces a DR that can be represented thus:

$m_1(8)$

u	v
.	.
Pedro owns a donkey	
u = Pedro	
u owns a donkey	
donkey(v)	
u owns v	

Once again there is no choice for the anaphoric antecedent of either *he* or *it* in the second sentence of (8). So the complete DR of (8) becomes:

$m(8)$

u	v
.	.
Pedro owns a donkey	
u = Pedro	
u owns a donkey	
donkey(v)	
u owns v	
He beats it	
u beats it	
u beats v	

(8) is true in the model M provided there is an element d of U_M such that $\langle F_M(Pedro), d\rangle$ belongs to both $F_M(own)$ and $F_M(beat)$; and furthermore d

is a donkey in M—formally $d \in F_M(\textit{donkey})$, if we assume that common nouns are interpreted in the model by their extensions. This condition is fulfilled if there is a map g from $U_{\mathrm{m}(8)}(= \{u, v\})$ into U_M which preserves all conditions specified in $m(8)$. Note that $g(v)$ is not required to be the bearer in M of some particular name, but only to belong to the extension of the noun *donkey*.

Before turning to the donkey sentences (1) and (2) of Section 1 let us take stock of some principles applied in the construction of the DR's which we have encountered so far:

1 Certain singular terms, among them proper nouns and indefinite descriptions, provoke the introduction of items into the DR that function as the "referents" of these terms. We shall later address the question which singular terms give rise to such introductions and whether these introductions are obligatory or optional.

2 Other singular terms, viz. personal pronouns, do not introduce elements into the DR, instead they can only refer to items which the DR already contains.[5]

2.2. *Conditionals*

Our next aim is to construct a representation for the "donkey sentence" (1), which for convenience we repeat here:

(1) If Pedro owns a donkey he beats it.

Before we can deal with (1) however, we must say something about conditionals in general.

The semantic analysis of natural language conditionals is a notoriously complicated matter, and it seems unlikely that any formally precise theory will do justice to our intuitions about all possible uses of sentences of this form. The literature on conditionals now comprises a number of sophisticated formal theories, each of which captures some of the factors that determine the meaning of conditionals in actual use.[6] Although these theories differ considerably from each other they all seem to agree on one principle, namely that a conditional

(9) If A then B

is true if and only if

---

[5] As we have already observed, this is not quite correct, since a pronoun can be used deictically, in which case the referent need not belong to the DR; we shall, however, ignore the deictic use of pronouns in the course of this paper.

[6] See for example Lewis (1973), Turner (1981), Veltman (1976), Kratzer (1979).

(10) Every one of a number of ways in which *A* can be true constitutes, or carries with it, a way of *B*'s being true.

Up to now this principle has generally been interpreted as meaning that *B* is true in, or is implied by, every one of a certain set of *relevant possible situations* in which *A* is true. (This is true in particular of each of the theories mentioned in the last footnote.) The analysis of truth in terms of DR-embeddability, however, creates room for a slightly different implementation of (10).

Where *M* is a model and *m* a DR for the antecedent *A* there may be various proper embeddings of *m* into *M*, various ways, we might say, of showing that *A* is true in *M*. This suggests another interpretation of (10), viz. that each such way of verifying *A* carries with it a verification of *B*. In what sense, however, could such a way of verifying *A*—i.e. such a proper embedding of *m*—entail a verification of *B*? To verify *B*, in that sense of the term in which we have just been using it, we need a representation of *B*; but as a rule the content of *B* will not be represented in the DR *m* of *A*. To verify *B* in a manner consistent with some particular verification of *A* we must therefore extend the DR *m* involved in that verification to a DR m' in which *B* is represented as well. Thus we are led to an implementation of (10) according to which the conditional (9) is *true*, *given* a pair (m, m'), consisting of a DR *m* of *A* and an extension m' of *m* which represents *B* as well, iff

(11) every proper embedding of *m* can be extended to a proper embedding of m'.[7]

This is not yet an explicit statement of the truth conditions of (9), for it fails to tell us anything about the target structures of the verifying embeddings, and about their relation to the situation, or model, with respect to which (9) is evaluated. Here we face all the options that have confronted earlier investigators. We may elaborate (11) by stipulating that (9) is true in a model *M* iff every proper embedding of *m* into *M* is, or is extendable to, a proper embedding of m' on *M*. Or we may insist that (9) is true in the possible world *w* iff every proper embedding of *m* into any of the (models representing the) nearest *A*-worlds induces some proper embedding m' into that world. Indeed, any one of the existing theories could be combined with the principle conveyed by (11).

[7] (11) is akin in spirit to the game theoretical analysis of *if ... then ...* sentences proposed in Hintikka & Carlsson (1979), according to which a winning strategy for the defender of *if A then B* is a function which maps every winning strategy for the defender of *A* onto a winning strategy for the defender of *B*.

Here we shall, primarily for expository simplicity, adopt the first of the options mentioned:

(12) Let m be a DR of A and m' an extension of m which incorporates the content of B. Let M be a model. Then *if A then B* is true in M, given (m, m'), iff every proper embedding of m into M can be extended to a proper embedding of m' into M.

For conditionals in which there are no anaphoric links between antecedent and consequent, (12) boils down to the truth conditions for the material conditional. But where such a link exists its implications are somewhat different. To see this let us apply the condition to (1). We have already constructed DR's of the kind needed in the application of (12) to (1), namely, $m_1(8)$, and $m(8)$. According to (12), (1) is true in M given $(m_1(8), m(8))$, iff every function f from $U_{m_1(8)} (= \{u, v\})$ into U_M such that (i) $f(u) = F_M(Pedro)$. (ii) $f(v) \in F_M(donkey)$, and (iii) $\langle f(u), f(v)\rangle \in F_M(own)$, can be extended to a function g from $U_{m_1(8)}$ into U_M such that $\langle g(u), g(v)\rangle \in F_M(beat)$. Of course, in the present case $U_{m(8)} = U_{m_1(8)}$ and consequently there is no question of *extending* f to g. So the above condition reduces to the stipulation that every f as described has the additional property that $\langle f(u), f(v)\rangle \in F_M(beat)$. Clearly this condition is equivalent to the truth in M of the formula (3) which we adopted in Section 1 as giving the truth conditions of (1).

It is easy enough, however, to come up with examples which do involve the extension of embeddings, e.g.:

(13) If Pedro owns a donkey he lent it to a merchant.

If we extend $m_1(8)$ to a DR which incorporates the content of the consequent of (13) we get something like:

$m(13)$

u	v	w
.	.	.
Pedro owns a donkey		
u = Pedro		
u owns a donkey		
donkey(v)		
u owns v		
he lent it to a merchant		
u lent it to a merchant		
u lent v to a merchant		
merchant(w)		
u lent v to (w)		

In relation to $m_1(8)$ and $m(13)$, (12) requires that every mapping f of the kind described in the preceding analysis of (1) can be extended to a function g from $\{u, v, w\}$ into U_M such that—if we assume for simplicity that *lent to* is interpreted in M as a set of ordered triples of members of U_M - (i) $g(w) \in F_M(merchant)$; and (ii) $\langle g(u), g(v), g(w)\rangle \in F_M(lentto)$.

2.3. *Universals*

One of the important insights that went into Frege's discovery of the predicate calculus was that the restricted quantification typical of natural language is expressible in terms of unrestricted quantifiers and truth functions. Our handling of indefinite descriptions, which formal logic treats as expressions of existential quantification, harmonizes with this insight. For, as can be seen for instance from $m_1(8)$, the introduction of a discourse referent u for an indefinite term is accompanied by two conditions, one to the effect that u has the property expressed by the common noun phrase of the term, and the other resulting from substituting u for the term in the sentence in which it occurs.

I wish to propose a treatment of terms of the form *every* α that is in similar accord with Frege's analysis of restricted universal quantification. Again it will be easier to illustrate the proposal before I state it. Consider:

(14) Every widow admires Pedro.

A representation for (14), like those for conditional sentences, involves a pair of DR's. The first of these states that some "arbitrary" item x satisfies the common noun *widow*, the second extends this DR by incorporating the content of the condition *x admires Pedro*. Thus we obtain:

$m_1(14)$

x
.
widow(x)

$m_2(14)$

x	u
.	.
widow(x)	
x admires Pedro	
u = Pedro	
x admires u	

The truth value of (14) in M is to be determined by $(m_1(14), m_2(14))$ in precisely the same way as that of (1) is determined by $(m_1(8), m(8))$. Thus (14) is true iff every correlation of x with an element a of U_M such that $a \in F_M(widow)$ can be extended to a proper embedding of $m_2(14)$, i.e., to

a function g such that $g(u) = F_M(Pedro)$ and $\langle g(x), g(u)\rangle = \langle a, g(u)\rangle \in F_M(admires)$. Clearly this confers upon (14) the intuitively correct truth conditions.

In the same way

(15) Every widow admires a farmer.

licenses the construction of the following pair of DR's:

$m_1(15)$

x
.
widow(x)

$m_2(15)$

x	u
.	.
widow(x)	
x admires a farmer	
farmer(u)	
x admires u	

Again the condition that every association of x with an object a that is a widow in the sense of M can be extended to a proper embedding of $m_2(15)$ gives the correct truth conditions of (15); or, to be precise, the truth conditions it has on what is generally considered its most natural reading.

Consider now the second donkey sentence of Section 1:

(2) Every farmer who owns a donkey beats it.

Sentence (2) gives rise to the following pair of DR's:

$m_1(2)$

x	v
.	.
farmer(x)	
(x) owns a donkey	
donkey(v)	
(x) owns (v)	

$m_2(2)$

x	v
.	.
farmer(x)	
(x) owns a donkey	
donkey(v)	
(x) owns (v)	
(x) beats it	
(x) beats (v)	

So (2) is true in M iff every f such that $f(x) \in F_M(farmer), f(v) \in F_M(donkey)$, and $\langle f(x), f(v)\rangle \in F_M(own)$ has the additional property that $\langle f(x), f(v)\rangle \in F_M(beat)$. This is exactly as it should be.

Our treatment of conditionals and universal sentences gives—for the cases, at any rate, that we have thus far considered—intuitively correct conditions of truth. But it seems at odds with the *general* definition of truth

which I put forward earlier, according to which a discourse is true in M, given some representation m of it, iff *there is some* proper embedding of m into M. The semantic analyses of the sentences we have considered in this section refer to pairs of DR's rather than single DR's and involve conditions on *all* proper embeddings of a certain kind, instead of demanding the existence of at least one proper embedding.

To resolve this apparent conflict I must say a little more about the intuitive ideas behind the DR construction of which we have now seen a few instances. Essential to the analysis of the majority of our examples was the way in which we have treated indefinite descriptions. It would be quite unsatisfactory if there were no other justification for that treatment than the observation that, combined with additional principles for DR construction they give the truth conditions that speakers in fact associate with the sentences we have sampled. There is, however, a reason why we should *expect* a construction principle for the indefinites such as we have applied, but no direct analogue of it for phrases of the form *every* α. Let us go back to the first sentence of (8). What justifies us in adding to the partial DR of (8) the element v as a "referent" for *a donkey* is this: as I already argued, the DR of a sentence functions as a partial description of how the world ought to be if the sentence is true. To fulfill that role the DR must represent whatever information has been encoded into it in such a way that the significance of that representation is unaffected when one extends it to incorporate further information—or, what comes in this connection to much the same, when the DR is identified as a certain structure of a larger "real world" model via some proper embedding. The conditions $u = Pedro$, $donkey(v)$ and u *owns* v which make up $m_1(8)$ clearly satisfy this requirement. They convey precisely the same information in any extension of $m_1(8)$ as they do in $m_1(8)$ itself.[8] The content of an existential sentence has been exhausted once an individual has been established which satisfies the conditions expressed by the indefinite description's common noun phrase and by the remainder of the sentence.

But a universal sentence cannot be dealt with in such a once-and-for-all manner. It acts, rather, as a standing instruction: of each individual check whether it satisfies the conditions expressed by the common noun phrase of the universal term; if it does, you may infer that the individual also

[8] The fact that "existential" quantifier phrases can be represented in this manner is closely related to the familiar model theoretic proposition that purely existential sentences are preserved under model extensions.

satisfies the conditions expressed by the remainder of the sentence. This is a message that simply *cannot* be expressed in a form more primitive than the universal itself. The universal is thus, at the level of the DR to which it belongs, *irreducible*. The same is true of conditionals. *If A then B* functions as an instruction to check, and keep checking, whether the antecedent *A* has been satisfied, and to infer, when this is found to be so, that the consequent *B* must also hold. This too is a piece of information that cannot be represented in any more elementary form.

This means that when we form the DR of a universal sentence, such as (14), or of a conditional, such as (1), we cannot decompose the sentence in some such fashion as we were able to decompose, say, the first sentence of (8) when constructing $m_1(8)$. So the DR for (14) cannot itself be elaborated beyond the trivial initial stage:

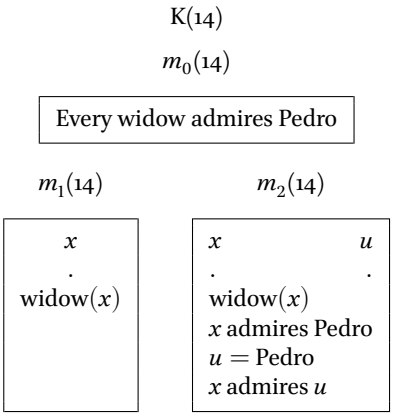
$m_0(14)$

Every widow admires Pedro

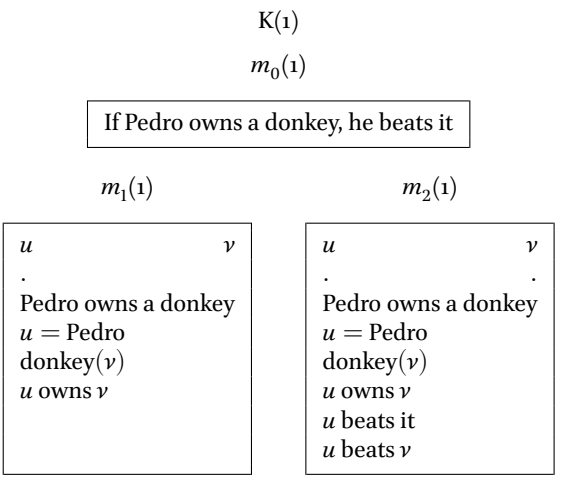
in which the sentence (14) occurs as a condition, but nothing else does.

There is however, another way in which we can represent the internal structure of (14), namely by constructing separate DR's for its components, and by integrating these DR's into a structure in which their connection reflects the syntactic construction by means of which these different components are amalgamed into the complex sentence. This is, in fact, essentially, what I did when constructing the DR-pairs I earlier presented for (1), (14), (15), and (2).

But these pairs do not provide, by themselves, the structural representations to which we can apply our general definition of truth. To obtain such a representation for, say, (14) we must combine the pair $(m_1(14), m_2(14))$ with the DR $m_0(14)$. This gives us the following structure:



Similarly the complete representation for (1) will now look thus:



It may appear as if something is still missing from these structures. For what tells us that the subordinate DR's $m_1(1)$ and $m_2(1)$ represent the antecedent and consequent of a conditional, while $m_1(14)$ and $m_2(14)$ represent the components of a universal? The answer to this is simple: the necessary information is provided by the sentences in $m_0(1)$ and $m_0(14)$ whose components are represented by the subordinate DR's $m_1(1)$, $m_2(1)$, and $m_1(14)$, $m_2(14)$. In fact we shall assume that with each syntactically well-formed sentence is given a particular syntactic analysis of it, which specifies unambiguously its immediate components and the construction which forms the sentence out of these. (For the fragments we shall study in Section 3, this condition will be automatically fulfilled as each of its wellformed expressions has a unique

syntactic analysis.) The role which, say, $m_1(1)$ and $m_2(1)$ play in the representation of (1) can thus be recognized by comparing their relevant entries, viz., *Pedro owns a donkey* and *he beats it*, with the syntactic analysis of the sentence (1) to be found in $m_0(1)$. All this will be discussed in detail in Section 3.

A representation of the sort just displayed, which involves structured families of DR's, will be called a *Discourse Representation Structure* or, for short, DRS. Each sentence or discourse includes the construction of such a DRS, and only where the sentence or discourse is comparatively simple will the DRS consist of a single DR only. Among the DR's that constitute a DRS there will always be one which represents the discourse as a whole. (In the two DRS's we displayed these are, respectively, $m_0(14)$ and $m_0(1)$. This DR will be called the *principal* DR of the DRS.

Once we assign to (1) the DRS K(1) the earlier conflict between the general definition of truth and our particular account of the truth value of a conditional can be resolved. We slightly modify the truth definition to read:

(16) A discourse *D is true in M, given* a DRS *K* of *D* iff there is a proper embedding into *M* of the principal DR of *K*.

Let us apply (16) to (1) and its DRS K(1). (1) is true given K(1) iff there is a proper embedding of $m_0(1)$ into *M*. Since the universe of $m_0(1)$ is the empty set, there is only one embedding from $m_0(1)$ into *M*, viz. the empty function, Λ. What is it for Λ to be proper? Λ is proper iff the conditions of $m_0(1)$ are true in *M* of the corresponding elements of U_M. In the present case however there are no elements in $U_{m_0(1)}$, thus no corresponding elements of U_M; and there is only one condition in $m_0(1)$, namely (1) itself. Thus Λ is proper iff (1) is true in *M*.

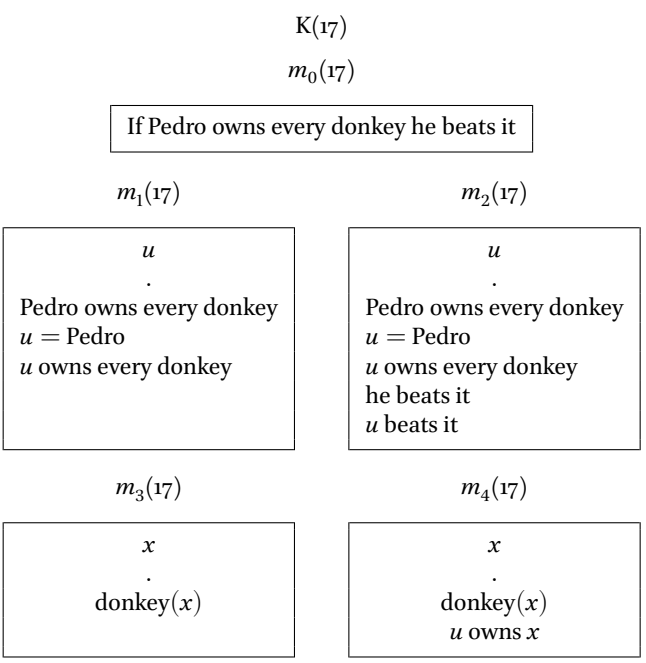
It might seem at this point that we are trapped in a circle. But in fact we are not. To see that we are not it is necessary to appreciate the difference between (i) asking for the truth value in *M* of (1), given K(1); and (ii) asking for the truth value in *M* of some condition that belongs to some member of K(1). This second question has, as we saw earlier, a straightforward answer when the condition has the form of an atomic sentence. For in that case it is directly decided by the embedding and the function F_M. But when the condition is a complex sentence, e.g., a conditional or a universal, which permits no further analysis *within the very DR to which it belongs*, the answer involves an appeal to certain members of the DRS that are *subordinate* to that DR. Thus the condition (1) of $m_1(1)$ is to be taken as true in *M* iff it is true, in the sense defined earlier, *given* the pair $(m_1(1), m_2(1))$ of DR's subordinate to $m_0(1)$; and in *that* sense (1) is true in *M*, we saw already, iff *M* verifies the first order formula (3).

To see more clearly how the various components of our theory are to be fitted together, we should look at a few more examples.

The next example shows why it is that certain anaphoric connections are impossible. In

(17) If Pedro owns every donkey then he beats it.

it cannot have *every donkey* for its antecedent. The reason for this becomes transparent when we try to construct a DRS which gives such a reading to (17):



We cannot complete this DRS as intended, for the discourse referent x, which we want to assign to the pronoun *it* of $m_2(17)$, is not available, as it occurs only at the level of $m_3(17)$, which is below that of $m_2(17)$. A similar explanation shows why *it* cannot be anaphorically linked to *every donkey* in

(18) Every farmer who owns every donkey beats it.

and also why *it* cannot be anaphorically linked to *every donkey* in

(19) If Pedro likes every woman who owns a donkey he feeds it

it cannot be co-referential with *a donkey*, whereas such a link does seem plausible in

(20) If Pedro likes a woman who owns a donkey he feeds it.[9]

These last examples give, I hope, an inkling of the predictive powers of what in particular linguists might think constitutes the most unusual feature of the theory I have so far sketched: the fact that it handles singular terms of the forms *a* β and *every* β in entirely different ways. I hope that these and subsequent illustrations will help to persuade them that the conception of a perfect rule-by-rule parallelism between syntax and semantics is one that must be proved rather than taken for granted.[10] In fact, the data here presented point towards the conclusion that this conception is ultimately untenable.

Another feature that distinguishes the present account from many, albeit not all, existing theories of reference and quantification is its entirely uniform treatment of third person personal pronouns. This has already been apparent from the examples at which we have looked. It is further illustrated by such sentences as:

(21) Every farmer courts a widow who admires him.

Occurrences such as that of *him* in (21) have been put forward as paradigms of the use of pronouns as bound variables—an identification that is natural, and in fact well-nigh inescapable, when one believes that the logical forms of natural language sentences are expressions of the predicate calculus. Indeed several earlier theorists have perceived a real chasm separating these pronoun uses from those which we find exemplified by, say, *her* in (7) and *he* in (7) and (8); and looking at pronouns from this perspective, they have often felt helpless vis-à-vis the pronoun occurrences that have been of particular concern to us in this section, viz. those exemplified by (1) and (2). Forcing these either into the mold that had been designed for uses such as that in (7), or into that measured to fit occurrences such as that of *him* in (21) turned out to be hopeless enterprises. Evans (1977, 1980) gives conclusive evidence against the latter of these two; but his own suggestions, which go some way towards assimilating the problematic pronouns to definite descriptions, do not appear to be fully satisfactory either.[11]

Note that the more unified treatment of these pronoun uses given here is possible partly because the same construction rule for pronouns operates both at the level of the principal DR's and at subordinate levels. Thus the

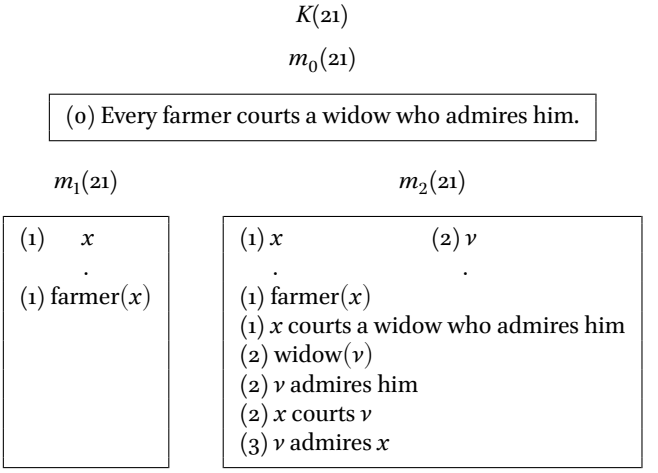
---

[9] I have found at least one speaker for whom (20) is distinctly less acceptable than for instance (1).

[10] See for example (Carlson 1976: Chapter 1), which warns against this prejudice in similar terms.

[11] Proposals similar to that of Evans can be found e.g. in Cooper (1979) and Hausser (1974). These suffer in my view from similar shortcoming.

DRS for (21) is constructed as follows (the numbers in parentheses which precede discourse referents and conditions indicate the order in which the operations are carried out; we shall often use this notational device):



The rules for pronouns applies here in just the same way to the *him* of *v admires him* in $m_2(21)$ as it does for example to the *he* and *it* in the DRS construction of (8) or the *it* of (1) in the construction of the DR of (1).

3. The Formal Theory

3.1. *Syntax*

The time has come for a more formal and systematic presentation. We shall consider a fragment of English for which I shall give an explicit syntax and explicit formal rules for DRS construction. Our fragment will be exceedingly simple to start with, much simpler even than that of Montague (1973).[12] The syntax adopted resembles Montague's, but the resemblance is rather superficial; for the syntactic analysis of a sentence will play a much more modest role in the determination of its interpretation than it does in Montague grammar. In presenting the syntax I shall presume some familiarity with Montague grammar, specifically with Montague (1970a) and Montague (1973). Our fragment, to which I shall refer as L_0, contains expressions of the following categories with the following basic members:

[12] The two fragments have roughly the same quantificational powers. But the present fragment lacks adjectives, prepositional phrases and intensional contexts.

1	*T* (Term)	:	*Pedro, Chiquita, John, Mary, Bill, ...; he, she, it*
2	*CN* (Common Noun phrase)	:	*farmer, donkey, widow, man, woman, ...*
3	*IV* (Intransitive Verb phrase)	:	*thrives ...*
4	*TV* (Transitive Verb phrase)	:	*owns, beats, loves, admires, courts, likes, feeds, loathes, ...*
5	*S* (Sentence)	:	– –
6	*RC* (Relative Clause)	:	– –

Formation Rules

FR 1. If $\alpha \in$ TV and $\beta \in$ T then $\alpha\beta' \in$ IV where $\beta' = him$ if $\beta = he$, $\beta' = her$ if $\beta = she$ and $\beta' = \beta$ otherwise.

FR 2. If $\alpha \in$ IV and $\beta \in$ T then $\beta\alpha \in$ S.

FR 3. If $\alpha \in$ CN then (i) $a(n)\ \alpha$, and (ii) *every* α are in T.

FR 4.k If $\phi \in$ S and the k-th word of ϕ is a pronoun then $\beta\phi' \in$ RC, where ϕ' is the result of eliminating the k-th word from ϕ and β is *who, whom, which*, according as the pronoun is *he* or *she, him* or *her*, or *it*, respectively.

FR 5. If α is a basic CN and $\beta \in$ RC then $\alpha\beta \in$ CN.

FR 6. If $\phi, \psi \in$ S then *if* ϕ, ψ and if ϕ *then* $\psi \in$ S.

Some Comments

1 The rule schema FR 4.k is defective inasmuch as it allows for wh-movement out of forbidden positions. Within the present fragment there are only two sorts of noun phrase positions to which wh-movement may not apply, those inside relative clauses and those inside the antecedents of conditionals. It is not difficult to modify the syntax in such a way that these restrictions are observed. For instance we could stipulate that each time a relative clause is formed all pronouns it contains are marked, and that the same is done to those occurring in the antecedent of a conditional at the time when antecedent and consequent are joined together. The rule of relative clause formation can then be altered so that it applies to unmarked pronouns only. Such a solution is rather ad hoc, so as it would moreover complicate the syntax as a whole, I have refrained from incorporating it. I must beg the reader to keep in mind that the syntax of this section is intended as no more than a convenient basis for the definition of DRS-construction rules, and that it has no pretentions of capturing important syntactic generalizations.[13]

2 The present fragment differs from most familiar versions of Montague grammar in that it contains neither variables nor indexed pronouns.[14] Conse-

[13] One might have hoped that a theory of semantic processing such as the one attempted here could provide an explanation of why island-constraints exist and why they operate in precisely those linguistic contents that are subject to them. I have not succeeded, however, in finding such an explanation.

[14] See Montague (1970a, 1970b), Partee (1975), Thomason (1976), Cooper & Parson (1976), Cooper (1979).

quently the syntactic analysis of a sentence of the present fragment tells us nothing about anaphoric relations.

3 Every well-formed expression of L_0 has a unique syntactic analysis. This is a feature that is bound to be lost at some point as we extend the present fragment. It allows us, however, to omit, while uniqueness of syntactic analysis obtains, all explicit references to syntactic analyses in discussions and, particularly, in definitions where such reference becomes essential as soon as well-formed strings do not unambiguously determine their analyses.

4 When defining the process of DRS construction we shall have to specify the order in which various parts of a given sentence are to be treated. What we need here is, in essence, a specification of scope order. I shall assume in this paper that the scope relations within a sentence are directly determined by its syntactic construction. Thus the subject term of a simple clause will always have wide scope over the object term; the *if* of a conditional sentence will always have wide scope over the terms occurring in antecedent and consequent, etc. Let us call the formation rule which is applied last in the construction of an expression γ the *outermost rule* of γ. Where γ is a sentence and the outermost rule is FR 6, γ is called a *conditional* (*sentence*). If the outermost rule of γ is FR 1 or FR 2 and this rule forms γ by combining some IV or TV with the term α, α is said to *have*, or to *be the term with, maximal scope in* γ. If the outermost rule is FR 1 and α begins with *every*, γ is called a *universal IV*; similarly, if the outermost rule of γ is FR 2 and α begins with *every*, γ is called a *universal sentence*.

By eliminating Montague's rule of substitution and quantification we have dispensed with one natural way of distinguishing between alternative scope relations—such as, for instance, the two possible relations between *a widow* and *every farmer* in

(22) A widow admires every farmer.

Sentence (22) can be generated in only one way and according to that generation the subject has wide scope over the direct object as it enters the construction of the sentence at a later stage. No syntactic analysis would thus appear to convey upon (22) the reading given by

(23) $(\forall x)(\text{farmer}(x) \rightarrow (\exists y)(\text{widow}(y) \wedge \text{admires}(x, y)))$.

It might be thought that the construction of a DRS which imposes this latter reading upon (22) involves an order of application of the construction rules which contravenes the scope relations implied by the syntax. This problem too must be left for another paper.

5 We shall refer to the basic terms *Pedro, Chiquita, John, Mary, ...* as the *proper names* of L_0 and to *he, she, it* as the *pronouns* of L_0. Terms of the form *every* β will be called *universal terms*.

6 I have admitted only compound common noun phrases consisting of a common noun and *one* relative clause. It would of course be possible to relax FR 5 so that it can attach several relative clauses to the same head noun. Many of

the resulting expressions, however, seem marginal at best. I have decided to cut the knot and keep such complex common nouns out of the fragment altogether.

3.2. *Models and Discourse Representation*

By a *model for* L_0 we shall understand a structure of the form $\langle U, F \rangle$ where (i) U is a non-empty set and (ii) F is an interpretation function which assigns an element of U to each of the proper names of L_0, a subset of U to each of its basic CN's and basic IV's, and a set of pairs of elements of U to each of its basic TV's.

We must now address ourselves to the main task of this section, the formulation of the rules of DRS-construction and of the definition of truth for L_0. To state the rules we shall have to decide on a format for DR's and DRS's. In choosing such a format I have been partly guided by considerations of notational convenience. In particular it is just a matter of convenience to specify (as I have already done in the examples discussed in the preceding section) that one or more discourse referents satisfy a certain predicate by adding to the relevant DR a sentence which is obtained by combining that predicate with, in the appropriate positions, these referents themselves; using them, that is, autonymously (a policy against which there can be no objection, given the symbolic nature which must be attributed to the discourse referents in any case). Almost all other features, however, of the DR-format I have chosen are determined by empirically significant aspects of the rules of DRS-construction.

Let V be a denumerable set of entities none of which is a basic expression of L_0 or a string of such expressions. V is the set from which the elements are drawn that make up the universes of the DR's. We shall often refer to the members of V as *discourse referents*. For any subset X of V let $L_0(X)$ be the result of adding the members of X to the set of basic terms of L_0.

As all our earlier examples showed, the introduction of a discourse referent is always accompanied either by a condition which identifies it as the referent of a proper name or else by one which stipulates that it satisfies some common noun. These conditions cannot be expressed in $L_0(X)$; so we must slightly extend the notation which that language provides. We shall allow in addition to what $L_0(X)$ contains already, sentences of the form $u = \alpha$ where α is a proper name and $u \in X$, to express the former, and sentences of the form $\beta(u)$ where, again, $u \in X$ and $\beta \in$ CN, to express the latter type of condition. We shall refer to the language obtained from $L_0(X)$ through these additions as $L_0'(X)$.

We shall limit ourselves here to the simplest type of discourse, that of a discourse constituted by a finite sequence of declarative statements, made by one and the same speaker. Formally we shall identify—as in fact we already did implicitly in Section 1—such a discourse with the sequence of the uttered sentences. So let us, where L is any language, define an *L-discourse* to be any finite string of sentences of L.

The examples we considered in the preceding section were carefully chosen so that the same singular term would never occur more than once. This made it unnecessary to distinguish between different occurrences of the same expression. In general, however, different occurrences must be kept apart. The need for this is most obvious in connection with pronouns—it is only too common a phenomenon that the very same pronoun occurs twice in a bit of discourse, but each time refers to a different individual, as e.g. might be intended by someone using the sentence.

(24) If Bill courts a widow who admires him then Pedro courts a widow who admires him.

But in longer stretches of discourse other expressions are liable to recur as well. Although the DRS construction rules defined below only require us to keep track of the individual occurrences of certain expressions, little if anything would be gained by introducing a mechanism for distinguishing just *those* individual occurrences. In fact probably the simplest way to distinguish the individual expression occurrences is this: Let $D = \langle \phi_1, \ldots, \phi_n \rangle$ be an L_0-discourse and let $\langle \tau_1, \ldots, \tau_n \rangle$ be the sequence of the (uniquely determined) syntactic analyses of the sentences of D. It is easy to formulate an algorithm which assigns a unique index,—say, a positive integer—to each of the nodes of these analyses, and, by proxy, also to the expressions formed at any such node. For instance we enumerate first all the nodes of τ_1, in some order fixed by its structure, then those of τ_2, etc., until we have dealt with the entire discourse. There is no point to go into greater detail here. We shall simply assume that one such algorithm has been fixed. By an *occurrence* of an expression α *in* D we shall understand a pair $\langle \alpha, n \rangle$ where n is the index of a node of the syntactic analysis of one of the sentences of D to which α is attached.

The relation which holds between two expressions α and β if α is a subexpression of β has an obvious counterpart between expression occurrences: $\langle \alpha, n \rangle$ is a "suboccurrence" of $\langle \beta, m \rangle$ if $\langle \alpha, n \rangle$ occurs as part of the syntactic analysis of $\langle \beta, m \rangle$. I shall often speak, by a minor sleight of hand, of one expression occurrence being a *subexpression* (*subformula*, etc.) *of* some other occurrence. No confusion should arise from this.

The construction of a DRS for D does not only require the separate identification of particular occurrences of expressions of L_0; we must also be able to keep track of different occurrences of the same expressions of $L_0'(X)$. However, as our examples have already indicated (and we shall soon make this fully explicit) the expressions from $L_0'(X) \setminus L_0$ which enter into DR's are always derived from corresponding expressions of L_0. To be specific, they result either (i) through one or more substitutions of members of X for singular terms in some sentence of L_0; or (ii) from placing a member of X in parentheses behind a CN of L_0; or (iii) from combining a member of X with $=$ and a proper name of L_0. In the first case we can label the $L_0'(X)$-sentence occurrence unambiguously with the index of the occurrences of the L_0-sentence from which it is obtained through successive substitutions; in the second case we assign the index of the relevant occurrences of the common noun; and in the third the index of the relevant occurrence of the proper name. In each of these cases (i), (ii), and (iii), we shall say that the sentence of $L_0'(X)$ is a *descendant* of the relevant expression of L_0, and similarly that the occurrence of the $L_0'(X)$-sentence is a *descendant of* the corresponding occurrence of an expression of L_0. Formally we shall represent any occurrence of such an expression also as a pair consisting of the expression together with the appropriate index.

There is one other notion which we have already defined for L_0 but which must also be extended to cover certain expressions of $L_0'(X)$ as well. This is the notion of the *outermost rule of* an expression. We shall need to refer to the outermost rule only of those sentences of $L_0'(X) \setminus L_0$ which result from making in sentences of L_0 one or more substitutions of members of V for occurrences of singular terms of $L_0(X)$. Any such substitution leaves the syntactic structure of the sentence in which it takes place essentially inviolate: it can only lead to some "pruning" of the syntactic tree, viz. where the replaced singular term occurrence is itself complex. In that case the subtree dominated by the node to which the singular term α is attached is deleted and replaced by a single node to which is attached the inserted (basic) term u. The outermost rule FR i of the resulting sentence should *not* count as outermost rule of the syntactic analysis of the substitution result. For FR i is the rule which combines u with the remainder γ of the sentence, and this is a syntactic operation which, unlike the analoguous operation that combines the replaced singular term with γ, should give rise to no further step in the DRS construction (the singular term α has after all just been dealt with!) Thus we should identify as the *outermost rule of* the substitution result rather the outermost rule of γ. Since, as we already observed, each of the $L_0'(X)$-sentences in question results from a finite

sequence of such substitutions the above stipulation defines the outermost rule of each such sentences.

Having extended the concept of the outermost rule of an expression to certain sentences of $L_0(X)$ we can now also apply the notions *conditional* and *universal sentence* to those sentences. Moreover, we shall call *atomic* those sentences of $L_0'(X)$ which consist either (i) of a discourse referent followed by an IV; or (ii) a TV flanked by two discourse referents; or (iii) a CN followed by a discourse referent in parentheses; or (iv) a discourse referent followed by $=$ and a proper name of L_0.

Here is the definition of the "format" of Discourse Representations I have chosen, as well as of some related notions which we shall need in later definitions:

Definition 1 Let D be an L_0-discourse.

1 A *possible DR* (*Discourse representation*) *of D* is a pair $\langle U, \text{Con}\rangle$, where
   (i) U is a subset of V; and
   (ii) Con is a set of occurrences in D of sentences of $L_0(U)$.

2 Where m and m' are possible DR's for D we say that m' *extends* m if $U_m \subseteq U_{m'}$ and $\text{Con}_m \subseteq \text{Con}_{m'}$.

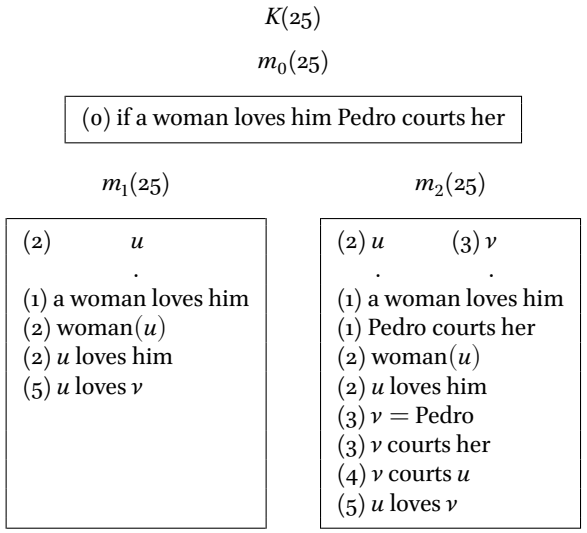
3 Let m be a possible DR for D. A sentence $\phi \in \text{Con}_m$ is called *unreduced in* m iff Con_m contains no descendant of ϕ. m is called *maximal* if each unreduced member of Con_m is either (i) an atomic sentence, (ii) a conditional, or (iii) a universal sentence.

We have seen in Section 2 that in general we must associate with a given discourse a Discourse Representation *Structure*, i.e. a partially ordered family of DR's, rather than a single DR. As it turns out the partial orders of those DRS's which our rules enable us to construct can always be defined in terms of the internal structure of their members. This makes it possible to define a DRS as a set of DR's.

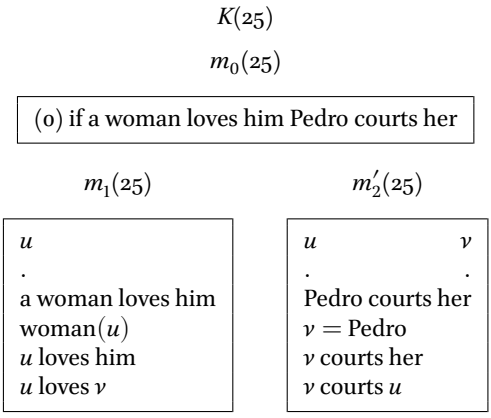
To show how the partial order can be defined in terms of the structure of the DR's that make up the DRS we have to make explicit the structural relationship that holds between a DR m which contains a conditional or universal sentence ϕ and the pair of DR's which must be constructed to represent the content of ϕ. But before we can do that we must first discuss, and introduce, a slight modification of the schema for representing conditionals and universals that we have used in our examples. So far we have represented a conditional *if A* (*then*) *B* by a DR m_1 of A together with an extension m_2 of m_1 which incorporates into it the information contained in B. There can be no objection to this schema as long as the information contained in A can be fully processed in m_1 before one extends it by processing B. It is not always possible, however, to proceed in this way, as illustrated by (25).

(25) If a woman loves him Pedro courts her.

The order in which the construction rules must be applied to yield a DRS which links *him* with *Pedro* and *her* with *a woman*, is indicated in the following diagram:



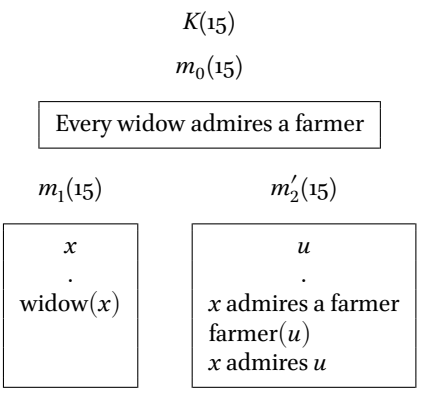
Not only is there duplication of the conditions which occur both in $m_1(25)$ and $m_2(25)$ but some of the operations have to be performed simultaneously *and in the same way*, on the identical entries of these two DR's. It would be possible to characterize DRS-construction so that such entries are treated simultaneously in all the DR's in which they occur, and give rise in each of these DR's to the same descendants. But this is awkward, particularly where the treatment produces new subordinate DR's. It is easier to introduce into the second DR of the pair representing a conditional only the information conveyed by the consequent. In the case of (25) this will lead to a DRS of the form:



Similarly, we shall represent a universal sentence by a pair of DR's into the second of which we enter the information that the remainder of the sentence is true of the discourse referent which stands in for the singular term *every* β in question. For example the DRS K(15) for

(15) Every widow admires a farmer.

now becomes



Evidently the second members of the representing pairs about which we have been speaking up to now can be reconstructed from these new pairs: where (m_1, m_2) is the old pair and (m_1, m_2') the pair which replaces it according to the present stipulation, m_2 is the union of m_1 and m_2', where the *union* of two DR's $\langle U_1, \mathrm{Con}_1\rangle$ is the DR $\langle U_1 \cup U_2, \mathrm{Con}_1 \cup \mathrm{Con}_2\rangle$—thus, in particular, $m_2(15)$ is the union of $m_1(15)$ and $m_2'(15)$, and $m_2(25)$ that of

$m_1(25)$ and $m_2'(25)$. Note that the truth clause (12) for conditionals and its analogue for universal sentences are not affected by this change.

Let us now describe how we can recognize two DR's m_1 and m_2 as representing a conditional or universal sentence that occurs among the conditions of the DR m. We first assume that m contains the occurrence $\langle \phi, k\rangle$ that ϕ is a conditional and that its antecedent and consequent are, respectively, $\langle \psi, r\rangle$ and $\langle \chi, s\rangle$.[15] We say that the pair of DR's $\langle m_1, m_2\rangle$ *represents* $\langle \phi, k\rangle$ iff:

(i) $\langle \psi, r\rangle \in \mathrm{Con}_{m_1}$ and every member of Con_{m_1} is a descendant of a subexpression of $\langle \psi, r\rangle$;
(ii) $\langle \chi, s\rangle \in \mathrm{Con}_{m_2}$ and every member of Con_{m_2} is a descendant of a subexpression of $\langle \chi, s\rangle$.

Now suppose $\langle \phi, k\rangle$ is a universal sentence. Here it is convenient to distinguish between the case where the term with maximal scope is of the form *every* β, where β is a basic CN and that where it has the form *every* $\beta\gamma$ with β a CN and γ a RC. Let us begin by considering the first of these. We say the pair $\langle m_1, m_2\rangle$ *represents* $\langle \phi, k\rangle$ iff for some $x \in V$ (i) $x \in U_{m_1}$; (ii) $\mathrm{Con}_{m_1} = \{\langle \beta(x), i\rangle\}$; (iii) $\langle \phi', k\rangle \in \mathrm{Con}_{m_2}$ and each member of Con_{m_2} is a descendant of a subexpression of $\langle \phi', k\rangle$, where i is the index of the occurrence of β in the term (occurrence) *every* β in question and ϕ' is the result of replacing that term occurrence in ϕ by x.

Now consider the case where the term with maximal scope has the form *every* $\beta\gamma$, where β is a common noun and γ a relative clause. In this case $\langle m_1, m_2\rangle$ *represents* $\langle \phi, k\rangle$ iff for some $x \in V$ (i) $x \in U_{m_1}$; (ii) $\langle \beta(x), i\rangle, \langle \delta, r\rangle \in \mathrm{Con}'_{m_1}$ and every member of Con_{m_1} other than $\langle \beta(x), i\rangle$ is a descendant of an occurrence of a subexpression of $\langle \delta, r\rangle$; and (iii) $\langle \phi', k\rangle \in \mathrm{Con}_{m_2}$ and every member of Con_{m_2} is a descendant of an occurrence of a subexpression of $\langle \phi', k\rangle$—here i and ϕ are as above, r is the index of the occurrence of γ in the relevant occurrence of *every* $\beta\gamma$ and δ is determined as follows: let ζ be the sentence from which the relative clause has been formed through "wh-movement"; δ is obtained by substituting x in ζ for the pronoun occurrence which was eliminated in the transition from ζ to γ.

Next we must give the definition of *partial Discourse Representation Structures*.

Definition 2 A *partial DRS* (*Discourse Representation Structure*) *for D* is a set K of possible DR's for D such that whenever m is a member of K and Con_m

[15] With the *occurrence* $\langle \phi, k\rangle$ are associated, of course, particular *occurrences* of antecedent and consequent.

contains a conditional or universal sentence $\langle \phi, k \rangle$ then there is at most one pair of members m_1 and m_2 of K which represents $\langle \phi, k \rangle$.

We say that a member m' of K is *immediately subordinate to* m iff either (i) there is a conditional or universal sentence occurrence $\langle \phi, k \rangle \in \mathrm{Con}_m$ such that m' is the first member of a pair which represents $\langle \phi, k \rangle$; or (ii) m is itself the first member of such a pair and m' is the second member of that pair. m' is *subordinate to* m iff there exists a finite chain of immediate subordinates connecting m and m'.

The rules for constructing DRS's will guarantee that they will always have a principal member. If the partial DRS K contains such a member it will be denoted as $m_0(K)$. Where K and K' are partial DRS's we say that K' *extends* K iff there is a 1–1 map f from K into K' such that for each $m \in K f(m)$ extends m. For $m \in K$ we denote as $K^{\geq}(m)$ the set consisting of m and all the members of K that are superordinate to m. We shall also write "U_K" for "$\cup_{m \in K} U_m$" and "$U_K^{\geq}(m)$" for "$U_m \cup \bigcup \{U_{m'} : m' \in K$ and m' is superordinate to $m\}$". We say that a partial DRS K is *complete* iff (i) every member of K is maximal; and (ii) whenever m is a member of K and Con_m contains an occurrence of $\langle \phi, k \rangle$ of a conditional or universal sentence, K contains a pair which represents $\langle \phi, k \rangle$.

We can now proceed to give a precise statement of the rules for DRS-construction. It is they, I must repeat here, that carry virtually all the empirical import of the theory. Their exact formulation is therefore of the greatest importance. Instead of trying to do justice to all relevant linguistic facts at once, I shall begin by stating the rules in a fairly simple manner. This will then serve as a basis for further exploration.

For the fragment L_0 there are five rules, one for proper names, one for indefinite descriptions, one for pronouns, one for conditionals and one for universal terms. The effect of applying a rule to a particular condition in some members of a DRS is always an extension of that DRS.

Only the rules for conditionals and universals lead to the introduction of new DR's. But this does not mean that the effect of each of the other rules is confined to the particular DR m which contains the condition to which the rule is applied. Thus, for instance—and this is a point we have so far neglected in our examples—the application of the rule for proper names will always result in the introduction of a new discourse referent into the principal DR of the DRS, even if the condition to which the rule is being applied belongs itself to some other member of the structure. (I shall argue below that the rule for proper names *must* operate in this fashion.) Directly connected with this is the need to refer, in the statement of the rule for pronouns, not just to the universe of the DR m that contains the relevant

condition, but also to the universes of certain other members of the DRS—in fact, as it turns out, of all those members which are superordinate to m.

To state the first three rules let us assume that K is a partial DRS, that $m \in K$, that $\langle \phi, k\rangle \in \mathrm{Con}_m$ is an unreduced member of m, and that $\langle \alpha, i\rangle$ is an occurrence of a term in $\langle \phi, k\rangle$ which has maximal scope in $\langle \phi, k\rangle$.

CR 1. Suppose α is a proper name. We add to $U_{m_0(K)}$ an element u from $V \setminus U_K$. Furthermore, we add to $\mathrm{Con}_{m_0(K)}$ the occurrence $\langle u = \alpha, i\rangle$ and to Con_m the occurrence $\langle \phi', k\rangle$, where ϕ' is the result of replacing the occurrence of α in $\langle \phi, k\rangle$ with index i by u.

CR 2. α is an indefinite singular term. (a) α is of the form $a(n)\beta$, where β is a common noun. We add to U_m an element u from $V \setminus U_K$ and to Con_m the occurrences $\langle \beta(u), r\rangle$ (where r is the index of the occurrence of β in $\langle \alpha, i\rangle$) and $\langle \phi', k\rangle$, where ϕ' is as under CR 1. The other members of K remain unchanged. (b) α is of the form $a(n)\beta\gamma$, where β is a basic common noun and γ a relative clause. We add $u \in V \setminus U_K$ to U_m and expand Con_m with $\langle \beta(u), r\rangle$, $\langle \phi', k\rangle$ and the pair $\langle \delta, s\rangle$ where δ is determined as in the definition of *represents* given above, and s is the index of the occurrence of γ in $\langle \alpha, i\rangle$.

CR 3. Assume α is a pronoun. Choose a "suitable" member u from $U_{K^{\geq}(m)}$. Add $\langle \phi', k\rangle$ to Con_m, where φ' is as under CR 1.

NB. I have given a deliberately "fudgey" formulation of this rule by inserting the word "suitable". To state what, in any particular application of the rule, the set of suitable referents is, we would have to make explicit what the strategies are that speakers follow when they select the antecedents of anaphoric pronouns. In the applications we shall consider below the restriction to "suitable" referents that I have built into CR 3 will never play an overt role (although I will occasionally ignore, without comment, readings of the sampled sentences which would impose anaphoric links that are ruled out by various factors that enter into these strategies, such as e.g. the principle of gender agreement). Nonetheless, I have included "suitable" in the formulation of CR 3, as a reminder that the rule is incomplete as it stands.

To state the last two rules let us assume that K and m are as above, that $\langle \phi, k\rangle$ is an unreduced member of Con_m and that ϕ is either a universal sentence or a conditional.

CR 4. Assume $\langle \phi, k\rangle$ is a conditional with antecedent $\langle \psi, r\rangle$ and consequent $\langle \chi, s\rangle$. We add to K the members $\langle \emptyset, \{\langle \psi, r\rangle\}\rangle$ and $\langle \emptyset, \{\langle \chi, s\rangle\}\rangle$.

CR 5. Assume $\langle \phi, k\rangle$ is a universal sentence and the term with maximal scope is $\langle \textit{every}\ \beta, i\rangle$ with β a basic CN. We add, for some $u \in V \setminus D_K\langle \{u\}, \{\langle \beta(u), r\rangle\}\rangle$, and $\langle \emptyset, \{\langle \phi', k\rangle\}\rangle$, where r and ϕ' are as in the definition of *represents* 2 pages above. Similarly, where the term with maximal scope is $\langle \textit{every}\ \beta\gamma, i\rangle$ where $\beta \in$ CN and $\gamma \in$ RC the DR's that must be added are $\langle \{u\}, \{\langle \beta(u), r\rangle, \langle \delta, s\rangle\}\rangle$ and $\langle \emptyset, \{\langle \phi', k\rangle\}\rangle$, where, again, $u \in V \setminus U_K$ and s, δ, ϕ' are as in the statement of CR 2.

Note that if K is a finite DRS, i.e. a finite set of finite DR's, then a finite number of applications of the rules CR 1–CR 5 will convert it into a complete DRS. Any complete DRS obtained from K by a series of rule applications is called a *completion of K*. Clearly, if K has a principal member, then so does every completion of K.

We can at last define the notion of a *complete DRS for* a discourse D. The definition proceeds by recursion on the length of D.

Definition 3 (i) Suppose D is a discourse consisting of one sentence ϕ. Let k be the index of ϕ in D. A *complete DRS* (*Discourse Representation Structure*) *for D* is any completion of the DRS $\{\langle\emptyset, \{\langle\phi, k\rangle\}\rangle\}$. (ii) Suppose that D has the form $\langle\phi_1, \ldots, \phi_n, \phi_{n+1}\rangle$ and that the set of complete DRS's for the discourse $D' = \langle\phi_1, \ldots, \phi_n\rangle$ has already been defined. Let k be the index of the occurrence of ϕ_{n+1} as last sentence of D. Then K is a *complete DRS for D* iff K is a completion of a DRS of the form $(K' \setminus \{m_0(K')\}) \cup \{m\}$, where K' is some complete DRS for D' and m is the DR $\langle U_{m_0(K')}, \mathrm{Con}_{m_0(K')} \cup \{\langle\phi, k\rangle\}\rangle$.

NB. It follows from this definition together with earlier remarks that every set of possible DR's which is a complete DRS for some discourse D contains a principal DR.

3.3. *Truth*

Our next task is to define truth. Much has already been said about this in the preceding chapters. So we can proceed with the formal definition almost at once.

There is just one feature of the definition that might be puzzling without a brief preliminary discussion. The evaluation of conditionals and universals as a rule involves only embeddings that respect certain previously assigned values to some of the discourse referents in superordinate positions. In other words we keep, in the course of such evaluations, certain functions fixed and consider only embeddings compatible with these functions. This means that the recursive definition underlying the characterization of the truth in M must be of a concept which is sensitive not only to the information encoded in the DRS but also to some partial function from the discourse referents of that DRS into U_M. If a sentence contains several nested embeddings of conditionals or universals, the maps considered in the evaluation of deeply embedded constructions may have to agree with several functions that have been stored, so to speak, along the way down to the conditional or universal concerned. However, as these stored functions must also be compatible with each other we need consider only single

functions in this connection; intuitively these are the unions of the sets of different functions accumulated along the path towards the embedded construction.

Let K be a complete DRS for D and M a model of D. We shall give the definition of the *truth value of D in M given K* in two steps. The first stage will give a characterization, by simultaneous recursion, of two relations: (i) The relation which holds between a member m of K, a function f from U_m into U_M and a partial function g from U_K into U_M iff, as we shall express it, *f verifies m in M given K, relative to g*; and (ii) the relation which holds between m, an unreduced member $\langle \phi, k\rangle$ of Con_m, a function f from U_m into U_M and a function g from U_K into U_M iff, as we shall say, *$\langle \phi, k\rangle$ is true in M under f, given K, relative to g*. The second stage uses the first of these two relations to define truth:

Definition 4 Let D be an L_0-discourse, K a complete DRS for D and M a model for L_0. D is *true in M on K* iff there is a function f from $U_{m_0(K)}$ into U_M which verifies $m_0(K)$ in M, given K, relative to Λ. (Λ is the empty function!).

The recursive part of the definition is inevitably somewhat more involved.

Definition 5 Let D, K, M be as in Definition 4; let $m \in K$ and let g be a partial function from U_K into U_M.

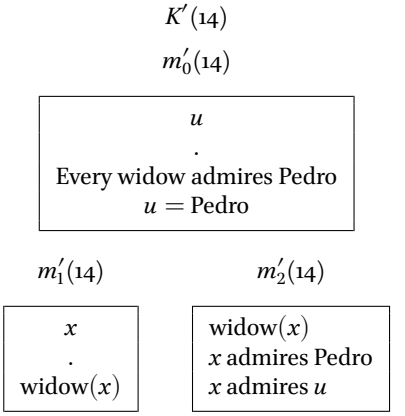
(i) *f verifies m in M given K, relative to g* iff each unreduced member $\langle \phi, k\rangle$ of Con_m is true in M under f, given K, relative to g.

(ii) Suppose $\langle \phi, k\rangle$ is an occurrence of an atomic sentence in Con_m. Then ϕ has one of the following four forms:
   (a) $u\alpha$, where $u \in V$ and $\alpha \in IV$;
   (b) $u\alpha v$, where $u, v \in V$ and $\alpha \in TV$;
   (c) $u = \alpha$, where $u \in V$ and α is a proper name;
   (d) $\alpha(u)$, where $u \in V$ and α is a basic common noun.
   The question whether *$\langle \phi, k\rangle$ is true in M under f given K, relative to g* splits up into the corresponding four clauses below (we omit the qualification "in M under f, given K relative to g"):
   (a) $\langle \phi, k\rangle$ *is true* iff $f(u) \in F_M(\alpha)$;
   (b) $\langle \phi, k\rangle$ *is true* iff $\langle f(u), f(v)\rangle \in F_M(\alpha)$;
   (c) $\langle \phi, k\rangle$ *is true* iff $f(u) = F_M(\alpha)$;
   (d) $\langle \phi, k\rangle$ *is true* iff $f(u) \in F_M(\alpha)$;

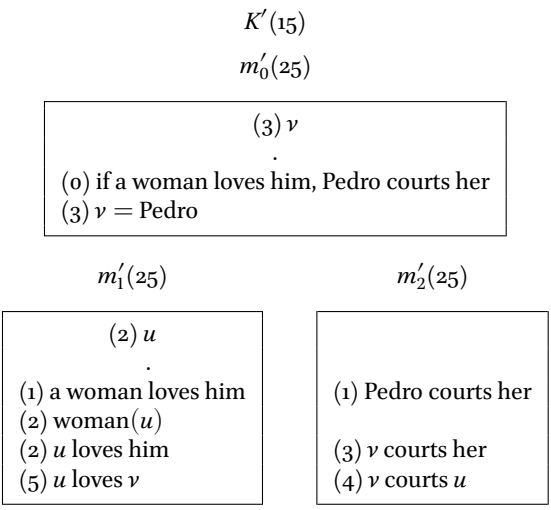
(iii) Suppose $\langle \phi, k\rangle$ is an occurrence of a conditional or universal sentence in Con_m. Then K will contain a unique pair $\langle m_1, m_2\rangle$ which represents $\langle \phi, k\rangle$. *$\langle \phi, k\rangle$ is true in M under f given K, relative to g* iff every map h from U_{m_1} into U_M which is compatible with $g \cup f$ and which verifies m_1 in M given K relative to $g \cup f$ can be extended to a function k from U_{m_2} into U_M which verifies m_2 in M given K relative to $g \cup f$.

We shall call a function which verifies $m_0(K)$ in M, given K, relative to Λ a *verifying* or *truthful embedding of K into M*. We shall also say of such a map that it *verifies D in M on* (*the reading provided by*) *K*.

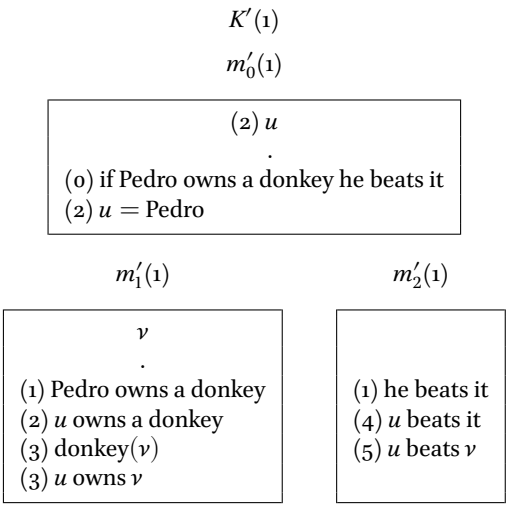
Many of the DRS's we have displayed earlier fail to be in complete agreement with the construction procedure as we have now formally described it. This is true, in particular, of the second representation I gave in Section 2.3 for (14). The DRS $K(14)$ violates the rule CR 1 in that the item u, which is introduced as the referent of the proper name *Pedro* should have been entered into the universe of $m_0(14)$ rather than into that of $m(14)$. Let us give the DRS for (14) once more, this time in its proper form.



The need to place the discourse referent introduced by a proper name into the principal DR is illustrated in (25) for which I gave a DRS in Section 3.2. This DRS is unacceptable by our rules as the referent u in $m_2(25)$ is not accessible from the position of *him* in $m_1(25)$, to which, at step (5) it was nonetheless assigned. This difficulty would not have arisen had CR 1 been properly applied in the first place. The correct DRS for (25) looks as follows:



Let us, for good measure, also give a corrected version of the DRS for (1), as the analysis of that sentence motivated so much of what I have been saying, and yet its earlier representation also contains a violation of CR 1.



We already saw in Section 2 how important it is that the discourse referents available to a given pronoun must all occur in the same, or else in some superordinate, DR. This, we saw, accounts for the fact that *it* cannot have *every donkey* as its antecedent in a sentence such as (17) or (18), or be

anaphorically linked to *a donkey* in (19). The reader will inevitably ask, however, why subordination is defined in the precise way it has been. Why, for instance is, where (m_1, m_2) represents a conditional or universal, m_2 subordinate to m_1 but not m_1 subordinate to m_2; or, to put it more directly, why may the elements of m_2 not serve as referents for pronouns in sentences belonging to Con_{m_1} while the members of U_{m_1} are admitted as referents for pronouns occurring in m_2?

That the elements of m_1 must be available for the pronouns of m_2 is too central an assumption of our theory to permit tampering: our analysis of the crucial sentences (1) and (2) depended essentially on that hypothesis. But what about referents in m_2 for pronouns in m_1? Here is an example which shows that the sets of possible referents must be as we have specified them:

(26) Every farmer who admires her courts a widow.

It is my intuitive judgement that in (26) *her* can be coreferential with *a widow*, but only if *a widow* has wide scope over *every farmer*. Such "wide scope" readings for indefinites that occupy positions which correspond to narrow scope according to our syntax are not discussed in this paper. A reading which (26) can *not* have is, according to my intuitions, the one given by

(27) $(\forall x)(\mathrm{farmer}(x) \rightarrow (\exists y)(\mathrm{widow}(y) \wedge \mathrm{admires}(x,y) \wedge \mathrm{courts}(x,y)))$.

To block this reading we must stipulate that the element ν of $m_2(26)$ is not available to the pronoun in $m_1(26)$:

$m_0(26)$

(0) Every farmer who admires her courts a widow

$m_1(26)$	$m_2(26)$
(1) x . (1) farmer(x) (1) x admires her	(2) ν . (1) x courts a widow (2) widow ν (2) x courts ν

Our theory seems to rule out a parallel reading for the conditional

(28) If a farmer admires her, he courts a widow.

It predicts, that is, that (28) cannot mean what is expressed by (27). Again, *her* in (28) can be understood as coreferential with *a widow* if the latter is taken to have wide scope—as it normally would in, say,

(29) If a farmer admires her he courts a certain widow I have dated and therefore know quite well.

(28) appears to have still another reading, in which *a widow* is taken as generic, a reading that is approximated by

(30) $\forall x \forall y[\text{farmer}(x) \wedge \text{widow}(y) \wedge \text{admires}(x,y) \rightarrow \text{courts}(x,y)]$.

Generics, however, are among the most recalcitrant constructions known to me. They will not be treated in this paper. Note also that

(31) If Pedro admires her he courts a widow,

though understandable, on the assumption that *her* refers to *a widow*, does not sound natural—barely, more natural in fact than do (26) and (28) on their wide scope reading, given by

(32) $\exists y[\text{widow}(y) \wedge \forall x[\text{farmer}(x) \wedge \text{admires}(x,y) \rightarrow \text{courts}(x,y)]]$.

The reason is that in order to get a reading of (31) in which *her* and *a widow* are coreferential we have to suppose—just as we must in connection with (26) and (28)—that *a widow* has wide scope over the subject Pedro. In another paper we shall have more to say about why such readings tend to be somewhat unnatural.

References

Bartsch, R., 1976. Syntax and semantics of relative clauses. In: J. Groenendijk, M. Stokhof (eds.), *Proceedings of the Amsterdam Colloquium on Montague Grammar and Related Topics*. Centrale Interfaculteit, University of Amsterdam, 1–24.

Bartsch, R., 1979. The syntax and semantics of subordinate clause constructions and pronominal reference. In: F. Heny, H. Schnelle (eds.), *Syntax and Semantics 10: Selections from the Third Groningen Round Table*. Academic Press, New York, 23–60.

Carlson, G., 1976. *Reference to Kinds in English*. Ph.D. dissertation, University of Massachusetts, Amherst, MA.

Cooper, R., 1975. *Montague's Semantic Theory and Transformational Syntax*. Ph.D. dissertation, University of Massachusetts, Amherst, MA.

Cooper, R., 1979. The interpretation of pronouns. In: F. Henry, H. Schnelle (eds.), *Syntax and Semantics 10: Selections from the Third Groningen Round Table*. Academic Press, New York, 61–92.

Cooper, R., Parson, T., 1976. Montague Grammar, Generative Semantics and interpretive semantics. In: B.H. Partee (ed.), *Montague Grammar*. Academic Press, New York, 311–362.

Evans, G., 1977. Pronouns, quantifiers and relative clauses. *Canadian Journal of Philosophy* 7 (3), 467–536.

Evans, G., 1980. Pronouns. *Linguistic Inquiry* 11, 337–362.

Geach, P., 1962. Reference and Generality. *An Examination of some Medieval and Modern Theories*. Cornell University Press, Ithaca, NY.

Hausser, R., 1974. *Quantification in Extended Montague Grammar*. Ph.D. dissertation, University of Texas, Austin, TX.

Hausser, R., 1979. How do pronouns denote? In: F. Heny, H. Schnelle (eds.), *Syntax and Semantics 10: Selections from the Third Groningen Round Table*. Academic Press, New York, 93–139.

Hendrix, G., 1975. *Partitioned Networks for the Mathematical Modeling of Natural Language Semantics*. Ph.D. dissertation, University of Texas, Austin, TX.

Hintikka, J., Carlsson, L., 1979. Conditionals, generic quantifiers and subgames. In: E. Saarinen (ed.), *Game Theoretical Semantics*. Reidel, Dordrecht, 179–214.

Karttunen, L., 1976. Discourse referents. In: J. McCawley (ed.), *Syntax and Semantics 7: Notes from the Linguistic Underground*. Academic Press, New York, 363–385.

Kratzer, A., 1979. Conditional necessity and possibility. In: R. Bäuerle, U. Egli, A. von Stechow (eds.), *Semantics from Different Points of View*. Springer, Berlin, 117–147.

Lasnik, H., 1976. Remarks on coreference. *Linguistic Analysis* 2, 1–22.

Lewis, D.K., 1973. *Counterfactuals*. Harvard University Press, Cambridge, MA.

Montague, R., 1970a. English as a formal language. In: B. Visentini et al. (eds.), *Linguaggi nella Società e nella Tecnica*. Edizioni di Comunità, Milan, 189–224. Reprinted in: R. Montague, *Formal Philosophy: Selected Papers by Richard Montague*, ed. by R. Thomason. Yale University Press, New Haven, CT, 1974, 188–221.

Montague, R., 1970b. Universal Grammar. *Theoria* 36, 373–398. Reprinted in: R. Montague, *Formal Philosophy: Selected Papers by Richard Montague*, ed. by R. Thomason. Yale University Press, New Haven, CT, 1974, 222–246.

Montague, R., 1973. The proper treatment of quantification in ordinary English. In: J. Hintikka, J. Moravcsik, P. Suppes (eds.), *Approaches to Natural Language: Proceedings of the 1970 Stanford Workshop on Grammar and Semantics*. Reidel, Dordrecht, 221–242. Reprinted in: R. Montague, *Formal Philosophy: Selected Papers of Richard Montague*, ed. by R. Thomason. Yale University Press, New Haven, CT, 1974, 247–270.

Partee, B.H., 1975. Montague Grammar and Transformational Grammar. *Linguistic Inquiry* 6, 203–300.

Partee, B.H., 1978. Bound variables and other anaphors. In: D. Waltz (ed.), *Theoretical Issues in Natural Language Processing 2* (*TINLAP–2*). Urbana, IL, 248–280. Reprinted in: B.H. Partee, *Compositionality in Formal Sematics: Selected Papers by Barbara H. Partee*. Blackwell, Oxford, 2003, 110–121.

Thomason, R.H., 1976. Some extensions of Montague Grammar. In: B.H. Partee (ed.), *Montague Grammar*. Academic Press, New York, 77–117.

Turner, R., 1981. Counterfactuals without possible worlds. *Journal of Philosophical Logic* 10, 453–493.

Veltman, F., 1976. Prejudices, presuppositions and the theory of conditionals. In: J. Groenendijk, M. Stokhof (eds.), *Proceedings of the Amsterdam Colloquium on Montague Grammar and Related Topics*. ILLC, Amsterdam, 248– 281.

Webber, A., 1978. *Formal Approach to Discourse Anaphora*. Ph.D. dissertation, Harvard University, Cambridge, MA.

CONTEXT, THOUGHT AND COMMUNICATION[a,*]

Hans Kamp

I

The sense of much that we say depends on the context in which it is said. Here is a simple example. The officer in charge of finding out what happened during the Defense Secretary's speech to the Student Union has called upon one of the informers (whom the police, with usual foresight, had placed among the audience). The informer selects a photograph from the pile in front of him and says to the officer:

(1) She was throwing tomatoes.

The are several ways in which this utterance depends on the context in which it is made. For a start let us note two of these.

(i) The pronoun *she* refers to the professor of mathematics whose photograph the informer has just chosen.
(ii) the past tense *was throwing* indicates some time preceding the time t_0 at which (1) is being made.

Thus (1) is true iff the person of the picture was engaged in an act of tomato throwing at some time before t_0. This illustrates the familiar fact that reference and truth may depend on such contextual factors as the time of the utterance and what deictic acts accompany it (pointing at an individual, staring at her, selecting a picture of her, and so forth). Evidently the interpretation of (1) requires context in other ways as well. For (1) to count as true it is not enough that the mathematics professor of the picture

---

[*] Meeting of the Aristotelian Society held 5/7 Tavistock Place, London WC1, on Monday, 17 June, 1985 at 6.00 pm.

---

[a] Editors' note: This article was originally published in the *Proceedings of the Aristotelian Society*, 85, 1984–1985, 239–261. It is reprinted here by courtesy of the Editor of the *Aristotelian Society* © 1985. We carefully edited the article, corrected some typos, added full bibliographical references and slightly adjusted the layout. We would like to thank Tillmann Pross for careful proofreading and helpful comments.

© HANS KAMP, 2013 | DOI:10.1163/9789004252882_015
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

threw tomatoes at *some* arbitrary time previous to t_0; she must have been throwing tomatoes on the particular occasion which the officer has been charged to investigate. It is to that event that the past tense *was throwing* must be understood to refer. And it is the context in which (1) is used which determines that it is that event to which the tense refers, and no other.

But let us return to the dependencies noted under (i) and (ii). These are instances of context sensitivity that are by now quite well understood. They received ample attention within the tradition of model-theoretic semantics, as it was established by Montague, Scott, Kaplan, Lewis and others in the late sixties and early seventies.[1] Typically, those studies recognized, the truth of an utterance depends on facts such as the time and place at which it is made, the identity of speaker and addressee, and accompanying demonstrative acts. Moreover, it was observed, much, perhaps all, such context-dependence can be located in particular context-sensitive, or 'indexical', linguistic elements, like pronouns, demonstratives, tenses and certain adverbs of space and time (e.g. *here, over there, yesterday, a week ago*, or *a week later*). And it appeared that a precise analysis of these dependencies could be given by recasting truth and reference (with respect to a given model) as functions taking the relevant contextual factors as arguments.

This work has been fruitful and influential. But it has ignored an aspect of context sensitivity which is crucial to the way language works. It is an aspect that affects both spoken and written language, but is of particular importance in connection with extended written passages, or texts. It is this. Not only do the utterances we produce, orally or in writing, often depend on context for their interpretation; they also do much to determine what the context *is*. The interaction between context and utterance is two-, not one-way.

In fact, the dependence of context on utterance is nearly as pervasive as its converse. To see this, note how easy it is to convert an example in which the relevant contextual parameters are determined by extra-linguistic factors into one where they are fixed by the preceding text. Compare for instance the case of (1) with the following bit of prose:

(2) On the 25th February the Defense Secretary addressed the Union. As expected the occasion was marked by serious disturbances and the Secretary, unable to finish his speech, left in a condition that called for a bath and a serious spot of

[1] See Montague (1974), Scott (1970), Lewis (1979b), Kaplan (1975).

> dry-cleaning. Later that week a police informer who had been present during the address selected from a group of photographs the picture of a professor of mathematics and told the officer in charge of the investigation that she threw several tomatoes.

The very last clause of this passage—the complement clause of *told*—depends for its interpretation on the contextual cues provided by earlier parts of the text, just as the utterance (1) depended on the contextual factors that were presumably fixed in other, non-linguistic ways. In particular, the pronoun *she* now refers to the mathematics professor spoken of in the preceding sentence and the past tense *threw* refers to the event of the secretary's visit which is mentioned at the outset, and remains the subject of the entire passage.

As it turns out, the vast majority of context-sensitive expressions have the property, illustrated by *she* and *threw*, that the contextual factors which determine the reference of their individual tokens can be fixed either by linguistic or by non-linguistic means. Precisely how they can be fixed by preceding discourse varies from one type of context-sensitive expression to the next, and a complete statement of the precise mechanisms that correspond to these various types is a far from trivial matter. But this is an issue that I do not want to elaborate on here.[2]

A typical speaker of English who reads (2) will, by the time he reaches the last clause, have acquired the contextual information needed to interpret that clause. His ability to extract this information from the antecedent text is, it would seem, an integral part of his ability to understand what the preceding sentences say. In other words, understanding a textual passage is not only a matter of grasping its truth conditions. It is also a matter of grasping the context it provides for what comes next.

It is in principle conceivable that the cognitive processes by which we come to grasp the truth conditions of a given bit of discourse are distinct from those by which we appraise its contextual significance. But one would not suppose there was such a separation unless compelled by observable facts of language and cognition; and what is known of such matters today

---

[2] To appreciate the differences that can be found among context-sensitive expressions compare words like *I* and *you* with demonstratives and third person pronouns. Or compare *now* and *today* with *then, two hours later*, or the tenses of the verb. Especially among context-sensitive elements that refer to time there is, it appears, a surprising plurality of types of context-sensitivity. For more on this see Kamp & Rohrer (1983); Kamp & Rohrer (1984/86); Kamp (1985).

suggests, as far as I can see, no such separation. I will assume, without bothering to supply evidence, that there is no separation. This assumption (which we may call the Principle of the Unity of the Context and Content) is fundamental to all that follows.

II

The assumption is one of the principal conceptual ingredients of an approach towards semantics which has come to be known as 'Discourse Representation Theory'—or 'DR Theory', for short. This paper is not the place to explore the technical and linguistic details of the theory. What I will say about them will be just what I need for present purposes.

Central to DR Theory is the concept of a Discourse Representation Structure, or DRS. Each text gives rise to a DRS, which is obtained by applying to it a certain (almost fully) algorithmic procedure, the *DRS-construction algorithm*. In the extant versions of the theory this algorithm takes the form of a set of rules which apply to the syntactic parses that some suitable theory of syntax assigns to the sentences that constitute the text. From a linguistic viewpoint the real challenge to the theory is to come up with an exact and concise statement of these rules; but this is an aspect of the enterprise that I must bypass here entirely.[3]

A DRS is a structure composed of two components, a set of 'reference markers' called the 'Universe' of the DRS, and a set of 'conditions', which act as predications on the markers in the first set. A DRS K, with universe U_K and condition set C_K, is defined as *true* (with respect to a given model M) if it is possible to assign an entity (from M) to each member of U_K in such a way that these entities satisfy all the conditions in C_K involving the corresponding markers. Thus, to consider a simple example, the three sentence text

(3) The Defense Secretary is addressing the Union. A professor of mathematics disapproves of him. She is flinging a tomato at him.

gives rise, when subjected to the construction algorithm of 'SID without Time or Questions',[3a] to a DRS of the form:

---

[3] For details see Kamp (1981), Kamp (1983a) or Kamp (1983b).

[3a] See footnote 3.

(4)

x	y	z	u
Defense Secretary(x) Union(y) x is addressing y			
		professor of mathematics(z) z disapproves of x	
			tomatoe(u) z is flinging u at x

The universe of this DRS consists of the marks x, y, z and u, which are displayed in the top line of (4); and its condition set is constituted by the conditions listed below that line. (4) is true (w.r.t. M) if we can find entities a, b, c, d (in M) corresponding to x, y, z and u which satisfy all these conditions (in M)—thus a must be the Defense Secretary (in M), b must be the Union, a must be addressing b, c must be a professor of mathematics, etc.

Actually (4) is a simplified version of what a DRS for (3) ought to look like. It ignores all aspects of temporal reference, as well as the distinction between definite descriptions, like *The Defense Secretary*, and indefinite descriptions, such as *a Professor of mathematics*. The problems of temporal reference are too complex to be explained here and in any case orthogonal to the issues I wish to address. I shall therefore leave them aside throughout what is to come. About the difference between definite and indefinite descriptions, however, I will have more to say later on.

When the construction algorithm is applied to a sequence of sentences (such as e.g. the sentence triple (3)) it deals with the sentences in order of appearance. Application to the first sentence gives rise to DRS K_1, in the present case consisting of the reference markers x and y and the first three conditions of (4). When the algorithm then turns to the second sentence it incorporates its content into K_1. Some of the construction rules that are applied during this stage of the procedure explicitly rely on K_1 for their application. For instance, the rule that deals with the anaphoric pronoun *him* must choose an antecedent for the pronoun from among the reference markers that are contained in the universe of K_1. (Exactly what is involved in such choices is a difficult question which has to do as much with assumptions about the nature of the real world as with grammar. A lot has been written on this subject, lately in particular in the context of Artificial Intelligence. This is one of many issues I cannot pursue here.) It is in

this manner that K_1 acts as context of interpretation for the second sentence in (3). The result of applying the algorithm to that sentence is a new DRS K_2 which comprises the markers x, y, z and the first five conditions of (4). K_2 now captures the joint content of the first two sentences and serves as context of interpretation for the third sentence. By applying the algorithm to this last sentence in the context of K_2 we obtain the DRS displayed in (4).

(4) exemplifies the simplest kind of DRS, in which all conditions take the form of a predicate (corresponding to a noun or verb from the original text) complemented with one or more reference markers. In general, conditions can be more complicated, and contain DRS's as constituents. (Thus the notions of a DRS and of a DRS-condition are recursive: DRS's can be used to form complex conditions, these can be the constituents of more complex DRS's and so forth.) Conditions of this more complicated sort arise where the treated sentence involves such logical notions as disjunction, negation, implication or universal quantification.

DRS's also tend to be more complicated than (4) in another respect. As a rule the reference markers a DRS contains will belong to different logical types or categories. Some will stand for physical objects, others for sets of those, or for events, or times, or spatial locations, etc. In fact, the explicit treatment of temporal reference (in the form in which it has thus far been worked out) always leads to reference markers for times and events. (Such markers would also have made their appearance in the DRS for (3) if we had taken its temporal aspects seriously.) However, as neither complex conditions nor typological variety of reference markers seem to be important for the issues to be discussed below, these are complications I will avoid. Thus I shall confine myself to DRS's of the simple sort exemplified by (4) and to examples of sentences and text that can, without serious distortion, be represented by such DRS's.

III

DRS's, I have just indicated, play the double role of identifying the content of what has been and the context for what is to come. Each of these two roles places certain constraints on DRS-structure. In particular, the role of DRS's as contexts requires them to specify a set of 'formal' individuals (the 'reference markers', as we have called them) which it makes available as antecedents for subsequent anaphoric noun phrases. It is this requirement which motivates the articulation of DRS's as consisting of a set of

reference markers and a set of conditions on those markers; and it is this also which motivates the principle that indefinite noun phrases, like *a professor of mathematics*, give rise to reference markers in much the same way as do definite noun phrases (i.e. proper names, demonstratives and definite descriptions). These theoretical decisions are illustrative of an important general methodological point, viz. that the contextual role of DRS's constrains their organization in ways that an exclusive concern with truth conditional content would not.

It turns out that the structural properties which answer to these constraints are not only instrumental in the description of the meaning and cohesion of discourse, but also contribute to a better understanding of linguistic communication, and of the formation of new beliefs as a result of such communication; and beyond that they appear to tell us something about the organization of cognition in general. This should not come as a surprise. For it is tempting to think of the DRS construction algorithm as modelling (in an admittedly idealized fashion) the interpretative processes which go on in the human recipient of a verbal input. And one who yields to this temptation would expect that the structure of the DRS's—which is essential to the theory's success in explaining how the earlier parts of a discourse impinge on the interpretation of its later parts—reflects some aspects of the form in which the information contained in those earlier parts is available to the human processor when he needs it to make sense of a newly received sentence or clause. We should recall in this connection our earlier observation that such context-sensitive elements as pronouns, demonstratives, definite descriptions, tenses and various adverbs of space and time are capable of picking up contextual cues whether those were established by the earlier discourse or by non-linguistic factors. On the reasonable assumption that such elements receive their context-dependent interpretations by the same processes irrespective of the cues' origin, this suggests that the contextual information on which the human interpreter relies, and on which those processes must operate, is available to him in much the same form whether he has derived it from the preceding discourse or from other sources (such as e.g. perception). The only difference between the case where the information has been obtained from verbal inputs and that where it has not would, on this hypothesis, have to do with *how* it gets moulded into the one form that can serve as basis for subsequent linguistic interpretation.

As it stands DR-theory only pertains to the processes by which this form is obtained from a textual source. About the perhaps even more intriguing question how similarly structured information can arise through sensory

perception it tells us nothing. But I confess to this without shame. For the problem of which this last question seems to be a special case—that of the conversion of sensory into propositional content—is one to which no extant theory seems to offer anything that could be called a solution. I will leave this question for what it is and stick to matters about which DR Theory has something to tell us.

Before proceeding further, let me summarise the position to which the reflections of the last two paragraphs lead us. They suggest the following hypothesis:

(5) The information that humans derive from verbal inputs (as well as, presumably, the information they derive from other sources), and which they may subsequently exploit in the interpretation of other verbal inputs, is available to them in a form similar to that of a DRS.

As it stands this does not seem much of an hypothesis. For the hedging '*similar to* that of a DRS' robs it of any bite. Let me explain. First, the types of DRS's to which I said we would confine our attention in this paper are in more than one way simplifications of the structures that would result from application of construction rules that are sensitive to details of meaning which we have decided to ignore. In fact, increasingly sophisticated algorithms will produce increasingly refined DRS's. It is plausible that the more refined a DRS the more it captures of cognitive organization. This leads one to suspect that while all types of DRS's capture some aspects of cognitive structure, none captures all. And certainly the impoverished DRS's that we are working with here will fail to make many structural features of cognition explicit.

Second, and this is the more important consideration, cognitive states are complex in the sense that they are composed of many different beliefs and other propositional attitudes. A single DRS cannot adequately represent such a state *in toto* because it has nothing to indicate how the state is divided into its different components. To capture this aspect of cognition a cognitive state should be identified not with a single DRS but rather with a family or network of DRS's.

Before we opt for such an alternative, however, one word of caution. Sometimes it seems quite arbitrary whether we describe a person as having one belief or two. Suppose John believes that Alison has divorced Fred and taken up with Bill. Does John have one belief or two? Perhaps there is nothing that is intrinsic to John's cognitive state which could settle this question. You may attribute a single belief to him, or a couple of beliefs, as befits your purposes. If you choose to describe him as having two beliefs

the division is yours, not his. A formal description of John's state need not capture this division, for it is not *there*.

The reply to this is: Yes, sometimes it is like that; but not always! Certain divisions *are* psychologically real, *are* somehow encoded in the structure of the state itself. This must be so for the simple reason that the owner of the state will as a rule associate different qualifications with different parts of it. Thus John may remember that he read about Alison's divorce in the Mirror some time in June and that he learned about her taking up with Bill by reading about it in the Mirror some time in April. In that case there is something in his state that keeps these two components apart as distinct beliefs.

There are all sorts of qualifications that a person may attach to components of his total belief state. Some of them have to do with the origin of the belief, some with the degree of confidence the believer attaches to it, some with his emotional attitude towards the content (e.g. whether the thought fills him with sadness or with joy). All such qualifications help to keep the demarcations between the different parts of the belief state in place. Moreover, cognitive states involve not only beliefs but also other attitudes such as doubts, hopes, fears, wants, desires and intentions. Clearly, it is an intrinsic part of the state which of its components act as beliefs, which as desires, and so on.

Ideally a theoretical description of cognitive states should be capable of specifying these various qualifications. But as the machinery needed for this is rather involved I will stay clear of it here. This is possible since we will only be concerned with beliefs (and not with any other attitude types), and the qualifications pertaining to origin, degree of certainty and emotional colouring will not be needed. It will be important that complex belief states are subdivided into distinct components, and so we need a way of representing belief states which renders this much of their structure explicit. But for that purpose it is enough if we identify belief states as sets of DRS's.

This then is the more specific form of (5) which we will explore in the remainder of this paper: A complex belief state is characterizable as a set of DRS's (whose universes may overlap). This proposal may seem simple-minded in the extreme. Yet it has one feature that helps in dealing with a number of problems which have proved recalcitrant to earlier treatments. It is this. Different members of a DRS-set that characterizes a given belief state may share the same reference marker. The intuitive significance of such sharing is this. Where two DRS's K_1 and K_2 share the marker x, the two beliefs which these DRS's identify act as 'being about the same thing'. They

act as being about the same thing in so far as the truth conditions which they jointly determine are fulfilled only if there is a single object c such that assigning it to x, and making further suitable assignments to the other markers in the universes of K_1 and K_2, results in satisfaction of all the conditions that belong to either K_1 or K_2. Note that this is a strictly *internal* sense in which the two beliefs are 'about the same thing'. It does not require that there *be* an object c in the outside world, which (possibly in combination with some additional objects) makes both beliefs true. On the other hand it is intuitively clear that believers often are in just this sort of position: they think of their two beliefs as being about the same thing. But of course their thinking so neither requires nor guarantees that there be something external that could be construed as the common object of those two beliefs. It is, as we shall see, quite useful to possess systematic means of representing this kind of internal identity.

Sometimes we form beliefs which, although resulting from our causal interaction with different individuals, nevertheless end up being about the same thing in the internal sense. Thus you may have, concerning a person you know under the name of John Williams, (i) the belief that he is an accomplished guitarist; and (ii) the belief that he is conductor of the Boston Pops. The first belief you formed when listening to a recital on Radio 3, the second on the strength of a televised concert transmitted from the USA. Somewhat foolishly you take yourself to have been listening to the same man on both occasions. In view of the causal origins of the respective beliefs there is a sense (external) in which they are about different individuals; yet, from an internal perspective they are about one and the same person. In cases such as this one, where there is a mismatch between external and internal identity, it is often difficult to say—at least in the informal terms we usually employ in belief ascription—what precisely it is that is being believed.

Puzzles of this sort have commanded a good deal of interest from philosophers of language, mind or logic. In most of them the mismatch is in the opposite direction: external identity matched by internal identity. They typically arise where the subject forms different opinions on two distinct occasions, but without realizing that on each of these he is confronting the same individual. The result is a pair of beliefs characterized by DRS's in which distinct reference markers stand in for this one real individual. Such is the predicament of Ralph, who has encountered Ortcutt on the beach, and also, without recognizing him, under suspicious conditions near the local power plant. It is also the predicament of Pierre, who in childhood learned of 'Londres' as a place of beauty and who now knows London, his present domicile,

as a place of squalor.[4] Here the two beliefs are, externally speaking, about the same object, and as such they are inconsistent; for they attribute to that object two incompatible predicates. Yet in an internal sense they are not about the same thing, and thus they do not appear inconsistent to the believer himself. No logical acumen on the part of the believer could therefore, as Kripke observed in relation to Pierre, lead him to detect such inconsistencies.[5]

IV

I have been speaking of an external as well as an internal sense in which two propositional attitudes can be about the same thing. The internal sense I explicated in terms of sameness of reference markers. But of the external sense I have only spoken informally.

To make it precise we need the concept of an *anchored* DRS. Anchored DRS's were originally introduced to account for direct reference.[6] The doctrine of direct reference holds that certain noun phrases contribute their referents directly to the content of the utterances in which they occur. This means that the proposition which such an utterance expresses simply attributes to the referent of the directly referring term the property identified by the remainder of the sentence uttered; in other words, the proposition is true just in case the referent has this property. The descriptive content of the noun phrase itself is not part of the proposition, but serves exclusively to pick the referent out.[7]

---

[4] Ewan Klein coined, some years ago, the apt phrase 'the problem of double vision' for the problem cases of this sort constitute for semantics and the theory of belief. See Klein (1979).

[5] See Kripke (1979). I should add that, as Kripke rightly observes, the major problem with such cases is not so much that of giving a coherent account of the sense in which Pierre's beliefs are inconsistent and the sense in which they are not, but rather that of explaining why in a case like Pierre's it is apparently so difficult to report the beliefs: The kind of sentence which we normally use for this purpose simply does not seem to work here. This is a problem which belongs not to the theory of belief simpliciter, but to the theory of belief reports, a subject that, for reasons I cannot go into here, must be carefully distinguished from the former. For more on this subject see Kamp (1984).

[6] See the report Kamp (1983b).

[7] Directly referring noun phrases are: proper names, pronouns, definite descriptions and demonstratives. Of these proper names and the first and second person pronouns always refer directly. Third person pronouns and demonstratives refer directly whenever they refer at all (i.e. whenever they refer to some particular external object). An example of such a directly referring use is the pronoun *she* as occurs in (1). However, these two types of noun phrases also have anaphoric uses. We encountered one instance of this in the *she* of (2). Some anaphoric uses, moreover, cannot be construed as referential. Consider e.g. the sentences:

A DRS that captures the truth conditions of such propositions must carry a built-in guarantee that the reference marker it has for the directly referring term can only be associated with the term's real referent: The only admissible associations to be considered in assessment of the truth conditions are those in which the marker is paired with that referent, and with no other object. There is no other way that I can see of establishing such a guarantee than to add it explicitly: When the construction algorithm deals with a directly referring noun phrase it is to introduce not only a reference marker for the noun phrase, but also, as a separate component, the pair consisting of that marker and the real referent of the phrase. We will call such a pair an *anchor*. We will also call the second member of the pair an *anchor for* the first member: and finally we will call a DRS one or more reference markers of which have anchors an *anchored* DRS. (Formally an anchored DRS is best identified as a pair consisting of a DRS of the sort explained in section 2, and a set of anchors for markers in that DRS. In that case DRS's of the kind we have been considering so far come to correspond to those 'anchored' DRS's in which the anchor set is empty. When using the term 'anchored DRS', however, I will always assume that the set contains at least one element.)

The semantic significance of an anchored DRS is, as I already indicated, that the only permissible associations of objects with markers are those which respect the anchors, i.e. which associate with each anchored marker that marker's anchor.

---

(1) Almost every student is worried about the first exam he has to take.
(2) When a burglar runs off with the loot of another burglar that second burglar will plot revenge.

In these examples there is evidently no particular student or burglar that can be identified as 'the' referent of the noun phrase *he* or *that second burglar*. Even in these cases however, the doctrine of direct reference applies in so far as the descriptive content of the term functions only to identify the grammatical antecedent of the anaphoric noun phrase.

Definite descriptions have, besides the two uses mentioned for pronouns and demonstratives, a third where the descriptive content of the term does enter into the content of the utterance in which the term figures. This is what Donnellan has called the 'attributive' use of definite descriptions.

A precise analysis of the various functions that the different types of noun phrases can fulfil is quite a complicated matter. Once more I must refer to the ms. Kamp (1983b).

Not everybody accepts that there is such a phenomenon as direct reference. Personally I am persuaded that there is. But this is not the place to argue for that position. All I am concerned with here is to point out the implications of the doctrine for a DR-theoretic account of truth conditions.

Just as in semantics we need anchored DRS's to capture the truth conditional contributions made by directly referential terms, so we need them in the analysis of belief to characterize those beliefs that are *de re*. A belief is *de re* with respect to an object *c* if it attributes some property to *c*: its content is simply that *c* has this property, and thus is independent of the particular way in which the believer has represented *c* to himself, or the criteria by which he might recognize *c*. In this regard *de re* beliefs are like the propositions expressed by utterances that contain directly referential terms. And anchored DRS's are needed to capture the content of either.

Just when a belief can be regarded as *de re* is not easy to say. There has been a good deal of debate on this point; but, as '*de re*' is a term of art, which moreover made its way into philosophy without the benefit of a fully explicit definition, any answer must be somewhat arbitrary. There is much more that ought to be said here. But this is yet another issue that I cannot pursue. Let me make just two points concerning the use of the interpretation of the term '*de re*' I favour. First, for a belief to count as *de re* with respect to an object *c* that object must have played a certain causal role in the formation of the belief—what kind of role is something I cannot go into. Second, there must be something internal to the belief that can be construed as a record that it was that object, *c*, which played this causal role. What form such a record can take is another question I must leave largely unanswered. But it appears that in many (and possibly in all) cases we may take it to be something like this: The reference marker *x* which, in the DRS that identifies the belief, represents (i.e. is externally anchored to) the object *c* also occurs within certain other components of the DRS-set characterizing the entire belief state; and these other components 'identify' *c*, in the sense that the only associations of objects with markers which yield satisfaction of all their conditions are those in which *x* is associated with *c*. Moreover, these additional DRS's must, in as much as they represent a record that is *internal* to the belief, all be unanchored.

I will refer to such an identifying DRS or DRS-cluster for a marker *x* as an *internal anchor for x* (and to avoid ambiguity I will call the anchors introduced above *external*). Internal anchors come in a wide variety. Sometimes they take the form of explicit descriptions expressible in the language *L* which the believer speaks. In other cases they are constituted by his ability to recognize the object, but he himself may have no way of articulating the criteria on which that ability rests, and perhaps these cannot be stated in *L* at all. There appear to be other possibilities as well. A typology of internal anchors would, I believe, be illuminating. But at present I have nothing to offer on this score.

Now that we have introduced the concept of an anchor we can rephrase the distinction which led to this excursus. Two beliefs are, we already explained, about the same thing in the internal sense if their characterizing DRS's share the same reference marker. They are, we may now add, about the same thing in the external sense if their characterizing DRS's contain markers that are anchored to the same external object.[8]

V

It should be reasonably clear how the puzzle cases I mentioned earlier can be described with the apparatus now at our disposal. Let me, by way of illustration, briefly apply the new concepts to one of them, the notorious case of Pierre.

The DRS-set B_P which characterizes Pierre's belief state contains two members, which identify his beliefs involving 'Londres' and 'London'. These DRS's have the forms:

(K_1)

x
x is beautiful

and

(K_2)

y
y is ugly

In addition B_P will have components which function as internal anchors for x and y. The first will contain as one of its conditions 'Londres(x)', signifying that Pierre associates the name 'Londres' with the internal object of the

[8] There is, we may note, a third sense in which two beliefs can be about the same thing: the characterizing DRS's may contain markers x and y whose *internal* anchors link them with the same external object. On my preferred use of the notion of '*de re*' there will be such internal anchors whenever x and y are externally anchored to the same thing. On the other hand the markers could be internally anchored to the same object without being also externally anchored to it. Someone who has two beliefs that are about the same thing in this third sense need not be able to detect that they are about the same object any more than he would be if the relevant markers only had external, but no internal, anchors.

belief identified by K_1. Similarly the other anchor will have 'London(y)'. The two anchors will have much besides these two conditions; in particular, the anchor for u will be continually fed and sustained by Pierre's daily experiences. But, as Kripke envisaged the case, all that information does not add up to the conclusion that x and y stand for the same city. Finally, in virtue of their causal origins, both x and y may be regarded as externally anchored to the real London. But these external anchors, important as they are in the characterization of the content of the two beliefs, are of no more help to Pierre in finding out that his beliefs pertain to the same place than are his internal anchors.

None of this addresses the problem (referred to in footnote 5) of explaining why the device we usually employ successfully to report propositional attitudes break down in the case of Pierre. I already said that I would not deal with that problem in the present paper. The main reason for that decision is that an analysis of the issues involved requires a much more refined description of the structure of cognitive states than the comparatively primitive method in terms of DRS-sets provides. We would also have to expand our theory in a slightly different direction. For the complexities of the report problem relate in part to the mechanisms of verbal communication and thus presuppose some formal account of how communication functions. In fact, such an account is needed for the analysis of a considerable variety of questions, both in semantics and in the theory of propositional attitudes. I will therefore, to conclude this paper, sketch how a theory of communication can be developed within the framework we have been using, and mention a couple of phenomena to which this theory can be profitably applied.

VI

To understand how verbal communication works we need some understanding of how language relates to thought. After all, to comprehend an utterance one hears or reads is to form a thought corresponding to what the utterance expresses; and, similarly, to express a thought in words is—well, just that: to produce an utterance which expresses the thought. Most verbal communication involves both these transitions: the speaker produces an utterance that expresses the thought he wishes to convey; the recipient 'decodes' the utterance and thereby forms a new thought whose content once again corresponds to that of the utterance. And thus, if all goes according to rule, the recipient's thought will match the one which the speaker wanted to communicate.

As it exists at the present time, DR-theory has something to say about the second of these two processes, that by which language is converted into thought. Let us focus our attention on the case where the speaker wants to impart some information to his audience and to this end utters a declarative sentence. As DR-Theory has it the recipient will, in response to the utterance, add to his cognitive state a new component characterized by a DRS which results through application of the construction algorithm to the sentence reaching him (relative to the DRS-complex that characterizes his total belief state at the time of processing). If he trusts the speaker this new component will be incorporated as a belief; if he doesn't it will get some other attitudinal status.

Nothing I have said about DR Theory pertains to the first process, that by which thoughts are put into words.[9] But the theory contains certain suggestions as to how it should be analyzed. What happens when a speaker puts a thought into words is the converse of what goes on when the recipient of an utterance distils from it a corresponding thought. Given all the commitments I have already made, this consideration leads us more or less inevitably to the hypothesis that this converse process is one of transforming a DRS into a sentence, or sequence of sentences. The algorithm responsible for such transformations must be a converse of the DRS-construction algorithm, in the precise sense that the discourse it produces from a given DRS should, when subjected to the construction algorithm, return the original DRS—or at any rate one of identical content and similar structure.[10]

With the limited machinery we have set up we can give a global description of what verbal communication involves. Suppose that, at time t, A wants to communicate a certain piece of information to B. Both A and B are at t in a certain belief state. These states are characterized by two sets of DRS's, B_A and B_B. Let us assume that K_0 is that member of B_A which identifies the piece of information A wants to convey to B and that B does not possess

[9] It is probably wrong to see all intentional speech acts as verbalizations of pre-existing thoughts. Often, or so at least it appears to superficial introspection, the thought crystallizes as the words are formed, and could hardly be said to have had an earlier, independent identity. Even in these cases, however, the formal relations between the thought and the linguistic expression of it will be as articulated here.

[10] This criterion certainly does not determine the verbalization algorithm completely, even where the construction algorithm is already fixed. For verbalization must choose between various possible strategies for ordering information in the way which the linear character of language makes necessary. The study of verbalization mechanisms within DR Theory has hardly begun. But for what I want to say the details of the algorithm do not matter as long as it is a converse of the construction algorithm.

this information already. Thus B_B has no member with the same content as K_0. The purpose of the communication is to bring it about that B_B comes to contain a DRS, K_1 say, which identifies the same information that is embodied in K_0. For the transfer to count as successful, K_1 must be much like K_0. It must be like K_0 at least in the sense of determining the same conditions of truth. Possibly it should resemble K_0 in other ways as well. Precisely in what ways the two should resemble each other is a delicate question I cannot pursue.[11]

VII

The first of the two applications I promised in V concerns a problem due to Geach. It has been known informally as the 'Hob-Nob problem', because of the example that Geach used to present it.[12] Geach makes us consider a community which still believes in witches and like manifestations of the supernatural. Recent events, unfortunate and without a ready explanation in terms of natural causes, have led the members of the community to the belief that some particular witch is responsible for those misfortunes. In particular two of the members Hob and Nob, have come to form beliefs which, though not identical, are nonetheless construable as both being about that same witch. In this situation it would seem possible to say, truthfully,

(6) Hob believes that a witch has killed Cob's cow and Nob believes that she has blighted Bob's mare.

In fact, (6) would appear to be acceptable even if the speaker himself does not believe that witches exist. This is puzzling from a traditional logical perspective, in as much as it looks as if, on the one hand, the phrase *a witch* must have wide scope, so that it can bind the pronoun *she*, while on the

[11] The explication of resemblance is a difficult and important task. For one thing the question how similar the two DRS's ought to be is closely related to the familiar, and notorious, problem of 'finding', as it has often been put, 'the correct identity criteria for the objects of the attitudes'. However, the present framework reveals that there is a complication to this problem which most traditional treatments obscure. Because of its connections with other components of the belief state of which it is part we cannot always meaningfully assess the truth conditions of a particular belief in isolation. Therefore even the 'minimal' criterion for resemblance, the equivalence of truth conditions, must be formulated much more carefully than I have done here—and, indeed, than it ever is. For more on this topic see the paper mentioned towards the end of footnote 5.

[12] See Geach (1962).

other hand it should have narrow scope since otherwise the sentence would seem to commit the speaker to the existence of witches, which apparently it does not. A complete solution of this logical problem cannot be given here (but see the paper mentioned in footnote 5). However, what I will say about it should make fairly clear how the solution will go.

It is clear that the sense in which Hob and Nob can be said to have beliefs about the same witch must be an internal sense, one which has to do with identity of reference markers rather than identity of external objects (which, in the present case, there aren't!) Relying on the framework we have already developed we may describe the situation as follows. Each member A of the community has his own set of beliefs 'about the witch'. This set is characterized by a set W_A of DRS's all of which contain some one common reference marker x_A representing the witch (in the sense that some members or member of W_A identify a belief to the effect that x_A stands for the unique witch that has been causing so much trouble). What justifies us in construing all the beliefs entertained by the different members of the community as about the same witch is the existence of certain links between the reference markers $x_A, x_B, \ldots, x_H, x_N, \ldots$ that represent the witch in their respective belief states $W_A, W_B, \ldots, W_H, W_N, \ldots$ All manner of interactions between the members can establish and sustain such links. Of paramount importance among those various interaction types is how the members of the community talk to each other about the witch.

The theory of communication which I have just outlined allows us to say something about how and why verbal exchanges cement such links. I'll give one example of this. Suppose Hob says to Nob:

(7) The witch has killed Cob's cow

Suppose that at the time when this is said Hob and Nob are in belief states B_H and B_N. B_H contains a member of the form:

x_H $\quad$ y $\quad$ z
cow(z)
z belongs to y
x_H has killed z

Here y has an internal anchor which involves, at the very least, the condition that it stands for an individual who goes by the name 'Cob'. (Quite probably y may also be construed as externally anchored to the real Cob, but that is of little importance now.) Hob is able to verbalize this belief as (7), relying

on, among other things, his internal anchor for *y*, which furnishes him with the name 'Cob'. Nob, in processing (7), will form a thought characterized by a DRS which has reference markers for each of the noun phrases (7) contains. We may assume that he takes Hob to be talking of the witch that he, Nob, thinks he already knows quite a bit about. Formally this means that he will identify the marker corresponding to the noun phrase *the witch* with the marker x_N which already figures in his belief state as the peg for all his opinions about 'the witch'. It is this process of identification which establishes or reinforces a link between the marker x_N that represents the witch for Nob, and the marker x_H that represents her for Hob and which led to the latter's use of the term *the witch*. The term acts as intermediary, serving to connect the two elements of their respective minds.[13]

It is not only the marker which *Nob* is to introduce for the phrase *the witch* that he is likely to identify with one that already figures in his belief state. The same may be expected to happen to the marker for *Cob*. For, as I have elaborated the story, Nob knows Cob only too well, and we may presume that he takes Hob to be talking about that Cob. And the same goes for the marker corresponding to *Cob's cow*.

Such identification—of the marker for an incoming noun phrase with one that is already part of the recipient's mental state—is very common when the noun phrase is *definitive* (i.e. a proper name, a demonstrative, a pronoun, or a definite description). Indeed, it appears that the need for such an identification is part of the meaning (or, if you prefer, semantic function) of definite noun phrases. This is, in essence, the so-called 'familiarity doctrine' of definiteness—the doctrine that a definite, as opposed to an indefinite, noun phrase carries the implication that its referential correlate is already familiar to the interpreter.[14]

---

[13] The links that are formed in this way, and the chains consisting of those links, are closely related to the *referential chains* discussed in Chastain (1975). In fact for all I know the chains that emerge from what I have said here and Chastain's referential chains are indistinguishable. I believe that DR Theory is well-suited for making formally precise much of what Chastain has to say about the structure of cognition, and about the interplay between thought and linguistic expression.

[14] This doctrine, which to my knowledge has received comparatively little attention from philosophers, is well-known, and widely accepted, among linguists. For a penetrating discussion of the doctrine, in a framework very similar to the one sketched in this paper, see Heim (1982); and Heim (1983).

The familiarity theory of definiteness should not be misconstrued as claiming that on all occasions where a recipient processes a definite noun phrase he will identify the marker for it with one that is already part of his cognitive state. Thus, you may hear me make a statement containing a proper name which is unknown to you and form the thought that the bearer of

Indefinite noun phrases carry precisely the opposite presupposition, viz. that the item corresponding to the phrase is *not* familiar. In our terms this means that the reference marker for the phrase should *not* be identified with one already available. Again, this should not be taken to mean that such identifications never take place. For instance, Hob could have said to Nob:

(8) A witch has killed Cob's cow

and Nob might, in reaction to this, have thought 'Oh, I know what witch he is talking about. It is the one who has made life so difficult for us all recently'. Here Nob would identify the new marker with the marker x_N that is already part of his mental baggage. But this is not something that the indefinite description *a witch* demands. Rather, it is, one feels, something that happens in spite of the intention that (8) seems to convey.

The difference between definite and indefinite noun phrases is the second of the two problems I announced at the end of section V. I claim no originality for these last observations.[15] But I have included them nonetheless, as a further, simple, instance of a question that would seem to belong squarely within linguistics and can be more easily, and more correctly, analyzed within the present framework than it can in any of the alternative formal theories of language with which I am acquainted. DR Theory is helpful here precisely because it offers an integrated account of language, thought and the ways they interact in communication. Many putatively 'linguistic'

---

that name—whoever he may be—has the property which my statement attributes to him. The DRS for this thought will contain a marker corresponding to the name which is not identifiable with any you already had. Yet, it would seem an unreasonable claim that, just for this reason, you have not understood what I told you.

In spite of such examples the familiarity theory is defensible. It must be understood as the thesis that definite noun phrases *presuppose* familiarity. In cases where the recipient does not satisfy the presupposition (as you don't in the example given) he will, as long as doing so is compatible with what he believes, make up for the missing presupposition. Where the presupposition is one of familiarity of a certain individual, making up for the missing presupposition simply takes the form of behaving as if it were familiar, i.e. as if one had a reference marker already with which the new marker can then be identified.

To see that proper names carry some such presupposition consider what you would say to a friend whom you want to tell something about a person, called *N* say, of whom you know that your friend has never heard. In such a case it seems more natural to refer to the person in some such way as 'someone called *N*' or 'someone by the name of *N*', rather than by using the name *N* directly. The reason is, I submit, precisely because you want to avoid the presupposition of familiarity implicit in simply using the name.

What I have said here about proper names applies, mutatis mutandis to the other types of definite noun phrase.

[15] In all significant respects they follow the account of the definite-indefinite distinction offered by Heim in the essays mentioned in footnote 14.

phenomena are like the distinction between definite and indefinite in that only such a comprehensive framework permits us to articulate both their grammatical and their cognitive aspects and to show how those interconnect; and only thus can we explain how they work or why they occur.

The general moral of all this is a lesson that some of us have taken too long to learn: Neither semantics nor cognition can be studied fruitfully on its own. Only by looking at them together can we hope to arrive, eventually, at a viable theory of either.

References

Chastain, C., 1975. Reference and context. In: K. Gunderson (ed.), *Language, Mind and Knowledge*. University of Minnesota Press, Minneapolis, MN, 149–270.

Geach, P., 1962. *Reference and Generality. An Examination of some Medieval and Modern Theories*. Cornell University Press, Ithaca, NY.

Heim, I., 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. dissertation, University of Massachusetts, Amherst, MA.

Heim, I., 1983. File change semantics and the familiarity theory of definiteness. In: R. Bäuerle, C. Schwarze, A. von Stechow (eds.), *Meaning, Use and Interpretation of Language*. de Gruyter, Berlin, 164–189.

Kamp, H., 1981. A theory of truth and semantic representation. In: J. Groenendijk, T.M.V. Janssen, M. Stokhof (eds.), *Formal Methods in the Study of Language, Proceedings of the Third Amsterdam Colloquium*. Mathematical Center, Amsterdam, 277–322. Reprinted in: J. Groenendijk, T.M.V. Janssen, M. Stokhof (eds.), *Truth, Interpretation and Information, Selected Papers from the Third Amsterdam Colloquium*. Foris, Dordrecht, 1984, 1–41, and reprinted in: K. von Heusinger, A. ter Meulen (eds.), *Meaning and the Dynamics of Interpretation. Selected Papers of Hans Kamp*. Brill, Leiden, 2013, 329–369.

Kamp, H., 1983a. Anaphoric links, comments on Reinhart. In: A. ter Meulen, E. Traugott (eds.), *Proceedings of the Conference on Conditionals*. Stanford University, Stanford, CA.

Kamp, H., 1983b. *SID without Time or Questions*, unpublished manuscript, CSLI, Stanford University, Stanford, CA. Available online: http://www.ims.uni-stuttgart.de/archiv/kamp/.

Kamp, H., 1984. Beliefs, belief formation and belief reports. In: J. Groenendijk (ed.), *Proceedings of the Fifth Conference on Formal Semantics*. University of Amsterdam.

Kamp, H., 1985. Deixis and discourse. *Logique et Analyse, Nouvelle Série* 28.

Kamp, H., Rohrer, C., 1983. Tense in texts. In: R. Bäuerle, C. Schwarze, A. von Stechow (eds.), *Meaning, Use and Interpretation of Language*. de Gruyter, Berlin, 250–269.

Kamp, H., Rohrer, C., 1984/86. *Temporal Reference in French*, unpublished manuscript, Bedford College, London/CSLI, Stanford University, Stanford, CA/University of Stuttgart, Stuttgart. To be published with CSLI Publications, Stanford, CA, 2013/14. Available online: http://www.ims.uni-stuttgart.de/archiv/kamp/.

Kaplan, D., 1975. *Demonstratives*, unpublished manuscript, Stanford University, Stanford, CA. Published as: D. Kaplan, Demonstratives. An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. In: J. Almog, J. Perry, H. Wettstein (eds.), *Themes from Kaplan*. Oxford University Press, Oxford, 1989, 481–563.

Klein, E., 1979. *On Sentences which report Beliefs, Desires and other Mental Attitudes*. Ph.D. dissertation, University of Cambridge.

Kripke, S., 1979. A puzzle about belief. In: A. Margalit (ed.), *Meaning and Use*. Reidel, Dordrecht, 239–283.

Lewis, D.K., 1979. General semantics. In: D. Davidson, G. Harman (eds.), *Semantics of Natural Language*. Reidel, Dordrecht, 169–218.

Montague, R., 1974. Pragmatics. In: R. Montague, *Formal Philosophy: Selected Papers by Richard Montague*, ed. by R. Thomason. Yale University Press, New Haven, CT, 95–118.

Scott, D., 1970. Advice in modal logic. In: K. Lambert (ed.), *Philosophical Problems in Logic*. Reidel, Dordrecht, 143–173.

DISAMBIGUATION IN DISCOURSE[a,*]

Hans Kamp

0. Introduction

This note is one in a series. The series consists of case studies concerned with the role that lexical information and lexically driven inference play in the construction of semantic representations of discourses. We hope that such studies will eventually lead to a clear perspective of the inferential mechanisms that are needed in text and discourse interpretation and of the kind of information which these mechanisms exploit as premises. There is—of this we are convinced—no quick and easy way to get at such a comprehensive picture; only a long and thorny path, leading past a large number of particular examples painstakingly scrutinized, is likely to lead us to such a comprehensive view. Each analysis along this path ought to produce some new insights into the nature of natural language interpretation; but it would be foolish to expect any one of them to reval the whole truth.

In this regard the present study is no exception. Like others we have carried out over the past two years, it yields some glimpses of the general principles that govern discourse interpretation. But a reasonably comprehensive understanding of those principles is, we fear, still a very long distance ahead.

The present study differs from our earlier case studies[1] in that the problems it considers have a direct bearing on (automated) translation. For the case we will consider concerns the role which context plays in the

[*] Most of what has gone into this paper arose out of joint work with Antje Roßdeutscher. Those familiar with that work will know the depth of my indebtedness.

[1] Kamp & Roßdeutscher (1992), Roßdeutscher (1992), Roßdeutscher (1993).

[a] Editors' note: This article was originally published in M. Aurnague, A. Borillo, M. Borillo & E. Bras (eds.), *Semantics of Time, Space and Movement—Proceedings of the 4th TMS International Workshop 1992*, IRIT Université Paul Sabatier, Toulouse, 1993, 281–308. We carefully edited the article, corrected some typos, added full bibliographical references and brought some formulas in line with now more common usage. We would like to thank Tillmann Pross for careful proofreading.

© HANS KAMP, 2013 | DOI:10.1163/9789004252882_016
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

interpretation of a lexically ambiguous word (the German verb *ausleihen*, which can mean either *borrow* or *lend*). When a German text is translated into, say, English, each occurrence of *ausleihen* must be disambiguated, for otherwise one would not know which of the two English verbs to use. Note, however, that it isn't just when we translate into English that the ambiguity of *ausleihen* needs to be resolved. A German speaker will typically assign a particular reading to any occurrence of the verb he encounters, whether or not he knows English, let alone is engaged in English translation. So the same ought to be the case for any text-understanding system.

Nevertheless this paper holds a special moral for the business of translating by machine. It is a moral that comes in two parts. First, it should be evident from reading this note that an MT system able to deal with the problems we will discuss must have inferential capacities quite unlike anything that can be found in the systems which are in operation at the present time. If one keeps in mind that those problems are still quite trivial when compared with those that cause genuine trouble to professional translators, it is not easy to preserve even moderate optimism about the MT enterprise. At the very least they support a sense of sobriety.

That a certain sobriety is called for becomes especially clear when one realizes—and here we come to the second moral—that translation systems which can deal with the kind of problem we will discuss must be able to disambiguate not just in those (arguably infrequent) cases where disambiguation is directly needed for the choice of a suitable target language equivalent. Disambiguations like the one we will look at depend crucially on the availability of an explicit and unambiguous representation of the discourse or text which *precedes* the ambiguous item in question. This means that whenever the translation system encounters an expression whose disambiguation is needed for translation purposes, it must either have such a representation of the antecedent discourse, or else it will have to construct it then and there. In order to secure that the representation is there when required, the system will have to build it as it goes along, much as a question answering system would do with a text about which it must be able to answer subsequent queries. A system that forgoes building such representations in the hope that they won't be needed, may be lucky in that no complicated context-dependent inferences will be wanted throughout the text with which it is dealing; but if it is not lucky, it will have to start all over or go back a considerable way; and so it might have been better off had it taken the more arduous course of building a proper discourse representation from the outset. There is no hope in either case that a powerful MT system can be significantly simpler than a fully-fledged discourse interpreter.

1.

We will concentrate on the following pair of examples:

(1) A: Du hast doch das Buch gekauft, oder?
B: Ich habe es gekauft. Aber ich habe es ausgeliehen.

(2) A: Du hast doch das Buch gekauft, oder?
B: Ich habe es nicht gekauft. Aber ich habe es ausgeliehen.

(1) and (2) can be translated as

(1.e) A: You've got the book, don't you?
B: I bought it. But I have lent it to somebody.

(2.e) A: You've got the book, don't you?
B: I didn't buy it. But I have borrowed it.

In (1) *ausleihen* translates as *lend*, in (2) as *borrow*.

We begin with an informal explanation of the difference between (1) and (2):

The two different uses of the verb *ausleihen*—one corresponding to *lend* and one corresponding to *borrow*—both involve change of possession. But the changes are in opposite directions: while the lender gets rid of the object, the borrower gets hold of it. The *lend*-use of *ausleihen* describes transitions from a state in which the subject is in possession of the object to one in which he isn't, and the *borrow*-use describes transitions from the latter state to the former. *Kaufen* too describes changes of this general type—changes from a state in which the subject does not have an object to one in which he does have it. So the result of the event described in the first sentence of (1) is a state in which the subject has the book. This situation is fit as a starting point for an event of the type described by the second sentence if *ausleihen* is understood as *lend*, but not when the verb is interpreted as *borrow*. This means that if we take *ausleihen* in this second sense, then we must assume that between the two events mentioned—the buying event and the borrowing event—there was a third event, which changed the state of the subject having the book into that of his not having it. When *ausleihen* is taken in the sense of *lend*, no corresponding assumption is necessary. So (simplifying considerably) the first interpretation yields the less complicated interpretation and therefore qualifies as the correct one.

With respect to (2) the argument is similar, but not altogether. We just saw that verbs like *ausleihen* and *kaufen* express transitions from states of

one type into states of the opposite type. However, the status of these two facts—of the initial presence of a state of the first type and the eventual holding of one of the second type—is different. The latter is part of what the verb is used to assert, whereas the former rather has the character of a presupposition: it tends to be seen as holding not only when the verb is used affirmatively, but also when it occurs in a denial. In particular, the first sentence of (2) is naturally understood as pertaining to a period at the outset of which the subject did not have the book.[2] Moreover, it claims of this period that it contains no event of the described type (i.e. no event of *B* buying that book). This time, therefore, it is the *lend* interpretation which leads to difficulties. Since at the beginning of the period that the first sentence is taken to be about the subject did not have the book, and since for all the first sentence has to tell us this state of affairs was not terminated, the precondition for the *lend* interpretation of the second sentence is not fulfilled—if it were fulfilled, this would have been in virtue of some event which the discourse does not mention. The *borrow* interpretation does not run into this problem here. Its precondition is fulfilled by the state of affairs presupposed by the first sentence; there is no need to assume the occurrence of some unmentioned event.

Summarizing, the intuitively preferred interpretation of *ausleihen* is in either case that which avoids assuming the occurrence of events that are not entailed by the antecedent text; the better interpretation is that which can make do with a minimum of happenings.

Put this way the principle of disambiguation in (1) and (2) is reminiscent of a number of theories about non-monotonic temporal reasoning which define non-monotonic consequence relations as preservation of truth by those and only those models of the premises in which the set of events and/or states of affairs is in some sense minimal.[3] This parallel, however, could be misleading in two ways. First, as noted explicitly in Lorenz (1993), there is an important difference between on the one hand the non-monotonic inferences relevant in the contexts for which most of those theories appear to have been designed—contexts of predicting a state of affairs that is still to come or of explaining an observed state on the basis of incomplete

---

[2] As is the rule with presuppositions, this species of presuppositions is easily cancelled. It is, for instance, in:

(3) Ich habe das Buch nicht gekauft. Ich besitze es schon seit Jahren.

[3] See for instance McCarthy (1986), Lifschitz (1986, 1987), Shoham (1988a, 1988b), Kautz (1986), Kowalski & Sergot (1986), Sandewall (1988), Schubert (1990), Lorenz (1993).

information—and on the other the non-monotonic inferences that often arise in discourse interpretation and that are exemplified by the reasoning apparently involved in making sense of (1) and (2). The justification for assuming minimality when explaining or predicting facts about the world is that *this* world—the one in which we live—tends on the whole to be fairly regular; events of the sorts that we are most ready to identify as causes of states or other events are the exception rather than the rule. They may be likely occurrences in situations of some particular kind, but are unlikely to occur otherwise. Thus the default inferences we tend to make in explanation and prediction and which involve the assumption that certain events did or will not occur are supported by the way the world is, by the statistical distribution of events of one type over situations of some other type; it has nothing to do with the ways we speak about what we know.

The justification of non-monotonic inferences like the ones we encountered in our gloss of examples (1) and (2) is of a different kind. In these examples the speaker in question, *B*, may be assumed to know what happened. If indeed he bought the book, then got rid of it or had it taken from him, and then proceeded to borrow it, this would have been a sequence of events he himself would have known about. And if he did know about it, then, in order to avoid possible misunderstanding, he could, and should have mentioned it. That he should have mentioned it is a convention of speech, one of the rules governing cooperative verbal communication. Given this convention, the very fact that *B*'s utterance does not mention that he came to be without the book after he had bought it may be taken as evidence that no such event occurred: Had the intention been that such an event occurred, then the event would have been explicitly mentioned. Because of this communicative convention, minimalization inferences in discourse interpretation tend to be very reliable. Such a convention is applicable especially to fiction, where the author can be assumed to know all there is to know about the world of his narration. Here inference by minimalization is usually a sound means of arriving at the intended interpretation (i.e. for arriving at a representation of the described world that matches the author's intentions).

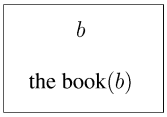
There is a second respect in which the parallel between the cases at hand and the minimalization approaches towards non-monotonic reasoning in the AI literature is less than perfect. As we will see in section 2, examples (1) and (2) are a good deal more complex than the informal analyses we have just given reveal. Thus, in connection with (1) we have so far ignored the possibility that the *ausleihen*-event described by the second sentence of *B*'s answer does not follow the *kaufen*-event of the first, but that the two events stand in some other temporal relation. For instance, could they not have

been simultaneous, or even identical? Another aspect we have ignored is the role played by the word *aber*, which figures in both (1) and (2). When we will eventually turn to these issues, we will find that event minimalization is anything but the whole story there is to tell about these tiny dialogues.

2. Representing *B*'s First Answer

But first we must put the given analyses on a firmer formal footing. We start with (1) and proceed as follows. We almost entirely ignore *A*'s question, since, on the account we will be formalizing, it is not directly relevant to the interpretation of *B*'s second sentence. The only information implied by it which we need to retain is that there is among the individuals shared between *A* and *B* some particular book (viz. the book to which *A*'s question refers; for a first pass at an analysis of shared information within a DR-theoretical framework see Kamp 1990). I will assume that this book is represented by the discourse referent *b* and will state the condition which guarantees that *b* does indeed stand for the book in question somewhat ineptly as *das Buch* (b). Thus we start with a DRS consisting merely of the following:

(3)



We now sketch the construction of a representation for the first sentence of *B*'s answer. After that we will show how the two different interpretations of *ausleihen* lead to two distinct extensions of this representation and how the "good" extension is simpler than the other one. Then having completed our account of (1), we will go through the same exercise for example (2).

Semantic representations are constructed as in DRT: Discourse representations (DRSs) are built sentence by sentence, with each new sentence yielding an extension of the DRS for the preceding text. By and large we assume that the construction procedure is as developed in Kamp & Reyle 1993, but there is one important difference. For present purposes it is crucial that DRS construction involves the use of a lexicon, which specifies the relevant syntactic and semantic properties of the words that make up the processed sentences. A sketch of how the construction procedure would exploit lexical entries for certain verbs can be found in Kamp & Roßdeutscher (1992). This proposal is also the basis for the procedure followed here.

To work out a lexically based construction procedure in full detail is a

major undertaking, with which we have so far only made a start. To become fully explicit, such a procedure would need to have a lexicon at its disposal that contains entries for all words that it encounters in the texts with which it is to deal. A lexicon of such size, but whose entries contain the kind of information that is exemplified in the few entries we will give below, is still a thing of the future. So we will have to be content as in Kamp & Roßdeutscher (1992) to make the construction-lexicon interaction explicit only in the case of the two words on which the present disambiguation strategy depends—the verbs *kaufen* and *ausleihen*—and to deal with all other words in the manner of earlier construction algorithms (such as the one in Kamp & Reyle 1993). That is, the expressions besides *kaufen* and *ausleihen* will be treated as 1-place predicate, 2-place predicate, predicate modifier, etc.

To show how a DRS construction procedure can exploit the lexical entry for a verb, it is best to first have a look at an example. Here is the entry we propose for the verb *kaufen*:

(4)

kaufen	Nom	Acc	*von* + Dat
e	x	y	(z)

e :

$e_c \quad e_e$
$e_c \quad Caus_{\textit{kaufen}} \quad e_e$
$Agent(e_c) = x$
e_e : $TRANSFER(y, z, x)$

Like other verb entries (4) consists of an upper and a lower part. The upper part, consisting of the top two lines of (4), specifies the *syntax-semantics-interface*. To be precise, it stipulates how the semantic arguments (e, x, y, z), which figure in the lower, purely semantic part of the entry, are syntactically realized when the verb is used in its "basic" (i.e. active) form. That is, when the verb is used in an active clause, its "non-referential" arguments $(x, y\ and, perhaps, z)$ are realized by phrases with the "case markings" indicated in the top line—as subject, direct object and as prepositional phrase with *von*, respectively. The parentheses around z indicate that this argument is syntactically optional (an active clause with *kaufen* as main verb need not contain the PP, whereas every such clause must have a subject and a direct object). Besides the arguments x, y and z, *kaufen* also has a "referential argument" e. e is the event which the given use of the verb describes. The lower part of the entry says that each such event is a "causal complex" consisting

of an action e_c (the *c* stands for "cause") by the agent x which stands in a special causal relation $Caus_{kaufen}$ to the process e_e (the *e* stands for "effect". For details see Kamp & Roßdeutscher 1992). This process is the transfer of the purchased item from the seller, z, to the buyer, x.

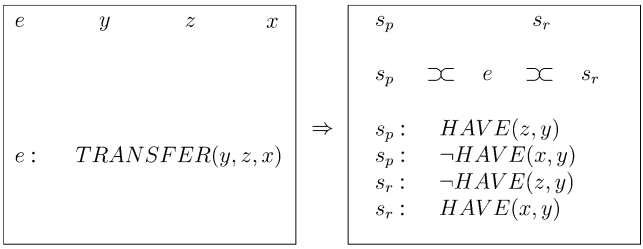
There are various other properties which buying events necessarily have and which one might expect the lexical entry for *kaufen* to make explicit, e.g. that such events require a certain kind of action on the part of the seller as well as on the part of the buyer or that buying always involves a transfer in the opposite direction of a certain sum of money. As these aspects of the meaning of *kaufen* play no role in the arguments below, we have chosen not to include them in (4); however, a comprehensive entry for *kaufen* ought to include them.[4] The binary predicate $Caus_{kaufen}$, the 1-place function *Agent* and the 4-place predicate *TRANSFER* are elements of a "concept language" in which the purely semantic part of the lexicon is formulated. The logical properties of and the relations between these concepts are made explicit in another component of the lexicon, to which we will refer as the (*set of*) *meaning postulates* or *lexical axioms*. One of these postulates says that a transfer of y from z to x leads from a state in which y is had by z and not by x to a state in which it is had by x and not by y:

[4] Entries like (4) make no attempt to distinguish between the two levels which in the work of Bierwisch and others (see e.g. Bierwisch 1983, 1988, 1989, Wunderlich 1992) is called *semantic form* and *conceptual structure*.

The idea behind this distinction is that lexical items have a variety of semantic properties, in particular those that have to do with the syntactic realizations of their arguments, which can be read of a certain comparatively uninformative level of semantic representation—that of semantic form—which will in general not suffice for the full identification of the truth conditions associated with sentences in which those lexical items occur. A lexical entry containing a semantic form which answers these desiderata would need no separate syntax-semantics-interface, since the information for which that interface is responsible could be gleaned from the semantic form in any case. On the other hand, the entry would have to either contain a separate conceptual structure or else the additional information needed to compute conceptual structure from logical form. Precisely how this juxtaposition or transition between semantic form and conceptual structure is to be conceived and represented appears to be still a topic of investigation.

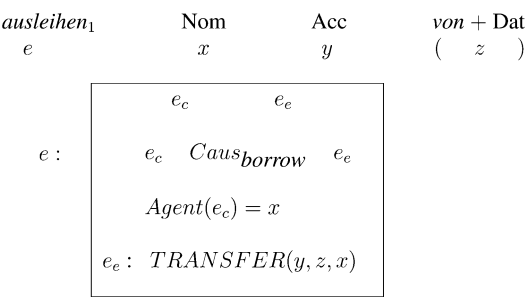
So long as this issue has not been clarified it is difficult to assess the exact relationship between the verb entries proposed here and in the other papers cited in fn. 1 and those proposed in the studies cited in the present footnote.

(5)

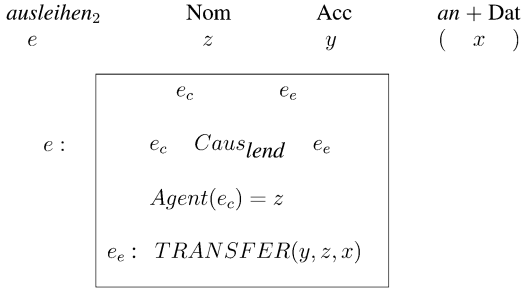


We assume that the verb *ausleihen* has two lexical entries, one for each of the uses we distinguished. They look much like the entry for *kaufen*:

(6)



(7)



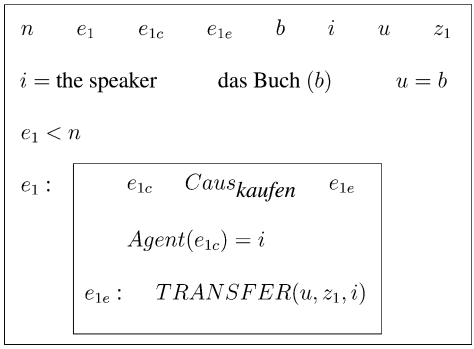
Note that the only point at which the entry for *ausleihen*$_2$ differs from that for *ausleihen*$_1$ are (i) the causal relations *Caus*$_{borrow}$ and *Caus*$_{lend}$ and (ii) the syntax-semantics interface; this time it is *z* which is realized as subject, while *x* has become optional and when it is realized at all, it is not as subject but as PP with *an*.

To make use of the entry for *kaufen* in constructing a representation of the first sentence of *B*'s answer in (1) the construction algorithm must be able to identify, via a preliminary parse it has built for the sentence, the set of argument phrases which qualify, in virtue of their configurational relations to the given occurrence of *kaufen*, as potential arguments of it. This set has to be checked against the syntax-semantics interface of *kaufen*'s entry

(4) to see if all the arguments which the entry specifies as obligatory are in the set (with the right case markings and/or prepositions) and further whether some of the syntactically optional arguments are present as well. If the match is successful, the variables marking the argument positions in the lexical entry are identified with the discourse referents introduced by the corresponding argument phrases and the structure that results from making these identifications in the purely semantic part of the entry is introduced at the place held by the verb in the DRS. In the present case the only identifications are between *x* and the discourse referent *i* introduced for the first person pronoun *ich* (see Kamp 1990) and between *y* and the discourse referent *u* for the pronoun *es*, which in turn is identified with the discourse referent *b* for the pronoun's anaphoric antecedent. For further details see Kamp & Roßdeutscher (1992).

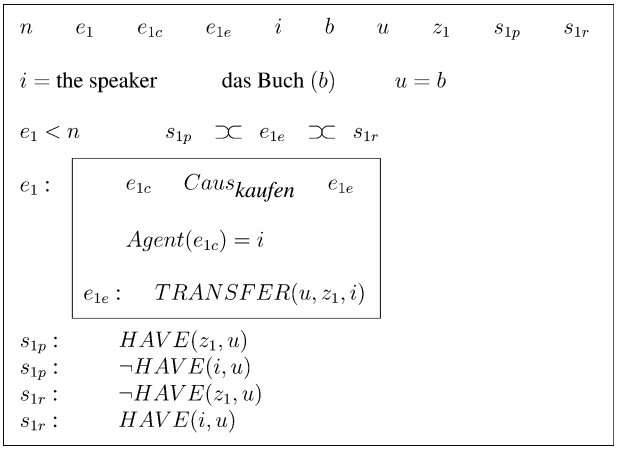
Combining this procedure with construction principles familiar from earlier accounts of DRS construction, we arrive at

(8)



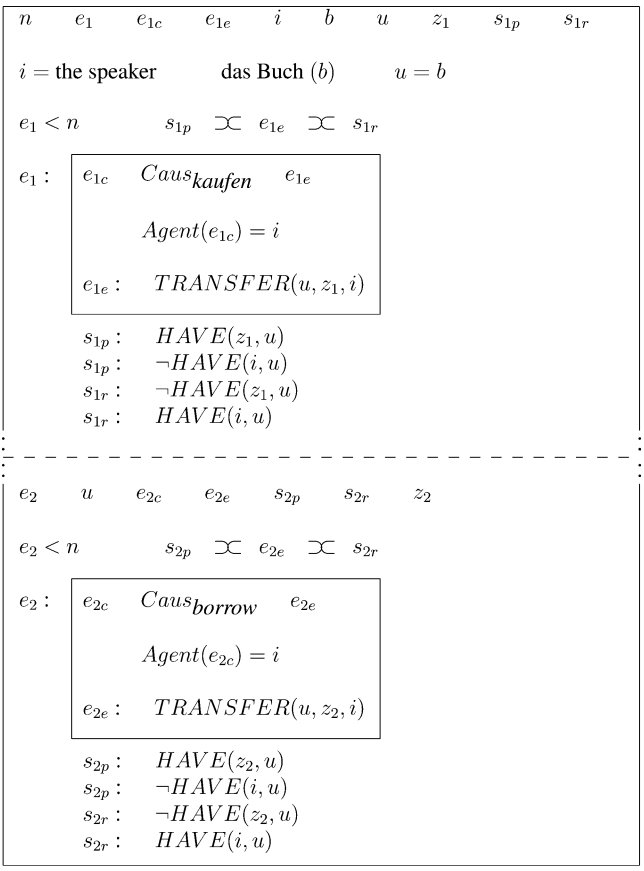
Axiom (5) permits us to expand (8) to (9):

(9)

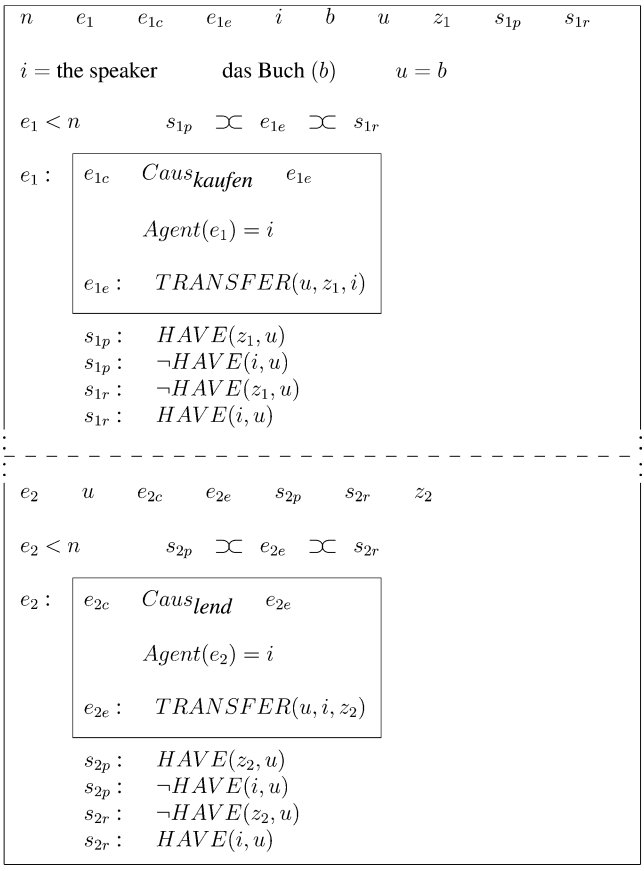


We must now process the second sentence of (1) in relation to (9). As we noted, the result will depend on which lexical entry for *ausleihen* the construction algorithm is going to select. (Note that argument matching does not decide the matter, since the complements of the present token of *ausleihen* include neither a *von-* nor an *an*-PP.) We present the two outcomes as (10) (corresponding to $ausleihen_1$) and (11) (corresponding to $ausleihen_2$)

(10)



(11)



To evaluate and compare (10) and (11) we need to make an additional assumption about the temporal relation between e_{1e} and e_{2e}. In our informal analysis we assumed that $e_{1e} < e_{2e}$. For the time being we simply adopt this assumption. We will turn to the question of its justification in Section 4.

First consider (9). Simple temporal reasoning allows us to infer from the premisses

(i) $e_{1e} < e_{2e}$
(ii) $e_{1e} \supset\subset s_{1r}$
(iii) $s_{2p} \supset\subset e_{2e}$

the conclusion

(vi) $(s_{1r} < s_{2p}) \vee (s_{1r} \bigcirc s_{2p})$

Together with the following two conditions from (10):

(v) $s_{1r} : HAVE(i, u)$
(vi) $s_{2p} : \neg HAVE(i, u)$

we can infer, first, that the second disjunct of (iv) is impossible: since the two states satisfy contradictory predicates, they cannot overlap. Consequently only the first disjunct remains:

(vii) $s_{1r} < s_{2p}$

But in order that a state of one type (here: $\neg HAVE(i, u)$) can follow one of the opposite type (here: $HAVE(i, u)$) there must have been an intervening event which effected the change. There are various ways in which this principle can be formulated. Here is one:

(12)

$$\boxed{\begin{array}{l} s_1 \quad s_2 \quad x_1, \ldots, x_n \\ \\ s_1 < s_2 \\ s_1 : \quad P(x_1, \ldots, x_n) \\ s_2 : \quad \neg P(x_1, \ldots, x_n) \end{array}} \Rightarrow \boxed{\begin{array}{l} e \\ \\ s_1 \quad \supset\subset \quad e \quad < \quad s_2 \end{array}}$$

Using (10) we can infer that there must have been such an event between s_{1r} and s_{2p}. So, restricting ourselves to the relevant part of the "temporal structure" of (10), we get

(viii)

$$\boxed{\begin{array}{l} n \quad e_1 \quad e_{1c} \quad e_1 \quad s_{1p} \quad s_{1r} \quad e_{2e} \quad s_{2p} \quad s_{2r} \quad e_3 \\ \\ e_{1e} \quad \supset\subset \quad s_{1r} \quad \supset\subset \quad e_3 \quad < \quad s_{2p} \quad \supset\subset \quad e_{2e} \quad < \quad n \end{array}}$$

As noted above, this structure is problematic from the perspective of text coherence, as the event e_3, which must have occurred between s_{1r} and s_{2p}, is not mentioned in (1). Moreover, as we will see presently, no such extra assumption is forced upon us by (11).

To see this, note that corresponding to the condition (vi) of (10) we have in (11) the condition

(ix) $s_{2p} : HAVE(i, u)$

Thus the characterizing predicates of s_{1r} and s_{2p} coincide. Consequently it is possible to assume that s_{1r} and s_{2p} are (parts of) one and the same state, so that the corresponding temporal structure implied by (11) has the form:

(x)

$$\boxed{\begin{array}{l} n \quad e_{1e} \quad s_{1p} \quad s_{1r} \quad e_{2e} \quad s_{2p} \quad s_{2r} \\ \\ e_{1e} \quad \supset\subset \quad s_{1r} \quad \supset\subset \quad e_{2e} \quad < \quad n \\ \\ s_{1r} = s_{2p} \end{array}}$$

Thus (11) does not force us to assume the existence of events that are not mentioned in the text, and thus its temporal structure is simpler than that of (10). So, in accordance with the minimalization principle it should be (11), not (10) that represents the correct interpretation of (1).

We already stressed that the role of minimalization in discourse interpretation should not be conflated with the one it plays in causal explanation and prediction. Indeed, for future reference it will be useful to summarize what is involved in the present use of minimalization once more.

(i) The premisses of the argument contain the claim that
  (a) the occurrence of *ausleihen* in the second sentence of (1) is either an instance of $ausleihen_1$ or of $ausleihen_2$.

This premiss combines with

  (b) the DRS (8) for the first sentence of (1)
  (c) the syntactic structure of the second sentence which functions as input to the construction of (10) and (11) and
  (d) the (relevant) principles defining the construction algorithm

to yield the disjunction consisting of the DRSs (10) and (11). The models of this disjunctions are either models of (10) or models of (11). All models of (10) have at least the events and states represented by the discourse referents of (10) together with the discourse referent e_3 occurring in (viii). Among the models of (11) there are also models in which the only represented events and states are those represented by the discourse referents of (11). These models are simpler than any of the models of (10). So all minimal models of the disjunction verify (11).

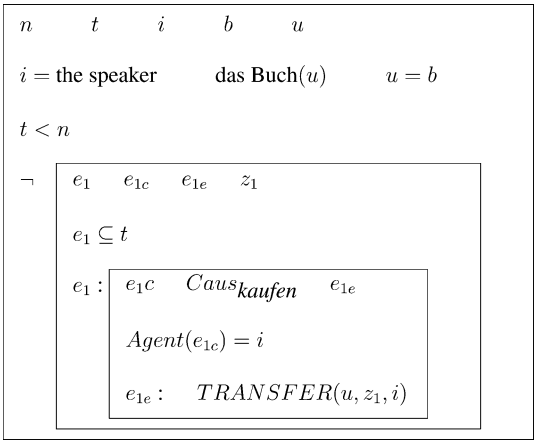
As we intimated before, there is a good deal more to say about (1) than the present analysis reveals. But before turning to those other issues, we first present a similar analysis for the second answer of *B*.

3. *B*'s Second Answer

After all that has been said about *B*'s answer in (1), we will be able to deal with his answer in (2) fairly swiftly, for most of the relevant issues have already been discussed. But there are two points connected with (2) that are new. The first has to do with the representation of negations of past tense (and future tense) event sentences. Such sentences assert of some period of time that within it no event of the sort described by the unnegated sentence occurred. In some cases this period is indicated explicitly by a temporal adjunct, in others it has to be reconstructed from the context, or simply guessed at. The case at hand is one of the latter kind: By his use of the

simple past tense *B* suggests that there was some interval of past time during which aquiring the book was an issue. It is of this interval that he asserts no purchase took place. In (13), a DRS for the first sentence of (2), the interval is represented as *t*:

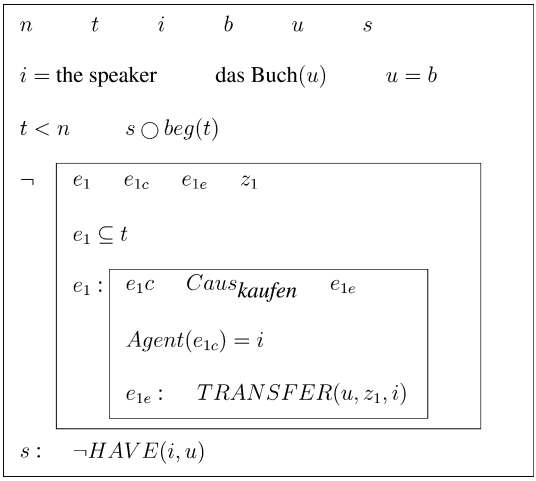
(13)



The second point concerns the presuppositional aspects of the verb *kaufen*. Many telic verbs (i.e. verbs expressing actions with well-defined result states), *kaufen* and *ausleihen* (in both senses) among them, *presuppose their prestates*. The occurrence of an event that such a verb describes presupposes that there be a certain state of affairs when it starts, a state of affairs that is opposite to the result state type associated with the verb—this must be so, for otherwise there would be no way for the event to *bring about* its result state. The way in which we interpret negated clauses with telic verbs shows that the presence of the prestate is a matter of presupposition in the technical sense of the word: unless the assumption is explicitly canceled, the interpreter will assume that the prestate holds at the outset of the period *t* over which it is being claimed that no event of the indicated type occurred. Thus, the first sentence of *B*'s answer in (2) implies that at the beginning of *t B* was in a state opposite to that which would have resulted from his buying the book, that is, in the state of his not having the book. In (14)—which is in all other respect just like (13), the presupposition about the prestate has been explicitly incorporated:[5]

[5] The rule responsible for the introduction of the state *s* in (14) can be stated as follows. As we saw in (5), a telic verb like *kaufen* induces a postulate to the effect that each instantiating event is flanked by a state of the associated result type ate the one end and a corresponding,

(14)[6]



For how long does the prestate associated with a telic verb persists? This is a separate question, with answers that depend on how the verb is being used. When the verb occurs in a positive clause, which asserts the occurrence of an event of the given type, then this event will terminate the prestate, by turning it into the result state. If the clause is negative, as it is in the present case, then the state will be assumed to persist, though the extend of this persistence needs to be determined by some other mechanism. This

opposite prestate at the other. When such a verb occurs in a negative clause, which gives rise to a sub-DRS *K* governed by ¬, which denies the existence of an event of the given type over a certain period *t*, then information must be added to the DRS which contains *K* as sub-DRS to the effect that the prestate held at the beginning of *t*. This information takes the form of adding a new discourse referent for the state, a condition which says that the represented state is of the type of a prestate for the given verb and a condition expressing overlap with the beginning of *t*.

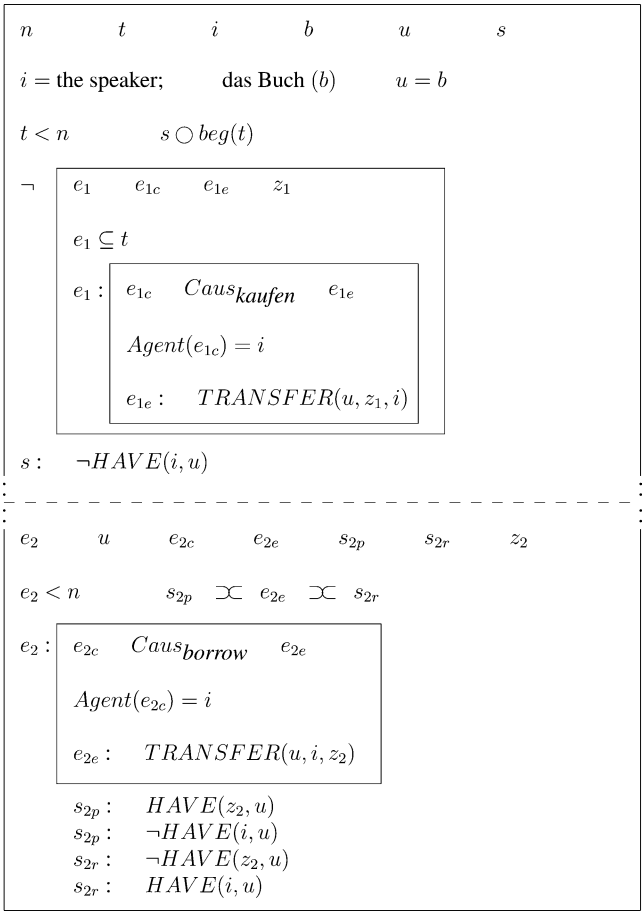
Described in this way the rule seems rather ad hoc. However, the need for it will disappear in a DRS-construction account which would incorporate a component dealing with presupposition in general. Such a theory would have to provide principles to handle presuppositions whose triggers occur within the scope of negation, to make sure that these presuppositions are incorporated into the representation unless canceled explicitly. Given such principles, all that would be needed to insure the transition from (13) to (14) is a specification that the presence of a prestate is, unlike the subsequent occurrence of a result state, something which the verb presupposes.

[6] The relation between *s* and *t* which is expressed here is that where *s* must have been in existence when *t* began—*s* and *t* overlap, but *s* "sticks out beyond *t*" in the direction of the past. The relation can be straightforwardly defined in Allen's system of temporal relations, but we won't bother.

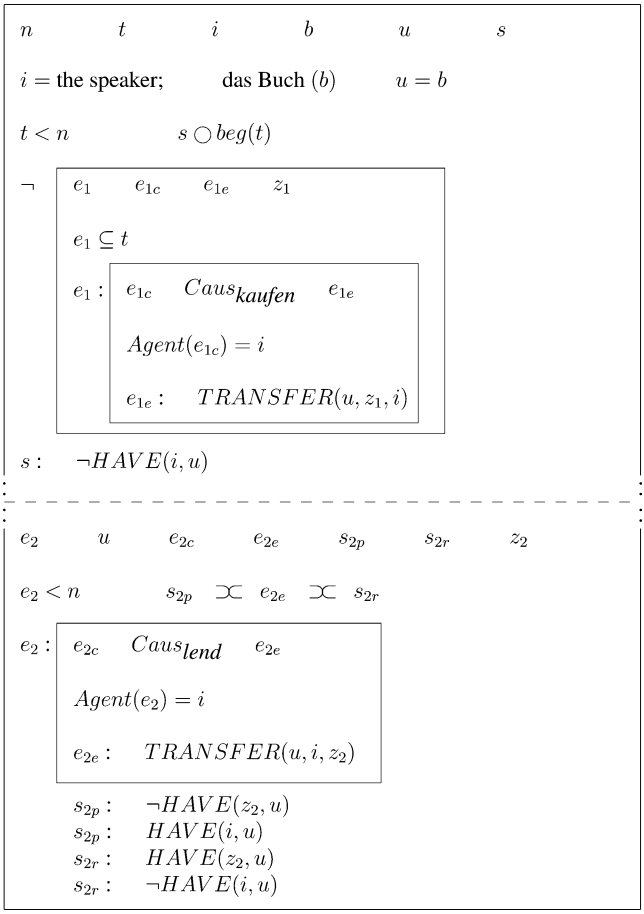
principle of (conditional) persistence enters into the ambiguity resolution in (2) just as the persistence of the result state of *kaufen* did in our analysis of (1).

To complete the present analysis of (2) is now straightforward. Once more we extend the representation of the first sentence—that is (14)—twice, once on the assumption that *ausleihen* is used in the sense of *borrow* and once on the assumption that it is used in the sense of *lend*. The resulting representations are given in (15) and (16). This time it is the *borrow*-sense of *ausleihen* which leads to the simpler DRS, and thus qualifies as the preferred interpretation.

(15)



(16)



4. Things are Not as Simple as They Seem

Things are not as simple as they may have seemed so far. That they aren't has a couple of reasons.

The first reason has to do with the temporal relation between the *kaufen*-event e_1 and the *ausleihen*-event e_2 in (1). In our treatment of (1) we assumed without argument that the second event occurred after the first. But what is the justification for this assumption?

It has often been noted that there is a natural tendency for speaker to report events in the order in which they actually happened. In particular, when a text contains a sequence of past tense event sentences $S_1, S_2, \ldots$

the assumption that the event e_1 described by the first sentence preceded the event e_2 described by the second, etc. is often a good bet. But it is not much more than that. Exceptions to the principle abound and one would have to be remarkably obtuse not to see them. Indeed, the early literature on tense and aspect within DRT, which adopted the principle as a first approximation for past tense narrative, acknowledged the existence of such exceptions, although it didn't offer any formal proposals for how they could be recognized and assigned the correct interpretations.[7] The last few years have seen substantial progress on this point, mostly through the work of Lascarides and Asher and others, which exploits the systematic connections between on the one hand the temporal relations between states and events and on the other the rhetorical relations between the sentences which describe them.[8] This work has done much to sharpen our perception of the complexity of the problem, but precisely because the problem is so very complex, we have at this point still no more than fragments of a full theory. Moreover, the applicability of what theory there is to the examples of this paper is especially problematic, since the theory has so far concentrated on the interpretation of monologue. Even within single-speaker texts the principles governing temporal and rhetorical interpretation appear to be varying substantially with genre—whether the text is a narrative, a joke, a scientific or argumentative piece of prose, a servicing or operating manual, etc. When we turn to dialogue, the matter is even more different than that; and dialogue has, from the perspective that is here relevant, barely been looked at.

In view of this an analysis of our examples that takes rhetorical as well as temporal relations into account would require substantial groundwork, which would reach far beyond the limits of the modest aims of this paper. So we leave this task for another occasion.

As it turns out, in the present case all new possibilities that arise when the assumption is dropped can be eliminated by the kinds of considerations that we used in the preceding sections. Suppose for instance that in (1) e_1 and e_2 were simultaneous. There are two cases to be considered: e_1 and e_2 might be one and the same event, or they might be distinct but overlapping events. First consider the case of their being identical. If *ausleihen* is interpreted as *borrow*, we would have a single event that would at the same time be a

[7] See for instance Kamp (1981), Kamp & Rohrer (1983), Partee (1984), Hinrichs (1986).

[8] See for instance Lascarides & Asher (1991), and the already cited Asher (1993). Also relevant are Caenepeel (1989), Caenepeel & Sandström (1993), Sandström (1993).

buying and a borrowing of the book. Clearly that cannot be, for buying and borrowing are incompatible kinds of actions.

But the second possibility, that where e_1 and e_2 are distinct but overlapping events, is ruled out also. Once you are engaged in buying a given object, you cannot then also borrow it. For—we have made similar points repeatedly before—you would first have to get rid of the thing again; only then could you proceed to borrow it from its new owner. By the same token, once you have set about borrowing the thing, you cannot then at the same time go about buying it.[9]

The situation is no better when we interpret *ausleihen* as *lend*. First, lending something and buying it cannot be the same event. Moreover, as we have seen repeatedly, to lend an object you have to have it and to buy it you have to not have it. So the events of buying an object and lending it out cannot overlap.

So we can rule out the possibility that e_1 and e_2 are wholly or partly simultaneous. What about e_2 being before e_1? The assumption that an event of *B*'s borrowing the book preceded the event of his buying it runs into the same difficulties we already encountered: there should have been an intermediate event to change the result state of the borrowing back into one where *B* did not have the book. This speaks against reading *ausleihen* as *borrow*.

Now suppose that *ausleihen* is interpreted as *lend*. Here the type of incompatibility argument we used in the last paragraph to eliminate the *borrow*-interpretation won't work. As we represented *ausleihen*$_2$ and *kaufen* in Section 2, the result state of *ausleihen*$_2$ would seem to be precisely the precondition required by *kaufen*. So by the argumentation we have been using so far the present interpretation ought to be all right. But intuitively it isn't, and the question is why? We suspect that this is a fact that can be explained in more than one way. The explanation we offer here has to do with the contractual aspects of *buy* and *lend*. When you buy a thing, you not only have the use of it; you have actually become its owner. This is not so when you borrow something. You may use what you have borrowed, but ownership remains with the one who has lent it to you. Turning things around, if you first buy a thing, you become its owner and so are in a condition that allows you to do all manner of things with it. For instance, you can sell the thing again, or you can lend it to someone. But if you first lend something out, this

[9] But see below!

does not create a condition in which it is possible for you to buy the thing. You still own it and buying presupposes that you do not.[10]

We repeat emphatically: this is not a definitive analysis. Nonetheless it should have given an impression of how far it is sometimes possible to go with the comparatively simple analytical tools we have been using. It would be in line with the way we have been proceeding so far to check what these tools yield in the case of (2), but this is something we will leave as an exercise. Instead we will devote the remaining pages to a point which so far has been entirely ignored, but which changes the look of things considerably when it is taken into account.

The point is this. We have been assuming all along that there was one object *b*, the book in question, which could be bought or borrowed, but not both. There are many contexts in which such an assumption is warranted. Consider e.g. the following English example, which closely parallels (1).

(17) A: Did you finally get that house on Lakeshore Drive you liked so much?
B: We did buy it. But we rented it.

(18) A: Did you finally get that house on Lakeshore Drive you liked so much?
B: We didn't buy it. But we rented it.

(17) and (18) present the same disambiguation problem as (1) and (2). But there is one important difference between the English examples and the ones we have been considering so far. (17) and (18) can only be understood as concerning one unique physical object, the house on Lakeshore Drive. If *B* bought that house then he cannot at the same time have rented it from someone. But not so with books. When *A* asks whether *B* has a certain book, he may be asking about some particular copy, e.g. the one that he, *A*, had previously lent to *B*. But *B*'s replies in (1) and (2) suggest that that wasn't the context in which the question was asked. From what *B* is saying in response to *A*, it may reasonably be inferred that "the book" in *A*'s question was a type—a text, of which there are presumably many copies. And when "the book" is understood this way, the claim that you cannot at the same time buy and borrow it holds no water. You can buy one copy and borrow another. Nevertheless this possibility does not affect the way in which (1) and (2) are interpreted. In fact, all those whose judgement we have tested did get

[10] Note that the impossibility of buying what you have just lent derives from the contractual properties of these verbs. This information is missing from the entries for *kaufen* and *ausleihen*$_2$ which we gave in Section 2. Complete entries for these verbs should of course contain this information.

the interpretations we informally described in Section 2; and when queried, they all testified to having understood the book to be a book in the type, not the token sense. In the light of what we have been saying this is puzzling: Why is it that we cannot get a reading for *B*'s answer to the effect that he has now got the book twice over—one copy borrowed and one copy bought?

The solution to this problem crucially depends, it turns out, on the role of discourse particles. To see this compare the following possible answers to *A*'s question:

(19) A: Du hast das Buch doch, oder?
B: (a) Ich habe es gekauft. Aber ich habe es auch ausgeliehen.
(b) Ich habe es gekauft. Ich habe es auch ausgeliehen.
(c) Ich habe es gekauft. Und ich habe es auch ausgeliehen.
(d) Ich habe es gekauft. Und ich habe es ausgeliehen.

In cases (a)–(c) the clearly preferred interpretation is that *B* bought one copy and borrowed another, in other words, precisely the one that is not available for (1). (d) seems to allow for both interpretations.

The moral of the comparison between (1) and (19) is clear: A proper analysis of texts such as (1) and (2) presupposes a detailed theory of discourse particles such as *aber* and *auch*. In particular, the analysis should be able to explain the crucial difference between (1) and (19.a)—i.e. the difference between *aber* + *auch* and *aber* on its own. The analysis we have been able to offer falls well short of this (even if it almost succeeds dealing with the problem presented by (17) and (18)).

We conclude with a hint of how the discourse particles in (1) and (19) produce the effects we have observed. When the verb *ausleihen* is interpreted as *lend*, then the second sentence becomes closely parallel in semantic structure to the first: both describe ways of *B*'s acquiring the book. It is a curious but important property of discourse that when such parallels are sufficiently transparent, then the second of the two parallel expressions must be explicitly marked for this parallel; one device for marking parallels is the German word *auch*, or, in English the words *also* and *too*.[11] But there are also other devices to mark such parallels. For instance, among the particles that can serve this purpose there are also *even* (German *sogar*) and emphatic *and* (or, in German *und*). When these are present, there is no need for *also* or *too* no matter how palpable the parallel. *Und* is of particular interest in the present

[11] The insight into this aspect of the way in which many discourse particels function is due to Sæbø. See in particular Sæbø (1990, 1993).

context; its capacity for marking parallels explains why (d) is ambiguous: even without the support of an accompanying *auch*, *und* counts as a formal mark of the parallel which results when we take *ausleihen* to mean *borrow*, so here this interpretation is not ruled out. In contrast, when *und* is replaced by *aber*, as it is in our original example (1), *auch* must be present to render the parallel interpretation legitimate. For *aber* does not have this capacity.

But even if this points in the right direction, it is a long way from anything that could be called a systematic treatment. In a sequel to the present paper we will make an attempt to deal with some of the phenomena noted in this last section in a less impressionistic manner.

Bibliography

Asher, N., 1993. *Reference to Abstract Objects in Discourse*. Kluwer, Dordrecht.

Bierwisch, M., 1983. Semantische und konzeptuelle Interpretation lexikalischer Einheiten. In: R. Růžička, W. Motsch (eds.), *Untersuchungen zur Semantik*. Akademie Verlag, Berlin, 61–99.

Bierwisch, M., 1988. On the grammar of local prepositions. In: M. Bierwisch, W. Motsch, I. Zimmermann (eds.), *Syntax, Semantik und das Lexikon*. Akademie Verlag, Berlin, 1–65.

Bierwisch, M., 1989. Event nominalizations: Proposals and problems. In: W. Motsch (ed.), *Wortstruktur und Satzstruktur*. Akademie Verlag, Berlin, 1–73.

Caenepeel, M., 1989. *Aspect, Temporal Ordering and Perspective in Narrative Fiction*. Ph.D. dissertation, University of Edinburgh.

Caenepeel, M., Sandström, G., 1993. A discourse level approach to the past perfect. In: M. Aurnague et al. (eds.), *Semantics of Time, Space and Movement—Proceedings oft the 4th TMS International Workshop 1992*. IRIT Université Paul Sabatier, Toulouse, 167–182.

Hinrichs, E., 1986. Temporal anaphora in discourses of English. *Linguistics and Philosophy* 9, 63–82.

Kamp, H., 1981. Événements, représentations discursives et référence temporelle. *Langages* 64, 39–64.

Kamp, H., 1990. Prolegomena to a structural theory of belief and other attitudes. In: C.A. Anderson, J. Owens (eds.), *Propositional Attitudes*. CSLI Publications, Stanford, CA. 27–90. Reprinted in: K. von Heusinger, A. ter Meulen (eds.), *Meaning and the Dynamics of Interpretation. Selected Papers of Hans Kamp*. Brill, Leiden, 2013, 513–583.

Kamp, H., Reyle, U., 1993. *From Discourse to Logic. Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer, Dordrecht.

Kamp, H., Rohrer, C., 1983. Tense in texts. In: R. Bäuerle, C. Schwarze, A. von Stechow (eds.), *Meaning, Use and Interpretation of Language*. de Gruyter, Berlin, 250–269.

Kamp, H., Roßdeutscher, A., 1992. *Remarks on Lexical Structure, DRS Construction*

and Lexically Driven Inferences. Arbeitspapiere des Sonderforschungsbereiches 340, Sprachtheoretische Grundlagen der Computerlinguistik, Vol. 21. University of Stuttgart, Stuttgart.

Kautz, H., 1986. The Logic of Persistence. In: *Proceedings of the Fifth National Conference on Artificial Intelligence* (*AAAI–86*). Association for the Advancement of Artificial Intelligence, Philadelphia, PA, 401–407.

Kowalski, R., Sergot, M., 1986. A logic-based calculus of events. *New Generation Computing* 4, 67–95.

Lascarides, A., Asher, N., 1991. *Discourse Relations and Common Sense Entailment*. DYANA Deliverable 2.5B, Tech. Report. Centre for Cognitive Science, University of Edinburgh.

Lifschitz, V., 1986. Pointwise circumscription: Preliminary report. In: *Proceedings of the Fifth National Conference on Artificial Intelligence* (*AAAI–86*). Association for the Advancement of Artificial Intelligence, Philadelphia, PA, 406–410.

Lifschitz, V., 1987. Formal theories of action. In: F.M. Brown (ed.), *The Frame Problem in Artificial Intelligence. Proceedings of the 1987 Workshop*. Morgan Kaufmann, Los Altos, CA, 35–57.

Lorenz, S., 1993. *Temporales Schließen unter Standardannahmen bei der Verarbeitung natürlicher Sprache*. Ph.D. dissertation, University of Stuttgart.

McCarthy, J., 1986. Applications of circumscription to formalizing commonsense knowledge. *Artificial Intelligence* 28, 89–118.

Partee, B.H., 1984. Nominal and temporal anaphora. *Linguistics and Philosophy* 7, 243–286.

Roßdeutscher, A., 1992. Der Umtausch (the exchange). A case study of anaphora resolution in lexically-based Discourse Representation. In: P. Bosch, P. Gerstl (eds.), *Discourse and Lexical Meaning*. Arbeitspapiere des Sonderforschungsbereichs 340, Sprachtheoretische Grundlagen der Computerlinguistik, Vol. 30. University of Stuttgart, Stuttgart.

Roßdeutscher, A., 1993. *Using Lexical Information to Construct Discourse Representation Structures: A Case Study in Anaphora Resolution*. Arbeitspapiere des Sonderforschungsbereichs 340, Sprachtheoretische Grundlagen der Computerlinguistik, Vol. 33. University of Stuttgart, Stuttgart.

Sæbø, K.J., 1990. *Discourse Particles*, unpublished manuscript, University of Tübingen, Tübingen.

Sæbø, K.J., 1993. Presuppositions and Anaphora. In: H. Kamp (ed.), *Presuppositions and Anaphora*. DYANA–2 Deliverable R2.2A Part II, Tech. Report. ILLC, Amsterdam, 135–192. Revised version published in: *Linguistics and Philosophy*, 19, 1996, 187–209.

Sandewall, E., 1988. *Non-monotonic Entailment for Reasoning about Time and Action, Part I–III*. Research Report LiTH-IDA-R-88, Nos. 27–29. Dept. of Computer and Information Science, Linköping University, Linköping, Sweden.

Sandström, G., 1993. *When-clauses and the Temporal Interpretation of Narrative Discourse*. Ph.D. dissertation, University of Umeå.

Schubert, L., 1990. Monotonic solution of the frame problem in the situation calculus: An efficient method for worlds with fully specified actions. In: H.E. Kyburg, R.P. Loui, G.N. Carlson (eds.), *Knowledge Representation and Defeasible Reasoning*. Kluwer, Dordrecht, 23–67.

Shoham, Y., 1988a. Chronological ignorance: Experiments in nonmonotonic temporal reasoning. *Artificial Intelligence* 36, 279–331.
Shoham, Y., 1988b. *Reasoning about Change*. The MIT Press, Cambridge, MA.
Wunderlich, D., 1992. CAUSE and the structure of verbs. *Linguistic Inquiry* 28, 27–68.

PART FIVE

PRESUPPOSITION

INTRODUCTION TO PRESUPPOSITION

Hans Kamp

The two papers in this part are from the same time and they deal with different aspects of the same issue. The issue is one that arises in connection with semantic presuppositions, associated with particular lexical items or grammatical constructions. A competent speaker, who expresses herself in accordance with the rules of her language and makes use of such a construction or word, must *ipso facto* have the intention to impose the associated presupposition.

The problems that the two papers address both have to do with the following facts about the use of presupposition-triggering words and constructions. First, it is extremely common for the sentences we produce in ordinary speech and writing to contain presupposition triggers. What's more, our sentences often contain not just one presupposition, but two or more. Furthermore, among the words that act as presupposition triggers, there are some that give rise to several presuppositions and not just to one. Many of the utterances we produce thus generate a combination of presuppositions, and one of the tasks incumbent on the interpreter of such an utterance is to verify that all these presuppositions are justified by the context in which the sentence is used. Justification of multiple presuppositions is an intriguing phenomenon for several reasons. First, there is the notion of presupposition justification, as it applies also to single presuppositions. David Lewis was the first to state explicitly that when the context available to the interpreter of a sentence does not verify a presupposition that is triggered by the sentence, then he will accommodate the presupposition. He will assume that the presupposition is part of the context after all, or at least that the speaker thought so. By and large, where matters of presupposition are concerned, verbal communication operates on the principle of 'speaker knows best'. This simple view of accommodation, however, which suggests that you can never go wrong in using presupposition triggers as speaker, is no more than a rough first approximation, because accommodation isn't always possible. Secondly, when accommodation, in the sense of context adjustment, is possible, it often follows, and seems required to follow, particular strategies for which Lewis' formulation doesn't give any clues. For the most part presupposition accommodation isn't simply a matter of adding the given

© HANS KAMP, 2013 | DOI:10.1163/9789004252882_017
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

presupposition to the context as it is, or as what one takes it to be. As pointed out in particular by David Beaver, it often involves taking more new information on board than strictly necessary to get the context to entail the presupposition. What interpreters try to do is to come up with a plausible story about what the speaker could have taken the context to be, if she took, as she apparently did, the presupposition to be verified in it. And she must have taken the context to be like that, for otherwise her use of the presupposing word or construction would have been infelicitous. While the accommodations that interpreters actually make often strengthen the context more than is strictly speaking required for entailment of the presupposition, the accommodated information will, on the other hand, often be too little to entail the presupposition on its own. All that matters is for the presupposition to be entailed by the context, once the accommodated bit of information has been added to it. Kamp & Roßdeutscher (1992, ***Remarks on Lexical Structure***) coined the term 'justification' to cover all the possible ways of making sure that a presupposition is satisfied: (i) verify that the presupposition is contained in, or, synonymously, is entailed by, the context as it is given to him; (ii) accommodate the presupposition wholesale by adding it to the given context; (iii) accommodate some bit of information that has the effect that the presupposition is entailed by the context updated with this bit, though not necessarily by the bit on its own. Much of the presupposition literature can be read as implying that the only options for dealing with presuppositions are (i) and (ii). But many actual accommodations are instances of (iii) without being instances of (ii). Hence (ii) is a special case of (iii); some accommodations take this particular form, but many do not.

When two or more presuppositions are involved in the production of an utterance, further complications can arise. One presupposition may contain other presupposition triggers in its turn, so that what the interpreter faces is a structure in which some presuppositions are nested within other presuppositions. Typically this produces a kind of presupposition duplication effect. For example, consider the sentence *Emma was in love with a different young man as well*. Both *as well* and *different* are presupposition triggers, and each present their presupposition to the recipient. But the *different* presupposition is embedded within the *as well* presupposition. The *as well* presupposition is to the effect that someone else was in love with a different young man. But that means that both Emma and this other person must satisfy the presupposition associated with *different*: the young men they were in love with must both have been different in some context-related way. This carries various implications for the justification of these presuppositions. To consider just one possibility, suppose that the context suggests who the other

person was who was also in love with a '*different young man*' and thereby justifies the *as well* presupposition. But in order that the *as well* presupposition can be justified in this way, the sense in which that young man is different must be similar, or parallel to that in which Emma's young man is different. For instance, if the young man that this other person was in love with was different in the sense of being different from the young man (s)he was in love with at some earlier time, whereas the young man that Emma was in love with was different in the sense of being different from the young man that this other person was in love with, that just won't do. *The Importance of Presupposition* explores a couple of such cases of nested presuppositions in depth, revealing hidden ambiguities in such sentences, which can lead to distinct interpretations depending on what resolution strategy is chosen.

Another type of problem arises when a bunch of presuppositions are triggered by different parts of a single comparatively compact phrase, e.g., the constituents of a single clause, or even of some smaller unit, such a single noun phrase. Such clusters of presuppositions may sometimes be justifiable through the accommodation of a seemingly small piece of information, which will make all presuppositions in the cluster fall into place, when it is added to the given context. Moreover, where that is so—that assuming this small piece of information is all it takes to justify the presuppositions—there is often a kind of necessity to this kind of accommodation: the context must be accommodated in *this* way, and in no other. The net effect of this necessity is that such sentences strongly convey the impression that they entail the missing bit of information whose accommodation makes the presupposition or presuppositions verifiable.

Presupposition Computation and Presupposition Justification: One Aspect of the Interpretation of Multi-Sentence Discourse investigates one such case in considerable depth, showing how the different presuppositions that make up the bunch can be represented and how one natural assumption—of the 'missing piece of information'—leads to the justification of all of them. But the paper is like *The Importance of Presupposition* in that it does not go beyond the particular examples to which it is devoted. No general account is given of why the apparent entailment effect, for this case and a few others like it, is as strong as it appears to be.

Both papers in this section are thus more exploratory in nature than presenting well worked out general accounts of the phenomena they bring to attention. Each pinpoints a problem that, for all I know, had not been explicitly identified before. I suspect that it is the general approach to presupposition within DRT—in which explicit representations of the presuppositions are first computed and the question of justification is dealt with only after

that has been done—that makes it possible to uncover these problems and recognize them for what they are. But the papers don't go any farther than that. I still see these problems as real, interesting and in need of a more systematic investigation. But I am not aware that a great deal of progress has been made towards more general treatments, or even towards a larger catalogue of individual cases, since the papers were first published. So far, they seem to have been something of a cry in the wilderness.

PRESUPPOSITION COMPUTATION AND PRESUPPOSITION JUSTIFICATION: ONE ASPECT OF THE INTERPRETATION OF MULTI-SENTENCE DISCOURSE[a]

Hans Kamp

1. Introduction

As a rule discourse meanings are more than plain conjunctions of sentence meanings.[1] And this "more" is often the effect of interpretation principles that are an integral part of linguistic knowledge, and thus legitimate objects of linguistic study. This observation has been the main driving force behind dynamic theories of discourse semantics such as Discourse Representation Theory (DRT) and its extensions S(egmented) DRT and U(nderspecified) DRT.

[1] The substance of this paper goes back several years. I have used the example with which the paper is exclusively concerned repeatedly in oral presentations, going back as far as 1996, as a simple illustration of the inferential effects of presupposition justification, and of the interaction between different presuppositions triggered within one and the same sentence. I know that over the years I have benefited from feedback of more people than I now recall. On pain of offending some who ought to be mentioned here, I would like to thank some of them by name: Peter Krause, Uwe Reyle, Antje Roßdeutscher and Michael Schiehlen have helped me not just to understand better the details of the example in question, but also to arrive at the general perspective of the role of presupposition in discourse of which the treatment proposed in this paper is a single illustrative instance. I am grateful also for the recommendations of an anonymous referee, which I have found very helpful and with which I have tried to comply to the best of my ability. Thanks, finally, to Peter and Antje for seizing this opportunity to wean me (almost) off Word 5.1 and throw me into the deep end of LaTeX.

[a] Editors' note: This article was originally published in M. Bras & L. Vieu (eds.), *Semantics and Pragmatics of Discourse and Dialogue: Experimenting with Current Dynamic Theories*, Elsevier, Amsterdam, 2001, 7–84, (Current Research in the Semantics / Pragmatics Interface 9—now published by Brill). We carefully edited the article and corrected some typos. We would like to thank Emar Maier for careful proofreading and constructive comments. Hans Kamp provided additional footnotes, indicated by Roman letters, which clarify some of the arguments in the original text.

© HANS KAMP, 2013 | DOI:10.1163/9789004252882_018
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

Our understanding of the mechanisms for computing discourse interpretations is still quite limited. But there has been significant progress, and one thing which has become much clearer in recent years is the role played by presupposition. Most natural language sentences come with presuppositions of one kind or another. These presuppositions must be justified in the context in which the sentence is used. In ongoing discourse or text, the relevant context is often the 'discourse context' established by the sentences or utterances preceding the one whose presuppositions are in question. In such cases presupposition justification is justification in the discourse context.

Often presupposition justification takes the uneventful form of finding the given presupposition or presuppositions satisfied in the given context. But not always. In many other cases the context does, as it stands, not quite measure up to the verification task. It doesn't verify the presuppositions as is, but needs adjustment—by 'accommodation', as linguistic parlance has it—to fit the requirements that the presuppositions impose. This doesn't mean, however, that whenever direct verification fails, the unverified presuppositions get accommodated lock, stock and barrel. There are many instances where the context, while failing to verify the presuppositions at issue, nevertheless contains much of what is needed for their verification; just a small bit of information is missing to make verification complete. In such cases it is not only possible to achieve accommodation by accommodating just this little bit; as a rule, when such a limited accommodation suffices, that accommodation is highly preferred or even mandatory: Even if other, more comprehensive accommodations are possible which also transform the given context into one in which the presuppositions are also satisfied, the rules of interpretation require the smaller, 'less costly' accommodation. As a consequence, the bit of information that gets accommodated will be perceived as one of the discourse's entailments.[2]

[2] To my knowledge the fact that presupposition justification often takes the form of exploiting as much of the information that is part of the context as given, while accommodating only those bits which are needed to make verification complete, and the importance of this fact for discourse interpretation, were first stressed explicitly in Kamp & Roßdeutscher (1994). The term 'presupposition justification' was introduced in that paper as a cover term which subsumes (i) the cases of straightforward presupposition verification (including cases of 'anaphoric binding', van der Sandt 1992), (ii) the cases of wholesale accommodation and (iii) the cases which can be seen as a mixture of verification and accommodation. More about this in Section 6.

An example of the effect that presupposition justification can have on discourse meaning is the following discourse:

(1) I gave the workers a generous tip. One thanked me. The other one left without saying a word.

Anyone who reads this little text is led to conclude that the number of workers mentioned in the first sentence was two. It is clear that this conclusion is available only after all three sentences have been processed.

It is not hard to see that this conclusion crucially depends on the interpretation of the subject NP *The other one* of the third sentence. In the following sections I will argue that it is the need to justify the different presuppositions triggered by this phrase in the context provided by the first two sentences of (1) which is responsible for it, and then I will analyse in detail how the justification process gives rise to it.

(1) also illustrates another aspect of presupposition justification. The inference that there were exactly two workers who got a tip depends on the interaction between several presuppositions. (In this case the relevant presuppositions are all triggered by different components of the NP *the other one*: the word *one*, the word *other*, the definite article *the* and the number feature *singular*). It is the requirement that the package consisting of these different presuppositions be satisfied as a whole that accounts for the inference. In fact, it is easy to see how closely this conclusion depends on the exact form of the subject NP of the third sentence: As we will see in Section 6, 'small' changes in this NP, such as from *the* to *a* or from singular to plural, yield different conclusions about the size of the set of workers.

2. General Assumptions about the Syntax-Semantics Interface: Construction of the Preliminary Representation of the First Sentence

In my reconstruction of (1) I will assume syntactic analyses of the individual sentences as given. The analyses of which I will make use conform (to the best of my knowledge) to assumptions made within generative linguistics, in particular those associated with the Theory of Government and Binding. However, I do not wish to make a definitive commitment to this or any competing theory of syntax.[3] On the semantic side, the representations I will be using are based on Discourse Representation Theory, more precisely,

[3] This paper is not intended as a contribution to the syntax-semantics interface in the

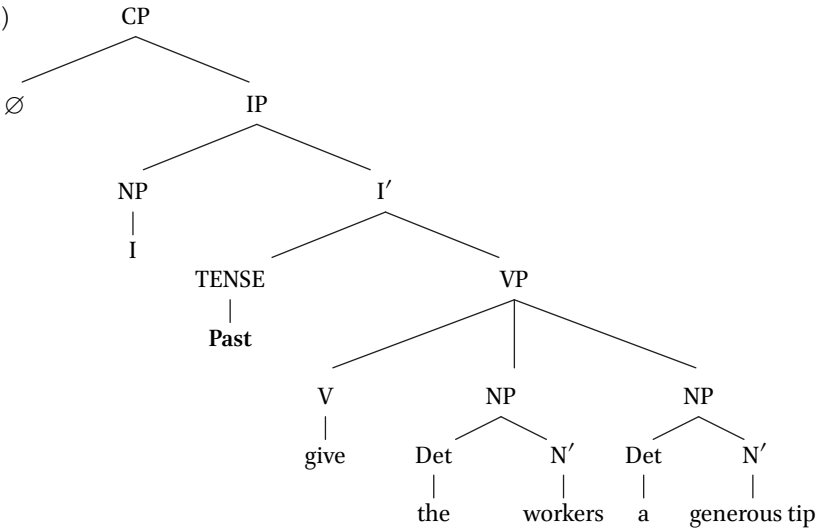
the representations belong to a DRT-based formalism closely related to that used by van der Sandt in his seminal presupposition paper (van der Sandt 1992). The use of a representation formalism that permits integration of the representation of the current sentence of a discourse or text into the representation of the preceding sentences is, for reasons that should be obvious in the light of what I have said about (1), essential to the kind of analysis the basic outline of which I sketched in section 1. Specifically, we will proceed by and large along the path which Van der Sandt and others working within a DRT-based framework for the treatment of presuppositions have broken:

(2) Given a representation K of the context provided by the sentences $s_1 \ldots s_{n-1}$ of the given discourse:
   (i) construct a *preliminary* representation of sentence s_n in which all the presuppositions generated by presupposition triggers in s_n are explicitly represented;
   and then
   (ii) integrate this preliminary representation, if possible, into the context representation K, thus transforming K into a new context representation K' for the next sentence.

We begin with the first sentence of (1), assuming the syntactic analysis given in (3):

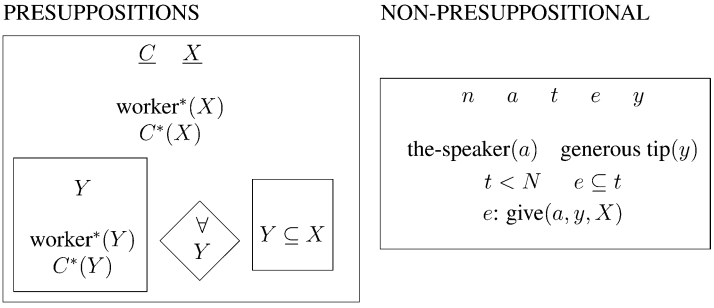
strict sense of the word—where the concern is to combine a specific, fully articulated syntactic theory with a systematic account of how semantic representations are derived from the syntactic structures the syntactic theory proposes. Rather, what I will present is to be seen as an attempt to identify some of the necessary elements that any syntax-semantics interface of the kind alluded to should incorporate. In this context, the syntactic analyses which I will assume should be regarded as incorporating sets of syntactic properties of the analysed sentences which any viable theory of syntax should make available in some form. Of course, this way of proceeding, which is now usually referred to as specifying an "abstract syntax-semantics interface", is meaningful only to the extent that there is agreement among syntacticians about the extensions of these syntactic properties, even if there may be substantive differences about the principles which determine these extensions. For more on the syntax-semantics interface see Kamp (2001).

(3)



The preliminary representation of this sentence which we will be using is that in (4)

(4) Preliminary representation for (1):



Here is a gloss for reading representations like (4).

(i) The representation consists of a non-presuppositional part on the right and a presupposition structure preceding it on the left. (In the present case this structure consists of one presupposition only, but we will later see representations involving sets of presuppositions and even cases of presupposition nesting.) The non-presuppositional part of the representation is the one that identifies the actual propositional content of the represented sentence. This part gets added to the context representation once the presuppositions in the presupposition structure have been justified in it.

(ii) Both the non-presuppositional part and the presuppositions in the presupposition structure are given in the form of (one or more) Discourse Represen-

tation Structures (DRSs).[4] Here is a rough indication of how these are to be interpreted. Consider in particular the non-presuppositional part of (4). This is a DRS which, like any other, consists of a DRS *universe*—the set of *variables* (also called "discourse referents") $n, \ldots, y$ at the top of the representation—and a set of *conditions*, entered below them. The variables represent entities of different types—n and t represent times, e an event and a and y ordinary individuals. Their presence in the universe of the DRS signifies that there exist entities of the corresponding types which satisfy the DRS-conditions, in which the variables occur as arguments. n acts as an indexical element; it stands for the utterance time of the represented sentence.

(iii) Most of the conditions of this DRS are self-explanatory. Just to be on the safe side: "e: give(x, y, z)" can be paraphrased as: "e is an event of the type of x giving y to z".

(iv) From these few remarks it should be clear that the non-presuppositional DRS represents a statement to the effect that there is an individual a who is the utterer of the given utterance of (1), a generous tip y and an event e of a giving y to a set X consisting of two or more workers.

(v) The presuppositional part of (4) consists of two underlined variables X and C, which make up the universe of this part, and a number of conditions. The conditions "worker$^*(X)$" and "$C^*(X)$" and the complex condition below them jointly express that X consists of all the individuals that (a) are workers and (b) satisfy the predicate C. (The first two conditions state that X consists of such individuals, and the last condition that X is the largest such set, including all others that satisfy the first two conditions.)

(vi) As regards the details of the presuppositional part:

  (a) Capital letters are used for variables whose values are sets of two or more elements. Alternatively, such variables can be taken to range over the non-atomic entities of a mereological universe in the sense of Link (1983).

  (b) The * in "worker$^*(X)$" and "$C^*(X)$" is an operator which turns a predicate of individuals into a distributively interpreted predicate of sets. For instance, "worker*" is that predicate of sets which is true of a set Y iff each member of Y satisfies the predicate "worker".

  (c) The third condition is *complex* in the technical sense of DRT in that it is built from component DRSs with the help of one or more logical operators. In this case the operator is the universal quantifier, which, in the spirit of Generalized Quantifier Theory combines a "restrictor" (the left DRS) with a "scope" (the right DRS). The interpretation of this condition should be intuitively clear: Every possible value of the quantified variable Y which yields a verification of the restrictor DRS

[4] Familiarity with DRT (e.g. with an introduction such as Kamp & Reyle 1993) will make access to this and later representations much easier, but I hope that the comments given in the text will make it possible to follow what is going on even for those who have little or no knowledge of the theory.

also allows verification of the scope DRS. In this paper we will be dealing with only one other type of complex condition, in which the logical operator is negation.

(vii) The presupposition of (4) is triggered by the plural definite description *the workers*. The intuition behind this is as follows:

(a) Like other definite NPs, definite descriptions carry the presupposition that their referent can be uniquely identified in the given context. The referent must either satisfy the descriptive content of the description; or, alternatively, it must consist of elements that satisfy that descriptive content. Moreover, the descriptive content may be assumed to identify the referent uniquely, but often it will do so only in conjunction with an additional restriction C which is to be recovered from the context. Thus the description imposes in general a double task of identification-in-context: of the referent itself and, as part of this, of the restricting predicate C. In certain cases the descriptive content will suffice for identification by itself; in such cases C can be identified with the universal predicate.

(b) The underlining of the variables C and X is used as an indication of the requirement that antecedents be found for these variables in context. The paradigmatic example of this kind of underlining is that of the variable introduced by an anaphoric pronoun, for which an anaphoric antecedent has to be found in the discourse context.

(c) Plural descriptions differ from singular ones in that any referent of the former must be a set of cardinality >1 (or: a non-atomic element of the mereology), whereas any referent of the latter must be an individual (or: an atomic element). Thus, both plural and singular descriptions generate presuppositions of (i) existence of something satisfying the relevant conditions and (ii) maximality vis-a-vis those conditions. Singular and plural, then, add the respective conditions of atomicity and non-atomicity. The conjunction of the three conditions existence, maximality and atomicity amounts to unique satisfaction—the classical presupposition associated with singular definite descriptions.

Remark: Arguably the displayed presupposition of (4) is not the only one that is generated by the first sentence of (1). First, there is a pre-state presupposition associated with the verb *give*: In order that an event e of the type "e: give(x, y, z)" can take place it is necessary that at the moment when e starts, x has y. In addition, it could be argued that the definite NP I carries a presupposition of reference. In practice this presupposition is always satisfied in the context in which I is used, insofar as use—that is: utterance—implies that there is someone who does the uttering, and thereby is the referent of that token of I. And, finally, the sentence generates several *selectional restrictions*. Selectional restrictions are presuppositional constraints associated with the argument positions of verbs and other lexical predicates.

They are constraints imposed by the lexical meaning of the predicate, to the effect that the entities occupying these argument positions must be of certain specified ontological types. As none of these other presuppositions has any relevance to the problem about (1) in which we are interested, they have not been explicitly represented. Whenever a presupposition is not explicitly represented, whether in (4) or in the preliminary representations of the next two sentences, the implicit assumption will be that whenever necessary, the presupposition is accommodated.

The principles by which (4) is derived from (3) I won't spell out in full formal detail.[5] But the following tips ought to be sufficient to see how the construction goes:

(a) The representation of the non-presuppositional part of the DRS for (3):

(i) The lexical entry for the main verb contains a characterisation of the type of the eventualities (i.e. events or states) which the verb can be used to describe.[6] Here we make use of a simplified version of the semantic component of the entry for *give*, for which we assume the form given in (5):

(5) e: give(x, y, z)

The contribution made by the occurrence of *give* in the sentence (3) is to introduce an instance of (5) into the semantic representation, where the variables x, y, z are arguments contributed by the corresponding argument phrases of (3)—see under (iii). In addition, the occurrence of *give* leads to a condition of the form "$e \subseteq t$" which says that the event e is located within the temporal interval t.

(ii) The variable t is further constrained by information attached to nodes higher up in the tree. Here we assume that this information is restricted to that provided by the simple past tense (in *gave*), and that this means that t lies entirely before the utterance time n. Moreover, both e and t are existentially bound at the level of CP. In the DRT notation used here this takes the form of introducing these variables into the universe of the representation of the non-presuppositional part of (3).

(iii) Each argument phrase introduces a variable which (a) gets inserted into the relevant argument slot of the semantic representation of the verb; (b) serves as argument to the descriptive content of the argument phrase itself; and (c) gets *bound* according to principles which depend on the form of the argument phrase, in particular on the kind of determiner (in those cases where the phrase has a determiner), as well as on

[5] See Kamp (2001) for further comments on the construction of preliminary sentence representations.

[6] A lexicon is needed both to support syntactic parsing and for the conversion of the syntactic parse into the preliminary semantic representation; I won't say much about the lexicon in this paper, leaving the diverse roles it has to play for the most part implicit.

other aspects of the function that the given occurrence of the phrase plays in its sentential environment.[7] Binding, as the term is used here, can take different forms. One of these is the quantificational binding familiar from standard logic. (This is the kind of binding involved in the interpretation of genuinely quantificational NPs, which begin with determiners such as *every* or *most*.) But variables can also be "bound" by being identified with other designators—the typical case being that of the variables introduced by anaphoric pronouns, which are "bound" by their anaphoric antecedents—or by the default binding familiar from the classical analyses of indefinite NPs of original File Change Semantics and DRT. (The binding of *e* and *t* spoken of under (ii) is an example of this third main type.)

(iv) We follow van der Sandt in assuming that the binding conditions for variables introduced by definite noun phrases always take the form of presuppositions: The binding condition for any such variable is a constraint on the context in which the NP introducing the variable is used; only when it is possible to find (or accommodate) an "antecedent" for the variable in the given context can interpretation of the sentence in question proceed further.

(v) Precisely how the binding conditions which the different types of NPs impose on their variables are to be dealt with depends in part on the order in which the construction of interpretations of the different constituents specified by the syntactic analysis of the sentence proceeds. The construction algorithm I am assuming in this paper operates on trees like (3) in a bottom-up fashion. This means that in the construction of the representation of an NP both the variable it introduces and the binding condition which it associates with that variable have to be carried upwards jointly with the representation of the NP and of the various larger constituents which contain the NP as a subconstitutent, until the point where conversion takes place (either into a proper binding or into a presupposition). In the case of the presuppositional binding conditions associated with definite NPs this means that they are to be carried all the way up to the top node of the syntactic tree, at which point the construction of the preliminary representation of the clause is completed and the binding conditions are turned into genuine presupposition representations, which then get prefixed to the non-presuppositional representation of the clause. In the treatment of (1) presented here we will not go through the successive steps of this "compositional" method of constructing preliminary sentence representations, however, and the additional complications that are connected with this procedure may be ignored.

[7] For instance, different uses of definite descriptions involve different binding strategies. The same is true for many other types of definite NPs and also for indefinite noun phrases.

(b) The representation of the presuppositional part of the DRS for (3):
  (i) The construction of the representations of the different presuppositions generated by a given sentence will start at the moment the construction reaches the triggers responsible for them. In (4) the only presupposition is that triggered by the definite NP *the workers*. In (vii) of the gloss of (4) I indicated the intuitive motivation behind this presupposition. As that discussion indicated, we are dealing strictly with three distinct sources of presupposition: (i) the identifiability of the additional predicate *C*, something which the definite article *the* shares with quantificational determiners like *every*, *most*, *many*, etc.; (ii) the maximality condition, which is the specific contribution made by *the*, and (iii) the condition of non-atomicity connected with the plural. We will simply assume here, without working out the details, that these different lexical and morphological elements introduce their respective presuppositions into the representation via appropriate stipulations attached to their entries in the lexicon.

Warning: The presuppositional part in (4) is an unusually simple example of what presupposition structures of sentences can be like. More often than not sentences generate presupposition structures of much greater complexity. We will encounter a somewhat more complicated case when we come to deal with the third sentence of our example, but it too is quite modest when compared with, for instance, many of the propositions triggered by particles like *too* or *again*.

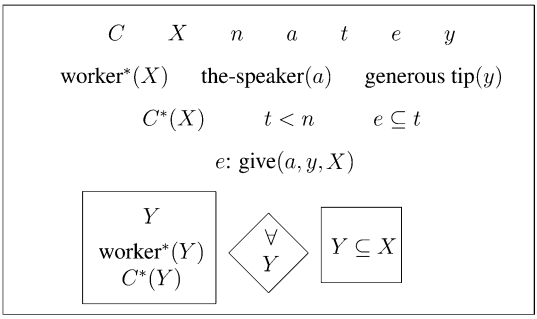
It ought to be clear from these remarks that the preliminary representation derived from (3) using the principles outlined is as in (4).

3. From Preliminary Representation to Discourse Representation. Presupposition Justification and Accommodation

Since in (4) we are dealing with the preliminary representation of the first sentence of a new discourse, there is no discourse context which could help us to justify the representation's presupposition. Thus, in this case accommodation is the only available strategy. Since this is always the situation at the start of a discourse, accommodation of the presuppositions of the first sentence tends to be the accepted thing. Wholesale accommodation, which the interpreter is wont to resort to in such a situation, has the effect of incorporating first the presupposition(s) into the (as yet non-existent) context, and then adding the non-presuppositional content to that.

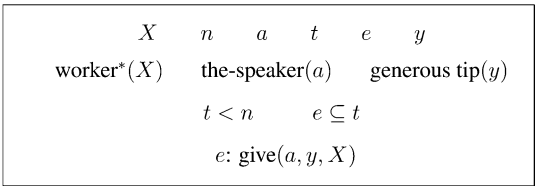
The result of wholesale accommodation in the case of (4) can be represented as in (6):

(6)



Since the accommodated predicate *C* has no other function than that of constraining the set of workers *X*, we obtain a simpler, logically equivalent representation if we eliminate all reference to *C*. The new representation is given in (7):

(7)



4. The Second Sentence

We now turn to the second sentence of (1), repeated as (8).

(8) One thanked me.

The part of (8) that requires our attention is the subject phrase *one*. The English word *one* has a number of different uses. In particular, there are different possibilities for the interpretation of the word when it occurs as the sole constituent of a noun phrase, as it does in (8). To be precise, there are, by my counting, three distinct interpretations of NPs of the form *one*: (i) as proper name of the first positive integer; (ii) as the impersonal pronoun (like French *on* or German *man*); and (iii) as a "dummy indefinite" NP, with a meaning that would be more transparently rendered by the archaic expression *a one*, where *one* acts as the stand-in for some common noun phrase, while *a* does its usual duty as indefinite article.

One problem which the interpreter of an instance of the NP *one* has to solve is which of these three different possibilities he is dealing with. It is intuitively clear that in (8) we are dealing with an instance of the third

possibility. But how is it that the other two possibilities are eliminated? The answer is quite different for each of the two possibilities (i) and (ii). (i) can be eliminated immediately because of the selectional restrictions associated with the subject argument of the verb *thank*: Numbers do not thank—not even in any readily conceivable metaphorical sense.

But the second possibility, according to which *one* would refer to a set of "relevant individuals", is not so easily dismissed. In the present context the "relevant individuals" would presumably be the workers mentioned in the first sentence, given that the interpretation of *thank* requires finding a subject that has something to thank the speaker for. In the context set by the first sentence only the workers mentioned in it seem to qualify as satisfying this condition. That the occurrence of *one* in (8) is nevertheless not to be assigned this interpretation has to do with the way in which the discourse continues. If *one* were to refer to the set of workers, rather than to one of them, then there would be no basis for interpreting the subject of the third sentence. For now I will set aside this additional complication. I will return to it in Section 7.

Once possibility (iii) has been chosen, the next problem is to interpret *one* according to the rules associated with that possibility. As noted, according to (iii) *one* is an indefinite NP with a dummy noun. Thus, its interpretation must involve the introduction of a new variable representing its referent, together with an identification of the nominal predicate that the dummy noun *one* stands for. The rule for finding this nominal predicate appears to be simply that of finding a common noun phrase from the antecedent discourse that is salient at the point where *one* is to be interpreted, with recency an important one among the factors by which the salience ranking is determined.[8]

---

[8] As in most other cases where salience appears to play a role, it is not easy to say exactly how the relevant salience ordering is determined. But that some sort of salience enters into the question how the common noun phrase is chosen is shown by examples where there seems to be an unresolvable tie between candidates which renders the use of *one* infelicitous. Thus in (9) neither taking the nominal antecedent to be *boy* nor taking it to be *girl* seems to make for a completely satisfactory interpretation:

(9) ? There were three boys and three girls in the room. One was asleep.

Probably (9) should not be regarded as really ungrammatical. Indeed, for all I know, some speakers seem to accept it, with the interpretation that there was one member of the set of three boys and three girls in the room that was asleep. But personally I cannot get this reading for (9).

N.B. Fully acceptable for me is a variant of (9) in which the subject of the second sentence is not *One* but *One of them*. But the interpretation of *one of them* is a quite different story. Here

As far as I can judge, the NP *one* cannot very well be used to pick an element from a previously introduced set if that set has not been given as the (contextually restricted) extension of a nominal predicate used in the antecedent discourse. Thus I cannot easily get *one* in (10).

(10) Fred, Carl and Lucy had arrived. One had brought a bottle of wine.

to mean "one of Fred, Carl and Lucy". I have a clear preference in such cases for *one of them* over the simple *one* (cf. fn. 8). However, if the set is given as the extension of some noun N, then *one* can be understood as referring to a member of that set, and not just to something or other falling under the noun generally. Thus one natural interpretation of *one* in (11.i) would seem to be that according to which one of the boys who had arrived lived in Reutlingen. But the less constrained interpretation, according to which *one* is short for "one boy" simpliciter, certainly exists too, as shown by (11.ii)

(11.i) Three boys had arrived. One lived in Reutlingen.
(11.ii) Three boys had arrived. One was still missing.

The interpretation of *one* in (8) that we want is of the type illustrated by the first interpretation of (11.i): The referent of *one* is one of the workers mentioned in the first sentence. The alternative, according to which *one* is understood as standing for some worker or other, but not necessarily one of those mentioned in the first sentence, is problematic because of the constraints imposed by the interpretation of the third sentence. We return to this question in Section 7.

(12) summarizes the reflections of the last two paragraphs on the possibility of interpreting *one* as an indefinite NP with dummy noun.

(12) (Interpretation of *one* as a one-word indefinite NP)
1. "Pure N" interpretation
   A. Add to the presupposition structure:
      (i) a presupposition expressing the requirement of finding in the discourse context a salient common noun phrase *N*
   B. Add to the non-presuppositional part:
      (i) a new (not previously used) variable *u*
      (ii) the condition "$N(u)$".

---

the plural pronoun *them* can be understood as referring to the set of the three boys and three girls introduced by the subject of the first sentence. Moreover, occurrences of *one* in explicitly partitive constructions such as *one of them* can be interpreted as referring to some member of the set referred to by the embedded partitive NP.

2. "Set of Ns" interpretation
   A. Add to the presupposition structure:
      (i) a presupposition expressing the requirement of finding in the discourse context a salient representation *U* of a set of two or more individuals for which the context contains a characterisation as the set of all *N* or as the set of all *N* which satisfy some further conditions, where *N* is a common noun phrase.
   B. Add to the non-presuppositional part:
      (i) a new (not previously used) variable *u*
      (ii) the condition "$u \in U$".

It is the second of the two options offered by (12) which we want in our preliminary representation of (8). Even though we have not spelled out the instructions of (12.2) in full formal detail, it should be clear that with this option the presuppositional part of the preliminary representation can be represented as in (13).

(13)

PRESUPPOSITIONS

$\underline{N}$ $\underline{U}$
$N^*(U)$

NON-PRESUPPOSITIONAL

n a t' e' u
the-speaker(a) $u \in U$
$t' < n$ $e' \subseteq t'$
e': thank(u, a)

As regards the non-presuppositional part of (13): The inclusion of *u* in the DRS-Universe and the condition "$u \in U$" among its conditions has already been accounted for in (12.2). For the other conditions, concerning the speaker *a*, the event e' described by the main verb and its temporal location in the past of *n*, see the comments on (4).

The context for (8) that is provided by (6) provides only one viable solution for the presuppositional task which (13) identifies: Identify *N* with *worker* and *U* with *X*. The result of this justification and the subsequent merge of (13) with (6) is given in (14):[9]

[9] No account has yet been given for the condition "$e < e'$". This condition has a somewhat different status than all others we have so far encountered, which is the reason I have placed it within parentheses—by way of a warning that the condition falls outside the scope of what is explicitly discussed in this paper.

The basis for the condition "$e < e'$" is the rhetorical structure of the discourse (1)—or,

(14)

X n a t e y t' e' u			
worker*(X)	the-speaker(a)	generous tip(y)	$u \in X$
	$t < n$ $\quad e \subseteq t$		
	e: give(a, y, X)		
	$t' < n$ $\quad e' \subseteq t'$		
	$(e < e')$		
	e': thank(u, a)		

(14) is the context for the interpretation of the third sentence of (1), to which we now turn.

5. The Third Sentence

We recall the third sentence of (1):

(15) The other one left without saying a word.

As we noted in our informal discussion of (1) in Section 1, the crux of the example is the interpretation of the subject NP of (15). The various elements of this NP each introduce their own presuppositional constraints and it is the joint resolution of these constraints in the context (14) which yields the result for which we are aiming: the conclusion that there were two workers. We list once more the relevant elements of the subject phrase and the associated presuppositions:

(16) (Presuppposition-carrying elements of *the other one*)
- (i) The definite article *the*.
  Presupposition: the referent of the NP is the maximal mereological element satisfying the conjunction of
  - (a) the descriptive content of the description; and
  - (b) (possibly) an additional contextual constraint
- (ii) The value "singular" of the NP's number feature
  Presupposition: the referent of the NP is an individual (atomic element of the mereology).

---

more precisely, of the initial segment of (1) which consists of its first two sentences. It is a typical (though not invariable) property of two successive event sentences in the simple past that the event reported in the second sentence is understood as coming after the event reported in the first sentence. It is this assumption of temporal odering between the two events which the parenthesised condition of (14) expresses. The general mechanisms which yield such conditions belong to a dimension of discourse interpretation that is not considered in this paper. For discussions see Kamp (1981), Kamp & Reyle (1993) and especially Lascarides & Asher (1993).

(iii) The dummy noun *one*.
Presupposition: Much as for the dummy noun *one* of the second sentence (8). For details see below.
(iv) *other*
Presupposition: See below for details.

First the case of *one*. As indicated under (16.iii), this is an occurrence of *one* that is similar to the one in the second sentence. But there are also differences. In (15) we are dealing not with a one-word NP, but with an NP consisting of a determiner, an adjective and what appears to be a clear case of a dummy noun (and nothing more). In this case the presuppositional condition connected with *one* is limited to the requirement of finding an antecedent for this dummy noun. Once more, however, we are confronted with the question what form this antecedent may take. I confess to being not entirely certain whether the facts are the same for this case as I have taken them to be for the NP *one* in the second sentence, but it seems likely to me that they are. So, formulated in the same format as we used in (12) to summarise our discussion of the contributions made by *one* to the interpretation of the second sentence, we get for the contribution made by *one* to the preliminary representation of (15):

(17) (Interpretation of *one* as a dummy common noun)
1. "Pure N" interpretation
   A. Add to the presupposition structure:
   a presupposition expressing the requirement of finding in the discourse context a salient common noun phrase *N*.
   B. Add to the non-presuppositional part:
   the condition "$N(\eta)$", where η is the discourse referent introduced for the NP of which *one* is the nominal head.
2. "Set of Ns" interpretation
   A. Add to the presupposition structure:
   a presupposition expressing the requirement of finding in the discourse context a salient representation *U* of a set of two or more individuals for which the context contains a characterisation as the set of all *N* or as the set of all *N* which satisfy some further conditions, where *N* is a common noun phrase.
   B. Add to the non-presuppositional part:
   the condition "$\eta \preceq U$",[10] where η is the discourse referent introduced for the NP of which *one* is the nominal head.

Somewhat more complex is the presuppositional condition connected with *other*. Intuitively the semantic contribution that a prenominal occurrence

[10] For an explanation of $\preceq$ see fn. 13.

of *other* makes to the interpretation of the NP γ to which it belongs involves implicit reference to something or some things from which the referent of γ must be disjoint. Moreover, the referent of γ and this something (or some things) must be of the same "kind". More precisely, that from which the referent is said to be distinct must fall under the concept denoted by the common noun phrase to which *other* is prenominal. I use the phrase "fall under the concept denoted by the common noun phrase …" advertently, instead of the simpler "satisfy the common noun phrase". For while this concept is sometimes just the one expressed by the common noun phrase as it stands (in which case the simpler formulation would do), in other cases it may be more restricted, e.g. the concept of belonging to a certain set of things to which the common noun phrase applies. It is this last case with which we are dealing in our example: Like the *one* of the second sentence that of the third sentence will end up being interpreted as the set X of workers introduced by the first sentence and the concept in question will thus be that of belonging to that set.

Using once more the format of (12), we get for the contribution of *other*:

(18) (Interpretation of *other* as a prenominal adjective)
Let C be the concept denoted by the common noun phrase to which *other* is prenominal, and η the discourse referent representing the referent of the NP of which this CNP is the head.[11]
(A) Add to the presupposition structure:
a presupposition expressing the requirement of finding in the discourse context an individual or set ζ, such that $C(\zeta)$.
(B) Add to the non-presuppositional part:
the condition "$\eta \# \zeta$".[12]

As I said earlier, the presupposition package of the third sentence is more complex than those connected with the first two sentences of (1). Not only are we now dealing with several presuppositions at once. These presupposi-

[11] As in Kamp & Reyle (1993) I am using Greek letters for variables whose range subsumes both the atomic and the non-atomic part of the mereology (or, in other words, which can take as values both individuals and sets). The need for using such "neutral" variables here will become more apparent as we go along.

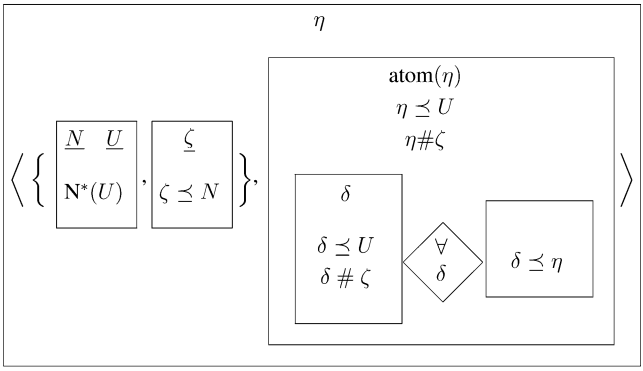
[12] The sign # is used to express disjointness between two elements of the mereological universe. Thus, when both elements are non-atomic, then # signifies disjointness in the familiar set-theoretical sense, if one is atomic and the other not, then # means that the first is not a member of the second, and if both elements are atomic, # simply means difference.

tions stand moreover in a kind of subordination relation. The definite article *the* gives rise to a "main" presupposition, which introduces a discourse referent η (representing the referent of the subject NP), for which an antecedent must be found in the context. The referent must satisfy the condition of being maximal with respect to the descriptive content of the NP. This content is given by the words *other* and *one*. The presuppositions which these words introduce in their turn have to be prefixed to the Condition part of the presupposition DRS introduced by *the*, the universe of which consists of the discourse referent η. In this way we get a nested presupposition structure. While this structure is more complex than those we encountered earlier, it is still quite simple when compared to what is found in other cases. (For examples see e.g. (3).)

In the presentation of the preliminary representation of (15) given in (19) below I have only represented the presuppositions connected with the subject NP:[13]

(19)[b]

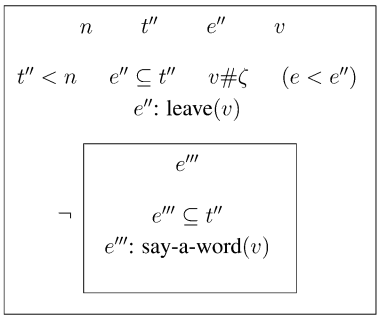
PRESUPPOSITION STRUCTURE



[13] $\delta \preceq \zeta$ means that δ and ζ stand in the mereological "part of"-relation. When both δ and ζ stand for sets, this amounts to inclusion, when one of them is atomic, it amounts to membership.

[b] The difference between underlined and non-underlined presupposed discourse referents is discussed in Section 6.

NON-PRESUPPOSITIONAL



After what has already been said about the presuppositional part of (19), this part should require no further comment. As regards the non-presuppositional part, note the complex condition in the lower right corner, consisting of the DRS whose universe is $\{e'''\}$ and which is prefixed by the negation operator ¬. This is our second example of a complex condition. The semantic significance of this condition ought to be intuitively clear: The condition is satisfied if there is no way of satisfying the embedded DRS—that is: no way of assigning an event to e''' such that the conditions of the DRS are fulfilled.[14]

What remains is the justification of the different presuppositions of (19) in the context given by (14). To a considerable extent this is a combinatorial problem: How can the anaphoric elements of the presuppositional part of (19) be identified with suitable elements of (14) so that the conditions in the presuppositional part are met? But as we have seen, the problem is actually more complicated than that, for no identification will meet the presuppositional constraints as things are. There is one possible combination of identifications, however, which comes close to this. It is stated in (20):

(20) $N \Rightarrow \text{worker}^*; U \Rightarrow X; \zeta \Rightarrow u; \eta \Rightarrow X \setminus \{u\}$.

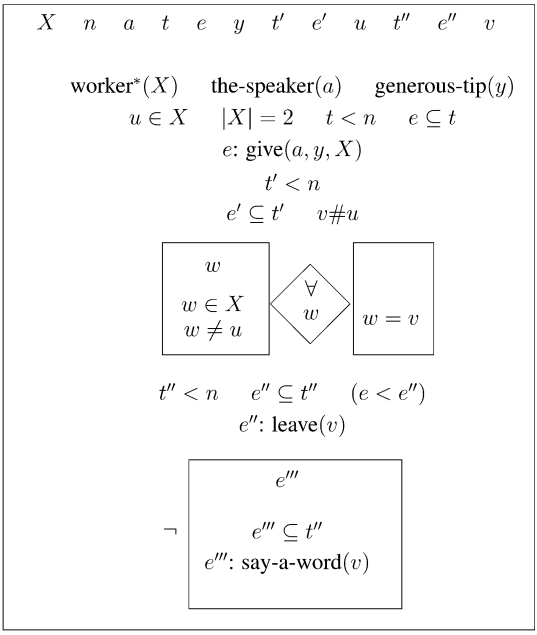
It is not difficult to see that given these identifications (14) satisfies all the presuppositional conditions of (19) except for the atomicity condition "atom(η)". (In particular, identifying η with $X \setminus \{u\}$ has the effect of satisfying both the disjointness condition "$\eta\#\zeta$" and the maximality conditions of the presupposition associated with *the*.) In order that the remaining

[14] There is the further question how this condition arises. In the present case it represents the *without*-phrase *without saying a word*. The representation rules for this and other non-finite constructions constitute a problem which needs careful attention. I have nothing to say about this problem here.

condition be satisfied, we have to assume that $|X \setminus \{u\}| = 1$, i.e. that $|X| = 2$. Thus, insofar as (20) constitutes the only nearly complete solution to the justification problem that (19) poses, accommodation of this last condition seems the best the interpreter can do to construct a coherent interpretation of the sentence. Hence the conclusion that the number of workers was equal to two.

To conclude this section, (21) gives the representation of (1) which results from resolving the presuppositions of (19) as indicated in (20), accommodating the condition $|X| = 2$ and then merging the non-presuppositional part of (19) with (14).

(21)[15]



6. Reflection: General Principles of Presupposition Justification

I am fairly convinced that this reconstruction of the interpretation of (1) is basically correct. Further evidence for this is provided by the way in

[15] The condition "$e < e''$" is like "$e < e'$" the result of processing the rhetorical structure of (1), an aspect of discourse interpetation not considered here.

which the conclusion licensed by (1) changes as a function of small textual changes. In (22) I have listed a few of these, together with conclusions about the set X which apparently follow from them. It should not be difficult, especially after the analysis of the preceding sections, to see how these different inferences are the results of changes in presuppositions that come with the modifications that turn (1) into these different variants.

(22) (i) I gave the workers a generous tip. One thanked me. Another one left without saying a word.
Conclusion: $|X| > 2$

(ii) I gave the workers a generous tip. Two thanked me. The other one left without saying a word.
Conclusion: $|X| = 3$

(iii) I gave the workers a generous tip. One thanked me. The other two left without saying a word.
Conclusion: $|X| = 3$

(iv) I gave the workers a generous tip. The first one thanked me. The second one left without saying a word.
Conclusion: $|X| \geq 2$

(v) I gave the workers a generous tip. One thanked me. Two others left without saying a word.
Conclusion: $|X| > 3$

(vi) I gave the workers a generous tip. One thanked me. The others left without saying a word.
Conclusion: $|X| \geq 3$

(Exercise: Adapt the analysis of (1) presented in the preceding sections to each of the examples (22.i)–(22.vi).)

In all examples of (22) the conclusion is perceived to be entailed by the text—in this respect the examples are all like (1). That presupposition justification can produce such apparently firm entailments is something of which Antje Roßdeutscher became aware in the early nineties, albeit in connection with presuppositions of a different type (triggered by the word *again*). As we have seen for example (1), these entailments rest in part on a kind of "economy" principle, according to which presupposition justification must make use of as much contextual information as it can get hold of, thereby minimising what remains to be supplied by accommodation. In the absence of a definition of what counts as more and less, however, terms such as "economy" and "minimise" are little more than idle metaphor. In Kamp & Roßdeutscher (1994), in which we drew attention to the existence of presupposition-based

entailments, Roßdeutscher and I noted that in presupposition justification there is a special premium on the accommodation of new discourse referents; such accommodation should be avoided if at all possible. But we pretty much left things there.

Here is a proposal for an explicit characterisation of when one accommodation is more economical than another. Suppose that the presupposition structure K_P must be justified in the context DRS K_C. Note that since accommodation involves adding either new discourse referents or new DRS conditions or both, the material that gets added to K_C can itself be regarded as forming a DRS. (In general this DRS will be an improper DRS, as the discourse referents occurring in its conditions need not occur in its universe.) Now let K_1 and K_2 be two possible accommodation DRSs which each verify K_P when added to K_C. Let, for $i \in \{1, 2\}$, $K_i = \langle U_i, Con_i \rangle$, and $U_i = X_i \cup Y_i$, where $X_i \subseteq U_{K_C}$ and $Y_i \cap U_{K_C} = \emptyset$. Then we say that *there is a reason for preferring* K_1 *over* K_2 iff there is a partial function f from Y_2 into $X_1 \cup Y_1$ such that (i) $Y_1 \subseteq f(Y_2) \cup Y_2'$ (where $Y_2' = Y_2 \setminus Dom(f)$) and (ii) $f(K_2) \models K_1$.[c] (Intuitively this means that K_2 involves at least as many new discourse referents as K_1—where f serves to identify certain discourse referents in K_2 with discourse referents in K_1 that can be seen as playing the same argument roles—and that, given this identification, the conditions of K_2 entail those of K_1.) And we say that K_1 *is preferred over* K_2 iff there is a reason for preferring K_1 over K_2 but not for preferring K_2 over K_1.

This definition seems to make the right predictions for the case of (1) and those listed in (22) and also for the cases considered in Kamp & Roßdeutscher (1994). But how well it will stand up when applied to a wider range of examples I do not dare to say at present.

---

[c] Emar Maier, who was asked to read the version for the present edition, notes that the definition of *there is a reason for preferring* K_1 *to* K_2 is problematic: The paper provides no definition of the term '$f(K_2)$', where K_2 is a DRS and f is the function spoken of in the definition—a partial function from the Universe of K_2 into the Universe of the DRS K_1. Maier observes that the definition of '$f(K_2)$' that will occur to anybody reasonably conversant with DRT is the following:

Given that $K_2 = \langle U_2, Con_2 \rangle$, $f(K_2) = \langle f(U_2), f(Con_2) \rangle$, where $f(Con_2) = \{f(\gamma) : \gamma \in Con_2\}$ and for each condition $\gamma \in Con_2$, $f(\gamma)$ is the result of replacing each discourse referent u occurring in γ by $f(u)$. Maier goes on to observe that this only works when f is defined for all discourse referents that can occur in conditions $\gamma \in Con_2$ and that in general the functions f spoken of in the definition of *there is a reason for preferring* K_1 *to* K_2 do not satisfy this requirement. One way to fix this is to say that $f(\gamma)$ is the result of replacing each discourse referent u occurring in γ by $f(u)$ *provided* u *is in the Domain of* f.

The question of accommodating discourse referents is important also in a somewhat different connection, where the issue is not that of comparing possible justification solutions in the sense of economy, but rather what forms of justification are permissible at all. There is compelling evidence that presuppositions vary with regard to the moves which are permitted to bring their justification about. This is particularly plain on a liberal view of what counts as presupposition (such as I have taken in this paper), which assumes that presupposition covers, among other things, the various forms of nominal anaphora, and treats all definite noun phrases as presupposition triggers, including pronouns and definite descriptions. For it is part of the traditional wisdom about these last two NP types that the justification rules for pronouns are quite different from those governing the presuppositions of definite descriptions, and also that this difference is first and foremost a matter of when accommodation of new discourse referents is allowed. The familiar theories of pronouns assert that interpreting a pronoun requires finding in the context either an entity which can be taken as the pronoun's referent (or a representation of such an entity). In particular, it is assumed that the interpretation of a pronoun in a text requires finding an entity representation in the discourse context which can serve as the pronoun's anaphoric antecedent.

In contrast, the classical presupposition-based account of definite descriptions has it that a description imposes upon the context in which it is used a propositional constraint to the effect that there is one and only one thing satisfying its descriptive content. If the context entails this proposition, then the description's presupposition is justified and the description can be interpreted as referring to what the context determines as the unique satisfier. In DRT terms this means that descriptions are capable of introducing new discourse referents to represent their denotations, whereas pronouns must rely on old discourse referents. Of course, it has been known for many years that descriptions usually do not conform to this account in a literal sense and that their behaviour often resembles that of pronouns quite closely. But it remains true nonetheless that at there are least some occurrences of definite descriptions that do fit the classical account, and at this point that is all we need. For the mere fact that such cases exist implies that there are two different types of presuppositions—the "entity-oriented" presuppositions of anaphoric pronouns and the purely propositional presuppositions triggered by the "classical" cases of descriptions—which involve clearly distinct justification rules.

In the representations we have used in the present paper the distinction between these two types has been indicated by underlining those discourse

referents in the universes of presupposition representations for which the context must supply an actual discourse referent as antecedent.[16] This simple distinction between underlined and non-underlined discourse referents suggests a corresponding binary distinction at the level of justification. By and large this is what I have been assuming in the present paper: Presuppositions without underlined discourse referents are treated as "purely propositional", i.e. as representations of propositions that the context must entail. In contrast, presuppositions with underlined discourse referents count as verified only if these underlined discourse referents are identified with discourse referents that already belong to the given context and the context entails the propositional representation into which this identification turns the original presupposition representation.

What if no antecedents for the underlined discourse referents of such a presupposition can be found? One of the conceivable possibilities would be that accommodation of antecedents for underlined discourse referents is simply prohibited in such cases. This would mean that the interpretation aborts and the discourse is judged as ill-formed. We have already seen, however, that interpretation isn't always as inflexible as that. Presuppositions generated by discourse-initial sentences tend to get accommodated as a matter of course, and as example (1) showed, this includes presuppositions with underlined discourse referents. There are also other facts indicating that accommodation of antecedents for underlined discourse referents is sometimes allowed, which have also been known for a considerable amount of time. Let me mention just two. First, as discussed at length in (Kamp & Reyle 1993: Ch. 4), there are intriguing differences between the presuppositional requirements connected with singular pronouns and those connected with plural pronouns. If an antecedent for a plural pronoun is missing from the context, it can sometimes be "accommodated", but only in the sense of being constructed from material that the context DRS already contains by applying to it a quite limited set of logical operations. A second type of example is given by "bridging descriptions"—definite descriptions

[16] The first two examples of this in the paper are the underlined discourse referents $\underline{X}$ and $\underline{C}$ in the presupposition DRS of (4). The reason for underlining X is that the NP *the workers* is a definite description that cannot be interpreted in a purely descriptive manner and thus must find an antecedent in the context. The decision to underline C is similarly motivated: The context must contain a salient predicate with which the predicate variable C can be identified. However, the details of this second case are a little different as they have to do with the treatment of discourse referents for entities of higher type (i.e. other than atomic and non-atomic entities of the type of individuals); and this is a matter that it would take us too far to go into here.

that are interpreted as referring to an entity that stands in some functional or quasi-functional relation to some other entity, which is part of the context as given. The principles which govern bridging (i.e. those principles which permit the introduction of an antecedent for the description) are also subject to special restrictions, though these are, if I am not mistaken, not as well understood as those which govern the interpretation of plural pronouns.

All these cases—and I suspect that by looking closely we will find others—show that the simple dichotomy between presuppositions which permit the accommodation of discourse referents and those which do not is a serious oversimplification. What we need is a much finer classification of presupposition types, together with, for each of them, an exact specification of the justification strategies it allows. At the present moment I have only a rough idea of what such a refined theory of presupposition justification might be like. But I am convinced that such a theory is badly needed, and also that it constitutes one of the major remaining challenges for a comprehensive account of presupposition.

7. Remaining Loose Ends Concerning The Subject *One* of The Second Sentence

In this section we return to a couple of loose ends which I left dangling earlier on, with the promise that they would be tied up eventually. Both have to do with the interpretation of the word *one* in the second sentence of (1). We noted in Section 4 that this one-word NP has, in addition to the interpretation which we pursued there, two further interpretations, which we provisionally set aside. The first of these is the 'impersonal pronoun' interpretation, according to which *one* stands for something like 'the relevant individuals'. According to the second interpretation *one* is construed as a dummy common noun and the NP's referent may be any individual that falls under some common noun phrase which the dummy is taken to represent. We'll take these options one at a time.

First the 'impersonal pronoun' interpretation. It is intuitively clear why the third sentence of (1) eliminates this option. For what could 'the relevant individuals' mean in the context in which *one* appears here? Relevant individuals, it would seem, should be individuals who had something for which to thank the speaker. The context offers some such individuals, viz. the workers mentioned in the first sentence, and indeed these are the only available candidates. Since all of the workers mentioned in the first sentence are equally relevant, the referent of *one* we get is the set consisting of all

these workers. If this is what the subject NP of the second sentence refers to, however, then there is no good way of resolving the presupposition package of the third sentence. In particular, the presupposition generated by *other* will now be incapable of justification. For that, we have seen, requires from the context a pair $\langle X, \nu \rangle$, consisting of (i) a set X of two or more elements and (ii) an element or proper subset ν of this set. If the subject of the second sentence is interpreted as referring to the set X itself, then the context in which the presuppositions of the third sentence are to be justified will contain no such pair.

Giving a precise formal account of this interpretation option is not easy, since it would require giving a operational meaning to the vague notion of 'relevant individuals'. I have no clear idea how this might be done.[17]

The second alternative we mentioned in connection with the subject NP of the second sentence would identify its referent as some individual satisfying the common noun phrase which is chosen as antecedent for the anaphoric common noun. It is clear that the context provided by the first sentence of (1) offers no other antecedents of the type 'Common Noun Phrase' than the noun *worker*. So the present option boils down to interpreting the subject of the second sentence as referring to some worker u or other, who might but need not be a member of the set X. It is not entirely clear what should be said about this option in relation to the requirements imposed by the presuppositions of the third sentence. One conceivable possibility is that in this case the difference set $X \setminus \{\nu\}$ cannot be formed, since it is not known whether $\nu \preceq X$. If this were so, then justification of the *other*-presupposition would fail, and with it the interpretation of the discourse as a whole. An alternative possibility might be that forming $X \setminus \{\nu\}$ is permissible, but only after the missing condition "$\nu \preceq X$" has been accommodated. The end effect of this would be the same overall interpretation of (1) as we obtained in (21). At this point I see no way of deciding between these last two possibilities, though I am inclined to think that the accommodation option is available. If this is so, then (1) allows for two converging interpretations, based on different interpretations for the second sentence.

[17] For the purpose of the example at hand it would suffice if our formalisation of the concept of 'relevant individuals' would be such as to entail that in cases where a set of relevant individuals are provided by the context, but there are no salient conditions which allow for a partition of this set into two or more smaller subsets, then only the set as a whole is available as a possible referent for *one*. Some definition of 'relevant' that carries this entailment could no doubt be concocted without too much trouble. But such a solution would be ad hoc and likely to run into trouble when we turn to other examples.

8. Concluding Remarks

This paper has been concerned with the analysis of one single example. I hope that the story which has been told about this example—how the meaning of the mini-discourse (1) is computed and in what way presupposition justification contributes to that—is reasonably convincing. However, the paper's real aims, which prompted the particular case study I have presented, are more general and more ambitious. The case study was not just meant as a possibly outlandish example of one of the various possible ways in which presupposition justification can affect discourse interpretation. It was chosen in the belief of its being truly representative—that the mechanisms by which presupposition justification increments discourse meaning in this example are often found in discourse, and that the analysis which I have proposed will carry over to a substantial number of other cases, including 'real life' ones, which are encountered in actual NLP applications. To what extent this belief is justified may not be decidable at present, but only as our understanding of the principles that govern expands and deepens.[18]

Another question that the paper leaves unanswered concerns the possibility of turning the semi-formal analysis I have given into one that is fully explicit or even into a working implementation. In reflecting on this question it is necessary to distinguish between formalisations or implementa-

[18] Here is one example from "real life". In Samuel Butler's "Erewhon", Ch. 9 (p. 66, 2nd paragraph, of the Wordsworth Classics Paperback Edition) we find the following: "Mr Nosbibor took me through several spacious rooms till we reached a boudoir where there were his wife and daughters, of whom I had heard from an interpreter. Mrs Nosbibor was about forty years old, and still handsome, but she had grown very stout; her daughters were in the prime of youth and exquisitely beautiful. I gave the preference almost immediately to the younger, whose name was Arowhena; …". In English the combination of the comparative adjective form with the definite article presupposes that the referent is to be understood as a member of a set of two elements. In the given passage (in which the wife and daughters of Mr Nosbibor are mentioned for the very first time) this rule applies to the NP *the younger*. The missing noun from this definite description carries another, "anaphoric", presupposition, to the effect that a common noun phrase must be found in the given context (just as we saw this to be the case for the word *one* in the subject NP *the other one* of our example). If we resolve this anaphoric constraint by choosing as CNP (*Mr Nosbibor's*) *daughter*, then the first presupposition leads us to conclude that the set of Mr Nosbibor's daughters consists of two elements. This conclusion gets confirmed almost immediately when in the next sentence of the text we find a reference to *the elder sister*. Evidently such an example does not show that the phenomenon that has been discussed in this paper is one we are likely to run into incessantly when dealing with actual texts. But at the very least it shows that the example is not wholly artificial.

tions which stick with the given example without making a serious attempt to generalise beyond it and those which deal with an open class of cases in which (1) is only one among a significant variety of others. Of course, any formalisation or implementation worth its salt must aim at some form of generalisation, however modest; and in relation to the example discussed in this paper covering the small family consisting of minor variations of (1) which we get when we replace the subject NP *the other one* of the third sentence by NPs like *another one*, *others*, *the others*, *two others*, etc. (as illustrated in Section 6) should be considered a minimal requirement for any credible formalisation or implementation of the analysis this paper has outlined. But if dealing even with this very restricted class will prove non-trivial, generalising to larger classes can be expected to be very much harder, if only because it will require additional lexical entries for presupposition-carrying morphemes or words. The development of a semantic lexicon that can be used effectively by formal or automated interpretation systems is a daunting task that formal and computational linguistics are facing generally. It is a daunting task not just because of the sheer size of the vocabularies of most natural languages, but also because there is so much that is connected with the meanings of particular lexical items and their representation which is poorly understood and for which a proper theoretical foundation is largely missing. The representation of the presuppositional aspects of individual words and morphemes is among those ill-understood aspects of the lexicon. A formalisation or implementation which is to deal with a larger class of presuppositions than has been considered here, will have to have access to entries for the (lexical) triggers of those presuppositions which contain the necessary instructions for how those presuppositions are computed and how they are to be justified. (I am assuming that, with few exceptions, presuppositions are triggered lexically.) This alone will add substantially to the overall effort that such a broader formalisation or implementation would involve.

In connection with the question of implementation there is also another distinction that is important. Formalising or implementing the analysis presented in Sections 2–6, which relied on a kind of oracle to select the intended interpretation of the subject *one* of the second sentence may be a non-trivial task as it is. But doing without such an oracle, while building into the formalisation/implementation the considerations of Section 7 instead, is much harder. The extra difficulties that have to be overcome when the implementation is to compute an actual resolution of this ambiguity rather than adopting the resolution by fiat deserve special emphasis. As we saw in Section 7, the ambiguity of *one* in the second sentence cannot be resolved

until we get to the third sentence. Thus the context which the first two sentences of (1) offer towards the interpretation of the third sentence is itself three-ways ambiguous (corresponding to the three possible interpretations of *one*). In Section 7 it was implicitly assumed that each of these three possible contexts is used as starting point for an interpretation of the third sentence, that the results of these three interpretation attempts are then compared and that on the basis of this comparison an interpretation is selected which gets away with the most economical justification of the third sentence's presuppositions.

An implementation which follows this recipe to the letter is no doubt possible in principle. But experience has taught that where ambiguity is an issue, operating on fully expanded disjunctions (such as the disjunction of the three representations of the context established by the first two sentences of (1) which result from the three different possible interpretations of *one* in the second sentence) tends to be inefficient, and the inefficiency quickly becomes prohibitive for larger discourses where potential ambiguities multiply. Computational tractability requires that ambiguities be kept as local as possible, for instance along the lines of U(nderspecified) DRT (Reyle 1993). To get a better sense of the computational complexities that are involved when ambiguity interacts with presupposition-driven interpretation in the way illustrated in (1), a UDRT-based treatment of this one example would seem a natural first step. But even this is something that has yet to be tried, and until it has been, there is no good way of estimating how much more complex the problems of this paper become when ambiguity is paid its proper due.

References

Kamp, H., 1981. A theory of truth and semantic representation. In: J. Groenendijk, T.M.V. Janssen, M. Stokhof (eds.), *Formal Methods in the Study of Language, Proceedings of the Third Amsterdam Colloquium*. Mathematical Center, Amsterdam, 277–322. Reprinted in: J. Groenendijk, T.M.V. Janssen, M. Stokhof (eds.), *Truth, Interpretation and Information, Selected Papers from the Third Amsterdam Colloquium*. Foris, Dordrecht, 1984, 1–41, and reprinted in: K. von Heusinger, A. ter Meulen (eds.), *Meaning and the Dynamics of Interpretation. Selected Papers of Hans Kamp*, Brill, Leiden, 2013, 329–369.

Kamp, H., 2001. The importance of presupposition. In: C. Rohrer, A. Roßdeutscher, H. Kamp (eds.), *Linguistic Form and its Computation*. CSLI Publications, Stanford, CA, 207–254. Reprinted in: K. von Heusinger, A. ter Meulen (eds.), *Meaning and the Dynamics of Interpretation. Selected Papers of Hans Kamp*. Brill. Leiden, 2013, 455–501.

Kamp, H., Reyle, U., 1993. *From Discourse to Logic. Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer, Dordrecht.

Kamp, H., Roßdeutscher, A., 1994. DRS-construction and lexically driven inference. *Theoretical Linguistics* 20 (2/3), 165–235.

Lascarides, A., Asher, N., 1993. Temporal interpretations, discourse relations and commonsense entailment. *Linguistics and Philosophy* 16, 437–493.

Link, G., 1983. The logical analysis of plurals and mass terms: A lattice-theoretic approach. In: R. Bäuerle, C. Schwarze, A. von Stechow (eds.), *Meaning, Use and Interpretation of Language*. de Gruyter, Berlin, 303–323.

Reyle, U., 1993. Dealing with ambiguities by underspecification: Construction, representation and deduction. *Journal of Semantics* 10, 123–173.

van der Sandt, R., 1992. Presupposition projection as anaphora resolution. *Journal of Semantics* 9, 333–377.

THE IMPORTANCE OF PRESUPPOSITION[a]

Hans Kamp

1. Generalities

1.1. *Introduction*

The title of this paper may seem presumptuous. But I have chosen it anyway, as a signal of what I see as the central insight that recent studies of presupposition in natural language have given us. Presuppositions are not what early discussions suggested and later ones have, I feel, still done too little to explicitly refute: obstacles to logical transparency, logical simplicity or well-definedness of meaning, and thus a thing that languages would have been better off without. On the contrary, presuppositions are valuable tools which language uses to guide the recipient of an utterance, discourse or text towards its intended meaning. In fact, given what we are beginning to understand about the ways in which presuppositions arise and the purposes they serve, it is hard to imagine how a natural language could exist without them.

This paper is part of a larger enterprise and the first of a series. So it would seem to be the right place for a general overview of the theory that has been the target of this enterprise and to say enough about it to give a clear idea of how it works, and how it differs from other presupposition theories. Unfortunately, this would have required a paper well in excess of what is acceptable in a collection like the present one. So I have had to temper my ambitions and to select some parts of the over-all account for special focus, while leaving out a good many others.

---

[a] Editors' note: This article was originally published in C. Rohrer & A. Roßdeutscher (eds.), *Linguistic Form and its Computation. Selected Papers from the SFB 340*, CSLI Publications, Stanford, CA, 2001, 207–254. We would like to thank the Center for the Study of Language and Information (CSLI) for its kind permission to reprint the article. We carefully edited the article and corrected some typos. We would like to thank Atle Grønn for careful proofreading and constructive comments.

© hans kamp, 2013 | doi:10.1163/9789004252882_019
This is an open access chapter distributed under the terms of the cc by-nc-nd 4.0 license.

After some hesitation what to retain of earlier, much bulkier drafts I decided to focus on two aspects of presupposition theory which seem to me to have received comparatively little attention in the literature so far. The first is one that one might describe as 'presuppositional multiplicity': Natural language sentences often give rise not just to one presupposition but to several presuppositions at once. It turns out that combinations of presuppositions triggered within one and the same sentence pose problems that go beyond what we see when focusing on individual presuppositions, as much of the existing presupposition literature has done. In fact, presuppositional multiplicity leads to complications both at the level of 'presupposition justification'—i.e. at the level where presuppositions are checked against the context in which the sentence generating them is being used—and at what I propose to call the level of 'presupposition computation'—the level at which the presuppositions due to the various triggers in a sentence are computed in a form which allows them to be checked.

The distinction between these two processing levels, that of presupposition computation and that of presupposition justification, may not be self-evident. But it is hard to see how a precise and computationally viable theory of presupposition could do without it. And as far as I can see, many current presupposition theories imply the distinction in some form. However, the level of presupposition computation is seldom discussed in detail, as if that part of the theory was too trivial to merit explicit articulation (or, alternatively, because articulating it is thought to be the business of someone else). As it turns out, even the computation of individual presuppositions poses problems that are far from easy.[1] But the true complexities of presupposition computation become visible only where different presuppositions interact.

A presupposition theory consisting of the two connected components of presupposition computation and presupposition justification requires a general framework in which language interpretation takes the form of establishing first, on the basis of syntactic (and prosodic) information, 'logically transparent' representations of individual sentences, and then subjecting these representations to further semantic and/or pragmatic operations, which relate them to the context in which the represented sentences are being used. In those cases where the sentence is part of a larger discourse,

[1] Cf. the growing literature on the presuppositions generated by words like *again, also, even, still* and many other 'discourse particles'. A good example of the problems that arise in connection with computing, according to linguistically 'clean' principles, the presuppositions generated by different occurrences of *again* is von Stechow (1996).

these operations should integrate the sentence representation with the discourse interpretation obtained thus far.

Given such a general architecture, it is possible to make sure that the representation of each individual sentence contains explicit representations of all presuppositions that the sentence generates. The process of integrating a sentence representation into the interpretation of the preceding discourse may then be made dependent on justification of all the represented presuppositions in their context, so that integration proceeds only when justification has been successful.

The term 'presupposition justification' is not a standard term of the presupposition literature. But it is one I use advisedly. Antje Roßdeutscher and I introduced the term in the early nineties to reflect our conviction that what we took to be the standard picture of presupposition checking is an oversimplification of what often actually happens.[2] There is general agreement that presuppositions should be satisfied in their contexts, but that satisfaction failure does not lead to breakdown of the interpretation process automatically. For there often remains the possibility of 'accommodating' the context so that it satisfies a failing presupposition after all. However—and this is where we saw our views as diverging from the standard view—the relation between satisfaction and accommodation is in general not a matter of 'either or': either a presupposition is satisfied, or it isn't, and then it has to be accommodated whole-sale, much as when a new proposition is added to a given theory. In general, presupposition satisfaction and presupposition accommodation do not behave as mutually exclusive alternatives. The process of making sure that the context satisfies a given presupposition often involves a mixture of some satisfaction and some accommodation—of finding contextual information that goes some way towards satisfaction of the presupposition and then accommodating what is missing to make satisfaction complete. It is this combination of 'partial satisfaction' and (partial) accommodation for which we coined the term 'justification'.

Presupposition justification too is a process that proves to be much more intricate when it involves clusters of presuppositions than when we confine our attention to individual presuppositions. A story about presupposition in which presuppositional multiplicity holds a central place is thus seriously incomplete if it doesn't have anything to say about its effects at this level. For this reason an earlier version contained an abridged version of Kamp (2001), which is devoted to an example of multiple presupposition

[2] See Kamp & Roßdeutscher (1992), Kamp & Roßdeutscher (1994).

justification. Unfortunately this abridged version had to go, and all that has remained is a mention, at the end of this introductory section, of the example itself. For anything more than that the reader has to consult the cited paper.

In its present version, then, the paper is devoted almost entirely to the computation of presuppositions.

Although I am inclined to see this as the less interesting half of an over-all account of presupposition, it is, as I have noted above, the one which is most in need of attention because it has thus far been so badly neglected. Moreover, without a proper account of presupposition computation, formally precise accounts of justification aren't really possible. For these must take their departure from the explicit articulation of presuppositions for which the theory of presupposition computation is responsible. Methodologically, it is presupposition computation which comes first and justification which comes second (and for all it is possible to tell, this is also the order in which things are done in the language-processing mind). So, if more than one paper is needed to tell the full story, computation would seem the proper end at which to begin it.

1.2. *A Brief Look Back*

The theory of presupposition has had a checkered history. This is not the occasion to retrace the often surprising gyrations of its winding path. But two of the milestones which mark its progression should be mentioned. Both of them signaled changes in focus and perspective which occurred about thirty years ago, when presupposition became the subject of active study within linguistics. First, linguists became conscious at that time of the wide variety of presupposition 'triggers' that are to be found in natural languages—that is, of the multitude of lexical items and grammatical constructions which give rise to presuppositions. As a consequence, the presuppositions generated by definite descriptions, which had been the exclusive concern of virtually the entire presupposition literature up to that point, now came to be seen as only one type of presupposition within a much larger class. Second, the famous paper of Langendoen & Savin (1971) drew attention to the fact that sometimes a presupposition that is generated within a constituent of a logically complex sentence gets 'filtered' out, in the sense that the sentence as a whole behaves as if the presupposition wasn't there. Thus the *Projection Problem* was born, the problem which of the presuppositions that are generated within complex sentences get filtered and which 'survive'. Speakers' intuitions about presupposition filtering (which presup-

positions get filtered and which do not) are on the whole quite sharp and consistent. These intuitions provide the linguist with a profusion of intricate linguistic data, many of which have proved remarkably robust. To the theoretically minded they present a natural challenge. It is not surprising, therefore, that so much of the literature on presupposition has been concerned with aspects of the Projection Problem ever since the time when the problem was discovered.

1.3. *Presuppositions in Dynamic Semantics*

A major advance in dealing with projection questions came with the development of Dynamic Semantics, initially in the two forms of File Change Semantics (Heim 1982) and Discourse Representation Theory (Kamp 1981). In her seminal paper Heim (1983), the author showed how the framework of Dynamic Semantics makes it possible to account for many of the most uncontroversial projection data (pertaining to the behaviour of presuppositions in conjunctions, conditionals and under quantification) virtually without need of any assumptions in addition to those which had already been made for independent reasons. In retrospect this is even less surprising than it was at the time. Dynamic Semantics is about interpretation in context. The projection data fall into place if for many presuppositions generated within parts of complex sentences it is assumed that the relevant context is provided by another part of the sentence itself. On that assumption the presupposition will get filtered just in case it is verified by its local, sentence-internal context alone. For in that case the presupposition will not interfere with the interpretation of the sentence as a whole irrespective of what the 'global' context, in which the sentence is being used, may be like.

Heim's paper paved the way for further studies which tackled some of the remaining, more recalcitrant, projection issues within a dynamic setting. Possibly the most diffcult among those problems are questions which have to do with presupposition accommodation.

1.4. *Presupposition within DRT*

Of particular importance for the theory outlined in the present paper is the influential van der Sandt (1992). This paper too is centrally concerned with aspects of the Projection Problem (and specifically with the related issues of accommodation). But for us its central importance lies elsewhere. Van der Sandt makes use of DRT. But he adopted a form of it which differs significantly from early versions of the theory, like the ones of Kamp (1981) or Kamp & Reyle (1993). Both in these earlier versions and in that of van

der Sandt the basic interpretation strategy is cumulative: To interpret a discourse $\langle s_1, \ldots, s_n \rangle$, begin by building a representation K_1 of s_1, then use this representation as context for the interpretation of s_2, which, when merged with K_1 yields a representation K_2 for $\langle s_1, s_2 \rangle$; and so on. But below this most general level there is a difference. In van der Sandt's version of DRT processing of the next sentence s_{i+1} does not take the form (as it does e.g. in Kamp & Reyle 1993) of direct integration into the context representation K_i, but proceeds in two stages. First, one constructs a 'preliminary' representation K'_{i+1} for s_{i+1}, in which the presuppositions generated by s_{i+1} are explicitly represented, and then, in a second step, this preliminary representation K'_{i+1} is confronted with the context representation K_i. This confrontation involves justification of all the presuppositions represented in K'_{i+1}. If this succeeds, then the non-presuppositional part of K'_{i+1} is merged with K_i, yielding the new context representation K_{i+1}.

This two-step DRT-architecture has proved desirable for a number of reasons, and I have been for several years among those who are making use of it as a matter of course. In particular, I see it as the only viable option for a DRT-based theory of presupposition.

Within such a framework the division of the theory of presupposition into a component dealing with presupposition computation and another one dealing with justification arises quasi inevitably. Computation of presuppositions is part of the process of computing sentence representations; presupposition justification is an integral part of linking the sentence representation with the context.[3]

One of the central motivations behind van der Sandt[4] was the conviction that the phenomena traditionally described as 'presupposition' and 'anaphora', respectively, are but different aspects of what should be seen as a single type of process, and one of its aims was to present a unified account of presupposition and anaphora which makes clear in what sense this is so. This emphasis on a unifying perspective has been illuminating and fruitful,

[3] The first of the two stages of the interpretation process just described might be said to embody what is usually understood by the *syntax-semantics interface*, whereas the second stage constitutes an interface between semantics and pragmatics. Thus presupposition computation is part of the syntax-semantics interface, whereas problems of presupposition justification are located at the semantics-pragmatics interface. Here semantic representations provide the input to mechanisms which traditionally have been counted as lying on the pragmatic side of the semantics-pragmatics divide.

[4] And similarly of Zeevat (1992), a proposal which has close parallels with van der Sandt and was developed roughly at the same time.

and it has been adopted also in the account presented here, which follows van der Sandt in many of its innovating features.

Of course, a unified framework for presupposition and anaphora should not make us lose sight of the real differences which nevertheless exist between the principles for interpreting anaphoric expressions and those for dealing with presuppositions. But in fact, the problem of variation is a deeper one than that. There are important differences not only between the categories of 'presupposition' and 'anaphora' as they have been traditionally conceived, but also within the category of presuppositions itself. In particular, presuppositions differ in the constraints which govern the possible ways of their accommodation. At the present time our understanding of these constraints is fragmentary at best. This is not only a matter about nothing will be said in this paper, but which there is, as things stand, little I have to say about.

1.5. *Preview*

The paper is structured as follows.

Section 2 makes a start with the problem how presuppositions are represented. It begins by outlining, in very general terms, the general architecture of the theory, going over some of the things that have already been touched upon in the introduction. The section goes on to look at the form of sentence representations. Here, as throughout the rest of the paper, I limit myself to the analysis of particular examples.

Section 3 looks at the construction of preliminary sentence representations for the examples of Section 2. To set the scene for this, the section starts by outlining a general bottom-up construction algorithm for DRSs (still without explicit representation of presuppositions) in which lexical heads introduce variables (or 'discourse referents') representing the 'referents' of those heads, whereas the corresponding functional heads are responsible for the binding of those variables. (Since binding often occurs much higher up in the syntactic tree than where the variable that is to be bound gets first introduced, a storage mechanism is needed, were the variable is to be kept until the point has been reached where its binding can be made explicit.) This algorithm is then modified to yield *preliminary representations* in which the presuppositions generated by the sentence are explicitly represented. (In this section, where the only presupposition triggers considered are definite Noun Phrases, these rules convert store elements introduced by such NPs into representations of the presuppositional constraints that these NPs impose.)

In Section 4 we extend the construction algorithm of Section 3 by first looking at the problem of computing presuppositions that are generated by *again* and then considering the problems that arise when *again*-presuppositions interact with presuppositions stemming from definite Noun Phrases.

Although the discussion takes the form of looking in detail at particular examples and falls well short of the ideal of a fully general formulation of representations and rules for a significant fragment, I hope that they exemplify enough of the problems that such an algorithm will have to solve to show the way towards such a formulation. Section 5 concludes with a somewhat depressing list of the many things that remain to be done.

Here, to conclude this introductory section, is the promised example of simultaneous satisfaction of interacting presuppositions. Consider the following three sentence discourse (1):

(1) *I gave the workmen a generous tip. One went out of his way to thank me. The other one left without saying a word.*

An obvious conclusion from this discourse is that the set of workmen to whom the speaker gave a generous tip had two members. This conclusion, however, can be drawn only after getting both the second and the third sentence, and it is evident that all three sentences are involved in it. In fact, the inference rests on the justification of a number of different presuppositions and the presuppositions generated by the subject NP of the last sentence play a particularly prominent part. (The different elements of this NP which impose presuppositional constraints on the context that all contribute to the inference are (i) the singular definite article *the* of the description *the other one*; (ii) the word *other* occurring in that description and (iii) the word *one* occurring in it. The presupposition triggered by the occurrence of *one* in the second sentence also plays a significant role.) In a nutshell the story is this. Justification of the presuppositions associated with *the other one* requires from the context (a) a certain set X of at least two elements and (b) a (possibly singleton) subset Y of X. The referent of *the other one* is then the difference $X \setminus Y$. Since the singularity of this NP (the fact that it is in the singular rather than in the plural) requires that its referent be a single individual rather than a set of two or more elements, $X \setminus Y$ must be a singleton. Since the only viable 'solution' to these context requirements that the context established by the first two sentences has to offer consists in the identification: $X =$ the referent of *the workmen* and $Y =$ the referent of *one*, and since the referent of *one* must be an individual, a coherent solution will emerge only if it is assumed that $|X| = 2$.

2. The Form of Preliminary Representations

2.1. *Over-All Architecture of the Theory*

Let me, before we start looking at any questions of detail, summarise once more the main features of the general theory. As said, in many respects the present theory follows earlier proposals within the framework of Dynamic Semantics, and there is a particularly close connection with van der Sandt (1992).

1. The analysis of linguistic presuppositions is to be divided into two stages:
   (i) a *computation stage*, during which a *preliminary* representation is constructed of the current sentence, in which all the presuppositions it generates are explicitly represented; and
   (ii) a *justification stage* at which the presuppositions of this preliminary representation are justified relative to the context representation, after which the sentence representation is merged with the context representation.
2. The representation of a presupposition generated by a given trigger is attached to that part of the preliminary representation which represents the sentence part in which the trigger occurs.[5] Thus each presupposition gets its representation in its 'local context'.[6]
3. Some treatments of presupposition invite the interpretation that presupposition justification either takes the form of the presupposition being verified in the relevant context, or else of its being accommodated. But in general justification is not simply either one or the other, but consists of a mixture of 'verification' and 'justification'.
4. Each presupposition has to be justified in its '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 (i.e. at the 'global' context as well as at contexts intermediate between the global context and the local one). Accommodation, when needed, is usually accommodation of the global context, but under special circumstances local or intermediate accommodation are possible too.

[5] E.g. a presupposition generated by (a trigger that is part of) the consequent of a conditional will be attached to the representation of the consequent rather than to the conditional as a whole, and analogously for presuppositions whose triggers occur within the scope of other scope-bearing elements.

[6] In this paper I will not give a precise definition of the notion of the local context of a given presupposition. The formal definition is closely connected with the notion of accessibility from DRT. (Cf. Kamp & Reyle 1993.)

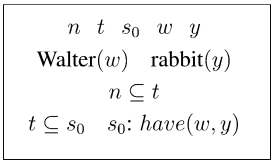
2.2. *Presupposition Representation*

We begin with a couple of examples drawn from familiar territory. (2) is verbatim from van der Sandt (1992).

(2) *Walter has a rabbit. His rabbit is white.*

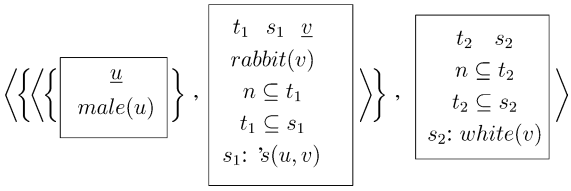
I assume that the first sentence of (2) gives rise to the DRS (3), which then functions as discourse context for the second sentence. Because of the importance which temporal information will play in later examples I have decided to make it explicit right from the start. The two conditions $n \subseteq t$ and $t \subseteq s$ entail that the state of *w* having *y* holds at the utterance time *n*.[7]

(3)



(4) gives what we will assume as the preliminary representation of the second sentence of (2).[8,9]

(4)



---

[7] The representation of temporal information used here employs the coding of temporal information introduced in (Kamp & Reyle 1993: Ch. 5).

[8] In official, set-theoretic notation the structure displayed in (4) has the following form: $\langle\{\langle\{\langle\{u\}, \{male(u)\}\rangle\}\rangle, \langle\{t_1, s_1, v\}, \{rabbit(v), n \subseteq t_1, t_1 \subseteq s_1, s_1 : 's(u, v)\}\rangle\rangle\}$, $\langle\{t_2, s_2\}, \{n \subseteq t_2, t_2 \subseteq s_2, s_2 : white(v)\}\rangle\rangle$.

[9] (4) differs from the representations given in van der Sandt in three respects. The first two of these—the explicit representation of temporal information and the underlining of the variables *u* and *v*, are discussed in the main text. Besides these there is a third difference, which, however, is more a matter of display rather than of substance: While van der Sandt puts presuppositions inside the boxes representing the sentence constituents that contain their triggers, we place them to the left of the representation of the relevant clause. The intuitive reason for this is that justification of the presupposition must logically precede assigning a definitive propositional content to the clause which generates it. In the present case, we see this twice over, as the presupposition generated by *his* is not just a presupposition of the second sentence as a whole, but in fact of the presupposition generated by *his rabbit*; in this regard too the present account follows van der Sandt.

Two comments. The first one concerns the representation of temporal information. The mode of representing temporal information in (4) is the same as in (3). But there is nevertheless one important difference, in that in (4) we find an explicit representation of the temporal dependence of a non-verbal predicate (viz of the possessive relation expressed by the pronoun *his*). In Kamp & Reyle (1993), to which we referred above when presenting (3), no provision is made for representing the temporal aspects of non-verbal predications. This obviously is a short-coming. In particular we need to be able to represent such information in order to be able to discuss many of the more interesting questions about the content and computation of presuppositions that will cross our path as we go along. The temporal aspects of non-verbal predications raise a number of separate issues, none of them altogether trivial. First, there is the matter of how the temporal aspect of non-verbal predicates (nouns, adjectives, adverbs, prepositions) is to be represented in the lexicon. This is a matter we will ignore here. Second there is the question how the temporal dimension of non-verbal predications is to be represented. For this we have adopted the same device that is also used to represent the temporal dimension of stative verbal predications, assuming that each non-verbal predication can be understood as the claim that a certain state of affairs, viz the predication in question, holds at a certain time.

Thirdly, and this is by far the most intricate issue, there is the problem how the time of a non-verbal predication is determined. This question was first investigated by Enç (1981), and then, more recently, by Musan (1995). Tonhauser (2000) studies the issue within the setting of DRT and it is on this study that the present treatment is based. Primary attention has been given so far to the non-verbal predications expressed by the lexical heads of noun phrases which act as arguments to verbs. Here the principal possibilities for determining the predication time are (i) the time of the eventuality described by the verb; and (ii) the utterance time,[10] in case this time differs from the time of the eventuality.

Of course this is only a small part of the full story, which also has to deal with the times of all other forms of non-verbal predication. The only other case that we will have to deal with in this paper is that of a predicatively used prenominal adjective phrase. (I am assuming that the possessive relation expressed by *his* also belongs to this category.) For this case the answer is, as far as I can see, quite simple: The predication time of the prenominal

[10] Or, more generally, the Temporal Perspective Time—see Kamp & Reyle (1993).

AP is always identical with that of the noun to which it is prenominal. This principle will play a role in the analyses of Section 4, even if it will be left more or less implicit in what we will have to say.[11]

In the representations that will be shown from now on I will sometimes represent non-verbal predications in the way this is done in the familiar versions of DRT—e.g. representing that y is a rabbit as $rabbit(y)$ rather than as s: $rabbit(y)$. I will do this, however, only in those cases where the time of the predication doesn't seem to matter—for instance, because the predication is 'eternal', or because it is understood as invariant over the entire period of time that is relevant to the given discourse.

A second aspect of (4) that requires explanation is the underlining we have used to mark the discourse referents u and v. This is meant as a way of indicating that the marked discourse referents have an anaphoric status, in the sense that justification of the presupposition representations which contain them as members of their DRS universe requires finding discourse referents in the context with which the underlined discourse referents can then be identified (and which thus act as their anaphoric antecedents). Note well that underlining is only a provisional device, which will have to be replaced eventually by a detailed statement of the principles of antecedent determination for the different types of expressions which give rise to 'anaphoric' presuppositions. The huge literature on the interpretation of different types of definite NPs (pronouns, definite descriptions, demonstratives, proper names) bears witness to the complexity of this task, even if our attention is limited to NP-triggered presuppositions alone.

Both presuppositions of (4) are justifiable in the context provided by (3). For the left-most presupposition justification involves identifying u with the discourse referent w which represents the bearer of the name *Walter*. This identification requires justifying that w represents a male person. We can see this either as supported by general (if defeasible) knowledge that *Walter* is normally used as a name for males, or as the result of accommodation. (The latter would have been a more likely strategy if the name had been *Hilary* or *Robin*). Given that u has already been identified with w, the justification of the second presupposition must necessarily involve identification of y

[11] The case is different for prenominal adjectives that act as operators, such as *former* or *future*. Here the predication time of the adjective is to be identified with that of the entire nominal predication, while the predication time of the noun itself is shifted by the operator. Thus x satisfies *former president* at t when x satisfies *president* at some time before t. This is an issue, however, with which we won't have to deal in the remainder of this paper.

with v.[12] To show that this identification is legitimate, we must verify that y satisfies the conditions associated with v. To do this we have to demonstrate (i) that y is a rabbit and (ii) that there exists a state s_1 surrounding the utterance time n which is to the effect that y is had by w. As it is, the context (3) contains both those conditions explicitly, so this second identification is legitimate as well.

Using these identifications we can integrate (4) into the context representation (3): We drop the presuppositions from (4), enter in their stead the 'resolution links' $u = w$ and $v = y$ and merge the resulting representation with the context DRS. Since u only figures in the presuppositional part of (4), it disappears altogether. The result is presented in (6):[13]

(6)

s	w	y	s_2	v
$Walter(w)$			$rabbit(y)$	
		$v = y$		
$n \subseteq s$			$n \subseteq s_2$	
s: $have(w, y)$			s_2: $white(y)$	

3. Presupposition Computation

In Section 2.2 we looked at the form of preliminary sentence representations. In the present section we consider how such representations are constructed from syntactic input.

Our primary interest in this connection concerns the principles for computing the representations of presuppositions. But since these computations are interwoven with the construction of the non-presuppositional

[12] There are a number of factors which militate against linking u and v to the same antecedent. Arguably the most compelling of these is that such a resolution would violate the Binding Theory, which forbids formal coreferentiality between the head of the phrase *his rabbit* and the possessive pronoun *his*.

[13] I have simplified this representation somewhat by leaving out the redundant location times t and t_2. A further simplification can be obtained by eliminating the discourse referent v, thus obtaining the representation in (5):

(5)

s	w	y	s_2
$Walter(w)$			$rabbit(y)$
$n \subseteq s$			$n \subseteq s_2$
s: $have(w, y)$			s_2: $white(y)$

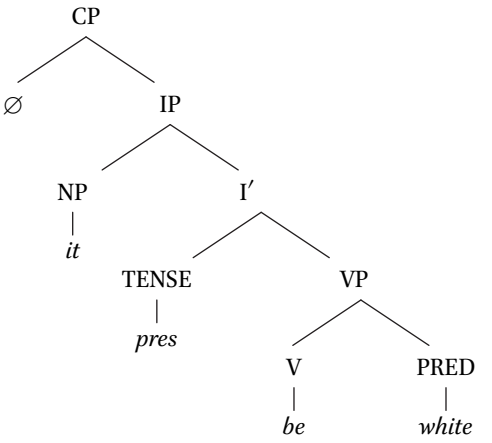
parts of preliminary representations, it will be necessary to say something first about this aspect of representation construction.

In the outline I will give of the algorithm, I will make use of syntactic structures that don't conform in every respect to the current demands of many syntacticians. I hope that none of the interface principles I will discuss, however, are seriously affected by this.

Let me show right away what syntactic trees we will be using. Before we look at the second sentence of (2) itself, we will first deal with the simpler sentence which we get when we replace the subject phrase *his rabbit* by the pronoun *it*. The simpler sentence is given in (7) and its syntactic tree in (8).

(7) *It is white.*

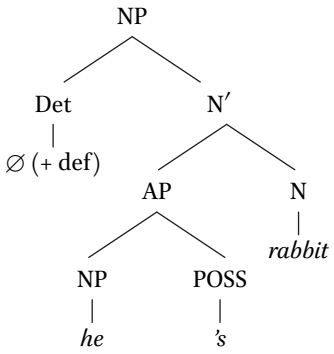
(8)



For the NP *his rabbit* we assume the structure given in (9). The syntactic tree for the second sentence of (2) results from substituting this structure for the subject NP of (7).[14]

[14] In this structure the possessive constituent is analysed as a prenominal adjective phrase consisting of the 'relational adjective' *'s* and to the left of it an NP representing the second argument of this relation. The other argument gets unified with the referential argument of the noun to which the AP is prenominal. The analysis of the possessive is thus reminiscent of those analyses of prepositional phrases according to which the preposition acts as a relation between the referential argument of the phrase to which the PP is adjoined and the referent of the PP's NP.

(9)



In the next subsection I have stated, in succinct form, the central principles for constructing DRSs from syntactic representations. These principles apply both to the construction of presupposition-free DRSs and to the preliminary representations of which we will see examples later on.

3.1. *General Principles of Bottom Up DRS Construction*

(CP1) Each argument phrase α introduces a variable x_α, the so-called 'referential argument' of α. This variable becomes part of three different components of the representation to be built:

(a) x_α gets inserted into the argument position marked by α.
(b) x_α becomes the principal argument of the condition or conditions expressing the descriptive content of α. The descriptive content is given by the 'nominal part' (or 'common noun phrase') of α. If the common noun phrase of α consists just of a noun ν, then there usually is just one descriptive condition, viz $\nu(\alpha)$; if the CNP is complex, then there will be several conditions, one corresponding to the head noun and the others to prenominal adjective phrases, relative clauses or PPs. For the most part the details of how such condition complexes are constructed need not worry us here.
(c) x_α also enters into a third component, the so-called Binding Conditions associated with α. More about this under (CP2):

(CP2) The *Binding Conditions*, as we understand this term, that are associated with a given variable, are instructions that determine the quantificational or referential role this variable is to play in the sentence representation into which it gets introduced. We distinguish three main types of Binding Conditions and make a corresponding three-fold classification of noun phrases that give rise to these Binding Condition types:

(a) Quantificational NPs;
(b) Indefinite NPs; and
(c) Definite NPs.

QNPs are those NPs where x_α gets bound by the quantificational operator that is introduced by the NP itself. Typical examples are NPs beginning with the determiners *every* and *most*.

From the perspective of the Syntax-Semantic Interface described here, QNPs are the simplest category. Their Binding Conditions are always unambiguous and they get implemented in a construction step which as a rule closely follows that in which the variable is introduced. Example: The subject NP of the sentence

(10) *Every prime number is odd.*

is assigned the representation (11)

(11)
$$\left\langle\, x\,,\ \boxed{\text{prime number}(x)}\ ,\ \text{BiCo}_{every}\ \right\rangle$$

This representation is then combined with the representation of the VP of (10), which we assume for now to have the simplified form in (12):

(12) $\langle \text{odd}(y_{subj})\rangle$

The result of this combination, in which the Binding Conditions of *every* are converted into a quantificational structure, is the complex DRS-condition (13)

(13)
$$\boxed{\begin{array}{c} x \\ \text{prime number}(x) \end{array}} \quad \forall x \quad \boxed{\text{odd}(x)}$$

The Binding Conditions of both INPs and DNPs differ from those for the QNPs in that they may have to be carried over by several construction steps before actual binding takes place. For INPs this is because their variables get bound by an operator introduced by some other part of the sentence, so the Binding Conditions will have to be carried along until this operator gets reached. The Binding Conditions that are of special interest here are those of the DNPs, as it is these which give rise to presuppositions. The point at which the Binding Condition of a DNP gets converted into a presupposition is that where a semantic boundary is reached, at which point all presuppositions generated below it have to be made explicit.[15]

[15] The notion of semantic boundary becomes important only where presuppositions may have to get attached to embedded positions. (These are the cases that give rise to the projection problem!) Since the section of the original manuscript which dealt with

In the diagrams below which show the different construction stages of preliminary DRSs I will, partly for reasons of space, abbreviate the Binding Conditions associated with variables x_β as $BiCo_\beta$, where β identifies the type of the given Binding Condition, either by mentioning the type of phrase that introduced the variable, or, if that does not suffice, in some other way. In some cases these labels hide much complexity of detail. This is true in particular of the Binding Conditions associated with the different types of DNPs. We need only recall the barely surveyable literature on definite descriptions and pronouns, or the extensive debates on proper names that have preoccupied philosophers of language for centuries (and especially for the last three decades). Most of what these publications and debates have been about concerns details of the Binding Conditions for different types of DNP's in the sense in which the term 'Binding Condition' is intended here (see the remarks about *Walter* in Section 2.2). However, in a theoretical framework like the one adopted here these details pertain primarily to the second part of the theory, which deals with presupposition justification, and so are of no direct relevance to the particular topic of of this paper. The way in which I will represent the presuppositions generated by DNPs will leave all such details to be filled in at some later stage.

(CP1) and (CP2) cover those aspects of the construction of sentence representations that are of the most immediate importance to the concerns of this paper. However, as I already hinted, the temporal aspects of predication also have a significant role to play in the examples we will be looking at later. So the construction principles which govern the representation of temporal relations are important too. For the most part I will be making implicit use, as far as the representation of time is concerned, of the combination principles contained in Kamp & Reyle (1993), Ch. 5. Still, the following two principles CP3 and CP4 may help the reader to better understand this aspect of the construction below.

(CP3) (CP1) stipulates that each NP α introduces a variable x_α, its 'referential argument'. A similar assumption is made for VPs: The referential argument of a VP (which strictly speaking originates with the verb that is its lexical head, just as the referential argument of an NP originates with its head noun) is always an eventuality—i.e. an event or a state, depending on the aspectual properties of the VP. This variable comes with a Binding

---

embedded presuppositions had to be cut, the notion is not important in what follows: all presuppositions we will encounter are attached to the highest possible position of the non-presuppositional part of the sentence representation.

Condition too, which gets converted into a real binding at the node which carries the information about the verb's tense features—or, in the case of certain intervening operators such as temporal quantifiers or negation, at the point where these operators intervene. We assume moreover that the introduction of the referential argument of the verb is always accompanied by the introduction of a *location time*, a variable ranging over times to which the referential argument stands in a certain relation of inclusion or overlap. Strictly speaking it is this location time, and not the referential argument itself, which gets located along the temporal axis by the information provided by tense (as well as, sometimes, other temporal constituents of the sentence). The simplest case, to which the examples with which we will deal in this paper conform, is that where no operators intervene and where the finite tense (here assumed, somewhat old-fashionedly, to be located at 'Infl') existentially binds the referential argument while relating it to the utterance time n in accordance with the tense of the sentence.[16] Note that temporally locating the eventuality described by the verb is a way of indicating the time of a verbal predication. Thus, by representing *John loved Mary* as involving the conditions s: $love(j, m)$, $t \subseteq s$ and $t \prec n$ the information is captured that the condition (or: state) of John loving Mary held at some time before the utterance time.

(CP4) As we argued in Section 2.2, not only verbal predications are dependent on time; this is true for many non-verbal predications as well. This means that non-verbal predicates must allow for the introduction of several variables, like verbs. Consider for instance the lexical head noun *student* of the argument phrase *a student*. If the predication it expresses of its referential argument x is represented as temporally dependent, i.e. by means of the conditions s: $student(x)$ and $t \subseteq s$, then this means that the noun must introduce three variables, x, s and t, each with its own Binding Condition. The Binding Condition for t must reflect the possibilities mentioned in Section 2.2; that for s should treat it as existentially bound, while the Binding Condition for x will be determined by the Determiner of the NP. A similar story is to be told about other non-verbal predications. This is true in particular for the prenominal possessive relation expressed by *his*. I refrain from further details.

---

[16] In case the sentence contains one or more temporal adverbials, these will impose additional constraints on the temporal location of the VP's referential argument. See Kamp & Reyle (1993), Ch. 5 for some details.

So much for the main construction principles. The representation construction, based on the principles (CP1)-(CP4), builds sentence representations by first attaching lexical representations to the lexical leaves of the syntactic tree.[17] The lexical representations attached to the leaves are then step-wise put together as we make our way up the tree, each time combining the representations of daughter nodes into a representation for their mother. Roughly speaking, when the representation that is attached to the node with a lexical head is combined with the information present at a functional node, the representation typically undergoes a certain transformation; and the combining of two representations of constituents with lexical heads takes the form of some kind of merge. Again we skim over a number of details, hoping that the reader will be able to fill in the missing bits if he wants to from what the examples will show.

One last remark before we look at some actual representation constructions. We need a general format for the various kinds of information that the representation pieces which get attached to the tree nodes in the course of representation construction must be able to carry. The following notion—of a *semantic node representation*—will, at least within the context of our present requirements, answer those needs:

(14) A *semantic node representation* is a pair consisting of:
(i) a set of one or more *store elements*; and
(ii) a content representation.
The content representation will always be a DRS.
The store element is a triple consisting of:
(a) a variable;
(b) a constraint, again a DRS; and
(c) a Binding Condition.

Two examples:

(15) (i) (Representation of an occurrence of the verb *love*)[18]

$$\left\langle \left\{ \langle t, , \text{BiCo}_{loc.ti.} \rangle, \langle s, \boxed{t \subseteq s}, \text{BiCo}_{m.ev.} \rangle \right\}, \boxed{s\text{: love}(v_{subj}, v_{d.obj})} \right\rangle$$

[17] These representations are semantic structures supplied by the lexical entries for the lexical items appearing at those leaves. I will say nothing here about the form of the lexicon and the way in which the information it contains gets transferred to the lexical leaves of syntactic trees.

[18] The subscripts *loc.ti.* and *m.ev.* are short for 'location time' and 'main eventuality'. For more on the rationale behind these annotations see the remarks on the following diagram (27) below.

I should add that the practice adopted in this paper of simply labelling the different Binding Conditions for temporal variables hides much complexity from view, just as this

In (15.i) the DRS on the right, consisting of the single condition: *s*: $love(\nu_{subj}, \nu_{d.obj})$ contains the contribution which the verb will make to the non-presuppositional part of the representation. (The designations ν_{subj} and $\nu_{d.obj}$ indicate that the argument places occupied by these variables are grammatically realised as subject and direct object. This information is part of the lexical entry of the verb.) In this case the store has two elements. The second of these consists of (a) the referential argument of the verb; (b) a constraint to the effect that the entity represented by the referential argument (the state of loving that the sentence describes) temporally includes its location time *t*; and (c) a reference to the Binding Condition for the referential arguments of verbs. The first store element concerns the location time *t* itself. This store element consists of (a) the variable *t*; (b) an empty slot, indicating that the verb imposes no constraints on *t*; and (c) a reference to the Binding Condition for location times of the referential arguments of verbs.

(16) (i) (Representation of an anaphoric token of the pronoun *it*.)

$$\left\langle \left\{ \langle x_{it} \ , \ \boxed{\text{non-person}(x)} \ , \ \text{BiCo}_{pron} \rangle \right\}, \boxed{} \right\rangle$$

In this case the second component on the right is empty. This is so for all definite noun phrases. It reflects the circumstance that all material in such an NP serves the purpose of identifying the referent and thus belongs to the presupposition. Second, the store contains just one element in this case, consisting of (a) the variable which is introduced by the pronoun and represents its referent; (b) the pronoun's descriptive content, to the effect that its referent must be a non-person; and (c) a reference to the Binding Conditions for pronouns (with all that involves).

These preliminaries should enable the reader to follow, at least in outline, the construction of the preliminary representations for (7) and for the second sentence of (2). To these we now turn.

---

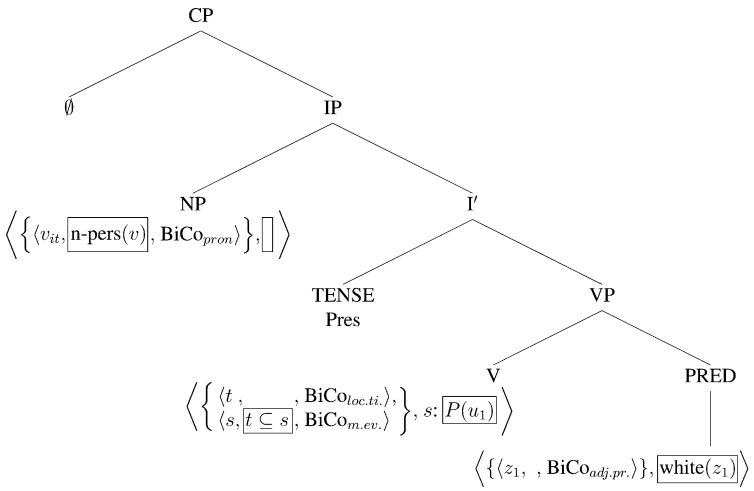
was noted earlier for the Binding Conditions associated with the variables introduced by NPs. Some of our most recent thoughts about the system of Binding Conditions associated with the different temporal variables that get introduced in the course of the construction of sentence representations can be found in Reyle et al. (2000).

3.2. *Construction of Preliminary Representations*

In this section we show some of the successive stages in the construction of preliminary representations. We begin with the construction of the representation of (7).

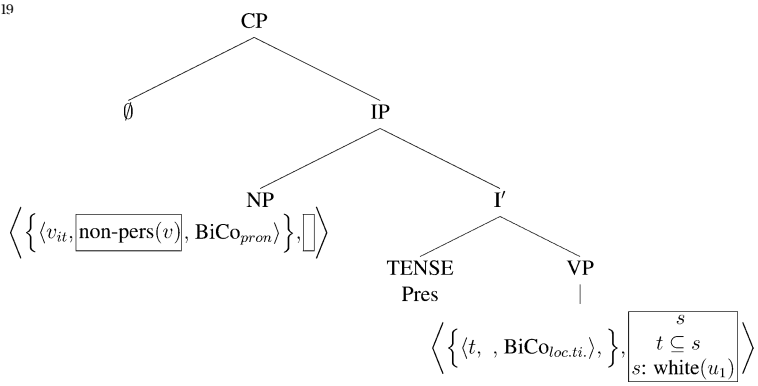
(17) presents the result of replacing the leaves of the tree (8) with the interpretations which the lexicon provides for the lexical items attached to those nodes.

(17)



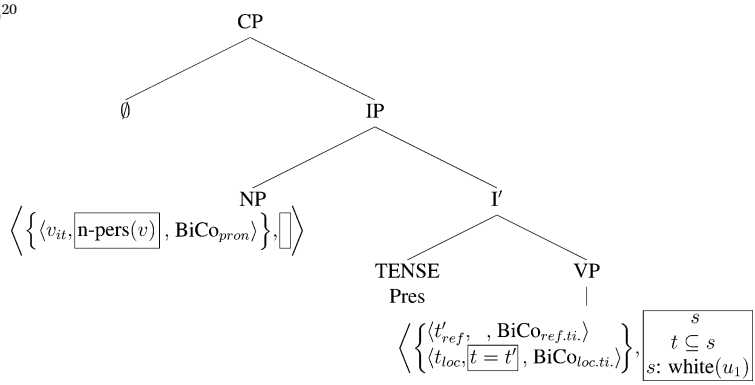
Putting the component representations together yields successively the structures (18)–(20).

(18)[19]

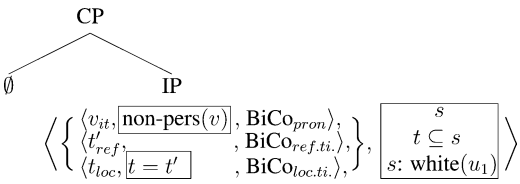


[19] Combining the representations for the copula and its complement takes the form of a

(19)[20]



(20)



---

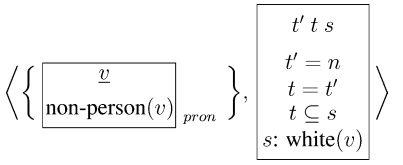
kind of 'unification', with the predicate variable *P* 'unifying' with the predicate constant *white* and the placeholders u_1 and z_1 get unified with each other. (It is immaterial which of these two variables is retained.) This identification process is to be distinguished from what happens at the one but next step, where the referential argument v of the subject phrase gets inserted in lieu of the placeholder u_1 for the argument position of the (verbal) predicate to which the phrase is syntactically linked. Here we have a case of genuine 'argument filling'. (The distinction between 'placeholders' and 'true arguments' to which this informal description alludes could have been made explicit by using different symbols for the two categories, but I have abstained from that in order not to overload what is a fairly complex notation in any case). (For copula constructions see Roßdeutscher (2000).)

[20] I have opted for a very simplified treatment of tense here. The two-dimensional analysis of Kamp & Reyle (1993), Ch. 5, has been reduced to one dimension (that of TENSE, leaving out that of the Temporal Perspective), and it is assumed that "reference time" (= TPpoint) according to the terminology of Kamp & Reyle (1993) always gets identified with n, when, at the level of CP node the sentence is recognised as main clause. For a more recent treatment of the temporal information conveyed by tense, also see Reyle et al. (2000). A matter that is often difficult to decide is at what stage the different temporal variables used in representing the temporal location of verbal predications (and others) are bound. I have here assumed that the eventuality variable (here s) gets bound at the level of VP, while both the location time and the reference time are bound only at the CP node. Binding the location time variable t_{loc} this late (i.e. after t_{loc} has been related to t_{ref}) may seem arbitrary. But it is necessary in certain cases we will encounter later on.

From (20) we obtain a preliminary representation (21) for (7) by

(i) interpreting the represented sentence as a main clause, through identifying the location time *t* with the utterance time *n*.
(ii) dealing with the other elements in the store in accordance with their Binding Constraints;[21] this turns the store element for v_{it} into a presupposition.

(21)



At last we are ready to consider the construction for the preliminary representation of the second sentence of (2). As this sentence is identical with (7) except for its more complex subject phrase *his rabbit*, our principal concern will be with this phrase. It is here that we encounter for the first time the problem of 'one presupposition inside another'. And the question we need to consider is how the nested presupposition structure that this NP should yield comes about? The answer I propose runs as follows. In processing the N′ part of (9) we also have to deal with the pronoun *he* (recall that we have analysed the possessive as decomposable into *he* and the possessive relation *'s*). The representation of *he* is like that of *it* in (21), except that its descriptive content is that of a male rather than a non-person. When this representation gets merged with that of *'s*, we obtain a representation in which the possessive relation ends up in the DRS on the right.[22] Moreover, the store of this representation contains, besides the two store elements for the relational state introduced through making the possessive predication temporally dependent on the location time for that state, also the store element coming from the representation of the pronoun. Combining the representation attached to the AP node which is obtained in this way with that of the noun *rabbit* then leads to a representation with a right hand DRS which contains two conditions: (i) the condition that the referential argument of *he* and that of

[21] We take this to be the information provided by the empty *Comp* node; it is assumed that the *Comp* of a subordinate clause always contains information which links it to some element of the clause in which it is embedded.

[22] To be exact, the right hand side DRS of the representation of the AP of (19) is the result of merging the right hand side of the representation of *'s* with the right hand side DRS of the representation of the pronoun. As the latter is empty, the result is simply the first DRS.

his rabbit stand in the possessive relation and (ii) the condition that the latter argument satisfies the predicate *rabbit*. (Once again, more should be said about the details of the combination principles than space permits.)

It ought to be clear from these remarks how the representation of *his rabbit* comes to contain a store element whose middle part is complex, involving a store of its own. The representation is given in (22).

(22)[23]

$$\left\langle \left\{ \left\langle v, \left\langle \left\{ \begin{array}{l} \langle u, \boxed{\text{male}(u)}, \text{BiCo}_{pro}\rangle, \\ \langle s'', \boxed{t'' \subseteq s''}, \text{BiCo}_{AP}\rangle, \\ \langle t'', \quad , \text{BiCo}_{adj.pr.ti.}\rangle \end{array} \right\}, \boxed{\begin{array}{c} s'' \\ s'' \text{: 's}(u, v) \\ \text{rabbit}(v) \\ t'' \subseteq s'' \end{array}} \right\rangle, \text{BiCo}_{d.d.} \right\rangle \right\}, \boxed{\quad} \right\rangle$$

The construction of the preliminary representation of the second sentence of (2) is largely similar to that of (20). In particular, the penultimate stage will be almost like (20) except that the representation of it is now replaced by (22). Conversion of this structure into the preliminary representation itself involves in particular dealing with the store elements which occur in (22). Of these the adjectival predication time t'' gets identified with the location time t of the main eventuality s and s'' is put into the universe of the non-presuppositional part of the preliminary representation.[24] That leaves the store elements connected with the referential argument v of *his rabbit* and the one associated with the referential argument u for *he*. Both these store elements are, according to their Binding Conditions, presuppositional, and thus have to be turned eventually into presuppositions. The conversion principle which is responsible for this is sensitive to the recursive structure of the store: the presupposition which corresponds to the descriptive component of a store element which is itself presuppositional gets attached as a satellite presupposition to the presupposition that results from that store element. In this way we obtain a nested pre-

[23] *AP* is shorthand for Adjective Phrase. *adj. pr. ti.* stands for 'adjectival predication time'. The Binding Condition for such location times is special insofar as the location time of a prenominal AP must be identified with the predication time of the head noun (which we have here left implicit). In turn the temporal location of the head noun predication is often to be identified with that of the verb to which the noun's NP is an argument, but not always. For discussion see Enç (1981), Musan (1995) and Tonhauser (2000). Because of the special requirements to which the temporal location of such predications is subject, it seemed right to make this explicit by introducing a special Binding Condition label.

[24] Here we see the need to retain t in store, so that it remains available as the variable with which t'' gets identified. Existential binding of t at the level of the I' node would have made this problematic from a logical-semantic perspective.

supposition structure which is essentially like the one proposed in van der Sandt. The preliminary representation which comes about in this way is given in (23).

(23)

$$\left\langle \left\{ \left\langle \left\{ \begin{array}{|c|}\hline \underline{u} \\ \text{male}(u) \\ \hline \end{array} \right\}_{pron}, \begin{array}{|c|}\hline s'' \ \underline{v} \\ t \subseteq s'' \\ \text{rabbit}(v) \\ s'' \text{: 's}(u, v) \\ \hline \end{array}_{def.desc.} \right\rangle, \right\}, \begin{array}{|c|}\hline t'\ t\ s \\ t' = n \\ t' = t \\ t \subseteq s \\ s \text{: white}(v) \\ \hline \end{array} \right\rangle$$

The set of the two nested presuppositions triggered by *his rabbit* and *his* is easily resolvable in the context provided by (3): *u* is identified with *w* and *v* with *y*. On these identifications the context (3) verifies the conditions of the presuppositions and after merging the non-presuppositional part of (23) with (3) we obtain, unsurprisingly, (24) as the discourse representation for the result of the two sentence 'discourse' (2).

(24)

$$\begin{array}{|c|}\hline s_0\ w\ y\ t\ s \\ n \subseteq s_0\ t = n\ t \subseteq s \\ \text{Walter}(w)\ \text{rabbit}(y) \\ s_0 \text{: have}(w, y) \\ s \text{: white}(y) \\ \hline \end{array}$$

4. Presupposition Interaction

As noted in the introduction, discussions of presupposition have tended to concentrate on cases where just one presupposition, or at best a couple, are being explicitly considered. But as a rule the number of presuppositions generated by a single sentence is bigger.

We have seen a comparatively simple form of this already in our discussion of van der Sandt's example *Walter's rabbit*. Indeed, the insight that presupposition nesting occurs is by no means new and goes back at least as far as van der Sandt. But often the complexity of the interactions and of the representations resulting from them goes well beyond what we have seen so far in this paper, and also, if I am not mistaken, beyond what can be found in the existing literature.

In the remainder of this paper we will be looking more closely at this sort of complexity. Our ultimate aim in this section is to compute the preliminary representation of (25):

(25) *Walter's rabbit is on the loose again.*

This sentence gives rise to presuppositions connected with (i) the definite NPs *Walter* and *Walter's rabbit* and (ii) *again*.[25] Before we can address the question how the computations of different types of presuppositions interact, however, we have to know how those different types of presuppositions are computed when taken in isolation. As regards the presuppositions triggered by *Walter's rabbit*, we have seen enough in the preceding sections. But presuppositions triggered by *again* haven't yet been considered in this paper, and it will be helpful to first look into their computation by looking at a sentence in which there is no interaction between the presuppositional effects of *again* and those connected with the definite Subject NP of (25). This sentence, in which the definite subject has been replaced by the indefinite *some rabbit*, is given in (26).

(26) *Some rabbit is on the loose again.*

First a few words about the syntax of *again* and its consequences for the computation of the presuppositions connected with it. One important difference between the presuppositions we have considered up to now and those triggered by *again* has to do with the question of the trigger's *scope*. For instance, (26) is ambiguous in that it can be interpreted as presupposing that the rabbit which is on the loose now has been on the loose before, but also as presupposing that it has been the case before that some rabbit was on the loose, but not necessarily the same one that is on the loose now. The difference between these two presuppositions can be accounted for as the effect of two different scope assignments to *again*: In one case the scope of *again* only includes the subject NP, in the other it doesn't. Underlying this account is the obvious observation that the presuppositions generated by *again* result from its interaction with other material from the sentence in which it occurs; this other material serves to identify an eventuality type (i.e. a type of event or state) that is true of the main eventuality *e* the sentence describes; the presupposition generated by *again* is to the effect that one or more eventualities of that type occurred before the time of *e*. Given this general characterisation, we can identify the scope of an occurrence of *again* with the sentence material which defines the relevant eventuality type. In this respect *again* is one of a larger class of presupposition triggers, which also includes the already mentioned *too*, its close alternative *also*,

[25] I am not claiming that these are all the presuppositions generated by occurrences of (25). However, for expository reasons they will be the only ones considered.

a number of temporal presupposition triggers such as *still* and *already*, and various other particles.

Sometimes the position of *again* in the sentence determines its scope unambiguously. This is so for instance with sentence-initial *again* (e.g. when we move the word from its sentence-final position in (26) to the front of the sentence). But in other cases the surface position of *again* is not decisive. I take (26) to be a case of the latter kind.[26]

An exact treatment of scope ambiguities like that of (26) involves questions of syntax which go beyond the scope of this paper. For the case of (26) I will assume that its sentence-final *again* is a right-adjunct and that its possible adjunction sites are VP and IP, with VP-adjunction giving rise to the narrow scope reading and IP-adjunction to the wide scope one. But more careful syntactic considerations will have to decide the tenability and generalisability of these assumptions.

4.1. *The Wide Scope Reading for* Again *in* (26)

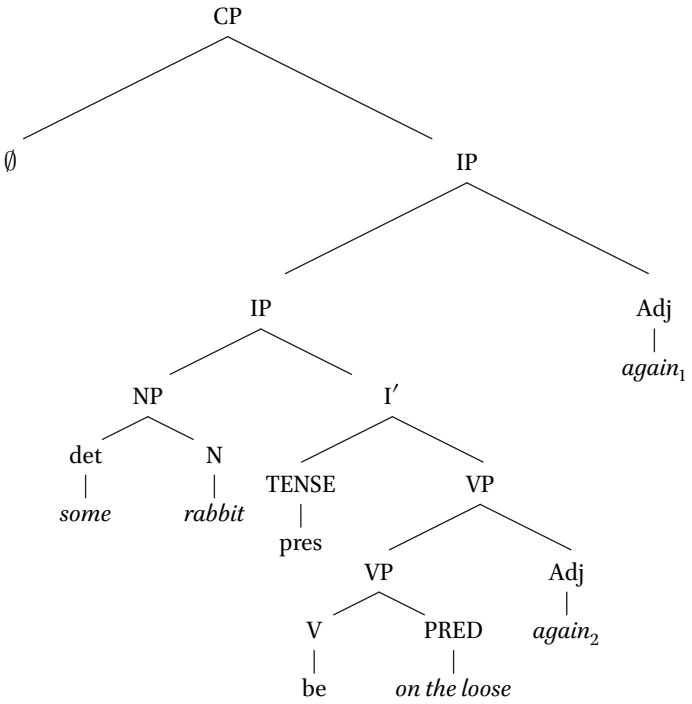
We just noted that the presuppositions triggered by *again* are claims to the effect that the time of the eventuality described by the sentence is preceded by one or more eventualities of the same type. So in fact, the problem of computing such presuppositions is in essence that of computing the representation of the relevant eventuality type. We will restrict our attention here to those cases in which the full content of the verb is part of the characterisation of this type.[27]

---

[26] Perhaps this is an arguable point. The narrow scope reading of (26) seems preferred and some people might perhaps want to argue that that is really its only reading. It appears that the relation between scope and surface position, for *again* and other presupposition triggers, has not yet been exhaustively investigated. As Roßdeutscher and I noted in our Kamp & Roßdeutscher (1994), the surface positions of the German equivalent of *again*, the word *wieder*, tend to be much more determinative of its scope than they are for *again* (even if *wieder* too is subject to certain scope ambiguities).

[27] I am passing over the notorious problem of the so-called restitutive interpretations of *again*, as one finds it in *The doctor cured the baby again*, where the baby had never been ill before contracting the disease from which the doctor is asserted to have cured it. Here the eventuality type of which the sentence is taken to describe a reinstantiation is that of a state of the baby being healthy, not that of an event of the doctor curing the baby.

(27)



(27) shows the two possible attachment sites for the *again* of (26). First we consider the construction of the preliminary representation corresponding to the higher site, where the adverb is adjoined to IP. Applying the bottom-up construction algorithm sketched in section 3 up to the level of the lower IP node we get, for the representation attached to this node, (28) (cf. (20)).

(28)[28]

$$\left\langle \left\{ \begin{array}{l} \langle y, \boxed{\text{rabbit}(y)}, \text{BiCo}_{indef} \rangle, \\ \langle t'_{ref}, \quad\quad\quad , \text{BiCo}_{ref.ti.} \rangle, \\ \langle t_{loc}, \boxed{t = t'} \quad , \text{BiCo}_{loc.ti.} \rangle \end{array} \right\}, \boxed{\begin{array}{c} s_{ev_m} \\ t \subseteq s \\ s\text{: on the loose}(y) \end{array}} \right\rangle$$

[28] *Some rabbit* is represented here as an ordinary indefinite, and no distinction is being made between *some* and *a*. N.B. BiCo_{indef} stands for the Binding Condition for indefinites, which requires that the referential argument be introduced into some appropriate DRS universe. In the present example this universe is the main universe of the non-representational part of the preliminary representation of the full sentence, and I am assuming that the discourse referent is placed in this universe at the level of CP. For more on this problem—and more generally on the scope of indefinites see in particular Abusch (1994) and also Kamp (2000b).

It is from this structure that the presupposition has to be computed which is triggered by *again*. As we noted, this presupposition involves the computation of the right eventuality type. In principle this computation is easy, as the type can be extracted from (28) straightforwardly. However, in order to do this we need to be able to recognise which element in (28) is the main eventuality; and we also need to be able to identify its location time. The source of this information is clear enough: The main eventuality is introduced as the referential argument of the main verb of the sentence, and the location time is introduced in tandem with it. Since this information has to be available when we need it, it has to carried up the tree. So the two variables must be somehow marked at the time of introduction, and these marks carried upwards. In fact, we already made this information explicit, when discussing the construction algorithm in Section 3, where the location time was explicitly marked as t_{loc} and the main eventuality variable got the subscript ev_m. (Cf. footnote 18 on p. 473).

Let us call the structure associated with the sister node of the given occurrence of *again* the *source structure* (for the presupposition triggered by *again*). The extraction operation can now be described as follows:

(29) (i) For each of the marked variables (here t_{loc} and s_{ev_m}) we introduce new variables (here, say, t^p and s^p).
  (ii) The representation of the content of the *again*-presupposition is obtained in the following way:
    (a) Replace in the source structure the marked variables by their new counterparts.
    (b) Introduce fresh copies also for all other variables occurring in the source structure.
    (c) Add to the store an element for the 'presupposed location time' t^p. This store element contains a new type of Binding Condition, specific to the location times of *again*-presuppositions. I will assume here that this Binding Condition is like that of indefinite phrases in that its variable eventually gets placed into a suitable DRS universe. But see the discussion towards the end of this section.
    (d) The store element for the reference time (here t_{ref}) gets deleted, and similarly the DRS condition $t = t'$ from the second component of the store element for t_{loc}, in which t'_{ref} is mentioned. The intuition behind this last requirement is that t'_{ref} plays a part in locating the asserted eventuality on the time axis, and plays no part in the interpretation of the presupposition structure generated by *again*. The need for deleting such information that has to do with the temporal location of the asserted eventuality is particularly prominent in conditions with adverbs like *yesterday*, *at 11:30 this morning*, etc. For instance, *Walter's rabbit was on the loose again yesterday* does not presuppose that there

was an earlier event of Walter's rabbit being on the loose *which also occurred yesterday*, but only that there was such an earlier event.

Note well that the resulting structure is a copy (modulo the modifications mentioned under (a)–(d)) of the entire source structure, including the store (as well as previously constructed full-scale presuppositions, should the source structure contain such). In this way an *again* -presupposition can produce, as we will see later in this section, 'reduplications' of other presuppositions.

The construction of an *again*-presupposition involves a further operation which affects the right hand DRS of the source structure. Somewhere it has to be made explicit that the presupposed eventuality (here represented as s^p) precedes the time of the main eventuality (here s). We add this condition, $t^p \prec t'$, in the DRS to the right in (30), which will eventually turn into the non-presuppositional part of the representation. (The condition should not be added to what will become the presupposition; for the presupposition should be justifiable independently of the content of the non-presuppositional part, which presents what the utterance asserts, queries, etc., since that content is defined only if and when justification has taken place. Thus the variable t is, from this point of view, not accessible from the perspective of the presupposition.)[29]

[29] The conditions $t^p \prec t$, $t^p \subseteq s^p$ and $t \subseteq s$ do not exclude the possibility that s^p and s overlap. It might be thought that disjointness is part of what *again* means and thus that failure to entail it is a fault in the analysis. Actually for states (as opposed to events) the matter of disjointness isn't straightforward, as shown by the following example, due to Manfred Bierwisch (p.c). I can say, watching the news, *Hilary is standing by Clinton again*, not wanting to imply that Hilary was standing by Clinton at some earlier occasion, then did not stand him for some time and now is standing by him again, but rather that the present occasion is another one at which she is standing by him, without suggesting that she ever ceased to stand by him. In this case the stative condition described in the sentence does not seem to have two temporally disjoint instantiations. In the light of examples like this one I have decided to ignore the disjointness problem here. Note that in the case of events temporal disjointness is guaranteed in our representations, since events represented as included in their location times $t^p \prec t'$, $e \subseteq t'$ and $e^p \subseteq t^p$ jointly entail that $e^p \prec e$.

(30) (Representation associated with the upper IP node of (27))[30]

$$\left\langle \left\{ \left\langle \left\{ \begin{array}{ll} \langle y^p, & \boxed{\text{rabbit}(y^p)} \rangle, \text{BiCo}_{indef} \rangle \\ \langle t^p, & \rangle, \text{BiCo}_{pres_{again}} \rangle \\ \langle s^p, & \boxed{t^p \subseteq s^p} \rangle, \text{BiCo}_{pres_{again}} \rangle \end{array} \right\}, \boxed{s^p\text{: on the loose}(y^p)} \right\rangle \right\},$$

$$\left\langle \left\{ \begin{array}{ll} \langle y, & \boxed{\text{rabbit}(y)} \rangle, \text{BiCo}_{indef} \rangle, \\ \langle t'_{ref}, & \rangle, \text{BiCo}_{ref.ti.} \rangle, \\ \langle t_{loc}, & \boxed{t = t'} \rangle, \text{BiCo}_{loc.ti.} \rangle \end{array} \right\}, \boxed{\begin{array}{c} s_{ev_m} \\ t \subseteq s \\ s\text{: on the loose}(y) \\ t^p \prec t \end{array}} \right\rangle$$

(30) prompts a further remark. It is plain that its presupposition representation, with its non-empty store, requires further operations, which empty its store while turning the Binding Conditions of the elements in that store into proper bindings or presuppositions. For the first store element, introduced by the indefinite subject *some rabbit*, conversion is straightforward: its binding condition is converted into a proper (existential) binding of the variable y^p.[31] The remaining two elements of the presupposition store in (30), which concern the presupposed eventuality s^p and its location time t^p, are more problematic. The question that needs an answer here is whether these store elements are to be converted into existential bindings, like the first element, or into (anaphoric) presuppositions.

This question arises for many store elements belonging to representations that will be turned into presuppositions eventually. Answers to this type of question are delicate. Here too I refrain from a detailed discussion of the issues and simply state the decision I have reached for this particular case: Both t^p and s^p are to be bound existentially; more precisely, they are to

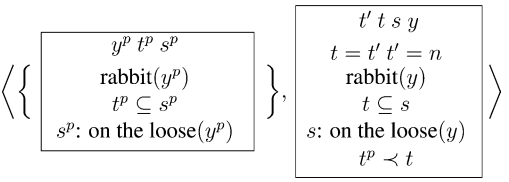
[30] It is merely for reasons of lay-out that the presupposition set is displayed above the 'non-presuppositional part' and not to the left to it, as in earlier diagrams.

[31] Strictly speaking conversion of this store element will take place when the right level in the bottom-up construction of the presupposition-generating sentence is reached. In the present case this point is the clause boundary marked by the CP node. At that point both the store element for y and that for y^p are eliminated and the two variables are introduced into the appropriate universes. The time of the conversion of the store element for y^p is thus determined (strictly speaking) by that part of the tree (27) which extends above the IP node at which *again* introduces the structure that will turn into the definitive representation of its presupposition by the time when we are done with the construction of (26)'s preliminary representation. We will presently see a more illuminating example of what happens when a presupposition gets introduced below the top node of the tree, when we consider the narrow scope interpretation of *again* in (26). I am assuming that the other two elements of the store of the presuppositional part of (30) are also taken out of it at the processing stage identified by the CP node.

be existentially bound at the point where the representation to which they belong is turned into the definitive representation of the presupposition it is destined to become.[32]

This concludes my description of the last round of operations that are needed in order to complete the construction of the wide-scope preliminary representation of (26). Applying these operations to (30) yields (31)

(31)



The content of (31) should be clear: Its claim is that at the utterance time *n* there obtains a state to the effect that some rabbit is on the loose; and this claim is made under the presupposition that a similar state—of some rabbit being on the loose—obtained at some time before *n*.

4.2. *The Narrow Scope Reading*

The narrow scope reading of (26), generated by the VP attachment site for *again*, leads to the creation of a non-empty presupposition set at the level of the upper VP. (32.i) displays the structure that application of the construction operation triggered by *again* causes to be attached to the upper VP node; (32.ii) shows that part of the syntactic tree which is still relevant to the remainder of the construction.

---

[32] The decisive issue is whether the presupposition whose representation is being built can be accommodated without any requirement that a specific antecedent be found for its distinguished variable in the given context. If this is so, then it indicates that the variable has an existential status within the presupposition. It appears to me that this condition is satisfied for the case we are considering: Inasmuch as accommodation of an *again*-presupposition is possible, it can always take the form of assuming that there was an eventuality of the required type at some time before the location time of the asserted eventuality. Questions of this sort tend to get obscured by the circumstance that in those cases where the context contains information which verifies the presupposition, this often has the effect of binding the variable to an element that is part of that information. E.g. an *again*-presupposition may be satisfied in the context because it contains information to the effect that there was an eventuality *s* of the required type at some earlier time *t*. In this case justification of the presupposition has the effect of identifying t^p with *t* and s^p with *s*. Observations of this sort might tempt us to think that t^p and/or s^p play(s) some sort of anaphoric role, but by themselves they are not enough to establish this.

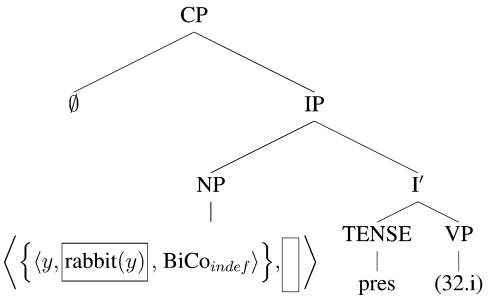
(32)

(i)

$$\left\langle \left\{ \left\langle \{\langle t^p, \quad , \text{BiCo}_{pres_{again}}\rangle\}, \boxed{\begin{array}{c} s^p \\ t^p \subseteq s^p \\ s^p\text{: on the loose}(u_1) \end{array}} \right\rangle \right\}, \right.$$

$$\left. \left\langle \left\{ \langle t_{loc}, \quad , \text{BiCo}_{loc.ti.}\rangle \right\}, \boxed{\begin{array}{c} t_{loc}\ s_{ev_m} \\ t \subseteq s \\ s\text{: on the loose}(u_1) \\ t^p \prec t \end{array}} \right\rangle \right\rangle$$

(ii)



The next steps are needed to (i) combine the result of this operation with the subject interpretation, (ii) relate *t* to the reference time *t′*, and (iii) anchor *t* to *n* and convert the remaining 'indefinite'-marked Binding Conditions into actual bindings.[33] The first operation is much like the instances of relating *t* to *t′* that we have seen before (but now with the new qualification that in the case of (32.ii) the operation only affects the non-presuppositional part, while leaving the presuppositional component untouched). The second operation is of more interest. Thus far, combining the representation of a predicate with the representation of one of its argument phrases took the form of inserting the variable representing the argument phrase into the relevant argument slot of the predicate. But now, when combining the representation of the subject phrase of (26) with that of the representation of the I′ node in the structure derived from (32.ii), we find ourselves in a different situation. For the placeholder u_1 now occurs in two places, once inside the non-presuppositional part and once inside the presupposition.

[33] Note that this time the remaining operations are connected with different nodes of the syntactic tree (those above the upper VP node, where the *again*-presupposition gets generated). Thus operation (i) must be performed to get the representation associated with the IP node while (ii) and (iii) are performed at the level of CP. Here we get a taste of the construction complexities connected with certain presuppositions that get introduced lower down in the tree.

Intuitively it is clear, however, that in this case substitution should apply to both occurrences of the placeholder. So substitution is for all occurrences of the relevant placeholder.

I have dwelt on this point because of the impact it has on the representation which will result. First, if the variable *y* gets inserted inside the presupposition as well as the non-presuppositional part of the VP-interpretation, then both should be within the scope of the store element introduced by the subject phrase. For otherwise we would end up, after conversion of this store element into a proper binding, with an unbound occurrence of *y*. This means that the representation structure associated with the IP node should now have the form given in (33).

(33)

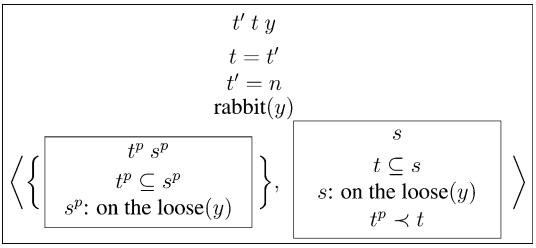
$$\left\langle \left\{ \langle y, \boxed{\text{rabbit}(y)}, \text{BiCo}_{indef} \rangle \right\}, \right.$$
$$\left\langle \left\{ \left\langle \{ \langle t^p, \quad , \text{BiCo}_{pres_{again}} \rangle \}, \boxed{\begin{array}{c} s^p \\ t^p \subseteq s^p \\ s^p\text{: on the loose}(y) \end{array}} \right\rangle \right\}$$
$$\left\langle \left\{ \begin{array}{l} \langle t'_{ref}, \quad , \text{BiCo}_{anch_{temp}} \rangle, \\ \langle t_{loc}, \boxed{t = t'}, \text{BiCo}_{loc.ti.} \rangle \end{array} \right\}, \boxed{\begin{array}{c} s_{ev_m} \\ t \subseteq s \\ s\text{: on the loose}(y) \\ t^p \prec t \end{array}} \right\rangle \Big\rangle \Big\rangle$$

In (33) the *again*-presupposition and the part that presupposes it are both within the scope of the Binding Condition on *y*, and so they will both be within the scope of the bound variable *y* when this Binding Condition gets converted.

To turn (33) into a preliminary representation of (26) we also have to deal with its other Binding Conditions. This business is more delicate than it might appear at first sight. As will be even plainer from the next example we will look at, the subject phrase may require temporal binding to t_{loc}, so t_{loc} must be available at the top level. Thus it, and with it the reference time t', have to be 'lifted' beyond the *again*-presupposition. Otherwise, the remaining operations produce the same effects as in earlier examples (t' identified with *n*, etc.).[34] The result is (34).

[34] As it stands, this is just a stipulation. I am unsure how a general principle governing such variables lifting should best be stated.

(34)

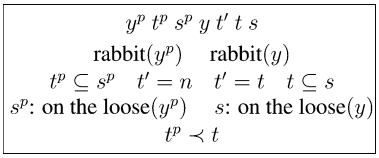


Note that the presupposition of (34) occurs not at the top level, but is within the scope of the implicit existential quantifiers represented through the presence of the discourse referents t', t and y in the main universe of the representation.

Here we encounter, for the first time in this paper, a presupposition in subordinate position, and therewith an instance of the Projection Problem. The problems connected with this particular kind of subordinate occurrence are wellknown from the literature and various suggestions for its solution exist.[35] The nature of the problem becomes clearly visible when we compare the question of justifying the *again*-presupposition in (34) with that of justifying the *again*-presupposition of (31). Accommodation of the presupposition of (31) is unproblematic. It involves the simple assumption that the presupposition holds in the given context, and the total gain in information, through presupposition accommodation and then adding the non-presuppositional information of (31), can be represented as in (35):

[35] The problem was first explicitly discussed in Karttunen & Peters (1979), who note a difficulty with their compositional account of presupposition within an extension of standard Montague Grammar. Karttunen and Peters observe that *Someone managed to succeed George III on the throne of England.* carries the presupposition that whoever succeeded George III must have made an effort in that direction. This presupposition is not satisfied if someone else tried to succeed George III, but failed, while the person who did succeed him didn't make any effort at all. Heim (1983) notes a similar problem for sentences such as *A fat man was pushing his bicycle. His bicycle* carries the presupposition that someone had a bicycle. But again, the presupposition is really that the man who was pushing his bicycle had a bicycle, and not somebody else. It is worth emphasising how easily and frequently such situations arise; the present case, where *again* is construed as excluding from its scope a phrase which introduces at least one variable, is one of many which require this type of binding into presuppositions. Because cases of binding into presuppositions are so very common, the problem of how the resulting presuppositions are to be justified is of the first importance.

(35)



Justification of the presupposition of (34) is not so straightforward. In fact, on the widely accepted theory that indefinite NPs (like the noun phrase *a rabbit*), convey novelty,[36] we seem to be faced with a real puzzle: If the discourse referent *y* (which is contributed by the indefinite *a rabbit*) stands for something unknown, how can the condition s^p: *on the loose*(y), in which *y* occurs as an argument, be presupposed? The solution of this puzzle involves, I believe, a reassessment of the notion of novelty, which draws a sharp distinction between what is new to the hearer and what is unknown, (that is, not uniquely identifiable) to the speaker. It is legitimate, and very common, for a speaker to refer to an individual about which he takes himself to have uniquely identifying knowledge by means of an indefinite NP, thereby conveying to the addressee that he takes the referent to be unknown to him.[37] Moreover, since such uses of indefinites are common practice, the addressee will also often be prepared to assume that there is a particular individual which the speaker is talking about. In connection with a representation like (34) this assumption allows the hearer to further suppose that the speaker is speaking about a particular rabbit, of which he knows that it had been involved in an earlier escape, at a time previous to the event that the statement (26) reports. In this way, and as far as I can see, only in this way, is it possible for the hearer to accommodate this presupposition.

A consequence of this analysis is that presuppositions like that of (34) always require accommodation. To be precise, whenever an indefinite with an existential interpretation—i.e. its referential argument ends up in the main universe of the representation—governs a presupposition in which it figures as an argument, then (a) on the one hand the indefinite must be interpreted as 'specific' (in the sense that the speaker must know which individual he is talking of) so that it is possible to impute knowledge of the presupposition to him; and (b) on the other hand, the individual spoken of and, a fortiori, the presupposition in which it figures as an argument, must

[36] See Heim (1982).
[37] See Kamp (2000a).

be unknown to the addressee, and thus it cannot be part of the context as it is available to the hearer at that point. Hence accommodation is inevitable.[38]

4.3. *Walter's Rabbit Is on the Loose Again*

We now turn to (25), repeated as (36)

(36) *Walter's rabbit is on the loose again.*

As with (26) there are two possible attachment sites for *again*, giving rise to two different *again*-presuppositions in the preliminary representation. This time we start with the narrow scope case (= VP attachment). The representation for this case is in most respects like that in (34). The only difference is that now the representation of the subject phrase also gives rise to presuppositions (or, more accurately, to a structure consisting of two nested presuppositions). These presuppositions end up in front of the representation structure that occupies the space below the top two lines in (34). (Of course, the discourse referent y and the condition $rabbit(y)$ have to be removed from (34)).

(37) gives the representation that gets attached to the subject node as part of constructing the intended preliminary representation of (36).[39,40]

---

[38] (34) is also compatible with a slightly different scenario, in which the speaker does not have uniquely identifying knowledge of some particular rabbit, but knows that the rabbit is one of a set of rabbits all of which have had previous escapes. It seems to me that in this case too the use of *some rabbit* (as opposed to, say, *one of the rabbits*) conveys to the addressee that the speaker does not take him to be familiar with the set in question. In that case it will be true once more that the presupposition is unknown to the addressee, so that accommodation will be necessary. Perhaps, however, the rules for use of the phrase *some rabbit* are not as strict in this case, and compatible with the assumption that the addressee knows the set over which the variable introduced by the phrase is intended to range. Then, if it is also known to the addressee that all rabbits in that set had previous escapes, accommodation would not be necessary.

[39] This time I have given the condition that y is a rabbit in its time-dependent form. This is not because of a sudden concern with the possibility that what is a rabbit at one time need not be one at another, but to allow for greater transparency of the process of temporal binding to which the interpretation of the subject term of (25) gives rise: the time t'' of the prenominal possessive predication gets identified with the time t' of the predication expressed by the lexical head *rabbit*, while t' can be bound either to the location time of the corresponding event or to the utterance time n.

[40] *nom.lex.h.* indicates that the Binding Condition thus annotated is that for the state of the predication of the nominal lexical head of an argument NP. *nlh. pr.ti.* annotates the corresponding predication time. Similarly, the annotations *AP* and *adj. pred. ti.* identify the Binding Conditions they label as those for the state of a prenominal adjectival predication and for its predication time. Except for operator-like APs this second time always gets identified with that of the lexical head. See the preceding footnote and the discussion of prenominal adjectives in Section 3.

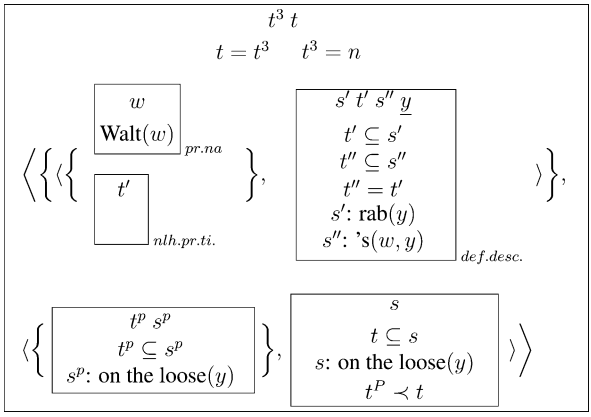
(37)

$$\left\langle \left\{ \left\langle y, \left\{ \begin{array}{l} \langle w, \boxed{\text{Walt}(w)}, \text{BiCo}_{pr.na} \rangle, \\ \langle s', \boxed{t' \subseteq s'}, \text{BiCo}_{n.lx.h.} \rangle, \\ \langle t', \quad\quad\quad, \text{BiCo}_{n.pr.t.} \rangle, \\ \langle s'', \boxed{t'' \subseteq s''}, \text{BiCo}_{AP} \rangle, \\ \langle t'', \quad\quad\quad, \text{BiCo}_{a.pr.ti.} \rangle \end{array} \right\}, \boxed{\begin{array}{l} s' \text{: rab}(y) \\ s'' \text{: 's}(w, y) \end{array}}, \text{BiCo}_{df.d.} \right\rangle \right\} \boxed{\quad} \right\rangle$$

When (37) is combined with the representation attached to the I′ node of the syntactic tree of (25), the location time and reference time for the non-presuppositional part are lifted to the top (cf. the transition from (33) to (34) above). Intuitively, the location time t' of *rabbit*(y), and with it the location time t'' of the possessive predication of the subject term, should be identified with the location time t of the main eventuality of the non-presuppositional part of the representation attached to I′. It is not entirely clear whether this identification is to be taken as the direct result of converting the Binding Condition for t' or whether conversion is to yield a presupposition first, which the identification of t' with t (or n) can then subsequently resolve. I have chosen the second option here, with an eye on the analysis of the other interpretation of (26) we will consider.

The considerations which have led me to choose the second option for t' do not apply to the state variable s' and I can see no good reason against converting its Binding Condition directly into an existential binding when the subject representation and the I′ representation are put together. The same applies to s''. Finally, I have treated the time t'' of the prenominal predication also as bound at this point, through identification with t'. Given these decisions, the preliminary representation of (26) comes to look as in (38):

(38)



If in (38) the presupposition involving t' is resolved by identifying t' with t and the other three presuppositions, connected with *Walter*, with *Walter's rabbit* and with *again*, are accommodated, while the present tense anchors the location time t once again to n, we get the representation in (39).

(39)

$t^p\ s^p\ w\ t\ s\ s'\ s''\ y$	
$t^p \subseteq s^p$ $\quad t = n$	$t \subseteq s$ $\quad t \subseteq s''$
s': rabbit(y)	Walter(w)
$t^p \prec t$	s'': 's(w, y)
s^p: on the loose(y)	s: on the loose(y)

4.4. *Narrow Scope for Walter's Rabbit*

It is the wide scope interpretation of *again* in (26), which gives a proper taste of the complexities to which presupposition interaction can give rise. The subject term is now within *again*'s scope and a copy of its representation will therefore appear within the representation of the *again*-presupposition. But this time the representation of the subject term involves presuppositions of its own, which now appear embedded within the presupposition generated by *again*. What consequences this has for the final representation depends crucially on the binding of the duplicates $t^{p'}$ and $t^{p''}$ of the predication times t' and t'' for the lexical head and the possessive AP. The first stage we display of the construction for the corresponding preliminary representation is the one reached after (i) combining the representations attached to the subject node and to the I′ node and (ii) transforming the representation of the lower IP node into that of the higher one, by performing the construction step dictated by *again*. The display is given in (40):

(40)

$$\left\langle \left\{ \left\langle \{\langle 1\rangle_{again}, \langle t^p, \quad , \text{BiCo}_{pres_{again}}\rangle\}, \boxed{\begin{array}{c} s^p \\ t^p \subseteq s^p \\ s^p\text{: on the loose}(y^p) \end{array}} \right\rangle \right\},$$

$$\left\langle \{\langle 1\rangle\}, \left\langle \left\{ \begin{array}{l} \langle t'_{ref}, \quad\quad , \text{BiCo}_{ref.ti.}\rangle, \\ \langle t_{loc}, \boxed{t = t^3}, \text{BiCo}_{loc.ti.}\rangle \end{array} \right\}, \boxed{\begin{array}{c} s \\ t \subseteq s \\ s\text{: on the loose}(y) \\ t^p \prec t \end{array}} \right\rangle \right\rangle \text{ where}$$

(2) is the structure (41.i) and (1) the result of replacing all the relevant variables by 'fresh copies' of the form α^p, displayed in (41.ii):

(41)[4] $\langle 1 \rangle$

$$\left\langle \left\{ \left\langle y, \left\langle \left\{ \begin{array}{l} \langle w, \boxed{\text{Walt}(w)}, \text{BiCo}_{pr.na.} \rangle, \\ \langle t', \quad , \text{BiCo}_{nlh.pr.} \rangle, \\ \langle s'', \boxed{t'' \subseteq s''}, \text{BiCo}_{adj.ph} \rangle, \\ \langle t'', \quad , \text{BiCo}_{adj.pred.ti.} \rangle \end{array} \right\}, \boxed{\begin{array}{c} s'' \text{: 's}(w, y) \\ s' \text{: rab}(y) \end{array}} \right\rangle, \text{BiCo}_{def.desc.} \right\rangle \right\}, \boxed{\quad} \right\rangle$$

(ii)

$$\langle 1 \rangle_{again} \quad \left\langle \left\{ \left\langle y, \left\langle \left\{ \begin{array}{l} \langle w^p, \boxed{\text{Walt}(w^p)}, \text{BiCo}_{pr.na.} \rangle, \\ \langle t^{p'}, \quad , \text{BiCo}_{nlh.pr.t.} \rangle, \\ \langle s^{p''}, \boxed{t^{p''} \subseteq s^{p''}}, \text{BiCo}_{adj.ph} \rangle, \\ \langle t^{p''}, \quad , \text{BiCo}_{adj.pred.t.} \rangle \end{array} \right\}, \boxed{\begin{array}{c} s^{p''} \text{: 's}(w, y) \\ s^{p'} \text{: rab}(y) \end{array}} \right\rangle, \text{BiCo}_{def.desc.} \right\rangle \right\}, \boxed{\quad} \right\rangle$$

Conversion of the Binding Conditions of (39) according to the same principles that we applied to (37) gives rise to (42):

(42)

$$\begin{array}{l} t^3\ t \\ t = t^3 \qquad t^3 = n \end{array}$$

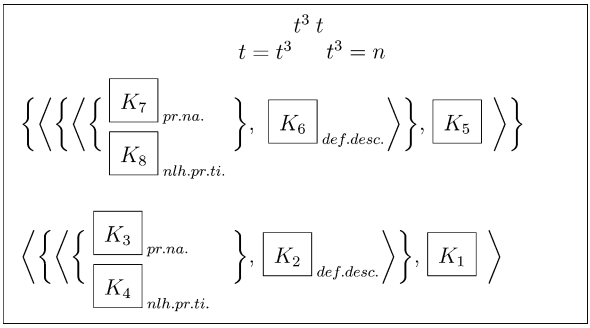
$$\left\{ \langle \{ \langle \left\{ \begin{array}{l} \boxed{\begin{array}{c} w^p \\ \text{Walt}(w^p) \end{array}}_{pr.na.} \\ \boxed{t^{p'}}_{nlh.pr.ti.} \end{array} \right\}, \boxed{\begin{array}{c} s'\ t'\ s''\ \underline{y^p} \\ t' \subseteq s' \\ t'' \subseteq s'' \\ t' = t'' \\ s' \text{: rab}(y^p) \\ s'' \text{: 's}(w^p, y^p) \end{array}}_{df.d.} \rangle \}, \boxed{\begin{array}{c} t^p\ s^p \\ t^p \subseteq s^p \\ s^p \text{: o-t-l}(y^p) \end{array}} \rangle \right\}$$

$$\langle \{ \langle \left\{ \begin{array}{l} \boxed{\begin{array}{c} \underline{w} \\ \text{Walt}(w) \end{array}}_{pr.na.} \\ \boxed{t'}_{nlh.pr.ti.} \end{array} \right\}, \boxed{\begin{array}{c} s'\ t''\ s''\ \underline{y} \\ t'' \subseteq s' \\ t'' = t' \\ t' = t'' \\ s' \text{: rab}(y) \\ s'' \text{: 's}(w, y) \end{array}}_{df.d.} \rangle \}, \boxed{\begin{array}{c} s \\ t \subseteq s \\ s \text{: o-t-l}(y) \\ t^p \prec t \end{array}} \rangle$$

The bracketing structure of this representation takes a while to unravel, so let me try to give an informal explanation. (To this end (43) repeats (42) in a schematic form, in which the different components of (42) have been replaced by labels $K_1, \ldots, K_8$.)

[41] This time I have given the condition that *y* is a rabbit in its time-dependent form. This is not because of a sudden concern with the possibility that what is a rabbit at one time need not be one at another, but to allow for greater transparency of the the process of temporal binding to which interpretation of the subject term of (25) gives rise: the time t'' of the prenominal possessive predication gets identified with the time t' of the predication expressed by the lexical head *rabbit*, while t' can be bound either to the location time of the corresponding event or to the utterance time *n*.

(43)



The non-presuppositional part of (43) is the DRS K_1; all else is presupposition. The top part of the diagram in (43), consisting of K_5–K_8, represents the presupposition generated by *again*. The non-presuppositional part of this presupposition is K_5; it is preceded by a single presupposition whose non-presuppositional part is K_6. K_6 is in turn preceded by a set consisting of two presuppositions, K_7 and K_8. The *again*-presupposition as a whole is attached to the structure consisting of K_1–K_4, which contains additional presuppositions, given by the part consisting of K_3, K_4 and K_2. The structure consisting of K_1–K_4 mirrors that of the *again*-presupposition: It assigns to K_1 a single presupposition, whose non-presuppositional part is K_2, preceded by the pair of presuppositions K_3 and K_4.

What is to be done with the presuppositions of (42)? Let us assume once more that (25) has been uttered in a situation in which the hearer's context is initially empty. In such a situation all presuppositions need accommodation, except for those connected with the predication times t' and $t^{p'}$. To see what representation will result when the remaining presuppositions are accommodated, we need to reflect on the justification principles for presuppositions generated by proper names and by definite descriptions.

First the presuppositions associated with w and with w^p. Though we are now dealing with two presuppositions connected with *Walter* rather than one, justification of these presuppositions (whether by direct verification or accommodation) must nevertheless be to a single individual which the context contains (or is assumed to contain). This requirement is a reflection of the now generally acknowledged fact about proper names that they do not refer through unique satisfaction of descriptive content, but by some other 'direct' reference mechanism. Given this constraint, we can be sure that the two eventualities of a rabbit being on the loose, the asserted one and the presupposed one, concern the same proprietor Walter.

The second point has to do with the presuppositions for y^p and y. In the context of this paper this point is of greater interest; and it is of greater interest precisely because the reference conditions for definite descriptions are different from those for proper names. Definite descriptions need not refer directly, but can also refer via unique satisfaction of their descriptive content (usually within some contextually demarcated domain).

For the two representations derived from the definite description *Walter's rabbit* that are contained in (42) the effect of descriptive reference is connected with the Binding Conditions of the predication times t' and $t^{p'}$ As we observed earlier, the presuppositions connected with t' and $t^{p'}$ can be 'internally' resolved, by identifying these variables with times represented within the preliminary representation itself. In fact, when discussing the 'narrow scope' interpretation of (26) I claimed that identification with the location time of the main eventuality was something of a default strategy for such predication times. In the setting presented by (42), however, it isn't entirely clear what this comes to. For t' only one internal identification is possible, that with the location time t. But what are the options for $t^{p'}$?

As far as I can judge, there are two possibilities in this case, (i) identification with t and (ii) identification with the location time t^p of the presupposed eventuality.[42] The possibility that is of special interest is identification with t^p. This identification has the effect that the two representations deriving from the subject NP have distinct descriptive contents—the one which contains t'' being to the effect that the rabbit in question is Walter's now while the one containing $t^{p'}$ is to the effect that the relevant rabbit was Walter's at the time t^p of the presupposed eventuality s^p. This difference in descriptive content makes it possible for the two representations to determine different denotations, (i) Walter's present rabbit and (ii) the rabbit Walter had at the earlier time t^p. Thus we get the effect that on the wide scope reading of *again* (25) can be understood as involving two distinct rab-

[42] This gives a maximum of 6 options in toto, 2 for t' times 3 for t^p. The two options for t'—identification with t and identification with n—come to the same thing in this case, and the same is true of two of the three options for identifying $t^{p'}$—with t and with n. As a matter of fact, I don't believe that all six options are admissible. It is well-known, in particular from the work on ellipsis, that the interpretation of duplicates (such as the one that was created in representing the *again*-presupposition in (42)) is generally subject to stringent parallelism constraints. Parallelism seems to play a role in the case before us, in that an utterance-time-bound interpretation of one of the duplicates requires the same binding for the other duplicate. With this constraint only three options remain of the original six. A further possible option—but it is one of which I am not sure that it really exists—is that where both t' and $t^{p'}$ are identified with t^p.

bits. In this (25) resembles what we saw in connection with the wide scope reading of (26), but the analyses that the two examples have been given here should have made it clear that the way in which this possibility arises for (25) = (36) is significantly different from the way in which it arises for (26).

It follows from what has been said that the wide scope reading is also compatible with the scenario of a single rabbit that has escaped twice. In fact, this compatibility can take different forms. First, the identification of $t^{p'}$ with t^{p} does not exclude the possibility of a single rabbit. If, as a matter of fact, Walter has had only one rabbit throughout the period which includes both the presupposed eventuality and the asserted one, then that rabbit will be the unique satisfier of both descriptions. Moreover, the options for resolving t' and t^{p} as both identified with t or n also lead to a unique rabbit, though this time as a matter of semantic necessity rather than real world contingency. For each of the two combinations: $t' = n, t^{p'} = n, t' = t, t^{p'} = t$[43] leads to identical descriptive contents for the two duplicates (according to either the denotation must be Walter's at n).

Note that there is nevertheless an important difference between these last two options and the one discussed above. The last two allow for the possibility that the rabbit in question wasn't Walter's at the time of the presupposed state of having escaped. (E.g. the rabbit first belonged to Susie and drove her nuts by its habit of getting out of its compound. So she gave it to Walter. But now the rabbit has done it again.)

There remains one further possibility, mentioned at the end of footnote 42: both t' and $t^{p'}$ are identified with t^{p}. This option would entail that the unique rabbit involved in both escapes was Walter's at the time of the first, but need not be his any longer now. While I think that there are contexts in which (25) allows for such an interpretation, I am uncertain that the interpretation strategy under discussion is really involved there. For if the context contains explicit information about an earlier escape by a rabbit of Walter's, then the NP *Walter's rabbit* might also be interpreted as anaphoric to the rabbit to which that information points.

This last observation shows how extraordinarily flexible interpretation is even for so simple a sentence as (25). Not only do we have to reckon with the possibilities which have been explicitly discussed in this section. As soon as definite descriptions come into play, we have to worry in addition about

---

[43] For the range of options see the previous footnote. As a matter of fact, even the options which were excluded there, will all lead to the same conclusion as the two options considered in the text. In particular, the dubious option: $t' = t^{p'} = t^{p}$ would lead to a single rabbit.

the possibilities that arise when these are interpreted as trans-sententially anaphoric. (And similar additional options are likely to arise for sentences containing other types of anaphoric expressions.) Thus there is at least one moral we can draw from this exploration of the principles of presupposition computation. That natural languages contain a very large amount of lexical ambiguity is by now a commonplace. Computational linguists have often and rightly stressed that this constitutes one of the principal challenges for a computationally viable account of compositionality. The experience with syntactic parsing has shown, moreover, that the ambiguity problem for parsers which operate on the basis of syntactic information only is hardly less daunting. What we have seen in this section suggests that at the level of semantics the range of structural ambiguities is very considerable as well, and significantly larger than some of us suspected. In the light of this the ability of human interpreters to understand what others are trying to tell them seems an even greater marvel than the one we knew it to be already. Formulating and implementing algorithms which emulate this extraordinary ability is, it seems to me, the central problem which computational linguistics has to solve. It is among the tasks of linguistic theory to identify the multiple ways in which ambiguities arise and the different forms they take. But that, surely, is only the lesser part of the story.

5. Conclusion

In more than one way this paper is just a beginning. In the light of all that has been said about presupposition over the past thirty years and the progress that has been made especially during the past decade, this may seem an odd thing to say. In fact, I hope I have managed to make clear to what extent the theory sketched here builds on what has been accomplished in the course of those many years by others. Nevertheless, the analyses presented in this paper cannot pretend to more than being pointers in a certain direction. Even if one assumes that the direction is right, it will have become evident that much more will have to be done before the sketch that has been offered here could be called a theory.

Such a theory should offer, for a non-trivial, well-defined part of at least one natural language that includes a fair sample of different types of presupposition triggers: (i) a general account of syntactic structure; (ii) a general definition of the "syntax" (i.e. the possible forms) of preliminary representations; (iii) a formulation of the rules which enable us to construct preliminary representations from syntactic structures; and (iv) the principles which

guide and constrain the justification of (the presupposition of) preliminary representations in context. Particularly urgent is a detailed articulation of the different Binding Conditions, which in the examples presented here only appear as labels, accompanied at best by informal descriptions of the actual binding rules they stand for. (This of course is a task which concerns not just the theory of presupposition computation I have outlined, but the over-all account of the syntax-semantics interface in which the theory is embedded.)

In addition, semanticists of a model-theoretic persuasion may want to see a formal semantics for the structures that are postulated by the account of presupposition I have sketched, in the first place for the preliminary representations, and, if possible, also for the node representations which arise in the course of building the preliminary representation from the syntactic analysis of the sentence. Unfortunately, the possibility of such a semantics is limited. To define a syntax of preliminary representations (for some reasonable fragment of English), which characterises them as the expressions of a given representation formalism (or as data structures of a certain form) is not too difficult. Moreover, for those preliminary representations (according to such a syntax) in which all presuppositions appear in the highest possible position (so that their local context coincides with the global context of the entire sentence) 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. Similar obstacles stand in the way of a plausible model theory for the node representations with which syntactic trees are decorated in the course of representation construction.

From a computational perspective this does not appear as a serious drawback. For the justification of presuppositions has to be subject in any case to rules which we can't expect to be derivable from the model-theoretic values of preliminary representations. (For instance, these rules must distinguish between the justification of "anaphoric" and that of "non-anaphoric" presuppositions. It is hard to see how such a difference could be recovered from the semantic values that are assigned to anaphoric and non-anaphoric presuppositions). But on the other hand there are many more computational issues that have to be addressed before a theory of the kind I have outlined can be of practical use to computational linguistics. To name just one such issue: The theory of presupposition will have to be integrated into a general theory of underspecified representation. I mention this matter, as it is one to which a fair amount of thought has already been given, for even in this

case we are still at a considerable distance from seeing our way through to a systematic solution. Other computational questions, e.g. those which have to do with the kind of inferences used in anaphora resolution and in presupposition justification generally, haven't made it to centre stage yet. But we may be sure they are waiting in the wings.

Acknowledgments

Special thanks are due to Kurt Eberle, Tim Fernando, Uwe Reyle, Michael Schiehlen, and, especially, to Peter Krause, who has implemented a significant part of the theory presented here and to Antje Roßdeutscher, without whom I might never have arrived at the general view of presupposition I now hold, nor managed to turn that perspective into a contribution to this volume.

References

Abusch, D., 1994. The scope of indefinites. *Natural Language Semantics* 2, 83–135.

Enç, M., 1981. *Tense without Scope*. Ph.D. dissertation, University of Wisconsin, Madison, WI.

Heim, I., 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. dissertation, University of Massachusetts, Amherst, MA.

Heim, I., 1983. On the projection problem for presuppositions. In: M. Barlow, D. Flickinger (eds.), *Proceeding of the Second West Coast Conference on Formal Linguistics*. Stanford University Press, Stanford, CA, 114–125.

Kamp, H., 1981. A theory of truth and semantic representation. In: J. Groenendijk, T.M.V. Janssen, M. Stokhof (eds.), *Formal Methods in the Study of Language, Proceedings of the Third Amsterdam Colloquium*. Mathematical Center, Amsterdam, 277–322. Reprinted in: J. Groenendijk, T.M.V. Janssen, M. Stokhof (eds.), *Truth, Interpretation and Information, Selected Papers from the Third Amsterdam Colloquium*. Foris, Dordrecht, 1984, 1–41, and reprinted in: K. von Heusinger, A. ter Meulen (eds.), *Meaning and the Dynamics of Interpretation. Selected Papers of Hans Kamp*. Brill, Leiden, 2013, 329–369.

Kamp, H., 2000a. *Representations of Propositional Attitudes and Indirect Discourse*. Tech. rep., University of Stuttgart, Stuttgart.

Kamp, H., 2000b. *Wide Scope and Intermediate Scope Interpretations of Definite and Indefinite Noun Phrases*. Tech. rep., University of Stuttgart, Stuttgart.

Kamp, H., 2001. Presupposition computation and presupposition justification. In: M. Bras, L. Vieu (eds.), *Semantic and Pragmatic Issues in Discourse and Dialogue: Experimenting with Current Dynamic Theories*. Elsevier, Amsterdam, 57–84. Reprinted in: K. von Heusinger, A. ter Meulen (eds.), *Meaning and the Dynamics of Interpretation. Selected Papers of Hans Kamp*. Brill, Leiden, 2013, 425–454.

Kamp, H., Reyle, U., 1993. *From Discourse to Logic. Introduction to Modeltheoretic*

Semantics of Natural Language, Formal Logic and Discourse Representation Theory. Kluwer, Dordrecht.

Kamp, H., Roßdeutscher, A., 1992. *Remarks on Lexical Structure, DRS Construction and Lexically Driven Inferences*. Arbeitspapiere des Sonderforschungsbereiches 340, Sprachtheoretische Grundlagen der Computerlinguistik, Vol. 21. University of Stuttgart, Stuttgart.

Kamp, H., Roßdeutscher, A., 1994. DRS-construction and lexically driven inference. *Theoretical Linguistics* 20 (2/3), 165–235.

Karttunen, L., Peters, S., 1979. Conventional implicature. In: C.K. Oh, D. Dinneen (eds.), *Syntax and Semantics 11: Presupposition*. Academic Press, New York, 1–56.

Langendoen, D., Savin, H.B., 1971. The projection problem for presuppositions. In: C. Fillmore, D. Langendoen (eds.), *Studies in Linguistic Semantics*. Holt, Rinehart & Winston, New York, 55–62.

Musan, R., 1995. *On the Temporal Interpretation of Noun Phrases*. Ph.D. dissertation, MIT, Cambridge, MA.

Reyle, U., Roßdeutscher, A., Kamp, H., 2000. Ups and downs in the theory of temporal reference. In: U. Reyle (ed.), *Presuppositions and Underspecification in the Computation of Temporal and other Relations in Discourse*. Arbeitsberichte des Sonderforschungsbereichs 340, Sprachtheoretische Grundlagen der Computerlinguistik, Vol. 164. University of Stuttgart, Stuttgart.

Roßdeutscher, A., 2000. *Lexikalisch gestützte formale Textinterpretation*. Arbeitspapiere des Sonderforschungsbereichs 340, Sprachtheoretische Grundlagen der Computerlinguistik, Vol. 157. University of Stuttgart, Stuttgart.

van der Sandt, R., 1992. Presupposition projection as anaphora resolution. *Journal of Semantics* 9, 333–377.

von Stechow, A., 1996. The different readings of "wieder" (again): A structural account. *Journal of Semantics* 13, 87–138.

Tonhauser, J., 2000. *Dynamic Semantics and the Temporal Interpretation of Noun Phrases*. Master's thesis, Institute for Computational Linguistics, University of Stuttgart.

Zeevat, H., 1992. Presupposition and accommodation in Update Semantics. *Journal of Semantics* 9, 379–412.

PART SIX

PROPOSITIONAL ATTITUDES

INTRODUCTION TO PROPOSITIONAL ATTITUDES

Hans Kamp

The most important of my formative years were without doubt the three I spent as a Ph.D. student at UCLA (1965–1968). As it must be for most people who end up spending their lives in academia, during those three years I came to feel and see myself as a member of a scientific community. That community may have shifted somewhat in the course of my subsequent career, but my sense of allegiance to what I perceive as my intellectual home remains as crystallized during my days as a graduate student in Los Angeles. The community into which I was absorbed, when I came to UCLA, was roughly what later (mainly, I believe, in the seventies and eighties) was sometimes referred to as 'Southern California Semantics'. In my recollection this 'community', if that is the right word for it, was never very large, but its impact has been very considerable, mainly through the work of Richard Montague, David Kaplan and David Lewis. For me, its impact was enormous and invaluable. It may seem strange for me to make these rather personal statements in the Introduction to one of the parts of this book, but there is a reason. It is this. One of the credos of this community can be fairly described, I think, as anti-psychologism. It was intellectually, and perhaps also emotionally, connected with a commitment to formalization as the royal road to conceptual clarity and philosophical salvation. That attitude wasn't, by the way, limited to philosophy at UCLA, or to the loose affiliation of logicians and philosophers pivoting around UCLA's Philosophy Department; it also included philosophical logicians of an extensionalist persuasion such as Willard van Orman Quine or Nelson Goodman, and arguably also Saul Kripke. The issue on which this anti-psychologism left its deepest mark was the semantics of attitude reports. Here its influence was what I came to see later as inhibitive, a moratorium on all attempts to say anything about this topic that smacked—if ever so slightly—of irresponsible speculation about the structure and content of the human mind. Part of the reason for this essentially negative positioning was no doubt a strongly felt resistance against 'psychologistic' attempts to deal with the semantics of attitudes and other vocabulary, such as the Semantic Marker theory of Jerrold Katz and Jerry Fodor, widely discussed at that time. To some extent that resistance was understandable: to many members of the SouCalSem

© HANS KAMP, 2013 | DOI:10.1163/9789004252882_020
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

community there was much in the Katz & Fodor account that they saw as question begging (as well as technically inept). But the emphatic rejection of approaches like this one reinforced an attitude towards issues situated on the borderline between language and mind that now strike me as curiously bloodless. It took a good part of my professional life before I would even admit such a thought and a good deal more to arrive at what I have come to see as a potentially viable alternative to those bloodless accounts, which accompanied me during those final years of intellectually growing up, and for many more years beyond. The two papers in this part are part of the larger effort to develop that alternative and are, indirectly, also witnesses to the struggle I had to make to free myself from these earlier dogmata.

From the perspective of that alternative, the accounts of the semantics of attitude reporting sentences that the SouCalSem community and others developed and advocated do not only seem rather anaemic and noncommittal; there also is a curiously defensive streak to them. The historical grounds for that are not hard to find. The problem one was facing was essentially Gottlob Frege's, who had appealed to the distinction between sense and reference as a way out of what could be called the 'equivalence trap': if your account of the semantic values of sentences (and other expressions) is too coarse, then there is no hope of getting the semantics of attitude-reporting sentences right. The argument is overly familiar: suppose that the sentences S_1 and S_2 get the same semantic value for any notion of semantic value your theory can supply. Then your theory must assign the same semantic values to *x believes that* S_1 and to *x believes that* S_2—if it is assumed that the semantic value of *x believes that S* is fully determined by the semantic values of *x* and *S*; but compositionality was taken for granted. So if the semantic values aren't fine-grained enough, S_1 and S_2 cannot be distinguished in terms of their values, even when it seems intuitively possible for *x believes that* S_1 to be true and at the same time for *x believes that* S_2 to be false. If the theory is to account for that possibility, it has to come up with values that make finer distinctions to S_1 and S_2 semantically apart. Developing well-motivated accounts of semantics that deliver values to individuate finely enough has proved very hard. A proof of how hard was involuntarily supplied by Alonzo Church in his first attempt to develop a mathematically precise formalization of Frege's sense-reference distinction, which, in spite of Church's proverbial care and precision, failed through an unintended collapse into the purely extensional.

It is partly in the light of such failures that Montague's Higher Order Intensional Logic (HOIL) can be seen for the very important achievement it

was. Here at last was a workable formalism, which captured what seemed to be most of the essence of Frege's distinction. It achieved this in a comparatively simple and surveyable way, providing philosophy and linguistics with a useful and much needed formal framework. But even so, it should not be forgotten that Montague's HOIL was a compromise from the start: it didn't ward off all unwanted identifications of the truth conditions of intuitively non-equivalent attitude reports. And so the equivalence trap kept nagging us, leading to all manner of attempts to develop hyper-intensional frameworks, in which the semantic values of expressions are more finely individuated than by intensions. None of these attempts seem to have been successful enough to end the debate. But what I want to stress is that, whether successful or not, all such attempts are in essence defensive; they all are attempts to escape the predicament of identification errors that are made by existing accounts and that those existing theories make because the principles that are built into them force them to make predictions about the semantics of attitude reports that are too strong. The central aim always is to find a way to *avoid* such commitments and, by doing so, to weaken the semantics of attitude reports even further.

DRT suggests a radically different view of these matters, because, unlike other approaches to formal semantics, it is built on the conception that a theory of linguistic meaning cannot be separated from a theory of language interpretation. DRT tries to give substance to this conception through the different things it has to say about the role, structure and semantic values of what it uses as logical forms, viz. its DRSs. On the one hand, DRSs have a well-defined syntactic structure, which determines their semantics—more specifically, their truth conditions—in the same systematic and transparent way that is distinctive of the formulas of predicate logic and the terms of HOIL (more specifically: of those terms that are of the type *t* of truth values). Moreover, DRT has ways of spelling out in precise detail how to assign DRSs to the sentences of a given natural language or natural language fragment as their logical forms, which capture the intuitively correct truth conditions of those sentences. Furthermore, DRS assignment proceeds systematically, by exploiting the syntactic structure of the sentences in much the same systematic way as in Montague Grammar. But this isn't all. The structure of a DRS is designed in such manner that it can also play a further role, that of *discourse context*: after a DRS has been assigned to one sentence as the representation of its content, it can serve as context in the interpretation of the following sentence. Here too the contributions that DRSs make are systematic: certain words, morphemes and syntactic constructions systematically exploit the structure of DRSs as discourse contexts, when relating the content of

the sentence that contains them with that of the preceding sentence, whose content is represented by the DRS that currently acts as context. When the DRS that has thus been obtained for the new sentence is then added to the DRS, that just served as discourse context. The resulting DRS will be a representation for the two sentences combined, capturing the semantic connections between them. In this way it is possible to build logical forms of cohesive sentence sequences in a systematic, step-by-step process. The picture that emerges—of a compositional sequential process, that assigns logical forms to successive sentences of a discourse or text, where each logical form will, once it has been constructed, go on to play an instrumental role in obtaining the logical form for the next sentence—is strongly suggestive that something very much like it must also go on when texts and discourses are processed by us humans. True, as it stands, this is only a suggestion, not anything that the theory gives us a clear license to infer. Apart from that, it is also far from clear what the suggestion exactly comes to: Which structural properties are psychologically real and in what sense, or at what level of psychological reality? Adopting the hypothesis that DRSs provide a good model for the mental representations that human interpreters construct from the linguistic input they get is hence clearly a matter of taking a plunge. The plunge will be even deeper, when we assume that DRSs are not only good models for representations so constructed out of linguistic input, but for the thoughts entertained by human agents more generally. For someone who sees the DRT approach to natural language semantics as basically correct, and also as so far remarkably successful in its detailed implementations, the plunge seems well worth taking. That plunge was the starting point for the project of which the two papers in this section are part.

In order to understand why and how this project differs, as much as it does, from other approaches to the semantics of attitudes, one point deserves special emphasis. It has to do with the nature of the evidence that DRT provides for the hypothesis that DRSs are good models for the structure of human thought. That evidence is independent from anything that pertains to the semantics of attitude verbs in any direct way. None of the natural language fragments to which DRT was originally applied contain any attitude verbs or such predicates. So none of their sentences are attitude reports. So the mental representations that our hypothesis is making claims about are, in the first instance, representations of thoughts that can be expressed by sentences *without* attitudinal predicates—that is, by those sentences that we find as complement sentences in the simplest, 'first level' attitude report sentences, in which no attitude predicate occurs in the scope

of another one. That means we can use a theory of the structure of thoughts built on the basis of our hypothesis as input to an account of the semantics of attitude reports.

In this way the problem of the semantics of attitude reports can be factored into two parts: (i) the development of an account of the structure and content of propositional attitudes and complex mental states and (ii) the formulation of the truth conditions of attitude reports, based on (i). The implementation of (i) that I and others are currently using, takes the form of a specification of a certain description language for mental states with a well-defined syntax and semantics. To make use of (i) in the execution of (ii) we need to bring the descriptions it provides of propositional attitudes and mental states into the DRSs, serving as semantic representations of attitude attributing sentences. In the version of the theory we are using, this has been achieved by adding to the vocabulary of the DRS formalism a special predicate, 'Att', with three arguments, one for the agent, whose attitude or mental state is being described, one for the kinds of descriptions of mental states that are provided by our implementation of (i), and a third slot, whose function doesn't matter for the present exposition. The central task for the account of the truth conditions of attitude reports is then to spell out the satisfaction conditions for DRS conditions in which 'Att' is the main predicate. Given how far we have come at this point, it seems only a small further step to admit DRSs that contain occurrences of the 'Att' predicate as thought representations as well. The thoughts modelled by these representations are attributions, of some attitude or mental state, the given agent makes to some other agent. This adds to the system the power of recursion it needs to deal with iterated attitude attributions and with the semantics of multiply nested reports of the *He thinks that she thinks that … that so-and-so believes that p* sort. The DRSs of such reports look a bit like *matryoshka*, the nested Russian dolls, or like 'Russian boxes', a less endearing remake of the beloved original.

The paper *Prolegomena to a Structural Account of Belief and Other Attitudes* was written at a time when much of the general picture it outlines was still much more of a research program than it is today. (I am leaving out of consideration here Asher's two impressive—and earlier—papers *Belief in Discourse Representation Theory* and *A Typology of Attitudinal Verbs and Their Anaphoric Properties*, which take a rather different line on some of the central issues that *Prolegomena* focuses on than is done here.) The *'Prolegomena'* in the title of the paper has found a certain justification in subsequent developments: In the meantime a theory has been put in place to which the paper can be seen as prolegomena to. While some of the

suggestions of the paper have since been worked out in explicit formal detail, other promissory notes are still to be made good. First, the model-theoretic semantics for the descriptions that the expanded DRT-formalism makes available for the description of mental states is an intensional one, and hence suffers from most of the shortcomings vis-à-vis the logical omniscience problem inherent in HOIL. In particular, the semantics provides no resolution for the particular logical omniscience problem discussed in *Prolegomena*'s opening pages. This problem is a notoriously hard and deep one, still crying out for a solution at a pitch as loud as ever. I do not think it can be solved just with the current methods of formal semantics, whether with DRT or without. What appears to be needed is a new alliance between model theory and proof theory; but that alliance has yet to be forged. The second promissory note concerns the intersubjectivity of reference in the absence of real referents, with Hob-Nob sentences as its earliest and arguably still most dramatic illustrations. The more recent literature has made plain, the intersubjectivity problem already arises at the level of verbal communication, in the semantic analysis of dialogue, and in other multi-speaker forms of verbal exchange. It is perfectly possible and can be entirely natural for two speakers to talk about one and the same thing, even when that thing only exists in their imagination, or is a hypothetical construct for the purposes of the given discourse. A truly satisfactory account of the nature of such kinds of purely intersubjective coreference is, to my knowledge, still outstanding.

The second paper in this part, *Temporal Reference Inside and Outside Propositional Attitudes*, is not only the last paper of the book, but also its most recent one. It dates from a time when DRT had not only been a well-defined framework for several decades, but had undergone a number of transformations. The two most important transformations in connection with this paper—the two-stage treatment of presuppositions introduced by Rob van der Sandt and Bart Geurts and the extension of DRT to a theory of propositional attitudes and a framework for the semantic analysis of attitude reports—have both been mentioned already, the first in the introduction to the part on presupposition and the second in the last few paragraphs above. The concerns that prompted these two changes in DRT come together in attitude reports whose complement clauses trigger linguistic presuppositions. As Robert Stalnaker may have been the first to point out explicitly, the interpreter of such reports is sometimes in the quandary of having to decide where to look for a justification of such presuppositions. Are they supposed to hold in the real world or only in the mind of the attributee? One of the reasons why such quandaries arise often and naturally is that

so much of our attitude reporting is not restricted to single sentences, but takes the form of little discourses on some individual's state of mind and, often, his changes of mind over time. In such discourse a picture of the state or states of mind is typically built up stepwise and cumulatively, much as we do when we use a fleet of sentences to draw up a picture of other aspects of our world. The interpretation of such discourses, Stalnaker saw, requires that two discourse contexts must be built in parallel. This pair of contexts gives the interpreter more options for justifying presuppositions generated within the complement clauses of attitudinal verbs, especially of those occurring in later parts of a multi-sentence attitude attribution. With these options also comes the predicament of having to choose between them, as different places where a justification of those presuppositions might be found or forced. Much of how interpreters deal with this 'predicament of choice' is still unclear. The DRT approach to the semantics of attitude reports gives a way of seeing more clearly what the problems are, but that, of course, is not the same thing as finding their solution.

Temporal Reference Inside and Outside Propositional Attitudes explores one aspect of this issue, but with a twist. Many complex attitude reports are not just about the subject's mental state at one particular time, but about how his propositional attitudes change as time goes on, i.e. about transitions from one set of attitudes to another one which in one or more ways is incompatible with its predecessor. When a later sentence *S* of such a complex attitude report, which has something to say about a later stage of the subject's mind, carries a presupposition, then the interpreter's difficulty is not only that of deciding where this presupposition should be satisfied—in relation to the outside world or the subject's state of mind—there also is another question that confronts him: Is it still possible, when trying to justify a presupposition by appeal to the agent's mental state at the later time *t*, to rely on the information that earlier parts of the discourse supplied about mental states the attributee was in at some time before *t*? Can this information still be assumed to apply to the attributee's state at *t* as well? There appears to be some evidence that, when an attitude reporting discourse gives rise to this kind of question, there is an implicit understanding that information provided earlier in the discourse about the subject's earlier state of mind does persist to the later time. If that were not so, the speaker should have given a clear indication of it. Like some other papers in this collection, among them the two in Part Five on Presupposition, this paper is exploratory in nature. It investigates one example in considerable detail, but with its eye on the more general phenomenon of which it is an example. An account of the general phenomenon is not attempted, nor, for that matter,

is there any attempt at compiling a longer list of telling examples. All that is left for future work. Here too the tasks appear to be as future now as they were, when the paper was written.

PROLEGOMENA TO A STRUCTURAL ACCOUNT OF BELIEF AND OTHER ATTITUDES[a,1]

Hans Kamp

1. Introduction

Calvin and Kevin, friends, and as sensible as boys of ten can be expected to be, are dropped by Kevin's father at the entrance of Disneyland. He makes them promise to be back at the entrance by 5:30, when he will come to collect them, and gives each twenty dollars for the afternoon. The lads stick together and spend their money on exactly the same rides. By 5:20 they are back at the appointed spot, tired and hungry. Both remember exactly which rides they took and what each ride cost.

Near the place where they are waiting for Kevin's father there is a hamburger stand. Hamburgers, a sign informs the potential costumer, can be had at $1.50 apiece. Both boys feel very much like a hamburger. But only one of them, Kevin, goes up to the stand to order one. When he wants to pay for the hamburger that the attendant hands him, he finds to his mortification that he has only 60¢ left and thus cannot pay for it. Calvin, who wants a

[1] This paper consists mostly of points and ideas that can be found in earlier work, most of it unpublished and some of it dating back to 1980, in which I have made various attempts to come to grips with the problems touched upon below. I have included some of the relevant papers and manuscripts in the bibliography. In recent years a number of papers by other authors have appeared, most notably by Nicholas Asher (1986, 1987) and Henk Zeevat (1989), which contain many of the proposals I will be making here too. As the similarities with their papers are numerous and obvious, it would be both onerous and tedious to point them out individually at every turn. My indebtedness to these authors should be evident however to anyone familiar with their work. I record it here once and for all.

[a] Editors' note: This article was originally published in C.A. Anderson, J. Owens (eds.), *Propositional Attitudes. The Role of Content in Logic, Language, and Mind*, CSLI, Stanford, CA, 1990, 27–90. We would like to thank the Center for the Study of Language and Information (CSLI) for its kind permission to reprint the article. We carefully edited the article and corrected some typos. We would like to thank Tyler Wilbers for careful proofreading and constructive comments.

© hans kamp, 2013 | doi:10.1163/9789004252882_021
This is an open access chapter distributed under the terms of the cc by-nc-nd 4.0 license.

hamburger just as badly, spares himself this embarrassment. He never goes up to the stand, for, unlike Kevin, he *knows* he has only 60¢ left.

There is nothing very remarkable about this story. Calvin and Kevin behave differently, but the difference is readily explained: Calvin is aware of something of which Kevin is, at the crucial moment, not aware. And the thing of which he is aware keeps Calvin from doing that which gets Kevin into trouble. Both act rationally, and their desires coincide. That, nevertheless, there is no coincidence of action is because the boys differ in what, at the crucial time, they are conscious of.

Indeed, Kevin's behavior is no less *rational* than that of his friend. The only point on which one might fault him is that he *could* have realized he had only 60¢ left, either by checking his pocket, or—and that is the consideration that matters here—by recalling what he paid for each of the rides he took and doing the necessary sums. For, as we assumed, both he and Calvin know exactly what they spent when and where, and a sensible boy of ten can, if he puts his mind to it, surely figure from this what is left of the amount with which he set out.

To the unbiased observer this account of the difference in their behavior would seem to be perfectly straightforward: One of the boys has the belief that he doesn't have the money to pay for a hamburger, and the other lacks such a belief. Surely a decent theory of belief should be able to make sense of such an explanation and of the role belief plays in it. A theory according to which the doxastic difference between Kevin and Calvin, which is needed in the explanation of their behaviour, does not really exist is surely too much at variance with our apparent use of the concept of belief to be acceptable!

Theories that analyze belief as a relation between the believer and a set of possible worlds (the set of possible worlds in which any sentence expressing the belief is true) are unacceptable in just this way. To be precise, if we attribute to Kevin the complex belief (call it B_1) that he started out with $20, spent x¢ on his first ride, y¢ on the second and so on, and identify this belief with the set of all possible worlds in each of which all this is true, then this set of possible worlds is exactly the same as the one in which all these things are true and in which, *moreover*, he has only 60¢ remaining (let us refer to this second proposition as B_2). For the conjuncts that make up B_1 logically entail that Kevin has only 60¢ remaining and so this last proposition is true in just those worlds which also verify B_1. Thus, if Kevin does have the belief B_1, he also has the belief B_2. For, according to the possible worlds account, B_2 *is the very same belief* as B_1. But B_2 is (*mutatis mutandis*) just the belief which, as I have told the story, kept the sagacious Calvin from going up to the

stand. So, if we are to trust the possible worlds account, there is no difference between the boys' beliefs that could explain why there is a difference in their behavior.[2]

The point is an overly familiar one: Theories of belief that identify beliefs with sets of possible worlds cannot differentiate them finely enough to do justice to our common understanding and use of the notion. As it is most often put: Possible worlds accounts entail that belief sentences—i.e. sentences of the form "*A* believes that *s*"—are truth-invariant under substitution of necessarily equivalents for the embedded clause *s*, and this does not appear to be the way in which belief sentences are actually understood. The trouble which the possible worlds account encounters with the case presented here is arguably of just this sort. For, as we have just seen, it boils down to the identification of B_1 and B_2, whereas it is precisely the difference between those beliefs that is essential to the explanation of why the two boys behaved differently. Nevertheless I feel that the case presented here illustrates the predicament in a particularly striking way. Let me explain.

Advocates of the possible worlds approach often argue against theories which identify beliefs with sentences of a natural language (or with "syntactic" structures belonging to some other formalism) on the ground that belief is *intentional*. Belief, it is argued, is always directed at a certain domain or subject matter, most commonly at the world in which the believer himself exists. The intentionality of belief (and other propositional attitudes) is especially important if we think about the relation between belief and action. Our beliefs can guide our actions in and on the world precisely because and insofar as those beliefs pertain to that same world. It is by virtue of the information they carry about that world that they enable us

[2] The difficulty which such cases present for the possible worlds approach can be brought out in a number of related ways. For instance, we could marginally extend the story by specifying that at the time when he goes up to the hamburger stand Kevin believes that he has enough money to pay for a hamburger. (As I told the story, it isn't clear that this belief should be ascribed to Kevin; it may be that he simply did not consider the question whether there was enough left). This belief is incompatible with the (true) beliefs he had about the money he began with and his various expenses during the afternoon. So the combination of these beliefs corresponds, according to the simple possible worlds account, to the empty set of worlds. Alternatively, suppose that Kevin almost went up to the stand, but, at the last moment, ran through the calculation which showed him that he wouldn't be able to pay; and so he checked himself just in time. There is a natural inclination to say of this case that what decides him against going up to the stand is a change in his beliefs. But according to the possible worlds theory there has been no change: For strictly speaking it does not allow for the possibility of acquiring new beliefs by means of deductive reasoning only; valid deduction never changes the set of worlds to which our beliefs commit us.

to infer how the actions we contemplate will affect it, and thus they guide us in choosing the action that seems best to us. Moreover, since we have an idea of how such choices are made, we can make predictions about the actions performed by others, provided we know enough about their desires and beliefs; and, similarly, explain some of the actions that we have already observed. Evidently this is one of the main reasons why we are interested in ascribing beliefs to others.

If it is one of the main points of attributing beliefs that it enables us to account for rational action as resulting, via certain processes of deliberation and decision, from those beliefs, then a theory of belief cannot pass unless it gives a proper account of its intentionality. The possible worlds theory does provide such an explanation, it has been claimed, in that it identifies a belief with the set of worlds that are compatible with it. This makes it possible to construe the having of a belief *P* about the actual world as a relation between the believer and the world, a relation which can (perhaps not very helpfully) be described as that of "taking the world to be one of those belonging to the set *P*". In contrast, "syntactic" theories of belief, which identify beliefs with sentences or other syntactic objects, have been accused of failing to explicate the intentional nature of belief.[3]

In a context where beliefs are attributed for the sake of explaining overt behavior their intentional character would thus appear to be of paramount importance. Yet, as we have seen, in the context of our example the identification of beliefs with sets of possible worlds causes the relevant explanation to collapse. What is needed therefore is a theory of belief which does justice both to (i) the intentionality of beliefs and (ii) to the fact that their identity conditions are stricter than those of the corresponding sets of worlds.

The second of these requirements is closely connected with a point which has been made, urgently and repeatedly, by Jerry Fodor (1987). Attitudinal states, Fodor has argued, must be computational in the sense of having some kind of finitary structure on which the inferencing mechanisms of the ratiocinating believer can operate. The importance of this point becomes evident as soon as we reflect on what a detailed theory might be like of the processes

[3] When it is addressed to theories which offer a merely syntactic analysis of belief, and have nothing to say about the semantic connections between the structures in terms of which belief is analyzed and the world, this accusation is justified. But no such objection can be leveled at theories which combine proposals concerning the syntactic structure of belief with an account of the intentionality of those structures.

by which belief guides action. The central part of such a theory ought to be an account of practical reasoning—an account, that is, of the cognitive dispositions which permit rational agents to form intentions in the light of their beliefs and assumptions, and of the ways in which these dispositions are applied to the task of determining actions that, within a realm of possibilities defined by the agent's beliefs, optimally realize his goals. These dispositions should be conceived as inference principles and rules, like those which we find in formal logic (at least, it is hard to see what else they could be like!). This entails that the items to which these dispositions are applied in individual cases of practical reasoning must, just as logical formulas, have a "syntactic" structure, which the agent's inferencing module can identify as suitable input for particular inference rules.

The desiderata for a theory of belief that are implied by these observations have long been recognized, and they have led to a variety of proposals that circumvent the embarrassments of the simple-minded possible worlds account, while satisfying the intentionality requirement. Within philosophy the most familiar—and probably the most successful—strategy has been that of identifying beliefs with structures of some natural or partly synthetic language for which there exists a truth definition that assigns to each of those structures an *intension*, viz. the set of all possible worlds in which the structure is true, or satisfied. A well-known example of such a proposal is the one due to Rudolf Carnap (1947) and David Lewis (1979b), according to whom beliefs might be identified with what Lewis (1979b) calls *meanings*, i.e. syntactic configurations the end nodes of which are intensions. Another proposal, often dismissed as naive and obviously wrong, but recently revived by, in particular, Igal Kvart, is to identify beliefs with sentences of a natural language such as English, while recognizing the need to explicate the intentionality of that natural language, presumably by providing a truth- or model-theoretic semantics for it.

As the suggestions I will put forward below bear a certain similarity to this last proposal, I want to take time out at this point to make a preventive strike against an argument that is sometimes cited against the identification of propositional attitudes and sentences of natural language. The argument goes back a fair way, and, to my knowledge, was first given by Church (1950). It rests on the simple observation that the beliefs of someone whose only language is e.g. English can as a rule be described just as easily in a language other than English as in English itself. Thus a German speaker could report Calvin's belief with the sentence: "Calvin ist sich bewusst, dass er nur 60¢ in der Tasche hat." Although Calvin himself would not be able to recognize this as a correct expression of something he holds true.

This observation is sometimes thought to show that belief cannot be construed as a relation between a believer and a sentence. But surely that does not follow. Whatever the observation shows, it certainly does *not* show that Calvin's believing that he has only got 60¢ in his pocket cannot consist in his standing in a "belief" relation to a certain natural language sentence. In particular, the observation does not tell against the claim that his belief consists in his standing in such a relation to the English sentence "I have only 60¢ in my pocket". At best, it shows that *if* it is true that believing is, or involves, standing in a certain relation *R* to a sentence; and that *if*, moreover, standing in that relation presupposes that one understands that sentence; then there will be cases in which the sentence to which one stands in the belief relation will not be identical with the complement clause of a statement used to report one's belief.[4]

I will return to this issue later. But before leaving it now, I want to make explicit something that is implicit in what I have just said about Church's argument, and which will be a central presupposition to the attitude theory that I am going to sketch. It is—I will argue this point below—crucial to distinguish carefully between:

(i) The question what form, structure or organization must be assumed for the beliefs and other propositional attitudes that are held by the various organisms capable of such attitudes; and

(ii) The question how an attitude report, i.e. a sentence in which an attitude is *attributed* to such an organism, must be related to the reported attitude for the report to count as correct.

Sometimes these questions are not distinguished; or, alternatively, only one of the two is being addressed. This is true in particular of those philosophical accounts of propositional attitudes which confine themselves to the study of the semantics of attitude reports.[5]

In many instances the refusal to try and do more than develop a semantics of attitude reports is born out of wisdom. At present there isn't much

[4] For reasons that will become evident later I do not think that the simple identification of attitudinal objects with sentences of a natural language like English can produce a theory of propositional attitudes that is particularly useful or plausible. But the argument just mentioned has next to nothing to do with the considerations that have led me to this conviction.

[5] Attitude reports come in a number of different syntactic forms; but the most prominent form in philosophical and logical discussions has been that which consists of (i) a subject term referring to the possessor of the attitude, (ii) a verb denoting the attitude (e.g. *believes, desires, doubts*, etc.) and (iii) a that-clause identifying the content of the attitude form. It is to this type of sentence that I will be referring henceforth when using the term "attitude report".

that we can say with any precision or certainty about the way in which the human mind represents and processes information (not to speak about the minds of the various kinds of animals which we also judge capable of belief and desire, and perhaps other sorts of attitudes as well). And there is little hope that this situation will significantly improve in the foreseeable future. So for the time being theorizing about these matters is bound to be speculative, and devoid of a proper empirical basis; it is something one had better stay away from. As regards the practice of *attributing* attitudes the situation is quite different. People have attributed propositional attitudes to other people (as well as to many kinds of animals) as long as anyone can remember; and those who have engaged in the practice have been no better informed about the inner workings of the mind than we are at present. Nonetheless, attitude attribution is a highly developed practice, about which many of us have firm and detailed intuitions. As such it is a proper and promising subject for philosophical analysis. And since the practice does not seem to require detailed knowledge of the mind's inner workings, a theory of it should have no need to refer to those inner workings either. Thus, a theory of attitude reports ought to be independent of any specific assumptions about the organization of mental states and the mechanisms which transform them.

Up to a point this is right. Yet, the methodological constraints that some have inferred from these considerations are, I think, not really workable—not even when the semantics of attitude reports is declared to be the only going concern. In particular, many of the outstanding puzzles relating to belief attribution—on which much of the recent philosophical debate about belief has focused—cannot be addressed within the austere frameworks to which these constraints condemn us. The difficulties those puzzles have been thought to present are only symptoms of the more fundamental problem we have already identified: Any attitude theory that fails to put forward substantive hypotheses about the structural properties of propositional attitudes thereby prevents itself from saying anything of substance about how attitudes are *cognitively active*. And so it cannot say much either about the ways in which attitudes issue in overt behavior (including that behavior, much of it verbal, against which the correctness of attitude reports is to be measured). Without certain assumptions about attitudinal structure we will never get very far.

I already noted that theorizing about attitudinal structure is a hazardous business, and hard-pressed for an empirical basis. Indeed, it has been surmised that a careful study of the medium in which we express our attitudes, i.e. natural language, provides perhaps the only significant lead that can

assist us in this effort.[6] Whether this is really so I do not know for sure; but I am inclined to believe it, and certainly I am like most linguists and philosophers in that I would not have anything else to go on.

I believe that it is possible to extract from the analysis of natural language important clues about the likely form and organization of our attitudes but that the aspects of language which are the most revealing in this regard have long been overlooked. It has only been in the course of the past decade that these aspects have become the subject of a more systematic and formal analysis. Their study has led to what might be termed a "dynamic" theory of the semantics of natural language, which was first developed, in at least two different forms, in the early eighties. One of these has come to be known as Discourse Representation Theory (DRT).[7]

In the next section I will give a brief sketch of DRT, focusing on those features of it that are relevant to the theory of belief and other attitudes which it has inspired. In Section 3 I will draw out the implications of DRT, as presented in Section 2, for a theory of the attitudes. Section 4 discusses direct reference and the related concept of an attitude *de re*, and then proceeds to discuss some issues relating to indexicality and the relation between linguistic meaning and verbal communication. In Section 5 the suggestions for a theory of communication of Section 4 are elaborated further and issue in a view about the form of shared information which involves a component of shared reference as well as common knowledge of the sort first treated in detail by David Lewis in his book *Convention* (Lewis 1969). Section 6 is no more than a preview of the treatment of three topics, (i) the semantics of the attitudinal structures proposed in the preceding sections, (ii) the analysis of attitude reports and (iii) the treatment of attitudes that are "self-reflective" in the sense of being about one or more of the subject's other attitudes (such

---

[6] I suspect that this judgment has somehow contributed to the austere view that all a theory of the attitudes could hope to accomplish is to formulate a semantics for attitude reports. Nevertheless, the two should be carefully distinguished. In fact, the austere view is compatible with the opinion that even the study of language will not yield any structural insights into the propositional attitudes of its users.

[7] The semantic approach which DRT exemplifies has been developed also in other ways, most notably in the form of Heim's File Change Semantics (Heim 1982), which was conceived at roughly the same time as DRT. (In fact, Heim's theory appears to have been worked out slightly earlier than the first explicit version of DRT saw the light). Since the early eighties a number of further variants of the approach have seen the light, most notably the relational approach of Barwise (1985) and the development of Dynamic Predicate Logic of Groenendijk & Stokhof (1991). If I am sticking in my discussion of the central ideas to the format of DRT this is partly out of habit and partly because I do not quite know how to state some of the points I want to make within the alternative frameworks.

as the belief that a certain belief you have rests on what you read in the *Times*, or the desire not to have a desire you do have and cant get rid of). Each of these three topics is of primary importance for the theory outlined in Sections 1–5, but spatial as well as temporal constraints prevented me from allotting them the treatment they clearly deserve.

2. Dynamic Semantics and Discourse Representation Theory

It has long been obvious to students of natural discourse that discourse understanding proceeds incrementally: What the hearer or listener has learned from the first n sentences of the discourse helps him in interpreting the $n + 1$st sentence; more often than not this help is essential.

In fact, it may be essential in two different ways. First, the sentences which make up a coherent discourse usually contain elements—demonstratives, definite descriptions and pronouns; tenses and temporal adverbs; "discourse particles" like 'however', 'hence', 'moreover', 'yet', etc.—which can only be interpreted with reference to the context which the preceding sentences of the discourse have established. Secondly, the intended interpretation of such elements typically links them, and therewith the content of the sentence in which they occur, with the interpretation of the sentences preceding it, thus creating a semantic "web" that cannot in general be equated with a simple conjunction of propositions expressed by the individual sentences.

One type of example that demonstrates the connectedness of discourse, and which has been pivotal in the original development of DRT, is illustrated in (1):

(1) Last month a whale was beached near San Diego. Three days later it was dead.

The information this sentence pair conveys is that there was (i) a time somewhere inside the month preceding the one containing the utterance time of (1), (ii) a place in the vicinity of San Diego, and (iii) a whale, such that that whale was beached at that time and died within a period of three days from that time. In (rough) predicate logic notation:

(1′) $\exists t \exists t' \exists p \exists x$ (Time(t) & Time(t') & Place(p) & PrevMon(t, n) & Whale(x) & Beach(x, p, t) & Near(p, sd) & $t \leq t' \leq$ day(t) + 3 days & Dead(x, t'))[8]

[8] Where "PrevMon" stands for the relation that holds between times t and t' iff they belong to successive months, "Beach(x,p,t)" stands for the three place relation which means

One could quarrel with certain details of this formula, in particular with the precise form of some of its conjuncts. But these do not matter here. What does matter is that the conjuncts which originate in the first, and those originating in the second sentence are all inside the scope of the same existential quantifiers. Thus the content of the two sentences together is that there exist times, a place and an individual which satisfy both the conditions imposed by the first sentence and those imposed by the second. This implies that the truth conditions of (1) cannot be represented as the conjunction of a pair of separate propositions, each one expressed by one of the two sentences.[9]

Such connections are largely responsible for the cohesiveness of discourse, which, discourse theorists have recognized for a long time, it is one

---

that x is beached in place p at time t, "$t \leq t' \leq t + 3$ days" expresses that the time t' lies between t and the end of the third day after the day containing t, and n and *sd* are constants denoting the utterance time of (1) and the city of San Diego, respectively.

[9] There is an ongoing debate over the question whether a sentence pair such as (1) entails that the time t and the whale x are uniquely determined. Arguably the subsequent reference, by means of the pronoun 'it', to the whale introduced by the indefinite description of the first sentence implies that there must be some particular whale that the two sentences are about. It need not be the case that only one whale was beached near San Diego last month, nor even that only one was beached near San Diego and dead within three days. But nevertheless, the argument goes, the utterance should be linked (in some appropriate causal way) to one particular such whale if it is to qualify as entirely felicitous. Similarly the time of the beaching, which is picked up by the adverb 'three days later', should be fixed by some such causal connection. This observation might be taken to imply that the content of (1) can be seen as the conjunction of two separate propositions, each expressed by one of the two sentences, after all. The first proposition is that the whale in question was beached at the relevant time, and the second is the proposition that that whale was dead within three days after that time. According to my own intuition this is true: Sentence pairs such as (1) presuppose some relation which links the utterance to a particular whale and time. However, this relation can be arbitrarily indirect. For instance, (1) might be said by a person who read a short note to this effect in the daily paper a few days earlier; and that note, though ultimately going back to an eye witness report, may itself have found its way into the newspaper in as roundabout a way as you like. I think that even in such cases the hearer of A's utterance is entitled to infer from the assumption that (1) was felicitously used that there must be some such link to a particular whale and time, even though he has no way of knowing, or even of finding out, what the link in this particular case is. It seems to me that in such cases the *information* which the hearer gets from the utterance is indeed correctly captured by (1′). But there is also a sense in which the utterance can be construed as carrying, in virtue of the *complex* causal chain that connects it with the particular whale and time, additional information; and when this additional information is taken into account as well, the total content can be factored as a conjunction of the two singular propositions about the whale and time that are expressed by, respectively, the first sentence and the second sentence. For discussion of the matter see in particular Chastain (1975), Evans (1980), Heim (1982) and Kadmon (1987).

of the principal tasks of discourse analysis to describe and explain.[10] But in more formal approaches towards semantics, in particular those which have followed Davidson and Montague in the methodological conviction that the principal task of a semantic theory is to articulate how syntactic form determines truth conditions, this kind of connectedness was ignored until quite recently. DRT arose as an attempt to fill this gap, by providing an explicit account of the kind of discourse connectedness which (1) exemplifies, while preserving the achievements of truth-conditional semantics.

Informal reflection about what goes on when someone reads or hears a sentence pair like (1) suggests the following: The recipient starts by taking in the first sentence. Let us assume he understands it. This means for one thing that he has grasped (in some appropriate sense) the conditions under which the sentence is true. But in addition it gives him a basis for making sense of the second sentence. His making sense of the second sentence involves linking its anaphoric elements to the interpretation he has already formed of the first sentence, in such a way that a new interpretation emerges, this time for the two sentences combined. The most natural way, it would seem, to analyze this interpretation process is to assume that the recipient builds, step by step, an interpretative structure of the discourse that reaches him, repeatedly relying on the structure he has built already in deciding what further pieces are to be added.

It is this model of natural language understanding that DRT has tried to develop in detail. To accomplish this one articulates a system of rules—the so-called *construction algorithm*—which maps discourses belonging to the language, or language fragment, under study into certain "interpretative structures". The formulation of the construction algorithm presupposes a certain syntax for the language, which assigns to each of its sentences a syntactic structure on which the algorithm can operate. The output structures are called "Discourse Representation Structures", or "DRSs". Each DRS represents certain truth conditions which are assigned to it by a general truth definition. The empirical adequacy of the theory is determined largely in terms of whether the construction algorithm transforms each syntactically admissible discourse segment *D* into a DRS whose truth conditions match those which are associated with *D* by competent speakers. The principal challenge of DRT, experience indicates, is the exact formulation of the construction algorithm.

---

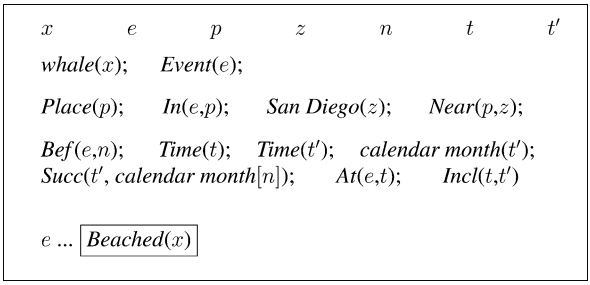
[10] See e.g. Halliday & Hasan (1976).

Since an important aspect of the construction algorithm has to do with how anaphoric sentence elements are linked with components of the previously constructed DRS (for it is this which accounts for the aspect of discourse cohesion that DRT was designed to explain), the construction of DRSs from syntactic inputs must be seen as an operation which always applies to a *pair* of input structures—on the one hand the syntactic description of the sentence under consideration and on the other the DRS as it has been constructed so far. This DRS functions, one might say, as the *context* in which the new sentence is being interpreted. This second role which DRSs play in the theory, i.e. as inputs, not outputs, of the interpretation process, imposes upon them certain special structural constraints which have no direct equivalents within earlier theories of natural language semantics.

Particularly important in this regard are *discourse referents*. Discourse referents are DRS constituents which represent individuals and other entities that are mentioned, or otherwise implied, by the processed discourse. For instance, the first sentence of (1) introduces, according to one of the construction algorithms currently in use, (i) a discourse referent corresponding to the indefinite description *a whale*; (ii) a discourse referent for the event of the whale's beaching which the sentence reports; (iii) a discourse referent for the place where this event takes place; (iv) a discourse referent representing the city of San Diego; (v) a discourse referent for the intended time of evaluation of the DRS, here (as normally) the utterance time of the represented discourse; (vi) a discourse referent for the time at which the event occurred; and (vii) a discourse referent for the calendar month preceding that of the evaluation time. In addition, the resulting DRS contains conditions, which specify certain properties of and relations between these discourse referents, and thus provide descriptive information about the represented entities. In the diagram-like form in which DRSs are most commonly presented the DRS looks like this:[11]

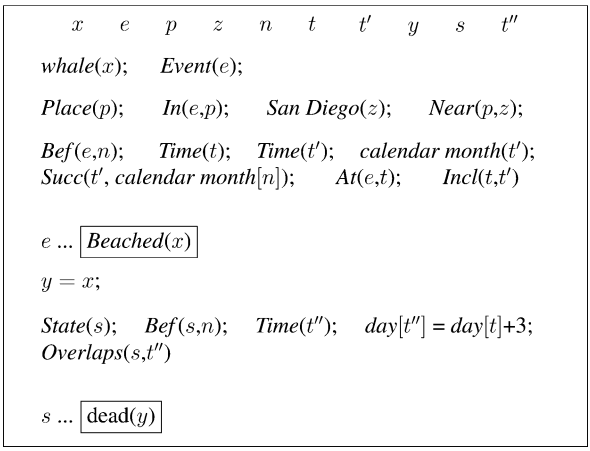
[11] The meaning of most of these conditions is, I trust, intuitively clear. "whale(x)" means intuitively that the discourse referent x stands for a whale, "Event(e)" conveys that e represents an event, and similarly for the predicates "Place" and "Time". "San Diego(p')" means that the discourse referent p' stands for San Diego, "In(e,p)" that the event represented by e occurred within the place represented by p, and similarly for the conditions that relate e to n and t to t'. The last condition specifies that e stands for an event of x getting beached. For our present purposes many of these details are not important. What is important is the general organization of the DRS into (i) discourse referents and (ii) conditions.

(2)



When the second sentence of (1) is processed in the context which (2) provides the construction algorithm can link the discourse referents introduced by certain elements of that sentence—the pronoun it, the past tense was and the prepositional phrase three days later—with the relevant discourse referents of the context-DRS (2). We will represent these links as conditions involving a new discourse referent, introduced by the anaphoric sentence constituent, and the already existing discourse referent with which that constituent is being linked. The resulting DRS is the following extension of (2):

(3)



The truth definition for DRSs has the following general form: A DRS *K* is true (with respect to a world *w*) iff it is possible to correlate entities (of *w*) with the discourse referents of *K* such that the conditions of *K* are satisfied (in *w*) by the entities correlated with the discourse referents occurring in those conditions. Thus (2) will be true provided there are individuals *a* and *b*, an event *E*, a place *P* and times *T* and *T′*, such that *a* is a whale, *b* is San Diego, *E* is an event of a getting beached, *E* precedes the utterance time *N*, *E* occurs in place *P* and at time *T*, *T′* is the month preceding that of *N*, and *T*

is temporally included in T'. Note that the definition does not require that the entities correlated with the discourse referents be uniquely determined by the associated conditions. In general these conditions only constrain the choice of the corresponding entities up to a point, without fixing them completely.

Much the same considerations apply to (3). (3) will be true if there are entities corresponding to its discourse referents which satisfy all its conditions. Note that the choice of entities correlated with, say, x and e is now more severely restricted than it was by the conditions of (2). For instance, the individual corresponding to x must now be a whale which not only was beached at the right time and place, but which moreover should have been dead within three days from the time at which it was beached. Still, the conditions need not fix this individual completely, any more than this need be so for the conditions of (2). It is for this reason that the truth conditions represented by (3) cannot be reduced to a conjunction of two sets of truth conditions, one for each of the two constituent sentences of (1). In particular, it seems impossible to find a separate proposition which reflects just the content of the second sentence. Evidently this impossibility is a consequence of the circumstance that the two parts of (3) which derive from the first and the second sentence, respectively, have certain discourse referents in common. (This is so for x, which occurs in the condition "$y = x$" of the second part, and for t which figures in the condition "$\mathrm{day}[t''] = \mathrm{day}[t] + 3$".)

Many of the details of these DRSs are of no relevance to what I will have to say about the attitudes.[12] What will be important is their organization into discourse referents and conditions—formally a DRS is defined as a pair consisting of a set of discourse referents and a set of conditions—and in particular, the fact that the discourse referents can act as links between conditions that originate in distinct sentences.[13]

---

[12] I should add that some of the details—especially those pertaining to the treatment of time and tense—are still a topic of debate: Thus, for all I can see at present, the temporal information conveyed by text segments such as (1) can also be represented in DRSs that look rather different from (2) and (3). So long as the methodological constraints under which DRT operates do not force a choice between such alternatives, the question whether attitudes might be organized in accordance with one of those alternatives or with another would, from the methodological perspective of this paper, be idle speculation. I have done my best to steer clear of such speculations.

[13] Many of the interesting issues in DRT have to do with the representation of universal and generic quantification and of sentence modifiers like negation, disjunction and conditionality. These notions require complex conditions, conditions which themselves have DRSs

3. First Suggestions for a Theory of Attitudinal Structure

The main reason why DRT appears to be a useful point of departure for a theory of attitudinal structure lies in its concern with the incremental character of language interpretation, and, therewith, of the assimilation of information by verbal means. Indeed, it is tempting (and, I think, up to a point legitimate) to see DRT as providing a model of the process by which the recipient of a discourse acquires new beliefs as he takes in its successive sentences. Consider someone who is reading a certain text—say, a newspaper report about a recent air show. He has full confidence in the paper's reliability and so accepts everything he learns. Each time he reads a new sentence, his information—or, if you prefer, his array of beliefs—about the show increases. DRT implies that the growth of information which the reading of each new sentence *s* produces corresponds to the change which processing *s* according to the construction algorithm in the light of the already erected DRS produces in that DRS. But it implies more—and here I recall the remarks about discourse coherence in Section 2: In general, we saw, the contribution made by a new sentence to the content of a discourse or text cannot be logically detached from the content of the preceding sentences. Therefore, insofar as that which the reader learns from the new sentence is to be considered a new separate belief of his,[14] we must be

---

as constituent parts. (The discourse referents of these constituent DRSs act essentially as bound variables!) In the present paper this aspect of DRT is entirely ignored. Nevertheless it is important for the attitude theory proposed here. For attitudes whose content involves the logical concepts (quantification, negation, etc.) for whose representation complex conditions are needed should, if the general approach of the present paper is correct, have structure which involves these same complex conditions.

[14] It would be implausible to hold that every new sentence of a discourse produces in the trusting reader a distinct new belief. Precisely how beliefs are to be demarcated and counted is a difficult question to which I doubt that any categorical answer is possible: whether we see a certain amount of information as representing a single belief or as being constituted of two or more distinct beliefs is a matter that may depend on the context of the particular use which the believer is, at any one time, making of the belief(s), and, as a reflection of this factor, also on the explanatory purposes for which we, who attribute the belief(s) to him, make these attributions. Nevertheless there are many cases where the presumption must be that a number of sentences give rise to a single new belief. However, the point is not affected by this qualification. For, first, a segment consisting of several sentences will often hang together with preceding segments in the manner illustrated by the example of Section 2, and second, there are at least some cases where a new belief derives from a single sentence and hangs together with earlier beliefs in the indicated way. One such example would be that where two people, *A* and *B*, are both present when (1) is uttered, but where *B* for some reason does not hear the second sentence, while *A* hears both. Assuming that they both believe all

prepared for the possibility that the content of this belief is inextricably interwoven with his other beliefs. The doxastic commitments which the reader makes by adopting the new belief impose additional limitations on the set of worlds compatible with his beliefs. But these commitments are associated with the entire *family* of beliefs he holds; there is no way of construing the set of worlds as the intersection of a number of world sets, each of which corresponds exactly to one of the beliefs in the family.[15] As we have seen in the preceding section, the formal reason DRT offers for this connectedness is that the (structures identifying) the different beliefs share certain discourse referents.

This brings us to two assumptions concerning beliefs and other attitudes which will be basic to all else I will have to say in this paper:

(A1) In general it is not possible to specify the truth-conditional content of each of the individual beliefs that are held by a given person; only the belief state as a whole (or, at any rate, substantial parts of it, which comprehend a large number of his beliefs) determine well-defined truth conditions. The same goes for attitudes other than belief. (For instance, desires are often inextricably connected with beliefs, as well as with each other.) Therefore, a theory of the attitudes must concern itself with analyzing the intentionality of such complex propositional states, and with the ways in which their constituent attitudes are interconnected.

(A2) The connectedness of different attitudes that are part of the same state is an effect of certain intentional elements which they share, and which play conceptual roles comparable to those played by discourse referents in the DRSs of DRT. Therefore, complex attitudinal states must be assumed to be organized "around" such elements. The elements represent entities in the world (or, more generally, in whatever domain that the relevant attitudes happen to be about). They are accompanied by "attributions" of certain properties and

---

they hear, it would seem correct to say that there is one belief that *A* and *B* share, but another which *A* has and *B* does not have. This second belief, moreover, is linked with the first belief of *A* in just the way the DRS (3) makes explicit. (This last example raises a further question about shared belief which I will take up in Section 5.)

[15] In a way the interconnectedness of discourses such as (1) does not capture the full extent of the interconnectedness of belief. In connected discourse there is an apparent asymmetry between the earlier and the later parts. In our analysis of (1) for instance we found that the second sentence could not be considered to determine a separate proposition that is logically independent from that expressed by the first sentence. But with regard to the first sentence this difficulty did not arise. In the case of belief, in contrast, the situation appears in many cases to be fully symmetric. For instance it may be true that the different beliefs of a connected family were acquired in a certain temporal order, but that the believer does not recall anything about this order. Here the earlier formed beliefs will have come to depend for their intentional significance just as much upon those formed later as the latter are dependent upon the earlier ones.

> relations, which act as attributions of properties and relations to the entities which the elements are taken to represent. These elements and the "attributions" which accompany them determine the state's intentionality.

Assumption (A2) appears to be little more than a prolix way of repeating what I had already said, viz. that attitudinal states are structured like DRSs. My reason for spelling matters out once more, and in this manner, is this: I am persuaded that some of the features of the DRSs that are used in current DRT correspond to important structural properties of attitudinal states. Yet I think we should not take it for granted that every feature which is needed in a semantic theory such as DRT must necessarily have its psychological counterpart.[16]

Before proceeding further I want to briefly raise a general methodological issue, which is related to what I just said. Someone persuaded that beliefs and other attitudes have a structure which resembles that of DRSs (or, for that matter, of any other class of formal objects) can express this thesis in one of two ways. He can either propose a theory which, in a certain sense, *models* the relevant doxastic phenomena, by positing that propositional attitudes are relations between individuals and DRSs (or whatever other formal objects). (Such a theory would be a direct alternative to any theory which analyzes propositional attitudes as relations between individuals and sets of possible worlds.) Alternatively, one may postulate that (cognitively capable) individuals are, at any one time of their existence, in certain cognitive states, and further that these states have structural properties analogous to, or exemplified by, DRSs (or other formal objects). On this second alternative the role of the formal objects is not that of being the attitudinal objects themselves, but rather of specifying complex structural predicates which given attitudinal states do or do not satisfy.

The two alternatives differ more in style than in substance, and there is not much to motivate a choice between them. If I have opted for the second alternative, this is primarily because I do not in general want to exclude, by ascribing some particular structure to a given state, the possibility that the state may have various other properties as well, either properties not yet revealed by current research or else properties that are irrelevant to the explanatory goals at hand. Strictly speaking, if one proceeds in the first way,

---

[16] It should be noted in this connection that the present state of development of DRT still leaves a range of questions pertaining to the form of DRSs unsettled. It is to be hoped that continuing research will provide answers to these questions. But as long as we do not know the definitive forms of DRSs, it is pointless to try and to assess whether all their properties can be meaningfully transferred to the psychological domain.

and *identifies* attitudinal states, or their objects, with DRSs, one thereby precludes the possibility of these states or objects having additional structural properties, which the DRSs do not display. Treating DRSs as complex predicates does not have this consequence.

On the second approach one cannot say literally of a given state that it contains discourse referents or conditions. Speaking exactly, it is only the DRS characterizing the state which can be said to have such constituents. The state only functions "as if" it had such constituents—for instance, if the state is a belief state (i.e. it consists exclusively of beliefs) then the person who is in this state will expect the world to be one of those in which the DRS holds. It will be both convenient and harmless, however, to speak informally of the state as *having* the kinds of constituents that make up DRSs.

What distinguishes the two approaches is hardly significant in comparison with what they have in common. It is important in this connection to note that they are similar not only in that they postulate similar attitudinal structures, but also in the obligations which they thereby incur. Either must, if it is to earn its keep, relate those structures in an explanatorily significant way to the phenomenology of belief and the other attitudes. In particular, it must be able to tell us how certain complex states consisting of beliefs, desires and possibly other attitudes lead to new beliefs, desires, intentions and plans.

As I intimated already in Section 1, a theory capable of this would have to include, at the very least, a detailed account of practical reasoning. This is more than any current theory of belief has accomplished yet, or, for that matter, that any could hope to accomplish in the near future. In fact, not many of the extant theories of propositional attitudes have made serious attempts to address this problem.

This is no less true of the theory presented here than it is of any others with which I am familiar. No detailed account of practical reasoning based on the structural proposals I will be making exists. However, even in the absence of such an account we should at least reflect on the minimal conditions which a structural description of attitudinal states must satisfy if it is to be capable of providing the foundation on which such an account would have to rest.

Perhaps the most obvious implication of this general requirement is that structural descriptions of complex attitudinal states distinguish between those components of the state that function as beliefs and those that function as desires. For it is obvious that in practical reasoning these two types of attitudes play entirely different roles. This is not a matter of their content. For instance, once a desire has been fulfilled, it typically makes place for a

belief with the same content. In such cases the cognitive role of the belief is utterly different from that of the earlier desire, notwithstanding the fact that their contents coincide. The difference between desires and beliefs is a matter of, as I will call it, *attitudinal mode*.

Among the attitudinal modes are, first and foremost, belief and desire. However, an explicit theory of cognitive processes will certainly have to distinguish many more. Doubt, hope, fear, regret are just some of many that come to mind. Another mode—which has received for less attention than I think it deserves, as it seems to me to be of paramount importance to a theory of cognition—is the one I will refer to as *wonder*. This is the attitude someone has towards a given content if he is aware of this content but totally uncommitted towards its *alethic status*—i.e. if he is entirely neutral as to whether it is true or false, or, for that matter, even whether it is possible.[17] The classification of attitudinal modes is a substantial task. It is one I could not undertake here. For the task of classifying cannot be separated from that of making explicit what role each mode plays in practical reasoning and other forms of reasoning, and, as I have already said, this is something that exceeds our present capacities.[18]

Even a superficial reflection on the question what factors influence the roles that the different parts of attitudinal states play in the processes of practical and theoretical reasoning through which they are continually transformed makes us realize that there are many more ways in which attitudes are discriminated than is suggested by the small list of attitudinal modes I just gave. Moreover, many of these factors do not seem to be a matter of mode, in the way in which the difference between desire and belief is. For instance, people frequently associate with the beliefs they hold further beliefs about how they came to hold the first beliefs, e.g. about who gave them the information the beliefs embody, or in what other way that information was picked up. Many such associations have an important part to play in determining the cognitive roles of the attitudes in question. Thus if you associate with a belief of yours assumptions concerning its origins, this

---

[17] By and large this is the attitude of one who asks the corresponding question; but not quite, for questions are often asked by people who already regard their content as possibly true, but want to know whether it is really true. The attitudinal mode of someone who asks a question regarding the content of that question and thereby manifests his ignorance concerning truth or falsity as well as his commitment to possibility, is to be distinguished from belief on the one hand and from mere wonder on the other.

[18] I suspect that subjective probability should also be seen as giving rise to distinct attitudinal modes, each one of which would reflect a particular probability commitment of the subject with regard to the relevant content.

will as a rule influence how confident you are in that belief: To a belief that as far as you can remember comes from a source you have learned to be reliable you tend to attach a higher degree of certainty than to one whose pedigree is more dubious. In their turn degrees of confidence are crucial inputs to our procedures of belief revision. For instance, if you receive new information which you regard as trustworthy, but which is incompatible with the conjunction of two beliefs you already had, and if you attach to one of these two beliefs a high, and to the other a low degree of confidence, then it is to be expected that you will give up the second of those two beliefs.

Similarly with wishes and desires. Desires come in varying degrees of intensity and they are accompanied by varying assumptions concerning the likelihood or the preconditions of their fulfillment; both intensity degrees and assumptions affect what plans we form for satisfying those desires. The examples can be extended *ad libitum*. There appears to be a vast spectrum of conditions which people associate with their various attitudes, and many of these influence the roles those attitudes play in the development of the attitudinal states of which they are part.

The conclusion we are forced to draw from this is that a complex attitudinal state will as a rule have a large part that is "higher order" in as much as it consists of conditions on the basic components of the state. This is an aspect of such states which DRSs of the sort presented in Section 2 are unable to represent. So, to arrive at an appropriate theory of attitudinal structure we will have to extend the structural descriptions which those DRSs provide in such a way that they capture this aspect as well.

The term "higher order", which I used in the preceding paragraph, suggests a fairly traditional solution to this problem, in which the "higher order" attributions are represented, in the spirit of higher order predicate logic, as conditions which combine attitudinal predicates with discourse referents that represent propositional attitudes. This is indeed how we will represent them. However, we must realize from the outset that by proceeding along these lines we will be able to provide only part of the solution. In particular, we cannot hope to distinguish desire and belief in this way. We cannot, for trying to do so would get us into an infinite regress.

To see that this is so we argue as follows. Suppose that a given state were to identify one of its parts as a belief by virtue of containing a condition consisting of the appropriate predicate—let us designate it as "Bel"—and a discourse referent, p say, representing that part of the state. Would this establish the relevant component as a belief? Well, we can't say yet. For we didn't say anything about the mode of the "second order" condition "Bel(p)". Is it a belief? A desire? Or what? Intuitively it seems to be more like a

belief than a desire. For in general people act as if they know—and thus believe—that their beliefs are beliefs, and not desires (and similarly that their desires are desires and not beliefs, etc). So let us suppose it is a belief (but the assumption is not essential; if we were to assume a different mode we would run into the same trouble). If, as we are assuming, attitudinal modes are identified via conditions on the attitudes of which they are the modes, then the belief whose content is the condition "Bel(p)" will require another condition, in which the predicate "Bel" combines with a discourse referent, q say, which represents the condition "Bel(p)". But then exactly the same problem will arise in connection with this new condition "Bel(q)": What determines its mode? Presumably a further condition. And so on *ad infinitum*.

In order to see how this regress might be avoided it is important to recognize that attitudinal *modes* like belief and desire are fundamentally different from the *attributions* we often make to our attitudes. Given that a certain attitude is part of your attitudinal state you can make it the subject of further predications. Thus when you remember of a certain belief of yours that you got it when you read yesterday's *Times*, you genuinely attribute to that belief a certain property.[19] But I have the intuition—although I would not know how to prove this—that such predications presuppose their subjects to be full-blown attitudes, each with a particular mode as well as a particular content. Your remembering that you got a certain belief from yesterday's *Times* can be analyzed as involving the belief that the first belief satisfies a certain condition. But knowing that the first belief is a *belief*—and not a desire, say, or an intention—is not like this. Its identity as a belief, i.e. as an attitude with that particular mode, is given as an inseparable aspect of the attitude itself.

These considerations and speculations have led me to a solution for the problem of "higher order" attitudinal structure which consists of two parts. The first part concerns the representation of attitudinal mode. Intuitively, what I would like to say is that complex attitudinal states have the structure of DRSs the components of which "come in different colors". Each such component represents one of the constituent attitudes of the state; its conditions and discourse referents identify the attitude's content while the color

[19] For the most part such second order attributions take the form of further beliefs. But it is also possible to make such attributions in desire. For instance I may desire of a given belief I have that you share that belief. And I, for one, desire of quite a few of my desires that I may be free of them.

determines its mode (e.g. blue for belief, red for desire, black for regret, and so on. But be free and pick your own). In the less colorful setting of set theory this proposal comes to the following:

(A3) Complex attitudinal states have the structure of sets of pairs, where each pair consists of (i) a mode indicator and (ii) a DRS. The different DRSs may share the same discourse referents, and this may be so even when their modes differ. Such sets of pairs of mode indicators and DRSs are called articulated DRSs.[20]

Eventually, I already surmised, a substantial range of distinct mode indicators will be needed, but only further research into the workings of practical reasoning and other cognitive processes can tell us what that range is like in detail. At the very least we will have to distinguish between belief and desire, and so we need at least different indicators for those two modes. I will use *Bel* as the indicator for belief and *Des* as that for desire.

Mode indicators are not constituents of the contents of attitudes. Therefore they cannot be expected to have meaning in the way of sentence constituents, or of the constituents of other truth condition representing structures. Nevertheless, modes have something partly similar to constituent meaning. I will refer to it as *logical force*. It is the logical force of each mode which determines the inferential role which attitudes with that mode play in the various processes of practical and theoretical inference. (Thus logical force is partly similar to the meanings of sentence constituents, insofar as the latter also (help) determine, via the contributions they make to the truth conditions of the sentences in which they occur, what inferential roles those sentences can play.) As the analysis of logical force evidently belongs among the tasks of the theory of practical reasoning, it falls outside the scope of this paper.

As (A3) explicitly says, different components of the same complex attitudinal state may be connected in that they share one or more discourse referents. In fact, this is an exceedingly common phenomenon, and it arises both in cases where the sharing components have the same attitudinal mode and in cases where they do not. Instances of the second kind are especially important, as they are crucially involved in virtually all processes of practical reasoning. Usually, when we move from a set of beliefs and desires to an intention (to act in such a way that the desires are likely to be optimally satisfied given what we believe the world to be like), the desires and the beliefs are connected in just the way that discourse sharing components of complex

[20] See also Asher (1986) in this connection. It should be noted, however, that Asher sometimes uses the term delineated DRS in the sense in which I am using articulated DRS.

attitudinal states seem to capture. Let me give a simple example. Suppose that I believe there is a cup of tea in front of me and that I have the desire to drink. I may then form the intention to pick up the cup and bring it to my lips. Here belief, desire and intention are connected in just the way that the sharing of discourse referents suggests: They all contain the same discourse referent, representing the cup, and it is only because the belief and the desire are connected in this way that the mechanisms of practical reasoning may be expected to yield an intention which is about the same cup of tea in virtue of containing this discourse referent as well. (A proper treatment of this and other cases of practical reasoning should spell out the inferential principles which exploit the connectedness of belief and desire; but I think the point ought to be clear enough even without these details.) Note that it would not do to try and reanalyze this inference by construing belief, desire and intention as all being about the same external object, the cup of tea that is actually standing in front of me. For it might be that I am hallucinating (or wrong for some other reason) and that there isn't any cup in front of me at all. In that case none of the attitudes in question can be construed in the way this analysis would require. Yet it is obvious that the processes of practical reasoning will evolve just as they would in a case in which there is a real cup; from a cognitive point of view—i.e. from the only viewpoint that matters to questions of reasoning, whether theoretical or practical—the two cases are indistinguishable. Only a theory of belief and desire which allows for a notion of *intended coreference*, such as that embodied in the communality of discourse referents, is in a position to explain how the two complex states can be intentionally identical, even though the one but not the other is directed at a real thing.

The first part to my proposal for dealing with higher order attitudinal structure does not deal with higher order as such, but only with the form of those "first order" components of states which may, though they need not, figure as the arguments of higher order conditions. It is only the second part of the proposal that deals with higher order attitudinal structure itself. This part posits that higher order conditions have for their arguments discourse referents that represent components of attitudinal states. Superficially this might seem an entirely straightforward matter, given what has already been said, in Section 2, about discourse referents and conditions generally. However, there are some subtle questions about precisely how the attitude discourse referents which this extension of the DRS formalism takes on board are connected with the attitudes (or, more accurately, with the components of attitudinal states) they are meant to represent. Note in this connection that in the cases which motivated the adoption of such

discourse referents—those where the subject attributes a property to one of his own attitudes—the discourse referent must be connected with a part of the attitudinal state to which it itself belongs. This situation is quite different from that exemplified by, say, the discourse referents x, t, and p of (2) and (3), which represent entities in the world, not components of the very same structure to which they themselves belong.

In fact, this is a problem which I will not be able to discuss in any depth in this paper. I will return to it only very briefly towards the very end, in Section 6.3; but a proper treatment will have to wait for another occasion. Instead of addressing it, we will, in Sections 4 and 5, look at questions which relate, some more, some less directly, to the problem of higher order attitudes and which need to be addressed before we can deal with the problem of higher order attitudes in the way I think one should.[21]

---

[21] The representation of attitudinal states as articulated DRSs bears on a question which I have not raised explicitly in the text. This is the question how many distinct attitudes are contained in any given attitudinal state. Formally articulated DRSs seem to provide a straightforward answer: The individual attitudes are its members (And so the number of them is its cardinality.) But this answer conceals a number of difficult issues which complicate the question. In fact, I doubt very much that the question permits more than a partial answer. Surely parts of an attitudinal state which either differ from each other in their modes or are the subjects of incompatible higher order attributions are thereby established as representing distinct attitudes. What is not so clear is whether an attitudinal state should sometimes be regarded as subdivided over and beyond the divisions thus established. This question is complicated by two separate factors. First, attitudinal states are subject to continual change, both as a consequence of new inputs from outside and of the inner processes of the mind—in short, because creatures with attitudinal states think. These changes affect in particular the ways in which bits of information are grouped together and/or separated. So what at one moment acts as a single belief may the next moment act as two or more, and this may happen even though there is no change in content over all. Secondly, whether we, as outside observers of a certain individual, should describe him as having, say, one single belief or rather as having two "smaller" beliefs whose conjunction is logically equivalent to the first may depend on our purpose in making such an attribution to him. (I suspect that there are important connections between these two factors, but I do not see my way through what they might be.) The following case may give an idea of the complications that worry me. Suppose you and I are waiting at a zebra crossing when a shiny convertible sweeps past. I recognize Emma at the wheel, and the car strikes me as a Ferrari. So I form the belief that Emma just passed us in a Ferrari. You too recognize Emma, but you also recognize the car as a Fiat. And so you now have the belief reflecting that impression. When I exclaim: "Gosh, did you see Emma in that Ferrari!" you retort: "That wasn't a Ferrari, you twit. That was a Fiat". Suppose you persuade me that I was wrong: Certain features of the car, which in fact I did notice, are never found with Ferraris but only with Fiats. So I adopt your version, and adjust my beliefs accordingly. Question: Have I changed the one belief I had, viz. that Emma passed us in a Ferrari, into the (incompatible) belief that she passed us in a Fiat? Or did I really have two beliefs—one that Emma just passed us in a car, and the other that that car was a Ferrari—the first of which I shared with you, while you succeed in talking me out of the

4. Direct Reference and Attitudes *De Re*

4.1. *Direct Reference*

In my discussion of the DRS (2) in Section 2, I emphasized that the real world correlates of the discourse referents x, e, p and t are not necessarily fixed. The content of (2) only requires that there be some set of entities a, E, P and T which satisfy the given conditions. If there happened to be two whale beachings near San Diego within the indicated period, then either one would provide a satisfying correlation of real world entities with these discourse referents. This is just another way of saying that these discourse referents function, from the perspective of traditional formal logic, as existentially quantified variables.

But this is not true for all of the discourse referents of (2). In fact it is not true for any of the remaining discourse referents n, p' and t'. The reasons are different for each of the three cases. We shall have to go over each of these in turn.

Before we address the problems which these discourse referents raise, let us begin by noting that there is one very straightforward type of case in which a discourse referent can be said to represent a unique entity. This is when the conditions in which the discourse referent figures are true of exactly one thing in the world. Such discourse referents have a status like that of definite descriptions which are proper in the sense of the classical description theories of Russell and Strawson. Indeed, one way in which discourse referents with this property arise is when a proper definite description is processed by the DRS construction algorithm, yielding a discourse

---

second? I just would not know how to answer this question and I doubt that an unequivocal answer could be found. One would expect a similar difficulty to arise in relation to questions involving simple existential or universal quantification over attitudes. Consider for instance, the claim that immediately after Emma raced past us, but before I could comment on the event, you and I had a belief in common concerning the event. If my only belief at that point was that Emma passed in a Ferrari the claim should be regarded as false; for that is a belief you did not have. If it is correct to attribute both to me and to you more than one belief on the subject, one of which is, for each of us, the belief that she passed in a car, then the claim should count as true. Interestingly, the intuitions about this claim seem pretty clear. Surely we did share the belief that she passed in a car. The fact that our apparent uncertainty how beliefs should be counted goes hand in hand with much clearer intuitions about claims involving purely existential or universal quantification suggests that belief is conceived more as a homogeneous than as a countable category. Independent evidence which points in this same direction is that in many languages, including for instance German and French, the closest translation of the English noun "belief" is a mass noun, not a count noun.

referent together with conditions reflecting the description's descriptive content, and thus satisfiable by exactly one thing.

In real life, however, pure cases of this sort are not all that common: Very few of the definite descriptions we encounter in ordinary use are proper in the strict Russellian sense. Mostly descriptions succeed in denoting uniquely only by virtue of the support they get from other noun phrases. And when one pursues the support chains to their beginnings one often finds noun phrases which act in a demonstrative or indexical capacity.[22] It is widely accepted nowadays that such indexical and demonstrative noun phrases succeed in picking out their referents in a non-descriptive manner; that they *refer directly*. The classical account of this is that an assertion containing such a noun phrase puts forward a so-called *singular* proposition, in which the property expressed by the remainder of the asserted sentence (i.e. by that which remains when the noun phrase is removed from it) is attributed to the individual which the noun phrase picks out. The noun phrase does not furnish the proposition with a description of the individual; rather, the individual is itself a constituent of the proposition. This does not mean that the noun phrase lacks all descriptive information. It only means that whatever descriptive information it has serves exclusively for the *identification* of the individual; it is only in this indirect way that the descriptive content of the description can affect the proposition the sentence expresses; it does not enter into the proposition directly.

By what means different types of directly referential noun phrases succeed in picking out their denotata is a subject about which there is a good deal to say and about which, in fact, much has already been said.[23] I will address certain aspects of this problem later. But for now we will leave the issue aside.

What does the doctrine of direct reference come to in DR-theoretical terms? The first point to observe is that DRSs of the sort considered in Section 2 cannot do justice to direct reference as is. For the information such DRSs contain is purely descriptive; and so their truth conditions are always those of strictly general propositions. What is missing is some direct

[22] The simplest sort of example illustrating what I have in mind is provided by descriptions which embed demonstrative or indexical noun phrases. Take for instance the description 'my father's second wife'. This description uniquely identifies a certain individual as the only one who stands in a certain relation—that of being second wife—to an other individual which is uniquely identified as the only one standing in some other relation—that of being father—to the individual denoted by the indexical pronoun 'my'.

[23] References are the familiar works of Kripke (1979), Kaplan (1989) and others.

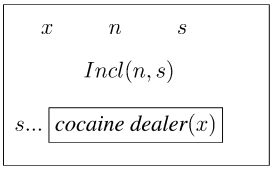
link with the world within which the DRS is situated—the world, that is, in which the sentence or discourse from which the DRS derives is uttered. To endow a DRS with the truth conditions of a singular proposition we must connect it with the appropriate entity or entities in the domain about which it intends to speak. Such connections are called (*external*) *anchors*. I will sometimes use the term *anchor* also for the individual with which a given discourse referent is connected, and, more specifically, refer to this individual as an *anchor for* the discourse referent, or *for* the DRS in which that discourse referent occurs, and I will call both the discourse referent and the DRS *externally anchored*.

Formally an external anchor is simply a pair consisting of a discourse referent and an entity. Besides unanchored DRSs, such as (2) and (3), DRT also admits *anchored* DRSs, which consist of an unanchored DRS together with one or more anchors. Given the motive for introducing them, the semantic role of external anchors should not be hard to guess. They serve to constrain the admissible correlations between discourse referents and entities in the world: Each of the anchored discourse referents may be correlated only with the entity to which it is already anchored. It is easily seen that this rule affects the truth conditions of the DRS in just the way the doctrine of direct reference requires. By way of an example, consider the case where someone says, pointing at a man standing on a street corner,

(4) That man is a cocaine dealer.

This utterance contains two directly referential elements. The first is the demonstrative noun phrase *that man*, which refers directly to the man at whom the speaker points; the second is the present tense of the verb; this is an indexical element, which refers directly to the time at which the sentence is uttered. So the DRS that represents this utterance has two anchored discourse referents, as shown in (5) below:

(5)



$\{\langle n, T\rangle, \langle x, a\rangle\}$

Here the expression on the extreme right specifies the two anchors. It is important not to misunderstand what this expression is doing. It is not part of the DRS itself, but should rather be seen as a reminder that its truth conditions are not those of an unanchored DRS, but the modified truth conditions that are obtained when T is the only permissible correlate for

n and a the only permissible correlate for x.[24] Thus, anchored DRSs express what Kaplan, following Russell, has termed singular propositions.[25]

Anchored DRSs capture the truth conditions which the doctrine of direct reference stipulates. But it is not clear how they fit into the picture of DRT that I sketched in Section 2. Recall that DRSs were supposed to result as representations of discourse, derived from it by a mapping (the construction algorithm) which uses as inputs (i) the (syntactic description of) the current sentence and (ii) the already constructed DRS. As I presented matters, the input DRS itself derives from the earlier parts of the discourse. At no point is there room, according to this picture, for the world that the discourse is about, to make its impact on the emerging DRS directly. Construction algorithms so conceived can only produce unanchored DRSs.

Part of the difficulty is that in the sketch of DRT which I gave in Section 2 the context in which a sentence is interpreted was identified with the result of interpreting the predecessors of that sentence in the same discourse. While information deriving from the earlier parts of the discourse tends to be needed for the correct interpretation, it rarely suffices on its own. As a rule interpretation relies on a host of additional information, reaching from the most general knowledge about the rules and laws which govern our world to knowledge of incidental facts about particular situations. Much of this information is acquired by non-verbal means. In particular, when

---

[24] I have chosen the terms, here '*T*' and '*a*', which are used to denote the entities to which the discourse referents are anchored, arbitrarily. Any terms (of the metalanguage in which the analysis of (4) is being carried out) which denote the right man and time will do here. Of course, I am not really in a position to provide the correct terms. For as the case has been presented, it leaves open when the utterance really occurred, or who the man was that the speaker pointed at.

[25] In fact, the structure of DRSs makes it easy to articulate in which respects singular propositions are singular. The proposition expressed by an anchored DRS is *singular with respect to* each of the individuals that anchor one of its discourse referents. Evidently, and as required by the accepted use of the term *singular proposition*, any proposition p that is singular with respect to the individual a can be understood as the attribution of a certain property P to the individual a—where p is expressed by the anchored DRS $\langle K, A\rangle$ and A contains the anchor $\langle x, a\rangle$, this is the property that, in any world w, is true of an object b iff the anchor $(A - \{\langle x, a\rangle\}) \cup \{\langle x, b\rangle\}$ can be extended to a correlation of the remaining discourse referents in K which satisfies all conditions of K in w (We say in this case that P is *expressed by* $\langle K, A\rangle$ *with respect to* x). If propositions are sets of possible worlds, then it is in principle possible that a singular proposition be expressed by an unanchored DRS. For instance, if p attributes the property P to the individual a, and there is a property Q such that (i) in each possible world w Q is satisfied by a and only by a, (ii) P necessarily entails Q, and (iii) P is the property expressed by K with respect to x, then K expresses p. Intuitively this means that the conditions of K fix the possible correlate of x to be a in all possible worlds. But all this is thoroughly familiar from the work of Kripke (1979), Kaplan (1989) and others.

we try to make sense of a sentence about our immediate surroundings we rely at least as much on what we can see of them as on what we have been previously told.

This points towards a reassessment of the model that, I said, DRT offers of the process of discourse interpretation. If this model is to stand a chance of being even remotely realistic, we must assume that the contexts to which the DRS construction algorithm is applied in sentence processing contain much more information than is supplied by the antecedent discourse alone. And it must allow for the possibility that even while discourse interpretation goes on, additional information may be introduced into the context via other than verbal channels.

Since some of the information that these so much more comprehensive contexts include must be expected to be of non-verbal origin, the hypothesis that such contexts are structured like DRSs reaches well beyond our earlier assumption that DRSs capture the structure of belief states acquired on the strength of verbal inputs alone.[26] Nevertheless I will stick to the hypothesis that contexts of interpretation have the form of DRSs. However much of an idealization this may well prove to be, I believe it will serve us well enough for the problems I will be discussing in this paper.

4.2. *The Anaphoric Dimension of Direct Reference*

Now that we have thus revised the notion of context, it is time to return to our example (4). Let us make the plausible assumption that among the contextual information available to the recipient of (4), just previous to the moment when she receives and processes the utterance, there is some pertaining to the man at whom the speaker points (because, say, the recipient had already noticed the man herself). In terms of DRT this means that the context DRS contains a discourse referent representing the man, together with certain conditions which reflect what properties the recipient sees the man to have (in the jargon of the philosophy of perception: What the recipient sees the man *as*).

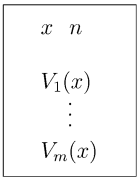
Against the background of such a context it is possible to provide a more pertinent analysis of the interpretation of (4). In particular, we are now in a position to see the processing of the demonstrative 'that man' as involving,

[26] In fact, I think the balance of probability is that visual information and verbal information are stored in very different forms, and that reasoning mechanisms typically exploit them in very different ways. How the two interact may well prove to be one the most challenging problems of cognitive science.

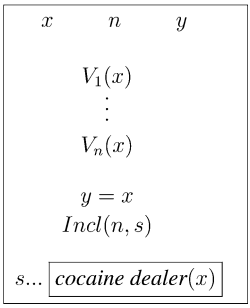
just like that of the anaphoric constituents of the second sentence of (1) in Section 2, connecting the new discourse referent with one that is already in the context: The discourse referent introduced by the demonstrative noun phrase is connected with the discourse referent in the context DRS which already represents the demonstrated man.

According to this analysis, the DRS-construction for (4) involves the extension of an existing DRS, not, as (5) suggested, the construction of one *ab initio*. Displaying of the context DRS only that part which matters (and that part merely schematically) we can indicate the difference before and after the processing of (4) by the following pair of diagrams:[27]

(6) Before:



After:



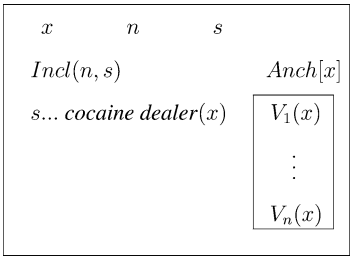
This seems to capture more of what goes on when an utterance like (4) is interpreted than did our earlier account, which led us to the DRS (5). Note however that the second DRS of (6) is *not* anchored. It is true that, as the direct reference account of demonstratives requires, the descriptive content of the demonstrative 'that man' has not become part of the content. Instead, the identity of the individual y represents is given via the conditions $V_1(x), \ldots, V_n(x)$, which reflect the recipient's antecedent observations about the man on the corner. But if we are to go by what was implied in Section 2, these conditions are once again descriptive. They represent the man as

[27] (i) Here the conditions $V_1(x), \ldots, V_n(x)$ are those which derive from the recipient's earlier observation of the man on the corner. (ii) I have ignored for the moment all questions which have to do with the discourse referent n. These will be addressed below.

standing in a certain place, wearing such and such clothes, etc. There is no direct, non-descriptive link between *y* and the man at whom the speaker pointed.

I can see two different ways in which this problem could be overcome. I shall discuss them both. According to the first, the construction algorithm not only introduces a discourse referent *x* for each sentence constituent which it recognizes as directly referential, but also attaches to this discourse referent a "signal" that *x* is to be treated as directly linked to the entity it represents. Which entity this is will, if all goes well, be determined by descriptive conditions on *x* of the sort with which we are already familiar. In the case of (4) the result will be a DRS looking like this:

(7)



Here $V_1(x), \ldots, V_m(x)$ are the conditions which fix (if all is well!) the anchor of *x*. *Anch*$[x]$ is the signal that *x* is intended to function as anchored, via the descriptive conditions $V_1(x), \ldots, V_m(x)$.

It is important not to confuse DRS constituents like "*Anch*$[x]$" with conditions such as "*Incl*(n, s)", "*s*... *cocaine dealer*(x)" or "$V_i(x)$". The latter conditions are directed towards whatever it is that is represented by the discourse referents that occur in them. But "*Anch*$[x]$" does not have this kind of significance. It does not make an attribution to whatever object the discourse referent *x* represents, but rather imposes a special constraint on the role of the discourse referent itself (viz. that of being rigidly and directly linked to whatever object the descriptive conditions associated with *x* determine). Constituents with this kind of function, that of saying something about how discourse referents relate to what they represent rather than describing the represented entities themselves, will be termed *formal* conditions. In contrast, conditions of the sort considered so far will be called *descriptive*, or *predicative* conditions. (Often I will refer to both predicative and formal conditions simply as *conditions*.) Graphically, I will distinguish formal conditions from predicative conditions in that the former will be represented with their arguments enclosed within square brackets, while in the latter they are enclosed within parentheses. If the condition "*Anch*$[z]$" is a

constituent of the DRS K, then we will say of the discourse referent z that it is *formally anchored in K*.

The semantic significance of conditions of the form "$Anch[\,..\,]$" becomes explicit in the truth definition for the DRSs in which they occur. Recall that a DRS of the kind considered in Section 2 is defined as true in a world w iff it is possible to correlate its discourse referents with entities existing in w which satisfy the corresponding (predicative) conditions of the DRS. To adapt this definition so that it produces the right truth conditions for DRSs with occurrences of "$Anch[\,..\,]$", two things are necessary. First, the truth relation can now no longer be simply a binary relation between DRSs and worlds. We need, besides the world with respect to which the DRS is evaluated, also a world which provides the anchors for the discourse referents. (For the evaluation, as true or false, of DRSs representing purely extensional contents the two worlds will coincide. But in general, when nonextensional information is represented, or when we are interested in the intension of an extensional DRS, this will not be so.) Second, we have to provide means that permit us to recognize those descriptive conditions which are ancillary to a formal anchor "$Anch[x]$" in that their role is to fix the entity to which x is externally anchored as playing this role. In (7) I have indicated this by placing those predicative conditions which serve to fix the anchor beneath the formal condition "$Anch[x]$". But we must make the division also formally precise.

What is needed here is something very similar to the articulated DRSs we found it necessary to introduce in connection with complex attitudinal states. Let us call the new structures *formally anchored* DRSs. The notion is defined as follows:

(8) A *formally anchored* DRS $\mathcal{K}$ is a pair $\langle K, A\rangle$, consisting of (i) a DRS K, which may contain one or more conditions of the form "$Anch[\,..\,]$"; and (ii) a function A which assigns to each discourse referent z that is formally anchored in K a DRS K_z not containing conditions of the form "$Anch[\,..\,]$".[28] $\mathcal{K}$ will be called the *principal* DRS *of K*.

[28] It is debatable whether the DRSs K_z should be permitted to contain formally anchored discourse referents in turn. If we were to permit formal conditions in these DRSs then they should be made into formally anchored DRSs themselves. In that case the notion of a formally anchored DRS would have to be defined recursively, with the definition given in the text as the base case.

While I think there are linguistic constructions whose analysis requires such more complex structures, I feel that the inductive definition would have produced a formal complexity that would not bring us anything of substance here, and only confuse matters.

The idea behind this definition is that the DRS K determines the content of the formally anchored DRS (modulo the individuals which the formally anchored discourse referents pick out), whereas the DRSs K_z have the function of picking out those individuals. Given what has been said about the truth conditions of such DRSs already, it should be clear that we can define the *truth value* of a formally anchored DRS *K with respect to* a world *w when anchored in* a world w' as follows:

(9) a. K is *true with respect to w when anchored in* w' iff

i. for each z such that "*Anch*$[z]$" occurs in the principal DRS $\mathcal{K}$ of K there is a unique entity a_z in w' such that there exists a correlation between the discourse referents of K_z and entities in w' which correlates a_z with z and satisfies all conditions of K_z in w'; and

ii. there is a correlation of entities in w with the discourse referents of K which correlates with each formally anchored discourse referent z of K the entity a_z, and satisfies all predicative conditions of K in w.

b. K is *false with respect to w when anchored in* w' iff

i. as (a.i) above; and

ii. there is no correlation as described under (a.ii)

Note that this is a partial truth definition. K will come out as true or false with respect to w when anchored in w' only if it is "properly anchored" in w'—only, that is, if the descriptive conditions specified by the K_z single out unique individuals in w' for each of the formally anchored discourse referents z.

What sorts of conditions go into the DRSs K_z? This is a question that I do not intend to discuss in depth. But, to get at least an idea of the issues involved, let us return to our example (4). It is part of the meaning of a demonstrative noun phrase like 'that man' that the individual to which it refers is determined by a combination of two factors—first, the descriptive content of the NP itself, and, second, what might be called the "information content" of the demonstrative act (if any) which accompanies the utterance.[29] It is neither a simple nor, I believe, a particularly rewarding enterprise to spell out precisely how these two factors interact to produce,

[29] Either of these factors may be vacuous. Thus demonstrative NPs can contain so much descriptive information themselves that there is no need for any disambiguating gesture. On the other hand there are the "pure" demonstratives *this* and *that* which have no descriptive content of their own and must rely entirely on the accompanying demonstration for referential success. (Actually this second claim isn't strictly correct, as even *this* and *that* contain a remnant of descriptive information: in cases where the referent is assumed to be a person *this* and *that* are incorrect, except in a few special constructions such as *That is a man*.)

in all those cases where demonstratives are used felicitously, the descriptive conditions that identify the referent correctly.[30]

It is important to understand the difference between anchored DRSs such as (5) and formally anchored DRSs such as (7). Evidently an anchored DRS like (5) is not formally anchored, for there is nothing internal to the DRS to indicate that the relevant discourse referent, *x*, is anchored. Conversely, a formally anchored DRS need not be anchored, in the sense that there in fact exists an external anchor for each of its formally anchored discourse referents. For it may well be that one or more DRSs K_z fail to identify unique referents. Ideally, of course, this situation should never arise in the context of actual language use. For ideally directly referential expressions should always be used properly. And when a directly referential expression is used properly, then the discourse referent *z* it introduces into the formally anchored DRS *K* which represents that discourse will receive through the DRS K_z associated with it a unique individual a_z in the world *w* in which the utterance takes place. Thus a formally anchored DRS deriving from a discourse in which all directly referring elements are used properly is, according to the truth definition for formally anchored DRSs, always either true or false in the world to which the external anchors for its (formally) anchored discourse referents belong. We can represent the truth conditions of such a formally anchored DRS *K* in the form of an externally anchored DRS consisting of (i) the principal DRS of *K* and (ii) the external anchors which map each of the formally anchored discourse referents of *K* to the associated entities of *w*.

4.3. *Formally Anchored Attitudes*

This first solution to the problem we encountered on p. 543 has been largely based on our earliest discussion of (4), which led to the DRS (5). The second solution, which we will discuss now, relates more closely to the discussion which issued in the DRSs given under (6). This solution, moreover, will at last lead us to the subject announced by the second half of the title of this section, *attitudes de re*.

---

[30] In the special case where the demonstrative act is a pointing, the descriptive condition should be something like "nearest thing of kind *C* lying in the indicated direction", where *C* gives the descriptive content of the NP and where "the indicated direction" is a shorthand for some descriptive condition that specifies the direction in terms of a coordinate system to which the recipient has cognitive access in view of the position he himself occupies. But this could only pass as a first approximation, even if pointings were all we care about. More could be said on this point, but I won't bother.

According to the account associated with (6) a recipient of (4) who is in the epistemological position which the use of the demonstrative *that man* presupposes, will connect the discourse referent *y* that the NP yields with another discourse referent, *x*, which is already part of the context provided by his background knowledge. By itself this does not supply *y* with an external anchor. But it will if there already is an anchor for the discourse referent *x* with which *y* gets connected. This, I want to suggest, is just what we have in the case of (6). The recipient of (4), who has been looking at the man standing on the corner, is thereby related to him in a way that puts her in an ideal position to have opinions that are directly about him, opinions which carry the commitment that it is *that* man—the one who is the actual cause of her perception, and no other—who has the properties attributed by those opinions. The content of such an opinion is, in the terminology of Russell and Kaplan, a singular proposition. Attitudes whose contents are singular propositions are sometimes referred to as *attitudes de re*. This is a term I will use also.[31] The proposal I made in Section 3 asserts that attitudinal states have the structure of articulated DRSs. That proposal did not allow for *de re* attitudes, and it is time to modify it in a way which makes room for these as well.

To arrive at a modification which does justice both to the psychological and the intentional aspects of *de re* attitudes it is important to keep in mind what was said in the last section about the difference between formal and external anchors. For it is crucial to distinguish between having, say, a belief that *is* about some external object (and whose content is thus a proposition that is singular with respect to that object) and having a belief that one *takes* to be directly about some external object. In particular, it is possible for someone to have a belief that he takes to be about some external object, and be mistaken about this. For instance he may be hallucinating, and thus be in a belief state which is psychologically indistinguishable from the state of someone who entertains a belief about an object he in fact sees, even though in his case there is no real object that he perceives.[32] When I say that in such a case the belief is psychologically indistinguishable from that of someone whose attention is directed towards an object which is actually there, I mean to imply in particular that the belief is qualitatively distinct from an existential belief—from a belief, that is, to the effect that there exists

---

[31] The notion should not be confused—not at any rate in the theory developed here—with that of a *de re* attitude report. There are, it is true, close connections between the two concepts. But in the present theory they are nevertheless crucially different.

[32] Recall the comparison of the case with and that without a real teacup in Section 3.

an object with such and such properties (including those which the subject hallucinates it to have) as well as the property of being seen by the subject.

This distinction—between believing that there is a thing with such and such properties and having a belief that the thing characterized by such and such properties, and to which, in particular, one stands in some special causal relation, has certain other properties as well—I propose to represent by means of formal anchors: Where there is no anchor the attitude is existential, where there is one the attitude carries, as a kind of presupposition, the commitment that one is in fact standing in some suitable relation to some object answering the descriptive conditions associated with the anchor. This having been said, the following proposal follows straightforwardly from the earlier proposal (A3):

(A4) Complex attitudinal states have the structure of *formally anchored articulated* DRSs, i.e. of sets whose members are pairs consisting of a mode indicator and a formally anchored DRS. In a DRS identifying such a state those and only those discourse referents are anchored that represent entities to which the possessor of the state takes himself to stand in some suitable causal relation.

Attitudinal states structured in the sense of (A4) raise a question of coherence which does not arise in connection with the simpler proposal (A3): It is possible for someone to be in an attitudinal state characterized by a formally anchored articulated DRS and fail to stand in some relation presupposed by one of the state's formal anchors. Where this happens—in other words, where a state has a formal anchor without a corresponding external anchor—the state strictly speaking misfires and lacks a well-defined truth-conditional content.[33] Of course this does not warrant the conclusion that, in such a situation, there is no attitude the subject could be said to have. Within the context of his own psychology the state is as real, i.e. as causally active, as it would have been had there no mismatch between formal and external anchoring.

We call an attitudinal state which is characterized by an articulated DRS that contains one or several formal anchors *formally de re*. Such attitudes, we have seen, may suffer from failure of presupposition. When there is no presupposition failure, and for every formal anchor there is a corresponding external anchor, the attitude will be called *truly de re*.

Whereas states that are formally but not externally anchored do in fact occur, states with external anchors unmatched by formal anchors do not, it

[33] Cf. the discussion of formally anchored DRSs and their truth conditions in the preceding section.

seems to me, represent a possible category. The notion of such a state strikes me as incoherent: It makes no sense to describe an attitude as *de re*, unless it is *taken to be de re* by the person who has it. Of course, it is possible to stand in the relevant causal relation to an object that does satisfy all predicative conditions associated with some discourse referent *x* of (the DRS identifying the content of) some attitude one has.[b] But that is not enough to make the attitude into an attitude *de re*. For instance, I may have a purely existential attitude that happens to be satisfied, maybe even uniquely satisfied, by an object to which I stand in such a relation—I may, for all I know, be looking at it. The attitude is no less existential for that.

As a matter of fact, this kind of case is not uncommon. I may be looking for something of a certain kind, believing that there are such things and that I will eventually find one. At the same time I may in fact be staring at the one and only thing that, as it happens, is of this kind, and yet I need not realize that this in fact is the thing I am after. In this instance the belief that I will find something of the kind I want surely is as purely existential in content as it would have been had I not been staring at the object.

Summarizing the last paragraph, we come to the following principle:

(A5) No external anchor without a formal anchor.

This principle does not play a central role in what I will have to say in the remainder of the paper. Nevertheless, it seems to me to express an important intuition concerning the relation between the psychological (i.e. strictly internal) and the intentional (i.e. world-oriented) side of cognitive content, important enough to be mentioned, even if it has no implications for what is to come. In fact, the issue is delicate enough to warrant a much more careful and extensive discussion. But this too must wait for another occasion.

As it stands, (A4) is incomplete. For I haven't said anything yet on the question *which* relations give rise to anchored attitudes. This is a notoriously difficult problem, about which there appears to be a good deal of uncertainty and disagreement. Paramount among the relations which are thought to give rise to attitudes *de re* is that established by direct veridical perception, and especially, for those of us who can see, visual perception: If there are any cases at all in which a person qualifies as attributing in thought a property or relation to some individual other than himself, certainly thoughts that are directed towards individuals he is actually looking at should qualify as such thoughts. For other perceptual modalities the matter

[b] That is, without there being any formal traces of this in the content representation of the attitude.

is perhaps not quite so clear. For instance, does hearing an alarm clock go off establish a direct link with it, in the way that (I think) looking at it would? Or is it more accurate to construe such a perception as involving an inference from the sound that one is hearing to the conclusion that there must be an alarm clock which is making it? I do not know how to answer such questions, and will have to leave them.[34]

Another type of relation which, I suggest, at least sometimes provides a sufficient basis for attitudes *de re* is memory. For one who has accepted that current perception provides such a basis, this seems to be especially plausible in the case of recent memories of perception: If a person can be said to have *de re* attitudes concerning an object while he perceives it, then it would seem unwarranted to deny he has *de re* thoughts about the individual as soon as he ceases to perceive it even if the perception is still vivid in his mind. But memory is a matter of degree, and moreover, its degree is subject to change. As English puts it so well, memories *fade*. How vivid or accurate the memory of an object must be if it is to qualify thoughts about it as *de re* is a question that probably cannot be settled without some measure of arbitrariness.

I want to mention one other kind of situation which has been argued to give rise to attitudes *de re*. This is a case which is of particular importance for the relation between language and mind. It is that where somebody forms a new belief (or other attitude) in response to an utterance which contains a directly referential expression. Note that, as with so many other concepts in the theory of language, there are two sides to the notion of direct reference. On the one hand there is the side of the speaker. He may use an expression e as directly referential, for instance when he intends to express the content of one of his *de re* attitudes—a certain proposition p which attributes the property P to the individual a—and to this end utters a sentence which contains e and the remainder of which expresses P. On the other hand there is the side of the hearer. He may take an expression e to have been used as a directly referential expression, and represent, in the belief which the utterance produces in him, the referent of this expression e by means of a formally anchored discourse referent x. In other words, his belief will be formally *de re*. And presumably, if all goes well, it will also be truly *de re*.

---

[34] Perhaps the answer should depend in part on how well the various modalities of the perceiver work. For instance, a congenitally blind person, who has to rely on hearing and touch as his primary perceptual resources, should perhaps qualify as being directly related to the alarm clock he hears, whereas someone who relies primarily on sight (as I believe most of us do), would not. But this is all highly speculative.

But when will it be truly *de re*? That depends. If for instance the hearer recognizes the referent of *e* as something from his immediate surroundings, something to which he has independent access (e.g. because he is looking at it himself), then there is as much reason for attributing to him an attitude whose content is directly about that thing as there is in other cases where we form beliefs about things which we are directly observing. And, as I have just said, if there are any situations at all which lend themselves to the formation of attitudes that are truly *de re*, it is those in which the relevant objects are directly seen. In fact, it seems reasonable to assume that in the case we are considering the predicative conditions which support the formal anchor for *x* in *K* include those which derive from the direct perception, and which, as they do in other such cases, suffice to identify the external anchor unambiguously.

But not all cases are like this. For instance, the hearer's line of sight may be blocked, so that he cannot see the thing at which the speaker is pointing; or it may be that the referent is not part to the immediate surroundings at all; and if it isn't, it might be something to which the hearer is related through memory; or it might be something with which he has never had any dealings before, so that the expression *e* mediates their first and only contact. (The list could easily be extended, but these few examples suffice for the point). It is not easy to decide in which of these cases the resulting belief is truly *de re* and in which it isn't. Again the problem is that there appears to be no general agreement over which relations between believer and object warrant that an attitude that is formally *de re* is truly *de re* as well.

It would be foolish to legislate in this matter, and I won't try to do so. But I want to make two observations relating to it, but which are also pertinent to some points that are to be discussed later. In the first and the last of the three cases mentioned in the previous paragraph the predicative conditions supporting the formal anchor for *x* will, if they are to secure unique identification of the referent of *e*, have to involve reference to the expression *e* and/or to some intention of the speaker (and possibly to some other parts of his attitudinal state as well). For instance, the anchor's descriptive support might consist of a condition to the effect that "*x* is the direct referent of the expression *e* (as used by the speaker)" or "the object to which the speaker is directly connected, via some discourse referent x' which gave rise to his use of the expression *e*"—or something to that effect; but without something of the sort unique reference cannot be assured. Thus, if it is possible in these cases for the hearer to acquire a belief that is truly *de re*, then, *ipso facto*, it is possible for the speaker to transfer his direct connection to external objects through the mere use of directly referential expressions.

In the above discussion I tacitly assumed that the hearer is correct in taking the expression *e* to have been used as a directly referential one. But of course, he could well be mistaken in this assumption—he may think the speaker used *e* in this way, although as a matter of fact he didn't. In some cases this mistake gives rise to attitudes which, though formally *de re*, are definitely not truly *de re*. This may happen, more specifically, when the speaker is using an indefinite description non-specifically, and the hearer misinterprets his use of it as specific. In that case the relevant discourse referent in the hearer's belief will be formally anchored, and that anchor will be supported by some such descriptive condition as "*x* is the object the speaker had in mind and to which he referred with his use of *e*". Under the circumstances this condition is not satisfied by anything; so it fails to associate an external anchor with the given formal anchor.

We may sum up this discussion as follows: It is difficult, and perhaps impossible, to come up with precise and yet non-arbitrary criteria for when an attitude that is formally *de re* is truly *de re* as well. But there exist on the one hand clear cases where attitudes are truly de re, and on the other also cases of formally *de re* attitudes which nevertheless are not *de re* in the external sense.

Let us briefly compare the two solutions that we have sketched in this and the preceding section of the problem that we encountered in connection with (4). The first solution, that of Section 4.2, offers an extended DRS formalism which allows us to express (i) the intended content of an utterance like (4), and (ii) the content which such an utterance has in those situations in which the presuppositions of direct reference are fulfilled. The second solution, of Section 4.3, involves the same distinction—between formally anchored and externally anchored DRSs—but it recasts the analysis offered by the first solution as one concerning the structural and intentional properties of the attitude which a recipient would form in response to (4).

However, there are some further differences as well. First, the construction principles needed to arrive at the structures posited by the second solution must in general accomplish more than those which the first solution requires. Thus the second solution demands, for instance, that in processing a demonstrative such as 'that man' as it occurs in (4) one not only introduce a new discourse referent and anchor that discourse referent formally, but also that, in those cases where this is possible, the predicative conditions supporting the anchor are exploited to link this discourse referent internally to one already representing the demonstrative's referent in the context.

Another difference, this one of a more general and methodological nature than the one just mentioned, relates to the concepts of linguistic meaning

(and, more specifically, of sentence meaning) that are implied by the two solutions. The first solution suggests, in line with what was said in Section 2, that the meaning of a sentence resides in its capacity for transforming truth condition bearing structures (DRSs) into other such structures. The second solution implies a somewhat different view, one on which the meaning of a sentence resides in its potential for transforming attitudinal states. This last difference carries a further implication which appears to be of particular importance for the theory of meaning and communication. Because different recipients of the same utterance may bring different attitudinal states—and thus different contexts of interpretation—to the same utterance, their interpretations of that utterance may yield significantly different results, even though both proceed in full accordance with the rules that govern the language to which the utterance belongs. Thus the second solution, but not the first, carries the clear implication that the meaning of an utterance, while itself independent of the individual properties of the actual or intended recipient, can nevertheless produce by equally legitimate means distinct messages in different receivers.

4.4. *The Now and the Self*

Up to now we have been looking in this section at direct reference by demonstrative noun phrases and by proper names.[c] The direct referents of such NPs are, we saw, objects that are external to both speaker and hearer. (At least, this was so for the few examples we considered; but, in fact, as further reflection on the matter will readily reveal, with NPs of these types this is the norm.) The present section is concerned with direct reference as well, but with a form of it that is nevertheless importantly different from that considered so far. This is *indexical* reference. Indexical reference, as I understand the term, is always to an entity to which either the speaker or the recipient stands in some special direct relationship. Most languages have set particular words, suffixes or constructions aside for the purpose of indexical reference. Prominently among these devices are, in English and most other languages, the first and second person pronouns.

---

[c] There is nothing specifically about proper names in this section; so mentioning proper names at this point was rather like writing a plainly bouncing check. Some remarks about proper names are scattered throughout Section 5, in particular in Section 5.3. For more recent work on proper names that builds on the ideas of 'Prolegomena' see R. Kamp (submitted).

We will look at two types of indexical reference, to the present and to the self. We begin with reference to the present, for the representation of which we already introduced a special discourse referent, viz. *n*. The discourse referent *n*, I said, represents the utterance time of the represented discourse. Its role in DRS construction is that of guiding the interpretation of temporal constituents of the discourse. It enters into the analysis of all tenses and of a large number of different temporal adverbs. (The details of this are complicated but do not matter here.) The truth conditional function of *n* is easily stated: Since it represents the utterance time it should always be correlated with that time.[35]

The discourse referent *n* also has a role to play in the characterization of attitudinal states. However, the semantic role which I propose that we assign to it when it acts as a constituent of an attitudinal state is somewhat different from the one just explained. For the possessor—call him *B*—of an attitude containing *n*, *n* is to represent, at any time *t*, that very time *t* (and not, say, the time at which *B* takes his belief to have originated). For instance, if a given person *B* has, at *t*, a belief characterized by the DRS (5) this is to mean that he believes the individual represented by *x* to be a cocaine dealer at time *t*. And this is to be the meaning even if *B* knows that he formed his belief at some earlier time t_0, the time, say, at which someone said to him "That man is a cocaine dealer". There are some intriguing questions to which this proposal gives rise. But since they would lead away from the central topics of the paper, it would be unwise to try and explore those here.[36]

The relation in which we stand to the present time is not the only one which has a strictly non-descriptive character. There is at least one other, one's relation to the self. This is the relation which Lewis (1979a) captured in his "Attitudes De Dicto and De Se" by making the contents of beliefs properties that are being self-attributed instead of propositions that are

---

[35] Formally we can proceed in the ways familiar from tense logic: We define first what it is for a DRS *K* to be true in a world *w* with respect to a time *t*. This will be the case iff it is possible to find a suitable correlation of entities in *w* with the discourse referents of *K* which in particular correlates *t* with *n*. We can then define a DRS *K* deriving from a discourse *D* to be true in *w* iff *K* is, in the sense of the previous definition, true in *w* with respect to the utterance time of *D*.

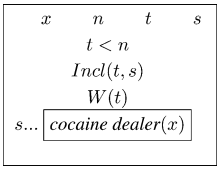
[36] There is an important problem about identity and change of belief through time which relates to this proposal. (I have been aware of it for a number of years, on the one hand through the work of John Perry 1977 and on the other through conversations with David Kaplan.) If someone who has formed a belief with the structure of (5) at t_0 retains the belief in this form the truth conditions to which he is thereby committed will in fact be subject to perpetual change. If on the other hand the truth-conditional content of the original belief is to be preserved, the belief must be "re-registered" as a belief about the past. Such a belief has a quite different form—according to DRT, something like:

attributed to the world in which one lives. It is a relationship that is perhaps more easily perceivable in attitudes other than belief, such as, for instance, desire. Most of our desires are desires to the effect that something happen to us, e.g. that it is *we* who get the things that seem desirable. That the content of such a desire does indeed involve a direct connection to the self, and not simply an attribution to an individual that we could not distinguish from ourselves descriptively, is notoriously hard to prove. However, the matter has been argued in detail by Perry (1977) and Lewis, and I have nothing to add to their subtle discussions

I propose that just like the present the self be represented by a special discourse referent, for which we will use the symbol "*i*". Like *n* this discourse referent comes with its own formal condition, which distinguishes it from all others, and which is implicit in the use of "*i*".

The explication I gave of the role of *n* distinguishes between its role in representations of the content of discourse and that which it is to play in representations of attitudinal states. I want to make a similar distinction in relation to *i*. This time, however, it seems better to reverse the earlier procedure and to explain first its role as a constituent of an attitudinal state and then to define the part it is to play in representations of linguistic contents.

The first role is easily explained, and indeed there is nothing new in what it seems to me ought to be said here: *i* always represents the possessor of the state of which it is part (and does this in the particular way that gives rise



where "$W(t)$" is meant to be some condition that identifies the past time at which *B* learned that the man on the corner was a cocaine dealer.

The upshot of this is that our beliefs about what is supposed to be the case now (i.e. at the very time of our having the belief in question) involve us in a perpetual tension between identity of form and identity of content. This tension gives rise, among other things, to the curious cognitive predicament that through mere intellectual apathy we may in actual fact allow the truth conditions of our beliefs to change, and thereby fall into error.

This is an intriguing topic in its own right, but not one to be pursued further here. I have mentioned it only to underscore my conviction that our relation to "now" is unlike any other cognitive relationship into which we enter with our environment. The special status of *n* is meant to reflect this. An important aspect of this relationship is that it is not mediated by any descriptive information. That *n* represents whatever time it happens to be is therefore a formal, not a predicative, condition. (I follow the existing practice in DRT of representing this formal condition by always using the symbol "*n*" for this discourse referent, and for nothing else).

to what Lewis (1979a) terms attributions *de se*). The implication of this for the truth conditions of DRSs containing i is straightforward: i must always be correlated with the possessor of the state of which it is a constituent.

It is important not to confuse what I have just said about i with the semantics of the pronoun 'I'. It has been suggested that all a semantic theory needs to say about 'I' is that the word refers in all cases in which it is used (except those where it appears inside direct quotation) to the person who uses it; and that is all there is to it. In a way this is clearly right. And yet I think I can see why such an account leaves some people dissatisfied. By relating 'I' to i we can, I hope, come a little nearer to the roots of that dissatisfaction. To explain what I have in mind I must say something about an aspect of language that we touched upon in passing in the last section, but about which I have not so far been explicit. This is its communicative aspect.

So far I have been concentrating on the interpretational side of language. But of course, as we noted already in Section 4.3, that is only one of its two sides. Before any bit of language can be interpreted, someone must have uttered it, must have chosen those words that make up the utterance to express the thoughts he wanted to convey. DRT suggests a very simple model for what goes on when someone puts his thoughts into words: The process is roughly the inverse of the interpretation process, one which turns DRSs into natural language rather than turning natural language into DRSs. I will refer to it as the *verbalization algorithm*.[37] Even so, the model of communication to which this simplifying assumption leads has turned out to be quite useful in a number of applications.

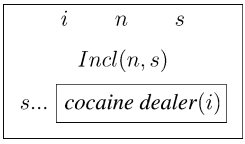
In simple situations, in which verbal communication involves just two participants, A and B, and in which A does all the talking, the model comes to this. Both A and B are in certain attitudinal states, characterized by articulated DRSs $K(A)$ and $K(B)$. Suppose A wants to communicate to B a certain bit of information. Typically this will be something that he himself believes; in other words, it will be some component of $K(A)$ of the form $\langle Bel, K_0 \rangle$. To communicate this belief to B, A has to find a sentence (or perhaps several sentences) which express it, i.e. which when processed according to the rules of the language in the context available to B will yield an addition to that context which captures the belief's content. If all goes well, B will, through processing the sentences that A speaks to him, add to

[37] The claim that verbalization and interpretation are inverses of each other is correct only in first approximation—there are many important asymmetries between the processes of language generation and language reception. In fact, one of these concerns the pronouns 'I' and 'you' and we will make its acquaintance shortly.

his attitudinal state a new part which has the same content as the belief that *A* tried to put into words. The reason why we may expect this to happen is that *B* will be using the same rules that guided *A* in his choice of those sentences—*A* and *B* employ the same rules, since they are speakers of the same language.[38]

The only additional assumption needed for the point I wish to make here is that both $K(A)$ and $K(B)$ involve i (this is plausible enough; everyone has attitudes *de se*) and that the information *A* wants to communicate to *B* concerns himself, that is, *A*. To be concrete, suppose that *A* wants to tell *B* that he is a cocaine dealer. In $K(A)$, we may presume, this information is represented in some such form as this:

(10)



and let us also suppose that the words he uses to convey this are

(11) I am a cocaine dealer.

In the last paragraph I said that to find the words which express this, *A* must apply a procedure, the verbalization algorithm, that is the inverse of the DRS construction algorithm. As a matter of fact, an analysis of the pronouns 'I' and 'you' shows that this cannot be quite right. Thus, for instance, the

[38] In most ordinary speech situations *B*'s context does not differ radically from *A*'s own. In such situations it will be good enough if *A*'s words produce, when interpreted in the context of the remainder of *A*'s own beliefs—i.e. what remains of his belief state after the given belief has been removed from it—the very same belief that was just removed. But of course this condition isn't always fulfilled. When *B*'s assumptions differ importantly from *A*'s own, *A* will have to take this difference into account and rely in choosing his words, not on his own assumptions but on what he takes *B*'s assumptions to be. If *A* is ignorant or mistaken about those assumptions, it may well happen that the message his words signify for *B* are quite different from the effect they would have had on *A* himself had he believed all he does believe except for the particular belief he is trying to transmit. Various kinds of communication failure are likely to result in such cases. Precisely what kinds of failure arise through what kinds of utterances in the face of what kinds of discrepancies between the beliefs of speaker and hearer is a topic that deserves careful investigation, and that ought to hold a prominent place in any theory of verbal communication, (and especially one along the general lines sketched here). For what I have to say about the attitudes in this paper, however, the topic need not be explored further. (It is important in relation to the analysis of attitude reports, but since I won't have room to deal with attitude reports in the present paper, and have relegated them to its sequel, the analysis of communication failure can wait for a later occasion too.)

verbalization algorithm must have a rule for 'I' which is not the inverse of the interpretation rule for 'I'.[39] To see this let us reflect on what the two rules should be like.

The verbalization rule for 'I' is extremely simple:

(12) 'I' may (again excepting occurrences within quotation contexts) only be used as designator for the individual represented by *i*.

In contrast, the interpretation (= DRS-construction) rule for 'I' is somewhat more complicated, and resembles fairly closely the processing principle for demonstratives which we discussed in Sections 4.1–4.3. To process an occurrence of 'I' in a sentence that reaches you, you must introduce a new discourse referent and anchor it by means of the descriptive information which the rule associates with 'I', viz. something like "whoever is, or was speaking (writing) what I am now hearing (reading)". Precisely how much descriptive information you will get out of this depends, as in the cases involving (4) which we discussed earlier, on the kind of access you have to the person who produced the sentence. If you face each other there will be a lot; when you read a letter or an autobiography there is often much less; and there are cases, as when you find a letter that was written by you know not who, nor to whom nor when, where it isn't even clear that the beliefs you form about the author on the basis of what you read could be construed as *de re*. But what matters is that in all these cases the new belief involves a formally and perhaps also externally anchored discourse referent for the presumed source of the utterance or text you interpret.

The asymmetry between the interpretation rule and the verbalization rule for 'I' suggests that there is more to the matter than the principle "'I' always refers to the speaker" reveals. To put it in Lewis' terms, the person who uses 'I' uses it to express self-attribution. But what an utterance containing 'I' conveys to the hearer is not a self-attribution; rather, it is an attribution to some external individual, albeit one to which he will typically stand in a relation that qualifies the attitude as *de re*. There is an intimate connection between the meaning of 'I' and the special access that we have to ourselves, but this connection is restricted to the context of language production. For the interpreter the word 'I' is much like a third person demonstrative such as 'that man' or the deictic use of 'him'. All these expressions lead to discourse

[39] This is one of the reasons why I qualified my claim, saying that the procedures are "roughly" inverses of each other.

referents which normally can be anchored descriptively, and often are connected with discourse referents that are already present in the processor's context DRS.

With 'you' the story is much the same, only reversed. 'You' also bears a special relationship to *i*. But here it is the construction rule, and not the verbalization rule that must exploit the special relation to the self. A minor difference between 'I' and 'you' is that the construction rule for 'you' is a little more complicated than the verbalization rule for 'I', since it must distinguish between the case where the interpreter identifies the utterance as addressed to him, and that where he takes it to be addressed to someone else.[40] It is the rule covering the first of these two possibilities that resembles the verbalization rule for 'I'.[41] It can be succinctly stated as:

(13) Represent the referent of 'you' by means of *i*.

The verbalization rule for 'you' resembles the construction rule for 'I' in that it is concerned with discourse referents other than *i*. It has some such form as:

(14) If *z* represents the person you are addressing then use 'you' for the individual *z* represents.

So much about the role of *i* in characterizations of propositional attitudes. But what about its role in representations of linguistic utterances? The policy that seems to suggests itself here is entirely parallel to the one we adopted for *n*: *i* always stands for the one who produced the represented utterance. This means that in the context of discourse representation the construction rule for 'I' should not be the one described above, but rather the inverse of the verbalization rule for 'I':

(15) Always represent the first person pronoun by means of the discourse referent *i*.

I trust that in view of what has been said already it is evident at this point that the formal similarity between this new rule and the earlier verbalization rule for 'I' is largely illusory.

Let me summarize what I have done in this section. In Section 4.1 we discussed external anchors as a device for capturing the truth-conditional

---

[40] I am speaking here only about the singular pronoun. The analysis of plural 'you' is somewhat more complicated.

[41] The rule for the other case treats the pronoun as some kind of demonstrative; I will make no attempt here to state it.

content of discourse involving direct reference. In Section 4.2 we noted that an expression can be directly referential in intention without actually succeeding in (directly) referring. We introduced formal as opposed to external anchors to express this distinction. This apparatus, comprising external and formal anchors, can be used, we found in Section 4.3, to represent a parallel distinction within the realm of propositional attitudes: An attitude can be, just like an utterance, directly referential in intention—or, as we put it, formally *de re*—without being directly referential in reality, or truly *de re*. Precisely when a formally *de re* attitude does succeed in being truly *de re* as well is, I surmised, a question to which there may be no conclusive answer.

Section 4.4 dealt with indexical reference. We discussed the roles of two "indexical DRS elements", *n* and *i*, each of which has its own unique cognitive status. I argued that the DRS constituent *i* must be carefully distinguished from the pronoun 'I'. The two are connected in verbalization, but not in interpretation. In fact, *i* has as much to do with 'you' as with 'I', and in interpretation it is the first, not the second of these two connections which matters. To be able to articulate these connections we needed to take a first step towards a DRT based account of verbal communication.

5. Communication and the Intersubjective Dimension of Meaning

5.1. *Internal Links*

Among the expressions of natural language that have been claimed to function in a directly referential manner there are its proper names.[42] Whether proper names do indeed give rise to attitudes that are formally *de re*; if so, whether these formally anchored attitudes will also be truly *de re*; and, if only some of them are truly *de re*, under what conditions; these are important questions. But they have, in one form or another, been subjected to massive discussion in the philosophical literature, and I do not want to make the mountain any bigger, as I would inflate it sideways, not upwards.

The feature of proper names that I do want to discuss is related to their directly referential character, but it is nonetheless a different one, which can be discussed only within a setting in which language is seen as a means

[42] In fact, the claim, first made in Kripke's *Naming and Necessity* (Kripke 1972), that names are directly referential seems to antedate similar claims about indexicals and demonstratives; but history is delicate in this particular domain, partly because it is sometimes difficult to decide whether certain claims really count as claims of direct reference.

of communication, not simply as a symbolic system which bears semantic relations to the world or subject matter that it is about.

So let us return to our simple-minded model of verbal communication, involving the two participants *A* and *B*, with their respective attitudinal states $K(A)$ and $K(B)$. Suppose that *A* utters a sentence containing proper names, e.g.

(16) Mars is more massive than Venus.

Let us assume that this bit of information is new to *B*, but that he takes *A*'s word for it, and thus that he adopts a new belief with the content of (16). According to our model this involves his adding to $K(B)$ a new component of the form $\langle Bel, K_1 \rangle$. I will assume that *B* already knew about Mars and Venus before the present exchange; in other words, that formally anchored discourse referents *x* and *y* for these two planets were already present in $K(B)$. Secondly, I assume that *B* understands 'Mars' and 'Venus' as proper names in Kripke's sense, so that K_1 contains formally anchored discourse referents to represent the bearers of these names. Finally I make the (DRT-internal, and conceptually unimportant) assumption that the discourse referents introduced for the occurrences of 'Mars' and 'Venus' in (16) are identified with *x* and *y*. From these assumptions it follows that *x* and *y* are components of the addition to $K(B)$ occasioned by (16) and that both are formally anchored.[43] So, in light of what has been said about language processing and attitude formation so far, the formally anchored DRS K_1 will be of the following form:

(17)

x y	$Anch[x]$	$Anch[y]$
x is more massive than y	bearer of the name *Mars*(x)	bearer of the name *Venus*(y)
	$V_1(x)$ ⋮ $V_r(x)$	$W_1(y)$ ⋮ $W_s(y)$

If one reflects a little more carefully on what goes on in utterance interpretation, however, it is plain that even verbal communications of this exceedingly simple sort give rise to a good deal more in the way of attitude

[43] The directly referential view of proper names seems to entail that the discourse referents *x* and *y* are also externally anchored (to the planets Mars and Venus). I am ignoring these external anchors as they are irrelevant to the central issue of this section.

formation than is captured by (17). To begin with there is the well-worn pragmatic wisdom that when *A* says something to *B* and *B* accepts this as (probably) true, that acceptance is, in all normal cases, grounded in two separate assumptions, (i) that *A* is sincere, i.e. that he has a belief corresponding to what he is saying, and (ii) that *A* knows what he is talking about, i.e. that his belief is likely to be true. Let us suppose that the communication involving (16) is of this kind, and let us concentrate on the first of the two assumptions that, according to received pragmatic doctrine, underlie the acceptance of what one is told; let us, that is, ask: What would *B*'s adopting the belief that *A* himself has a belief expressed by (16) come to within our model of communication?

It is clear that the theoretical vocabulary introduced thus far does not provide everything that a representation of this "second order belief" requires. So let us begin by filling this gap. We need to represent the belief content that *A* has a certain belief with a certain structure and content given by (16). I propose that we take this literally: There is an object of a certain kind, viz. a belief, which is held by *A* and which has certain properties relating to its structure and content. Glossed in this way the content requires for its representation:

(i) a discourse referent p representing a belief;
(ii) a discourse referent a representing the speaker *A*;
(iii) a condition to the effect that p is a belief of a; and
(iv) one or more conditions which specify the required structural and contentual properties of p.

Condition (i) demands that we add to our representational vocabulary discourse referents $p, q, r, \ldots, p_1, p_2, \ldots$ for attitudes; for us, for whom attitudes are components of attitudinal states, this means that the new discourse referents must be mapped onto such state components when the DRSs in which they occur are evaluated for truth or falsity.

Condition (ii) is unproblematic; discourse referents for persons have been available all along. We just note that in the case we are discussing *B* is bound to have a discourse referent for *A* which is both formally and externally anchored. Again, we will ignore the external anchor.

Condition (iii) will be given as $Bel(a, p)$. In general, we introduce for each mode indicator M which our theory admits a corresponding 2-place predicate M which takes as first arguments discourse referents representing beings endowed with consciousness and as second arguments discourse referents representing components of attitudinal states.

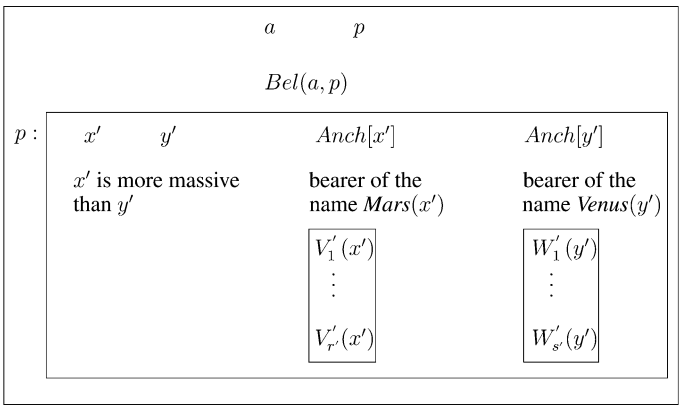
And as for (iv), we specify the structural and contentual properties of p in one fell swoop, viz. by specifying p to have the structure of a certain

formally anchored DRS. Graphically we do this by writing "$p : K$" where K is the specifying DRS. All that we have already said about the significance of DRSs in the specification of attitudinal content and structure can then be invoked to determine the truth conditions of conditions of this new form. More about this anon.[44]

Our earlier commitments determine virtually without remainder how this new notation is to be applied to the case under discussion. In particular, it seems clear that the DRS K which is to characterize the content of the belief which B attributes to A should be obtained from (16) by the same process as the representation of B's own belief of (16). Thus K will be isomorphic to that representation, for processing the sentence twice will yield DRSs that are isomorphic copies.

Proceeding along these lines we get for B's belief that A holds a belief he has expressed by (16) the representation:

(18)



Our account so far of the changes which B's reception of (16) occasions in $K(B)$ seems to come to this: B adds to $K(B)$ not only his own belief of (16), viz. $\langle Bel, (17)\rangle$; he also adds a belief to the effect that A holds a similar belief. According to what we have just said, this should be (18). Thus the total change will consist in adding both these pairs. It seems to me, however,

[44] Note that by adopting the notation "$p : K$" we are introducing an infinity of predicates in one go; in fact, since conditions of this form will themselves occur as components of DRSs, the addition of such conditions, in the sense in which we intend it here, leads to a recursive generation of DRSs containing such conditions on the one hand and on the other conditions in which such new DRSs occur on the right hand side. In this way we have the formal means of expressing iterated belief (and other attitudes) of any finite depth.

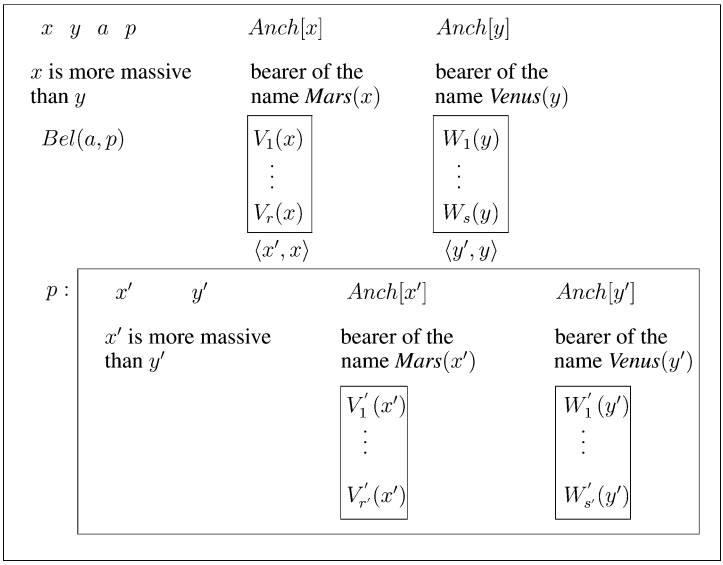
that this representation misses something important. The intention behind (18) is that *B* attributes to *A* a belief with the very same content as that which he himself adopts. The representation we have just obtained appears to warrant this, as the two contents are given by isomorphic DRSs (viz. (17) and the DRS following "p :" in (18)). And yet there is some aspect to the identity of these two contents that this isomorphism fails to capture. In situations such as this one, in which one adopts a belief that one also attributes to someone else, the belief one adopts and the belief one attributes are, I contend, conceived as the same also in the sense of being *about the same things*. The word "conceive" is crucial here. It is not just that the belief *B* adopts and the one he attributes to *A* are both about the actual Venus and the actual Mars, *B intends* that they be about the same things, and insofar as this intention is part of how he understands the two beliefs to be related, it ought to be captured by our representation of the over-all change to his attitudinal state.

The device we will use to represent this aspect of the identity of the two beliefs is that of what we call *internal links*. Formally an internal link is simply a pair $\langle x', x\rangle$ of two discourse referents which, roughly speaking, stipulates that these discourse referents represent the same individuals. In particular, with the use of internal links the change to $K(B)$ we have described can be represented as in (19). But what exactly could this stipulation, that x' and x (and y' and y) "represent the same individuals", mean? What, to put the question differently, could be meant by the intention that a belief of another person and one of one's own beliefs are about the same things?

The answer to this question that I want to propose embodies what is perhaps the most radical of the various theses this paper puts forward. It is a thesis which I cannot hope to prove, but which has come to impress itself upon me as a crucial component to any account of the relation between thought and communication. Let me state it in summary form and then try to elucidate it:

(A6) In many communicative situations we form beliefs and other attitudes that we intend to be about the same things as certain beliefs (or other attitudes) which we attribute to some other person. Such an intention carries the commitment that the content of the adopted attitude count as verified only in those situations which also verify the identifying conditions that the other person associates with the individuals he takes his attitude to be about.

(19)



To express what (A6) tries to say in a more precise and comprehensible form I need to make use of the theoretical machinery we have been developing. Suppose that the attitude adopted by *B* is characterized by a DRS K_1, that the one which is attributed to the other person, *A*, say, has a content characterized by K_2, that K_1 contains the discourse referent x and K_2 the discourse referent x', that x' has in K_2 the anchoring DRS $K_{x'}$, and that the individual represented by x for *B* is intended to be the same as that represented by x' for *A*. Then only those correlations *F* of objects with the discourse referents of K_1 will count as verifications of the content of *B*'s attitude which have the property that correlating $F(x)$ with x' yields a verification of $K_{x'}$.

Intentions of this sort render the contents of our own attitudes dependent on the attitudes of others, and so are responsible for a certain social dimension to the content of thought. In a case like the one we have been discussing in this section this effect is not visible. For, at least according to what we assumed about the case, the discourse referents x and y which represent Mars and Venus for *B* are formally anchored to these two planets, and the same is true for the discourse referents x' and y' with which x and y are linked and which, according to *B*, represent Mars and Venus for *A*. So *B*'s commitment that whatever satisfies his belief content must also satisfy *A*'s formal anchors for x' and y' does not add anything new to the truth conditions that his own belief would have had anyway. For the only correlations

that could have verified that belief would have had to associate Mars with x and Venus with y in any case.

But it isn't always like this. Often the type of commitment described in (A6) does make a real difference to the truth conditions of the attitude with which it is associated. Typical cases are those where A uses an indefinite description which B takes to have been used specifically. Suppose for instance that A says to B:

(20) There is a man in my village who collects fossilized anteaters.

More likely than not B will take A to be talking about some particular man he knows, one for whom he, A, has a formally and externally anchored discourse referent. If so, and if B adopts what he is told as a belief, he is likely to associate with this belief the commitment that it be about the man the speaker had in mind when he said (20). In this case the link between the discourse referent x representing the fossil collector in B's new belief and the discourse referent x' playing the analogous role in the belief B attributes to A will make a difference. Without it the content of B's belief could in principle be satisfied by any man in A's village with the mentioned hobby (For all we know the place may have a secret society of fossilized anteater collectors, a circumstance of which A, even though he is from there, is ignorant). But with the commitment that only the man whom A had in mind qualifies as a verifier of B's belief this possibility is excluded. For only the man A had in mind will fit A's formal anchor for that man, and the internal link then guarantees that this man will also count as the only verifier of B's own belief that there is such a collector in A's village.

But is this really right? We are bound to ask this question when we recall what was said at the end of Section 4.3. There I suggested that the recipient of an utterance will under certain circumstances be induced to attach to a discourse referent x a condition saying something like "x is what the speaker referred to with his use of the expression e" or "x is what the speaker had in mind when he used e". Would such conditions not have precisely that effect on the truth conditions of the newly formed attitude that are now being attributed to internal links?

The answer to this last question is of course: Yes. But that does not really detract from the central point I am concerned to make. For, obviously, the effect of conditions of this kind would be precisely that they make the attitudes of which they are part dependent on the attitudes of others. Indeed, the device of internal anchoring can be seen as a first and limited explication of conditions which, like those informally described at the end

of Section 4.3, create referential connections between one's own and others' attitudes.[45]

In Section 4.3 we were concerned with the question what kinds of utterance situations can give rise to formally anchored beliefs. In particular, we wondered whether there could be formal anchors which are supported by nothing more than identifying conditions which pass, so to speak, the burden of identification to someone else. By adopting (A6) I have in effect committed myself to a positive answer to the last question. For linking can produce precisely the effect of formally anchoring a discourse referent of the recipient's attitude via anchoring conditions that are directly accessible only to the speaker.

From the two illustrations of (A6) we have thus far considered it might appear that internal links always have this effect—that of making a formal anchor of one person available to another. But not all cases in which (A6) is operative are like this. It is also possible for *B* to establish an internal link with some discourse referent belonging to an attitude he attributes to *A*, but for which *A* (while having the attitude attributed to him) does not have a formal anchor (and for which *B* does not assume *A* to have such an anchor either). A typical sort of case where this seems to happen is that where *A* tells *B* of his future plans—say, his plan to buy a plot on the Riviera, build the bungalow of his dreams on it and then retire there. *B* may find this a wonderful idea, and start making plans of his own, for instance that of paying *A* an extended visit once he is properly set up. It seems to me plausible to maintain of this case that the intention which *B* forms, that of visiting *A* in his future bungalow, comes with a link between the discourse referent that represents the bungalow in *B*'s own plan with the discourse referent representing the bungalow in the plan that he now attributes to *A* (i.e. the plan that *A* has just revealed to him). *B*'s plan is, we might say, contingent on that of *A*, and this contingency manifests itself in part by his making the value of his discourse referent dependent on whatever may turn out to be the value of the corresponding discourse referent in the plan of *A*. Nevertheless it may be assumed that

[45] In fact, in a more careful and formal definition of the theory as we have (informally) developed it up to this point, in which one spells out in full detail what DRS conditions are admitted, conditions of the sort discussed towards the end of Section 4.3 would not be admitted: The only permitted mechanism for making the values of discourse referents dependent on the attitude of others would be the internal links. Refinements of that theory might extend this mechanism and introduce notation that could be seen as corresponding more closely to the problematic conditions.

A's discourse referent for his future bungalow is not formally anchored. In fact, there is nothing that it could be anchored to. The bungalow is at this point only a dream, a dwelling of no fixed place, size or form, something with no other identity than that which accrues to figments of the imagination.

In a case like this *B* evidently cannot make the value of his discourse referent x for the bungalow dependent on the formal anchor that *A* has for the corresponding discourse referent x', for there is no such anchor. So the dependence must take a different form. My intuition about the present case is that it is the conditions attached to x' in *A*'s plan that will provide the support for the link. After all, *B*'s plan presupposes the realization of that of *A*. But should all the conditions from *A*'s plan be taken on board? In general this would not seem necessary. For instance, *A* may have been quite explicit to *B* on one or two points of detail, and it is conceivable that on those points his plan will later come to be modified (e.g. because he will find that those aspects of the original plan are not practical, or through a simple change of heart). *B* may envisage this possibility and while conceiving his own plan as contingent on that of *A*, yet not conceive of it as contingent on those details. In fact, it appears as if the question precisely to which conditions he should see himself as committed need not be fully determined. Whether there are any general principles for identifying the conditions (some or all of them) that *B* should commit himself to in cases such as this one I do not know.[46]

5.2. *Sharing of Discourse Referents and Common Knowledge*

The instances of referential dependence which we discussed in Section 5.1 were all one-directional: One person, *B*, makes the values of some discourse referents dependent on the attitudes of some other person, *A*. In some situations in which this happens it is all that happens—more precisely, the dependencies of *B*'s discourse referents on the attitudes of *A*'s

[46] Another complication arising with the unanchored cases is that a good representation will in some way have to mark the conditions (from *A*'s attitude) which are to support the link *B* establishes. One way to do this would be to associate with any link $\langle x', x\rangle$ a DRS $K_{\langle x', x\rangle}$ *supporting* it. This DRS would have to be drawn in its entirety from *A*'s attitudinal state. In the cases where x' is formally anchored for *A* this DRS would be $K_{x'}$. In the cases where there is no formal anchor some more complicated procedure, which apparently could not be fully determinate, would be operative for culling $K_{\langle x', x\rangle}$s from *A*'s attitudinal state.

are the only dependencies of one person's attitudes on those of another to which the communication gives rise. More often than not, however, what happens is more than this. Not only do some of *B*'s attitudes become dependent on those of *A*, some of *A*'s become conversely dependent on attitudes of *B*. The principal reason for this has to do with a feature of face to face communication which I did not mention so far, but which is as familiar as the principle I offered as motivation for (A6) (the principle that the hearer typically adopts the belief that the speaker believes what he says). When, in oral communication the hearer does not overtly object to or doubt what the speaker has just said, then the speaker assumes by default that his words have been accepted and thus that the hearer now has a belief reflecting what he has just told him. Thus, in the standard case the speaker comes away with a belief that the hearer has a belief corresponding to what was said, and the belief that the hearer believes that he, the speaker, has a belief reflecting his own words is equally available to him. It is an insight of long standing that these are the first steps in an unbounded hierarchy of mutual attributions of ever more complicated, nested attitudes. It is a default assumption, operative in the standard case, that this entire hierarchy is in principle available and, moreover, that this assumption is itself one which speaker and hearer share.

There exist a number of theories which articulate the semantic and logical implications of this shared assumption, theories of, as it is usually called, *common knowledge*. These theories concentrate on common knowledge whose objects—in the sense of what is (commonly) known—are propositions. To that topic I have nothing to contribute here. However, there is a side to the phenomenon of shared information which relates directly to the issues that we have been talking about and which theories of common knowledge have so far ignored: The participants in a discourse may not only come to share the knowledge of, or belief in, certain propositions; the very identity of those propositions may be dependent on the very commitment which the sharing involves.

An example may help to explain what I mean. Consider the following case. *A* and *B* plan to write a joint paper. They once talk about this project, then go their own ways, and think more about what shape the paper should take. After that they talk about the project off and on. After a while their ideas about the paper diverge, but nevertheless it remains, in the minds of both, a joint project. Each knows that the other has a number of beliefs, desires and intentions relating to the paper. Or, put in the language of this essay, each attributes to the other a complex attitudinal state which contains a discourse referent representing the paper and which occurs as

argument in a variety of conditions, some but probably not all of which are known to his partner. It is, I claim, an essential feature of such a situation that the two participants take their respective discourse referents for the paper to be linked, in the sense that they have a joint commitment to count any correlation of one of those discourse referents with a real object (in this case with some real paper written at some later date) also as a correlation for the other discourse referent. Thus verification of *A*'s attitudes cannot be detached from verification of those of *B* nor vice versa. I will call this kind of joint commitment *sharing of* (the relevant) *discourse referents*.

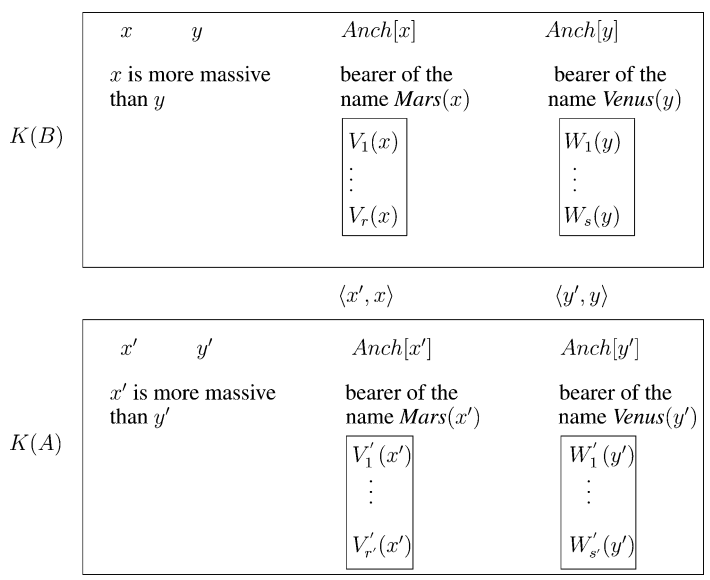
This is not to say that either *A* or *B* will have to see his own beliefs about the paper as verified only in those cases where his partner's beliefs would be fulfilled too. On the contrary, it is in virtue of the joint commitment which *A* and *B* share that they can see conflicting beliefs about the paper as actually incompatible. Suppose for instance that *A* believes the paper will be more than 30 pages long and that *B* thinks it will be less, and that they reveal these beliefs to each other. These two beliefs are obviously incompatible, and *A* and *B* will surely see them as such. But their incompatibility presupposes that they are about the same thing. And if there is any sense in which they are about the same thing this can only be because *A* and *B* are committed to construing them so.

The point is a perfectly general one. But for the possibility of sharing discourse referents we could rarely, or at any rate in far fewer cases than we actually do, see ourselves as holding views that disagree with those of others; there would be virtually no scope for disagreement and debate. Perhaps there is no better way of realizing how pervasive the phenomenon of sharing discourse referents must be.

The description of shared discourse referents cannot take quite the same form as that in which we have represented internal linking so far (as for instance in (19)). For that representation is, as the term "internal link" implies, internal to the attitudinal state of one particular individual. What we need is a way of saying of two or more individuals that certain discourse referents which figure in their respective attitudinal states are shared by them. As this is a claim that is made "from the outside" about the two attitudinal states together, our device for expressing it will have to be external to each of the two states too. Once this is realized, however, nothing seems to speak against using the same notation, involving pairs of discourse referents, to indicate sharing as well as internal linking; the place where the pair appears will make it clear how it is to be understood. The external claim (our claim) that *A* and *B* share the belief that Mars is more massive than Venus

(in the sense that they take their beliefs to be about the same planets) may be represented as:[47]

(21)



Here the pairs $\langle x', x\rangle$ and $\langle y', y\rangle$ stand outside the representations of both attitudinal states. In view of this they must be understood as external assessments of subjects and their states, which involve not only the individual psychologies of those subjects but also certain causal relations between them (having to do with their present and past communications and more generally with their interactions within the larger information exchanging community of which they are part), just as external anchors must be understood as claims based not only on a person's psychology but also on the causal relations in which he stands to the things towards which his thoughts are directed.

However, for reasons paralleling those which led us to the principle (A5) we may surmise that from a description such as that of (21) an essential bit is missing. What is crucial for sharing of discourse referents is that the participants *take* themselves to be sharing those discourse referents. Precisely how this psychological aspect of discourse referent sharing is best represented is something about which I have no definitive views, and what I shall propose in this regard should be seen as tentative.

---

[47] Again ignoring what else may be part of the attitudinal states of *A* and *B*.

In trying to decide how the psychological aspect of discourse referent sharing is best represented we do well to distinguish between two types of cases. The first, which is really a special subcategory of the second, is exemplified by the example we have been discussing and which found its (incomplete) representation in (21). Here the participants *A* and *B* take themselves to share a particular attitudinal content identified by a structure specifying the discourse referents that are taken to be shared. Both *A* and *B* have a commitment to this effect; and since these commitments may be assumed to be identical in form, let's concentrate on just one of them, say *A*'s.

It is clear that the condition expressing *A*'s commitment must specify (i) the content shared, (ii) the discourse referents occurring in this content that are shared, and (iii) who are doing the sharing. How these components should be represented follows more or less directly from the commitments already made: The content should be specified as a DRS, and the shared discourse referents can then be selected from that DRS. As regards (iii): One of the participants is *A* himself; in line with Section 4.4 he should be represented by the discourse referent *i*. The other participant will have to be represented by an "ordinary" discourse referent; in general *A* will, in the cases giving rise to shared discourse referents, have a formally and externally anchored discourse referent *b* representing his partner *B*.

What remains is the choice of a predicate that takes these different components as arguments. I propose the predicate "*Shared*". In general the condition expressing the subject's commitment to regard the content specified by the DRS *K* and the set *D* of discourse referents occurring in *K* as shared by him with the individual *b* will have the form

(22) $\mathit{Shared}(D, b, i) : K$

For the special case that was incompletely described by (21) the two relevant instances of (22) are as follows:

(23)

(A): $\mathit{Shared}(\{x'', y''\}, b, i)$:

x''	y''
x'' is more massive than y''	

(B): $\mathit{Shared}(\{x''', y'''\}, a, i)$:

x'''	y'''
x''' is more massive than y'''	

Evidently, somebody holding that he is sharing a certain attitude with someone else must have this attitude. This imposes an obvious constraint

on representations of attitudinal states in which conditions of the form (22) occur: Whenever the representation contains an instance of *Shared* (D, b, i) : K, then it must also contain a component $\langle M, \langle K', A\rangle\rangle$[d] where K' is an isomorphic copy of K.[48] Moreover, the representation should make explicit that K and K' represent the content of one and the same attitude. This means that each discourse referent in K' must be internally linked to the corresponding discourse referent of K.

Let us see what this means for the particular case of our example. To complete the description given in (21) in the spirit of what has just been said we must do the following: (i) we must add $\langle Bel, K_A\rangle$ to $K(A)$ and $\langle Bel, K_B\rangle$ to $K(B)$, where K_A is the DRS consisting only of the condition (23 (A)), and similarly for K_B. (ii) we must add all the internal links, i.e. $\langle x'', x'\rangle$ and $\langle y'', y'\rangle$ must be added to $K(A)$ and $\langle x''', x\rangle$ and $\langle y''', y\rangle$ to $K(B)$. If we once more ignore the mode indicators of the "sharing beliefs" (which in this as in earlier cases is unproblematic as all components are beliefs and thus have the same mode) we arrive at the following expansion of (21) shown in (24).

Just as we stipulated, in (A5), that the description of a complex attitudinal state is incoherent unless for each external anchor there is a corresponding formal anchor, so we stipulate now that there can be no coherent representation of a system of attitudinal states belonging to distinct persons which contains claims of discourse referent sharing (i.e. pairs $\langle x, x'\rangle$ occurring outside the representations of each of those persons' states) unless each of the states that make up the system contains a corresponding instance of (22). Note that this stipulation makes sense provided each claim $\langle x, x'\rangle$ concerns discourse referents x' for person A and x for person B such that A and B share an attitudinal content K in which these discourse referents occur. In that case the representation of A's state will contain a component of the form $\langle M, K_A\rangle$ and that of B a component of the form $\langle M', K_B\rangle$ where both K_A and K_B are isomorphic copies of K, and thus also of each other, and where the isomorphism between the K_A and K_B carries x' over into x. That A and B take themselves to share this content is then manifest through a corresponding instance of (22) involving further copies of K internally linked to K_A and K_B, respectively.

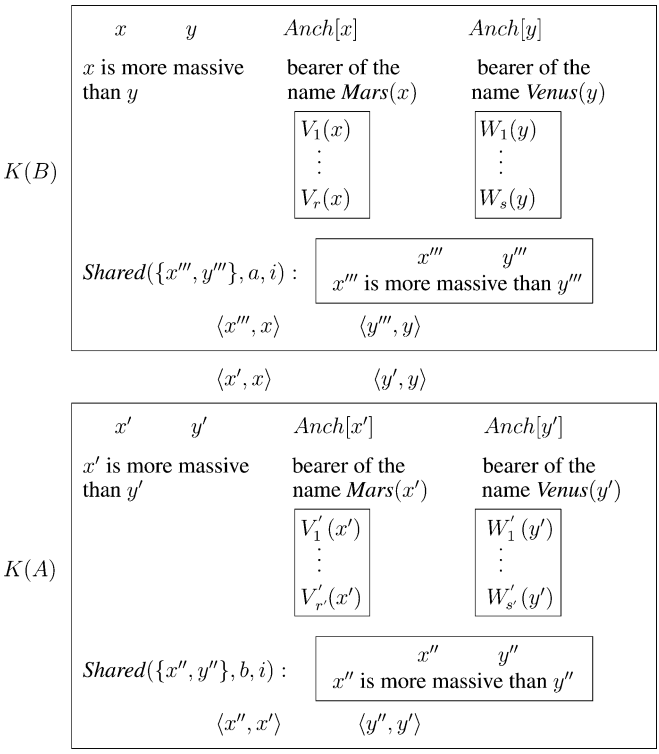
---

[48] By an isomorphic copy of a DRS K we understand any DRS which can be obtained from K by replacement of discourse referents.

---

[d] Here 'M' is a 'dummy' attitudinal mode operator. In applications 'M' will be 'Bel', 'Des',... depending on whether what is shared is a belief, a desire, etc. In the more specific context of the paper only the case where 'M' = 'Bel' is directly relevant.

(24)



The general case is that where *A* and *B* share discourse referents but where there need not be a corresponding sharing of an attitudinal content of which these discourse referents are part. What I said about disagreement suggests that such cases do arise and in fact that they are quite frequent. At the moment I do not see how best to handle these cases. In fact I am not persuaded that such cases actually occur. Perhaps a more careful analysis of sharing will reveal that where there is sharing of discourse referents there is also always a shared propositional content, even if the epistemically or pragmatically most prominent attitudes in which those discourse referents figure conflict.[49]

[49] Sharing lies at the heart of the notorious Hob-Nob problem: How are we to explain that the sentence "Hob believes that a witch killed Cob's cow and Nob thinks she has blighted Bob's mare" does not commit the speaker to the existence of witches? In a nutshell, the solution to this puzzle that I would vote for is along the following lines. Members of a given community, in the present instance Hob and Nob, may come to share a discourse referent for

5.3. *Sharing and Proper Names*

I have argued that sharing discourse referents is a pervasive feature of human thought and that it arises typically (if not exclusively) through verbal communication. But I have said very little about the mechanisms through which verbal exchanges give rise to shared discourse referents. To provide a detailed account of these mechanisms appears to be a difficult task, which on the one hand requires an explicit theory of common knowledge, and on the other careful analyses of the contributions made by particular parts of speech to the interpretation of the utterances that contain them. There is no question of carrying out such a program here. But there is one aspect of it about which I want to say a few words. This is the question—which I raised at the beginning of Section 5.1 but which then disappeared from sight—how sharing arises through the use of proper names.

There is only one example that we have discussed in any detail and that is directly relevant to this question, viz. (16). We concluded that the acceptance of (16) would, under normal conditions, lead to a belief in which the discourse referents for the names 'Venus' and 'Mars' are shared with the speaker. But it never became fully clear what role the names themselves played in the process which produces the sharing, because the discourse referents for 'Venus' and 'Mars' were subject to so many other constraints as well: formal and external anchoring, as well as internal linking to discourse referents already present. In fact, the sharing might, in this particular case, be simply a side effect of linking; the new discourse referents might be shared through being linked to discourse referents that were shared already.

It is not hard to see however that proper names yield shared discourse referents also where there can be no question of their inheriting this status through linking. The clearest cases are those where a name is new to the

---

a witch that they take to cause mischief in the region even if no such witch actually exists. This enables them to have beliefs "about the same witch", even if most of their beliefs concerning her aren't even the same. It is this kind of situation, involving two different beliefs concerning an individual that does not exist but a representation for which is nevertheless shared by the believers, that a sentence like the one above is designed to report. That a sentence of this particular form can be used to report just such a situation does of course not follow from these considerations. The principles which explain why the sentence might be used in this way, and precisely what sorts of situations can be described by such sentences, involve details pertaining to the interpretation of indefinites and the mechanisms of pronominal anaphora that we cannot go into here. See Asher (1986) for a way of spelling out these details.

recipient, and its referent unknown to him. Suppose for instance that you tell me about Thomlinson, a colleague of yours, who has, say, just won an important prize. I have never heard of Thomlinson, but I believe what you tell me. The belief I adopt in this case I will surely take to be about the very man that I take you to have the same belief about. And as this is exactly what you expect me to do, you will form the belief that I now have the belief you expressed, etc. This, it seems, is the general way with names. It is a default assumption that the recipient of a name will take the content of the utterance to be directly about the name's referent; consequently, if there is nothing to overrule this assumption, both speaker and recipient will adopt it and sharing is the result.

The power to produce sharing of discourse referents contributes importantly to the "social stability" of meaning. Thus the use of a proper name will produce among its users a shared presumption that they all use it to denote the same referent. Each time the name is used and gives rise to a shared discourse referent this presumption is extended or reinforced.

But it is not only proper names that exercise this stabilizing influence. Much the same appears to be true of common nouns. Indeed, traditional terminology ("proper name"—"common name"; "proper noun"—"common noun") has a point which the modern logical distinction between individual and predicate tends to obscure. Common nouns behave in many respects like proper names, and there are a variety of reasons for treating them as introducing discourse referents, representing the kind or concept that the noun denotes, just as proper names do. These discourse referents (which a more sophisticated version of DRT than I have sketched here would incorporate) are subject to sharing in much the same way as those which derive from proper names: A common noun too is used with the presumption that all use it to denote the same concept.

In fact, the sharing phenomenon may well be even more general and extend also to other concept words besides nouns, for instance to verbs or adjectives. For all such words sharing appears to be instrumental in maintaining an intersubjective uniformity of meaning; and for many a word, I suspect, this uniformity is about all the meaning it has.

6. Semantics, Attitude Reports and Self-Reflection[50]

6.1. *The Content of Attitudinal States*

One of the obligations of a theory of propositional attitudes is to account for their intentionality. I acknowledged this obligation in the introduction, and the proposals I made in the earlier parts of the paper do discharge it, inasmuch as the DRSs they involve come with an explicit truth definition (which I did not spell out in detail here, but which can be found in other work on DRT). The increasingly complex representations of attitudinal states, however, to which we were led as we went along, contain notational devices for which no standard truth-conditional treatment yet exists. This is one of the gaps that need filling.

I will not try to do this here, but limit myself to pointing out what appear to be the two main problems that make filling that gap non-trivial. The first of these has to do with the fact that connected components of complex attitudinal states often differ from each other in mode. Thus, as we saw in Section 3, desires are often connected with beliefs and with intentions. I have urged that when attitudes are connected, i.e. when they share one or more discourse referents, we cannot assess their truth conditions in isolation, but only those for the entire connected cluster of which they are part. I did not say, however, how that was to be done. I may have implied that it could be done along the lines of what was said in Section 2 about the truth conditions of the two-sentence discourse (1). But this really makes sense only when all the attitudes in the cluster are all of the same mode, e.g. if all of them are beliefs. Where we have, say, a connected pair consisting of a belief and a desire, truth conditions for their combined content, which treat them as if they were two beliefs—make little intuitive sense.

As a matter of fact this is not just a problem for an attitudinal theory like the one I have outlined. It is a problem that will have to be faced by any theory of practical reasoning which wants to base its logic on a plausible semantics. Where only belief and desire are involved, a sensible semantic

---

[50] For reasons of time as much as of space this and the next section have been kept much shorter than they ought to be. Properly expanded versions of these sections will be included in a sequel to the present paper.

([added 2013] Strictly speaking the 'sequel' referred to in this footnote never made it into print. But as indicated in the introduction to this section of the book, a number of suggestions that are made in the present paper were subsequently worked out in formal detail. See Kamp (2003) and Kamp et al. (2011)).

analysis appears to be that according to which the beliefs single out a certain set of possibilities, or possible worlds—those compatible with the beliefs—while the desires select from that set a certain subset, those that are compatible with the desires as well. But it is not obvious how such an analysis could be extended to incorporate further modes. To my knowledge this problem is for all its importance still very far from being solved.

The second obstacle relates to higher order attitudes. These we have been representing by means of the discourse referents $p, q, \ldots$ constrained by conditions such as $Bel(b, p)$ and $p : K$. The first problem here has to do with the discourse referents $p, q, \ldots$ These stand, I said, for parts of attitudinal states (I haven't been very careful in my use of the term "part"; but let us for definiteness' sake agree that a part of state is either a single component or a collection of components). So, it is attitudinal states that should be correlated with p, q, etc. in the evaluation of the DRSs containing them.

So we are committed to the existence of attitudinal states. I do not mind this. That people and other conscious creatures are in complex and ever changing attitudinal states—I would not know what it would mean to deny this. But even if we accept attitudinal states and their parts as real, and also accept that these states are structured in the way I have suggested, there still remains a problem, which is connected with the conditions of the form $p : K$. There are infinitely many such conditions for each of the discourse referents $p, q, \ldots$, one for each K. And for most of the conditions $p : K$, K will be complex. Suppose now that we want to evaluate some such condition $p : K'$ and that we have correlated with p the actual attitudinal state s. What is it for s to satisfy $p : K'$? The naive answer here is of course: "s satisfies the condition if and only if s in fact has the structure identified by K'". But this seems unsatisfactory; for in case K' is complex, one would like to reduce the claim that s has the structure it specifies to simpler structural claims; what one would like is a *compositional* account of the satisfaction of such state predicates.

How could such a reduction be accomplished? I do not know. If we were entitled to think of attitudinal states as literally put together from their structural constituents, just as a sentence is put together from its grammatical constituents, the reduction would therewith be given. But, as I said in the introduction, this is an assumption to which I do not think one is entitled. If on the other hand the justification for ascribing structure to states is strictly functional, then I do not see how a plausible, non-circular compositional account for the satisfaction of these predicates might be given.

6.2. *Attitude Reports*

Up to this point I have refrained from saying anything about an aspect of the theory of the attitudes that is often seen as the central part of such a theory, and that many see as the only feasible part. This is the semantics of attitude reports. As I explained earlier, I see the matter differently: Only when a sufficiently detailed theory of the attitudes is in place, can we tackle the theory of attitude reports with any hope of success. The practical consequence of this conviction for the present essay is that the treatment of attitude reports too must be relegated to its sequel. For now that the necessary foundations have been laid time and space are up! All I want to do in this subsection is (i) indicate in the briefest outline how the theory of attitude reports builds on what has been discussed here, and (ii) devote a few words to the notion of an attitude report *de re*.

On the first point I will be very brief indeed. The general strategy is to see an attitude reporting sentence, in the spirit of DRT, as a set of instructions to the recipient for constructing a representation of the reported attitude. Thus an attitude report will produce in the recipient an attitudinal state with the same kind of structure that we used, in Section 5, to represent beliefs which a recipient forms about the beliefs of the speaker. The details of the strategy concern the contributions which the grammatical constituents of the report make to the construction of this structure. Much of the interest of this enterprise derives from the possibility of showing that, as a rule, constituents make the same contributions when they occur inside attitude reports (more precisely, in the complement clauses of attitudinal predicates) as they do when occurring outside such contexts.

Particularly revealing in this connection are the construction rules for different types of NPs. These rules have a special importance within the analysis of attitude reports because they relate directly to the distinction between *de re* and *de dicto*. As this distinction has played such an important part in the philosophical literature on the attitudes, and since I have been using the term *de re* here in a somewhat different context, I want to make a couple of remarks to clarify the connection. Roughly speaking an attitude report *de re* is the report of a *de re* attitude. However, our analysis of the notion of an attitude *de re* showed that it is necessary to draw a number of distinctions which other approaches have overlooked. Since the attitude reported by a given attitude report can be *de re* in either of the senses we distinguished (i.e. truly *de re* or only formally *de re*) the same distinction is applicable to attitude reports. Moreover, we saw that besides the structural features that render attitudes *de re*, i.e. their anchors, there are

related features, linking and sharing. These too can be transferred from the attitudes themselves to the sentences which report them. Thus we find ourselves with a much richer repertoire of distinctions than the comparatively simple opposition between *de dicto* and *de re*. The resolution of many of the puzzles about belief that have bemused and amused the philosophical community during the past two decades turns on these finer distinctions. Another important ingredient in the treatment of those puzzles are the construction rules for the different kinds of NPs, which determine what kinds of attitudes (anchored, linked, shared) can be reported by a sentence whose complement contains that kind of noun phrase.

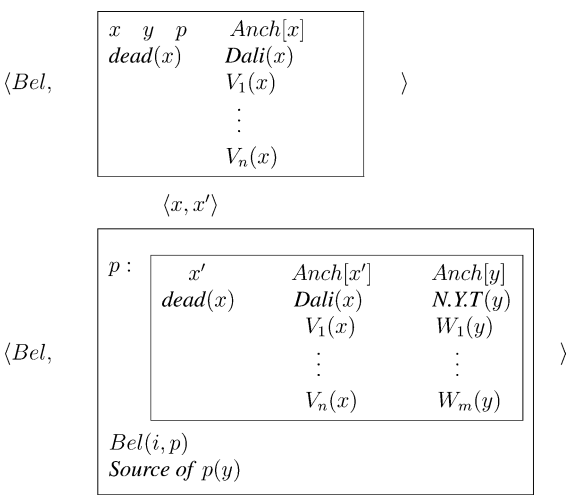
6.3. *Self-Reflection*

At long last we come to the second part of our proposal, back in Section 3, for representing the factors that partition attitudinal states into their components. Let us focus on an example we already considered in Section 3, that of someone, let us call him Bill, who believes that Dali is dead, and also that he got this belief through reading *The New York Times*.

It may well seem that all we need to represent these two beliefs in a form which agrees with the suggestions made in Section 3 is now in place: We have discourse referents for attitudes, DRS-conditions characterizing the structure and content of those attitudes and conditions to express the relation in which an attitude stands to the one whose attitude it is. With the help of these notational devices the relevant part of Bill's belief state may be represented as shown in (25).

I do not think, however, that this is quite the right representation. The problem with (25) is that it fails to convey the direct connection that I want to suggest exists between Bill's primary belief (that Dali is dead) and the second order belief about it. The connection is recoverable insofar as the structural specification of p matches that of the primary belief. It also seems reasonable to assume that even in the total belief state of Bill's the primary belief displayed in (25) will be the only one that matches the specified structure of p. And for all I can imagine it may well be a general feature of attitudinal states that they never contain more than one attitude with the same content. If this is indeed so, then representations like (25) will be adequate to the extent that they will always permit the unique identification of the primary attitude that the second order attitude is about.

(25)



Nonetheless, I can't help feeling that this would be a case of getting the right results by the wrong means. For I can't help feeling that the link between the second order belief and the primary attitude it is about has the same kind of directness that is distinctive of, for instance, the relation in which we stand to the present time—that relationship which informs our beliefs about what is happening now. In neither case is the connection between the believer and the relevant element that enters into the content of the belief secured descriptively.

A representation of second order attitudes which does justice to the directness of the connection should treat them as pointing directly to the primary attitudes they concern, as if one part of the complex attitudinal state to which both the primary and the second order attitude belong could "directly see" the other part. The simplest formal realization of this idea takes the form of a function which maps each attitude discourse referent that acts as an argument of a secondary attitude onto that component of the state which the secondary attitude is about. (This function fixes once and for all the attitude that might be correlated with the discourse referent in evaluation.) Thus we come to the final proposal of this paper.

First an auxiliary definition. Let *AS* be a complex attitudinal state, as characterized in (A5). By the *set of self-reflecting discourse referents of AS* we understand the set of those attitude discourse referents p such that (i) p occurs in the universe of a component of *AS* and (ii) p occurs in a condition of the form $\mathrm{M}(i, p)$ (where M is a mode indicator and i is as always the special discourse referent for the self.)

Our final proposal for the structural characterization then comes to this:

(A7) A complex attitudinal state can be represented as a pair consisting of (i) a structure *AS* as described in (A5) and (ii) a function from the set of self-reflecting discourse referents to components of *AS*.

As before, there is no reason why attitudinal states of this new, extended type could not be the subject of attribution (by one person to another). To do justice to this intuition we would have to characterize the structure of attitudinal states recursively, using (A7) as one of the clauses in that definition. It is intuitively clear, however, how such a definition would go and carrying out the formal details would not reveal anything of interest.

References

Asher, N., 1986. Belief in Discourse Representation Theory. *Journal of Philosophical Logic* 15 (2), 127–189.

Asher, N., 1987. A typology for attitude verbs and their anaphoric properties. *Linguistics and Philosophy* 10 (2), 125–197.

Barwise, J., 1985. *A model of the treatment of anaphora in situation semantics*. CSLI Working Paper No. 1. University of Stanford, Stanford, CA.

Carnap, R., 1947. *Meaning and Necessity. A Study in Semantics and Modal Logic*. Chicago University Press, Chicago, IL.

Chastain, C., 1975. Reference and context. In: K. Gunderson (ed.), *Language, Mind and Knowledge*. University of Minnesota Press, Minneapolis, MN, 149–270.

Church, A., 1950. On Carnap's analysis of statements of assertion and belief. *Analysis* 10, 97–99.

Evans, G., 1980. Pronouns. *Linguistic Inquiry* 11, 337–362.

Fodor, J., 1987. *Psychosemantics. The Problem of Meaning in the Philosophy of Mind*. The MIT Press, Cambridge, MA.

Groenendijk, J., Stokhof, M., 1991. Dynamic predicate logic. *Linguistics and Philosophy* 14, 39–100.

Halliday, M., Hasan, R., 1976. *Cohesion in English*. Longman, London.

Heim, I., 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. dissertation, University of Massachusetts, Amherst, MA.

Kadmon, N., 1987. *On Unique and Non-unique Reference and Asymmetric Quantification*. Ph.D. dissertation, University of Massachusetts, Amherst, MA.

Kamp, H., 2003. Einstellungszustände und -zuschreibungen in der DRT. In: U. Haas-Spohn (ed.), *Intentionalität zwischen Subjektivität und Weltbezug*. Mentis, Paderborn, 209–289.

Kamp, H., (submitted). Using proper names as intermediaries between labelled entity representation. *Erkenntnis*.

Kamp, H., van Genabith, J., Reyle, U., 2011. Discourse Representation Theory. In: D. Gabbay, F. Guenthner (eds.), *Handbook of Philosophical Logic*, 2nd Edition. Vol. 15. Springer, Berlin, 123–394.

Kaplan, D., 1989. Demonstratives. An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. In: J. Almog, J. Perry, H. Wettstein (eds.), *Themes from Kaplan*. Oxford University Press, Oxford, 481–563.

Kripke, S., 1972. Naming and necessity. In: D. Davidson, G. Harman (eds.), *Semantics of Natural Language*. Reidel, Dordrecht, 253–355.

Kripke, S., 1979. A puzzle about belief. In: A. Margalit (ed.), *Meaning and Use*. Reidel, Dordrecht, 239–283.

Lewis, D.K., 1969. *Convention: A Philosophical Study*. Harvard University Press, Cambridge, MA.

Lewis, D.K., 1979a. Attitudes *de dicto* and *de re*. *The Philosophical Review* 88, 513–543.

Lewis, D.K., 1979b. General semantics. In: D. Davidson, G. Harman (eds.), *Semantics of Natural Language*. Reidel, Dordrecht, 169–218.

Perry, J., 1977. Frege on demonstratives. *The Philosophical Review* 86, 474–497.

Zeevat, H., 1989. A compositional version of Discourse Representation Theory. *Linguistics and Philosophy* 12, 95–131.

TEMPORAL REFERENCE INSIDE AND OUTSIDE PROPOSITIONAL ATTITUDES[a]

Hans Kamp

1. Summary

In this paper I discuss one application of a theory whose general features have been presented elsewhere.[1] The core of this theory is a formalism, with well-defined syntax and (model-theoretic) semantics, for the representation of propositional attitudes, complex mental states consisting of several connected attitudes, attitudinal change (i.e. the change from one attitudinal state to another), and attitudes shared by several agents. A second, though closely related purpose of the formalism is that it can serve as part of a semantics of attitude attributing sentences of natural language. In designing the formalism special attention was given to the internal connections between the contents of different attitudes. In the context of attitude attribution these connections manifest themselves as problems for the semantics of discourses consisting of several attitude attribution sentences, which either attribute different attitudes to the same agent at the same time, or different attitudes to the same agent at different times, or attitudes to different agents at the same or different times.

The one example to which the present paper is devoted is a "mini-discourse" consisting of two sentences which ascribe attitudinal changes to the same agent at two distinct times. The example illustrates two recurrent features of multi-sentence attitude attributing discourses. On the one hand it exemplifies the by now familiar fact that in a sequence of two or more

[1] See Kamp (1990, 2003); Kamp et al. (2011).

[a] Editors' note: This article was originally published in K. von Heusinger & K. Turner (eds.), *Where Semantics Meets Pragmatics*, Elsevier, Amsterdam, 2006, 493–525 (Current Research in the Semantics / Pragmatics Interface 16—now published by Brill). We carefully edited the article and corrected some typos. We would like to thank Katja Jasinskaja for careful proofreading and constructive comments.

© hans kamp, 2013 | doi:10.1163/9789004252882_022
This is an open access chapter distributed under the terms of the cc by-nc-nd 4.0 license.

attitude attributions the later attributions tend to rely for their interpretation on the attributions which precede them. This is a phenomenon that also arises when the attributions concern one and the same time (as well as one and the same agent) and it is for such cases that it has been identified and discussed in the literature.[2] But in addition the example illustrates a number of issues that have to do with temporal reference and time. Some of these arise at the level of the single sentence. This is true in particular of the question how the tenses of the complement clauses of attitude attributing matrix verbs are semantically related to the tenses of the matrix verbs themselves. But there are also time-related questions that concern the way in which the two attitude attributing sentences are connected, as parts of a single cohesive piece of discourse. On the one hand these have to do with the circumstance that the sentences of our example attribute attitudinal *changes* and on the other with the fact that these changes are said to have occurred at different times. These temporal issues add a new dimension to the general problem how earlier attributions can provide interpretation contexts for later ones.

The theory which I will apply to the example of the paper is an extension of Discourse Representation Theory as it is presented in for instance Kamp & Reyle (1993). More accurately it is a combination of two extensions. One of these concerns the representation of attitudinal states and the semantics of that part of language which is used to describe such states; the other is presupposition. This second extension too has been presented elsewhere.[3] Abridged presentations of both extensions can be found in Kamp et al. (2011).

I will assume some basic knowledge of DRT, roughly corresponding to Chapters 1, 2 and 5 of Kamp & Reyle (1993). But in Sections 3.1 and 3.2 below I give brief introductions to the two mentioned extensions, which should give the reader enough to understand the treatment that follows in Section 4 of the example that this paper is about. The remaining sections of the paper are all quite short. Section 2 presents the example and lists the problems connected with it on which we will focus. Section 5 sums up and restates the principal morals.

---

[2] See Stalnaker (1988).

[3] Kamp (2001a, 2001b).

2. The Example and the Issues It Raises

The example which is the topic and focus of this paper is given in (1):[4]

(1) a. On Sunday Bill heard that Mary was in Paris.
b. On Tuesday he learned that on the previous day she had left.

The issues raised by this example that will preoccupy us in what follows all have to do with time. Temporal matters are relevant to the interpretation of (1) for three distinct reasons.

(i) Both attributions are attributions of attitudinal change—i.e. of a change from one attitudinal state to another—expressed by the verbs *hear* and *learn*. As opposed to verbs like *believe*, *want*, *intend*, *regret* and others which are used to attribute a single attitude to their subject at or over a single period of time, such attitudinal change verbs have not received much attention either in the philosophical or the linguistic literature on attitude attribution.

(ii) Like any other attitude attributing sentences with a matrix verb and a *that*-complement, the sentences in (1) raise the problem how the time of the event or state of the complement clause of a matrix verb relates to the time of the matrix event or state. It is a problem which strictly speaking arises for any combination of tenses in main clause and complement clause. But when both clauses are in the present tense, as has typically been the case for the examples discussed in the philosophical literature, the problem can be side-stepped by treating all predications as holding at the utterance time. In the linguistic literature, on the other hand, this problem has been discussed extensively. It is an issue to which I will have nothing new to add in this paper, and I will implicitly rely on the proposals that are documented in the cited publications.[5]

(iii) Each of the sentences of (1) attributes to Bill an attitudinal change: after the change Bill believes something which he didn't believe before it. As usual in cohesive discourse, the interpretation of the second sentence requires the interpretation of the first as context. Moreover, the interpretation of the second sentence must rely on the interpretation of the first in a way which is specific to sentences of this particular kind: The interpretation of the complement clause of the second sentence, which describes the content of the newly acquired belief, relies on the interpretation of the complement clause of the first sentence, which describes the content of the belief acquired on the occasion spoken of in that sentence. But in the case of (1) there is a further complication. The contextual background that is needed to interpret the second complement clause must provide information about the beliefs

[4] A closely similar example is also discussed in Kamp et al. (2011), though not as extensively as here.

[5] See for instance Ogihara (1994, 1996); Abusch (1997); Kratzer (1998); von Stechow (2003).

> which the subject Bill holds at the time—some time on Tuesday—when the belief described in the second complement clause is acquired. But the first sentence doesn't tell us anything about Bill's attitudinal state at that time; it only says something about his attitudinal state on the preceding Sunday. Thus, in order to obtain the kind of background information that can support the interpretation of the second complement clause it is necessary to extrapolate from what the first sentence tells us about Bill's beliefs on Sunday to what his beliefs may be expected to have been at the relevant time on Tuesday. Like other bridging inferences the ones we will be considering have a good deal of plausibility. But their plausibility rests on discourse-independent knowledge, and so they introduce into the analysis of examples like (1) an element which is not found in cases where all attitudes attributed to a given subject by a succession of attributing sentences are attributed at one single time.

Sentence combinations like (1) are interesting not simply because they illustrate these different temporal aspects of the semantics of attitude attributing discourse, but also because of the ways those aspects interact with each other as well as with non-temporal interpretation aspects. Such interactions between different interpretation principles are among the most difficult challenges for semantic theory. One of my aims in the discussion of (1) in Section 4 will be to make this interactional dimension of the example clearly visible.

3. Preliminaries

This section outlines the two extensions of "standard DRT" that were mentioned in the Summary. The extension required for the representation of attitudinal states will be outlined in 3.1, that needed for the treatment of presupposition in 3.2. The presentations will proceed by example and remain largely informal and illustrative. For formal definitions and further technical details the reader should consult the documents that are cited in the Summary.

3.1. *DRT-Based Representation of Attitudinal States*

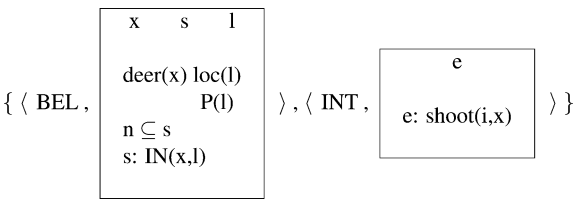
Consider the following scenario. Our subject Bill goes deer hunting in the company of his father-in-law. After reaching the forest where they hope to find what they are looking for they separate, taking different forks of the path which they have been following. At some point Bill sees (or at least he thinks he sees) a deer hidden in the undergrowth and barely visible. The effect of this visual experience is the belief that there is a deer in the location on which his eyes are focused; and this belief gives rise—not

surprisingly given the purpose of the expedition—to the intention to shoot the deer he sees (or thinks he sees). Belief and intention are propositional attitudes, each with its own "propositional content". But these contents are *referentially connected* in that they are targeted on the same individual—the deer that Bill takes himself to be seeing.

The connection between these two contents is reminiscent of the referential connections between the contents of successive sentences of discourse or text, where the later sentences may be anaphorically linked to the earlier ones. When two "propositional" contents are linked in this manner, it is in general not possible to understand them as each determining a self-contained proposition. In the original versions of DRT this difficulty was finessed by focusing on the propositions expressed by the conjunctions of sentences with referentially linked contents. But in the present context this 'solution' is not open to us: since the two attitudes are of different "mode"—one is a belief and the other an intention; thus they play distinct roles within the mental state of which they are components, and each plays the role it plays partly because of its own individual content—their contents must be separately identified. Kamp (2003) and Kamp et al. (2011) present a formal semantics for such representations in which linked content representations receive separate but nonetheless referentially connected semantic values. Here I will only show how the formalism is applied to the case of hunter Bill.

Combinations of propositional attitudes are represented as sets of representations of the individual attitudes, where the representation of each individual attitude consists of a 'mode indicator'—which determines whether the represented attitude is a belief or an intention etc.—and a DRS representing its content. Referential links between the attitudes that make up such a combination are represented by discourse referents that are shared between different content representations. For example, the combination of the described belief and intention of Bill can be represented as in (2)

(2)



Legenda: The content of the belief representation is that currently there is a deer x at a certain location l, which is further specified as satisfying certain characteristics P, about which more below. That the state of x being

in location l is a current state is indicated by the condition "n $\subseteq$ s", which says that the state s includes the time indicated by the indexical discourse referent n—as a constituent of the representation of the content of the given belief, n refers to the time at which this belief is being entertained. Note that it follows from this that if an agent entertains an attitude of such a form over some extended period of time, then the truth-conditional import of its content will change as time goes on. Suppose for instance that t1 and t2 are two distinct times within this period during which Bill entertains the belief represented in (2). Then at t_1 his belief has the content that the deer x is at the location l at t_1, whereas at t_2 its content is that x is in l at t_2. This phenomenon, of changing truth conditions determined by a constant representational form, will be important in the analysis of our example (1).

The content of the intention of (2) is that Bill, represented by the 'self-representing' discourse referent i, shoots x. The referential link between intention and belief is captured by their sharing the discourse referent x. In this case the semantics of Kamp (2003) and Kamp et al. (2011) treats the content of the belief as determining an independent proposition and the content of the intention as presupposing the belief content. The intuitive idea behind this is that it is only in combination with the belief that the intention has any real meaning.[6]

With regard to the visual experience which produces in Bill the two attitudes represented in (2) we can distinguish three possibilities: (i) there is a particular object he sees—which thus is the actual cause of his visual perception—and that object is a deer; (ii) there is an object which Bill is seeing, but that object is not a deer; (perhaps it is Bill's father-in-law, whose path has reconverged with Bill's and who is using the bushes for cover in pursuit of his own quarry); (iii) there isn't any object at all that is the perceptual cause of Bill's visual experience; the experience is an optical illusion, not a proper object perception in which the perceived object is at one end of the causal-perceptual chain and the resulting representation in the mind of the perceiver at the other end. Note that Bill's experience is in principle compatible with each of these three possibilities. But for someone who describes the case from his own, external point of view it may nevertheless be possible—and also important—to distinguish between them. (Which of

[6] In general, when two content representations K_1 and K_2 share a discourse referent α which occurs in the universe of K_1 while occurring in K_2 only in argument positions of conditions, then the content of K_2 is treated as presupposing the content of K_1.

the three possibilities is true can be quite important for what happens when Bill turns his intention into action).

Our representation formalism is designed to capture these distinctions. They are represented by means of *internal* and *external anchors*. In case there is an object which produces in Bill the experience of seeing it (through the kind of causal interaction that is typical of veridical visual perception), then the resulting discourse referent x is said to be *externally anchored to* this object; and in this case the content of any attitude which contains x as part of its content representation will be *de re with respect to* x—it will be a *singular* proposition, which attributes the conditions of the content representation in which x occurs to the object to which x is externally anchored. For example, if there is an external anchor for the discourse referent x in the intention representation in (2), then this representation determines the (doubly) singular proposition that there is an event of Bill shooting the object he is seeing. In the third case, where the experience involves an optical illusion, there is no external anchor.

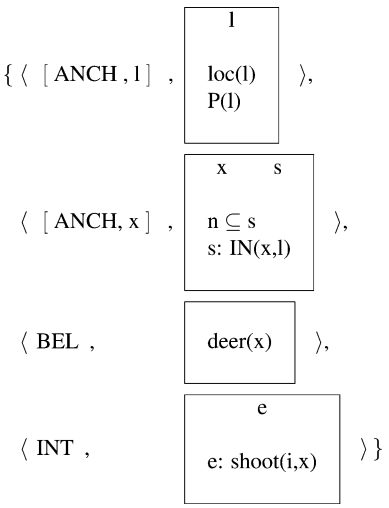
Bill's impression that he is connected by a relation of causal perception to an object which he thinks he is seeing is, we noted, compatible with each of the three mentioned possibilities. It is a further assumption of the present theory that whenever this impression is present,—i.e. irrespective of which of these three possibilities applies—it confers on the discourse referent which results from the perceptual experience a special psychological status, that of being an entity representation which is the causal product of a perception of the very object that it serves to represent. In the present theory this special status is represented in the form of internal anchors: discourse referents with this status are accompanied by internal anchors, those without this status do not. For instance, if as we have been assuming Bill attributes this special status to the discourse referent x which is the result of his visual experience, then his representation will involve an internal anchor for x (irrespective, to repeat, which of the possibilities (i)–(iii) happens to apply). In the cases (i) and (ii), where there is an external as well as an internal anchor, the propositional content of the belief will be *de re* with respect to the object to which x is externally anchored.

From a semantic point of view internal anchors act as presuppositions that their discourse referents are anchored externally. When all internally anchored discourse referents of a propositional representation have corresponding external anchors, the representation will determine the singular proposition described above. On the other hand, representations containing discourse referents with an internal but no external anchor are cases of presupposition failure and do not determine any real proposition at all. This

would be the case in particular for both the belief and the intention representation in (2) in the case of possibility (iii) (that of Bill's visual experience being an optical illusion).

As we have told the story of Bill's visual experience it is reasonable to assume that the resulting attitude complex involves internal anchors for both x and l. On this assumption, the representation of this complex in our revised formalism is the one given in (3):

(3)



Legenda: In representations of this sort internal anchors are treated as separate components of the represented attitudinal state. Each representation of an internal anchor specifies in its first component not only that what it represents is an anchor (as opposed, say, to a belief or an intention; this information is given by 'ANCH'), but also which discourse referent that anchor is an internal anchor for. The second component gives information about how the entity represented by the anchored discourse referent is conceived—as the object that is being perceived, in some particular way, and with the properties that are attributed to it as an intrinsic part of the perception. Exactly what information this second component should contain, and in what form, are questions about which I will say next to nothing here. To give a sense of the complexity of this issue let me mention just one particularly problematic case. It concerns the internal anchor for l. The conditions which make up that component should capture the spatial relations between the location represented by l and the perceiving agent Bill. How agents represent locations in their environment so that they are able to keep track of them when they move around themselves and could find their way to the repre-

sented location if and when they want to go there, is a matter with many intriguing cognitive aspects to it, few if any of which have been fully understood. I have finessed this particular problem here by assuming that in the internal anchor for l in (3) all this information is covered by the unanalysed predicate P.[7]

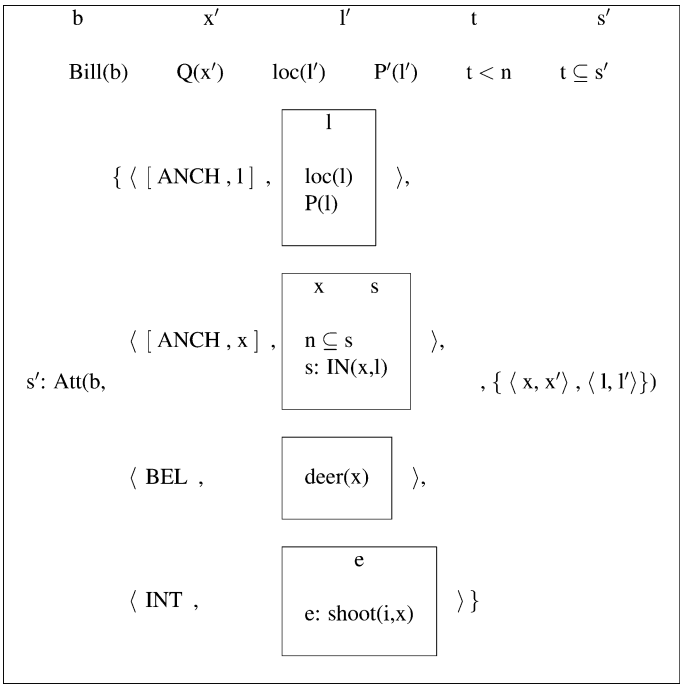
Attitude complexes like the one represented in (3) should be attributable to any agent at any time. So our formalism should have the possibility of expressing who it is that a given attitudinal state is the attitudinal state of, and when this individual is in the given state. In our formalism this is accomplished with the help of a 3-place predicate 'Att', whose arguments are (i) an agent; (ii) a part of the agent's attitudinal state, specified as a set of connected attitudes (e.g. the one given in (3)), and (iii) a set of external anchors for all or some of the internally anchored discourse referents which are part of the second argument. The time at which a given attitudinal complex is being entertained is captured in the way familiar from the treatment of temporal reference in DRT: 'Att' is a stative predicate; thus, any predication involving 'Att' takes the form of a condition 's: Att(α, β, γ)'. This condition says that s is a state which consists in α being in a mental state which minimally contains the components specified in β, and where all or some of the internally anchored discourse referents of β have the external anchors given in γ. The time at which α is in the mental state thus characterised can then be specified via further conditions involving s. In particular, that α was in this state at a given time t can be captured by the condition 't $\subseteq$ s'. Thus, the proposition that Bill was at t in the state described in (3), with external anchors for both x and l, can be represented as in (4), where x′ and l′

[7] The question what information should be represented in the content representations of internal anchors has many ramifications. For 'perceptual' anchors, as we find in (3), most of these ramifications are closely related to problems for the theory of direct perception. But as I have proposed elsewhere, internal anchors also arise for discourse referents whose origin is not one of (purported) perception, and the information content of such anchors raises questions of quite different sorts. The general problems of the use and form of internal anchors is closely connected with issues of direct reference, in thought and—derivatively—in language.

A more detailed account of the propositional content of internal anchors is needed before it will be possible to draw a motivated line between information that should be represented as part of the anchor itself and information which should be part of the belief or beliefs that come about as a result of the purported causal relation between the agent and the entity that he takes the anchored discourse referent to represent. In this paper I will distribute information between anchors and beliefs on an ad hoc basis. In some cases the division I adopt may be only one of several intuitively plausible alternatives. The matter will be of no importance to what concerns us in this paper.

represent (from a perspective external to Bill) the entities to which x and l are anchored.[8]

(4)



3.2. *Presupposition*

The second DRT extension we will need concerns the treatment of presuppositions. I here give just one example which will indicate the general features of this treatment. The sentence involved in this example—it is given in (5.a)—is closely similar to the *that*-complement of the second sentence of (1). Like most sentences of English and other natural languages it generates

[8] From the way in which I described the case represented in (4) the predicate Q cannot be determined. In fact, Q will vary depending on whether we are dealing with a case of possibility (i) or of possibility (ii). In the first case Q will be consistent with the property which Bill attributes to x′, whereas in the second case it won't be. (In that case Q could be, for instance, 'i's father-in-law'). In case of possibility (iii) there will be no external anchor for x and thus also no condition involving Q. Given the way in which I have described the mental episode represented in (4) it seems reasonable to assume that irrespective of which of the possibilities (i)–(iii) applies with regard to x, there will be an external anchor l′ for l. The predicate P′ is to be seen as short for one or more conditions which fix the location l′ in terms that are independent of the attitudinal subject Bill.

several presuppositions, and these presuppositions are ordered hierarchically in the sense that some of them presuppose others in their turn. The triggers of the presuppositions of (5.a) which are explicitly represented in the preliminary representation (5.b) are the following:

(i) The proper name *Mary*.

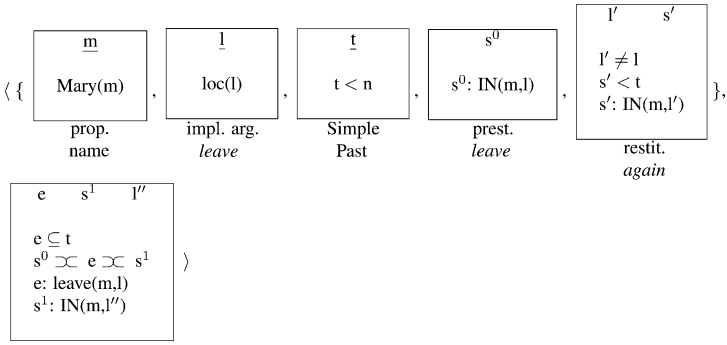
(ii) The verb *leave*. *Leave* triggers two presuppositions. One of these is connected with the implicit direct object argument of the ostensibly 'intransitive' use of *leave* in (5.a). The other presupposition is a "prestate presupposition", an instance of the general kind that is associated with all verbs of change. The prestate of an occurrence of *leave* consists in the subject being in the place that she is described as leaving and the prestate presupposition is to the effect that this state obtains at the start of the leaving event described by the given occurrence of the verb.

(iii) The past tense of *left*. In (5.b) I have followed the view according to which certain tense occurrences carry a presupposition to the effect that the time of the described state or event is identifiable from the given context. (One of the intuitions supporting such a presuppositional account of tense is that out-of-the-blue utterances of a sentence like (5.a) are perceived as strange or awkward because they lack a context which indicates what event time is intended.)

(iv) Finally, (5.a) contains the presupposition trigger *again*. The occurrence of *again* in (5.a) is one of those which allows for both a *repetitive* and a *restitutive* interpretation. In (5.b) I have opted for the restitutive interpretation, according to which the result state of the described event also held at some time before the event (and the event thus has the effect of 'restituting' that state). In the case of *leave* the restituted state is that of not being in the place that the described event is a leaving from.

(5) a. Mary left again.

b.[9]



[9] "≻≺" denotes *abutment* of two eventualities or periods of time. Thus, for instance, "s^0 ≻≺e" means that s^0 ends at the very instant when e begins.

Legenda: The top tier of (5.b) consists of the presuppositions of (5.a) while the lower tier is the non-presuppositional part. Note that some of the discourse referents occurring in the universes of DRSs which represent presuppositions are underlined. Such underlinings signify that justification of the presupposition involves finding an 'antecedent' for the underlined discourse referent. In DRT terms this means: finding a discourse referent in the context representation with which the anaphoric discourse referent is formally identified. In order that the chosen antecedent provide a proper justification for the presupposition containing the anaphoric discourse referent, the context must entail that it satisfies the condition or conditions contained in the presupposition. Presuppositions without underlined discourse referents have a purely propositional status. Such presuppositions are justified iff the context entails the propositions they express.

All presuppositions in (5.b) come with labeling subscripts. These are shorthands for constraints, determined by the type of their trigger, on the form their justification can take.[10] A proper taxonomy of justification constraints is one of the outstanding problems of current presupposition theory. In what follows I will, with few exceptions, ignore these constraints. In line with this the constraint labels will mostly be omitted.

The effect of presupposition justification on a 'preliminary' representation such as (5.b) is that the presuppositions disappear from it, though the contributions they make to the truth conditions of the representation remain in force, since these are now taken over by the context (either as it was given at the outset or altered by accommodation). The non-representational part of the preliminary representation, which is all that remains of the preliminary representation after justification, constitutes the new information which the represented sentence contributes to the discourse.

In (5.b) all presuppositions have been collected into a single set. This flat structure ignores presuppositional dependencies between the different presuppositions. Between the presuppositions of (5.a) there are several such

[10] In particular, constraints on presuppositions with anaphoric discourse referents may specify in what form their antecedents must be available in the context, or may be made available by it: Should the context explicitly contain the antecedent as it is given? Or is it enough if it can be expanded with entailments which introduce the antecedent discourse referent? In that case, what kinds of entailments are permitted for this purpose? But these are only some among the many different constraints on presupposition accommodation which our current knowledge of these matters suggests, but most of which have not yet been identified with any precision.

dependencies. For instance, the discourse referents m, l and t, all of which require resolution in context, also occur as arguments of conditions in the *again*-presupposition; so the m-, l- and t- presuppositions are not only presuppositions of the non-presuppositional part of (5.b) but also of the *again*-presupposition. These dependency relations could be made explicit by further structuring of the presupposition set. But for the purposes of this paper there is no advantage in this, and so we will make do with simple set-representations like that in (5.b).

Another simplifying feature of our example is that all presuppositions of (5.b) are located at the same attachment site of the representation. This entails that when it comes to presupposition justification, all presuppositions have access to the same contextual information. In general this is not so. Preliminary representations may also contain presuppositions that are found in different attachment sites, and the justification of these presuppositions can make use of information that is expressed by parts of the representation that have scope over these respective sites.[11]

4. The Example

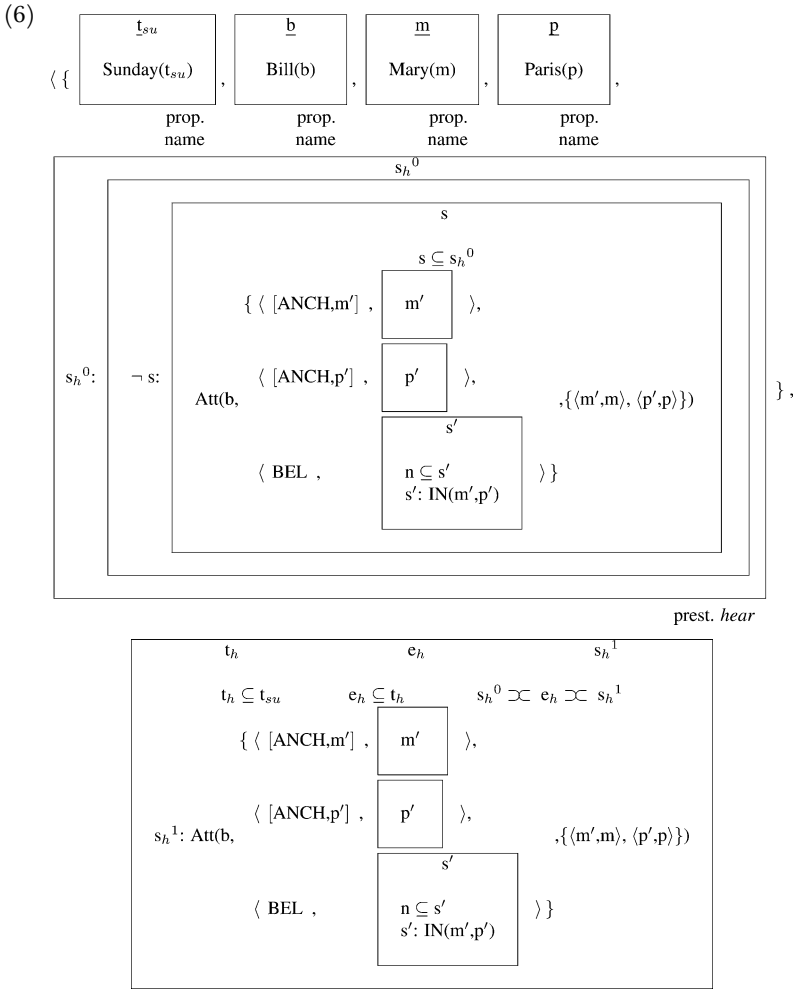
4.1. *The Representation of the First Sentence of (1)*

At last we are ready to tackle example (1). Our central concern will be the interpretation of its second sentence (1.b). The first sentence (1.a) is of interest primarily because of it provides the context for the interpretation of (1.b) and we will pass over most of the questions of detail that arise in connection with the construction of its representation. (However, many of these same questions will also come up when we discuss the second sentence in Sections 4.2 and 4.3. Having read those sections, the reader should be able to fill in most of the details which our analysis of (1.a) in the present section leaves out.)

I assume that the preliminary representation of (1.a) is that given in (6).

[11] It is along these lines that dynamic theories of presupposition account for projection phenomena: a locally triggered presupposition is not perceived as a presupposition of the entire sentence in which the trigger occurs provided it can be justified on the basis of 'contextual' information provided by the sentence itself. See e.g. Heim (1983); van der Sandt (1992); Beaver (1995) and Geurts (1995, 1999).

(6)



(6) represents Bill as having internally and externally anchored representations for the entities denoted by the proper names *Mary* and *Paris* which occur within the *that*-complement of (1.a)—that part which expresses the information of which (1.a) asserts that Bill acquired it on the given Sunday. This appears to be the only coherent interpretation of proper names occurring inside the descriptions of attitudinal states. It is important to note the difference between the two presupposition-like contributions of these name occurrences to (6). On the one hand they trigger—like any other occurrences of names, e.g. those of the proper name *Bill* and the name-like NP *Sunday* in (1.a)—the proper name presuppositions that are displayed in

the upper tier of (6). But they are special in that they also contribute the mentioned anchors. Put in more general terms: the occurrence of a proper name (or other 'directly referential' NP) in the description of the content of an attributed attitude has two distinct presupposition-like effects:[12]

(i) Like any other occurrence of a directly referential NP it carries the presuppositional requirement that the context provide identification of its referent.
(ii) As an occurrence within an attitude description it attributes to the possessor of the attitude an anchored representation of its referent.

I assume that all presuppositions of (6) can be justified in the context in which (1) is uttered, either directly or after accommodation, and that (7) is the representation which results from their justification. In (7) all information that is represented in the various presuppositions of (6) has been included as non-presupposed (and thus as on a par with the non-presuppositional part of (6)). The motivation is that after justification all information contained in the presuppositions will be available as non-presupposed information, either because it was part of the context already or because it has been made available through accommodation.

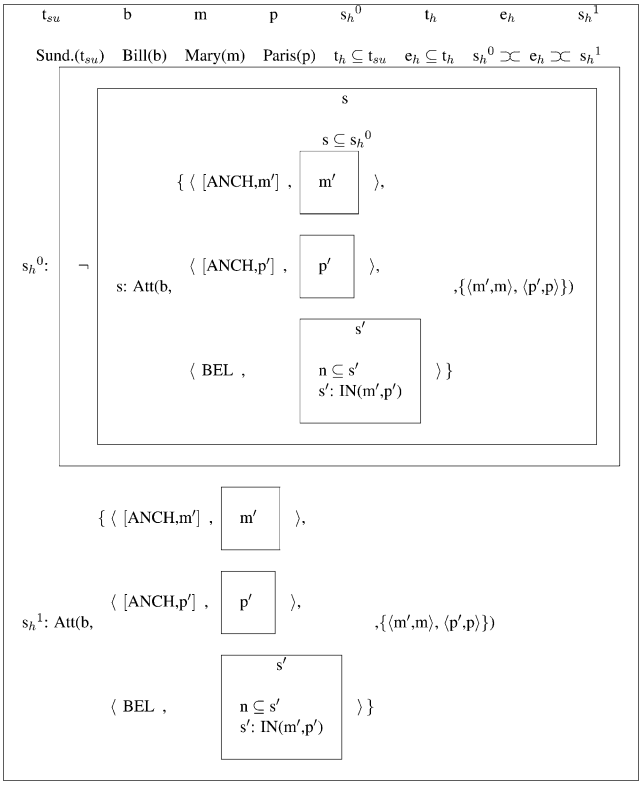
I already noted the problems connected with the propositional content of internal anchors. For the present application (as for many others) the exact content of the internal anchors hardly matters. For such applications a notation which abstracts from the content of internal anchors will be adequate. Such an alternative notation can be substantially simpler than the one we have used so far, provided there also is no need to worry about discourse referents which have internal but no external anchors—in other

[12] Comparison between (6) and (5b) might seem to show an inconsistency between their respective treatments of the past tense. In (5b) the past tense of *left* was treated as triggering a presupposition that the time of the described event can be identified in context. In (6) no such presupposition appears for either the past tense of the matrix verb or that of the embedded clause. The reasons for this deviation from what was assumed in connection with (5b) are as follows: (i) The main clause of (1.a) differs from (5a) in that it contains a 'locating adverb', viz. the temporal adverbial *on Sunday*. This adverb takes over the task of temporal location of the event described by the verb, which in our analysis of (5a) was left to the context. So the contribution which in (5b) is imposed on the context by the presupposition involving the anaphoric discourse referent t is made in (6) by the condition "$t_h \subseteq t_{su}$". (Of course, the referential NP *Sunday* gives rise to a presupposition of its own, but that is a different matter.) (ii) The tense of the embedded verb *was* is treated as anaphoric to that of the matrix verb. (See the references mentioned in fn. 5). In a more detailed analysis this anaphoric connection might be initially represented as involving a presupposition as well. For reasons of simplicity I have suppressed this further complication here.

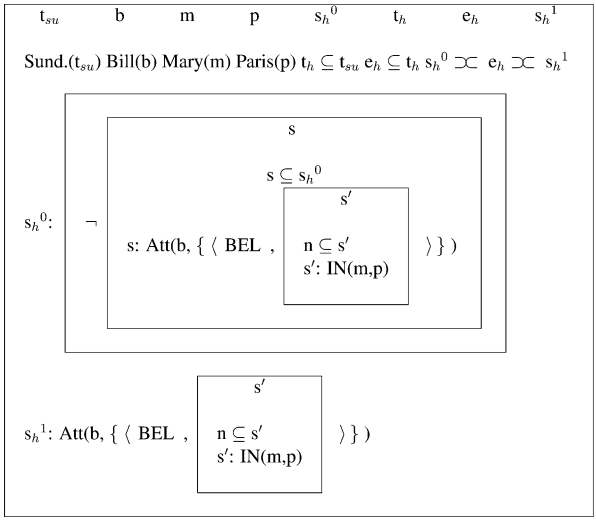
words, if it can be assumed that all internally anchored discourse referents are externally anchored as well. This is an assumption that can be made in connection with our example (1) without giving away anything of what makes the example of interest to us. In this alternative notation anchored discourse referents are eliminated, together with their internal and external anchors, and are replaced by the discourse referents representing the entities to which they are externally anchored. In other words, where the old notation has an anchored discourse referent x' with an internal anchor IA and an external anchor $<x',x>$, the new notation will have neither IA nor $<x',x>$, while the occurrences of x' in argument positions of conditions are all replaced by x. The new notation thus permits a discourse referent x which is 'declared' outside an Att-predication—i.e. which occurs in a universe that is not within the scope of the given occurrence of Att—to occur in positions that are part of this predication. The semantic import of such occurrences is essentially the same as the complexes of the old notation which they replace: Any attitude whose content representation contains occurrences of discourse referents that are declared in positions external to it is to be understood as determining a singular propositional content, which is *de re* with respect to the entities represented by these discourse referents.

Using the new notation we can simplify (7) to (8).

(7)


t_{su} b m p $s_h{}^0$ t_h e_h $s_h{}^1$
Sund.(t_{su}) Bill(b) Mary(m) Paris(p) $t_h \subseteq t_{su}$ $e_h \subseteq t_h$ $s_h{}^0 \prec e_h \prec s_h{}^1$
s
$s \subseteq s_h{}^0$
{ ⟨ [ANCH,m′] , m′ ⟩,
$s_h{}^0$: ¬ ⟨ [ANCH,p′] , p′ ⟩,
s: Att(b, ,{⟨m′,m⟩, ⟨p′,p⟩})
s′
⟨ BEL , n ⊆ s′ s′: IN(m′,p′) ⟩ }
{ ⟨ [ANCH,m′] , m′ ⟩,
$s_h{}^1$: Att(b, ⟨ [ANCH,p′] , p′ ⟩, ,{⟨m′,m⟩, ⟨p′,p⟩})
s′
⟨ BEL , n ⊆ s′ s′: IN(m′,p′) ⟩ }

(8)


t_{su} b m p $s_h{}^0$ t_h e_h $s_h{}^1$
Sund.(t_{su}) Bill(b) Mary(m) Paris(p) $t_h \subseteq t_{su}$ $e_h \subseteq t_h$ $s_h{}^0 \prec e_h \prec s_h{}^1$
s
$s \subseteq s_h{}^0$
s′
$s_h{}^0$: ¬ s: Att(b, { ⟨ BEL , n ⊆ s′ s′: IN(m,p) ⟩ })
s′
$s_h{}^1$: Att(b, { ⟨ BEL , n ⊆ s′ s′: IN(m,p) ⟩ })

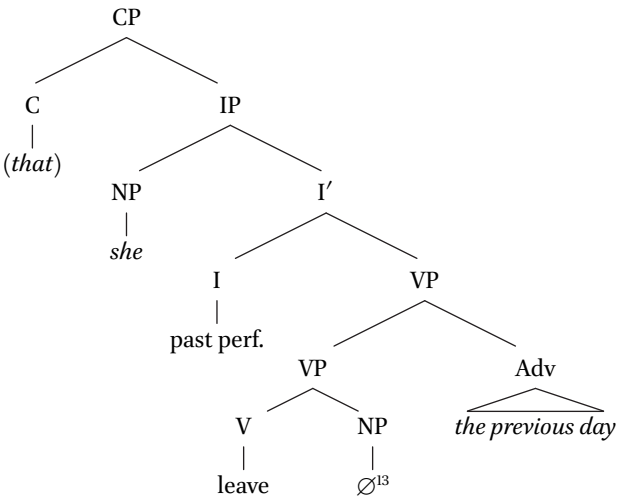
4.2. *An Interpretation for the* That-*Complement of* (1.b)

The *that*-complement of the second sentence of (1) is repeated in (9):

(9) (that) she had left the previous day

I assume the following syntactic tree for (9):

(10)



Our next task is to construct a preliminary representation of (9) on the basis of (10). More accurately, we will construct a semantic representation for the subtree of (10) whose root is the IP *she left the previous day*. As always when semantic representations are constructed from syntactic trees, the

---

[13] In (10) the 'intransitive' use of *leave* in (9) has been analysed as one that is syntactically transitive but involves a direct object that is phonologically null. At least on one occasion when I presented the material on which this paper is based this analysis has met with opposition. The opposition may well be justified: It may be that on balance we will obtain a better over-all theory of the syntax and semantics of 'intransitive' uses of verbs which also allow for a corresponding transitive use if we assume that these uses are genuinely intransitive in the sense that there is no direct object node in the syntactic structure of the clauses containing them. For my present purposes this question is not important. Anyone who prefers a syntactic analysis for (9) without a direct object node will find no difficulty in restating what I will have to say about the sentence below in a form which fits such an analysis.

Another decision which I have made here and which may offend syntactic persuasions is to describe the arguments of verbs (and other lexical predicates, although these play no role here) as 'NPs' rather than 'DPs'. In fact, the nomenclature matters little here since the internal structure of such phrases plays no role in the problems we will be discussing.

construction requires lexical entries for each of the words and semantically relevant features that occur at the leaves of the tree. So we begin by listing these entries. In doing so we follow the left-to-right order in which the lexical items occur in the IP of (10).

Our first item is the pronoun *she*. The entry which we will use is given in (11.a):

(11) a.

$$\langle \; \underset{\alpha}{\textit{she}}, \qquad\qquad \text{NP} \; \rangle$$

$$\langle \; \{ \; \boxed{\begin{array}{c} \underline{\alpha} \\ \text{pers}(\alpha) \\ \text{fem.}(\alpha) \end{array}}_{\text{pron.}} \; \}, \; \text{`}\lambda\text{P.P}(\alpha)\text{'} \; \rangle$$

A few comments before we proceed to the next entry. The upper tier of (11.a) states that *she* is an expression of category NP and that it introduces some discourse referent α. The semantics is given in the lower tier. It consists of a representation that is composed of (i) a set consisting of a single presupposition, and (ii) a non-presuppositional part. The presupposition is anaphoric with respect to its anaphoric discourse referent α which represents the pronoun's referent, and requires of α's antecedent that it be a female person. (We ignore the complications connected with the possibility of using English *she* to refer to ships, mares and certain other kinds of non-persons.) The subscript 'pron.' is short for constraints on the possible resolutions of personal pronouns. (Compare the remarks in 3.2.)

The non-presuppositional part of the semantic tier of (11.a) is a shorthand for something more complicated. The story that it abbreviates is too long to tell in full here. Since it is of marginal relevance to the central issues of the paper, I confine myself to a few hints. The condition "λ P.P(α)" as non-presuppositional component of the semantics of referential NPs such as pronouns would be just what is wanted in a theory of semantic form construction which treats the combination of a noun phrase and a predicate to which it is an argument as involving a succession of two λ-conversions, in the manner first proposed by Montague. In a theory of this form the meaning of the sister node of the NP-node (i.e. the node of the predicate which has the NP for an argument) would be assigned a semantics of the form $\lambda\beta$.K(β), with β a variable of the same type as α and $\lambda\beta$.K(β) of the same type as P. Such an account works fine so long as the outermost λ-bound variable in the semantic form $\lambda\beta$.K(β) of the sister to the NP node can be counted on to occupy within K the argument position of the NP. Making sure that this is

always so can be awkward, however, especially for languages in which there is much variation in clause-internal word order and argument phrases often occur in positions other than their 'canonical' or 'base' positions. Some form of *linking theory* is required (most evidently in such languages but also in others) to determine which NPs fill which argument slots of the predicates with which they cooccur. Once such a linking theory is in place, it can be (and has to be) used as a guide in the construction of semantic representations to determine in which slot of the argument frame of the predicate we should insert the variable (or discourse referent) that is introduced by the NP. Under such conditions it proves more convenient to represent the semantic contribution of the sister nodes of NPs as involving stores which hold the argument positions of the represented predicate that still have to be filled and to describe the operation which combines the two contributions as involving the insertion of the discourse referent (or variable) introduced by the NP into the argument slot indicated by linking theory, while removing the slot that is thereby filled from the store. A detailed account along these lines requires a fully explicit syntax as well as an explicit semantics. Sketches of a DRT-based syntax-semantics interface in which linking theory plays the indicated part can be found in several places, among them Kamp & Roßdeutscher (1994) and Kamp (2001a).[14]

Our next entry is for the past perfect. I have opted here for a treatment of the past perfect as a single tense form, instead of analysing it into smaller components (e.g. as the combination of a perfect operator and a simple past tense). The semantics of the past perfect proposed in this entry is based on the proposal for the essentially Reichenbachian treatment of the English tenses that can be found in Kamp & Reyle (1993), according to which the past perfect locates the predication expressed by the verb whose tense it is as having occurred at a time which is in the past of some "T(emporal) P(erspective) point"[15] and where this TP point is itself situated in the past of

[14] Had I assumed a syntax-semantics interface in which predicates and their arguments are combined in the manner of Montague the scare quotes around 'λP.P(α)' would have been unnecessary. Their presence is to remind us that the intended interface is strictly speaking not of this form. (I haven't bothered with scare quotes in the λ-terms in the following entries, even though similar caveats apply to their use in DRS construction. An elegant solution to the problem of "getting inside" a sequence of operators which prefixes a matrix representation is developed in Dekker (1993) within his version of Dynamic Semantics. But this is not the place for trying to adapt his ideas to the DRT-framework we are using here.)

[15] In this analysis of the past perfect the Temporal Perspective point plays the same role that is played by what Reichenbach (1947) calls "Reference Time" in his account of the past perfect.

the speech time n. The account assumes that the interpretation of an occurrence of the past perfect requires finding a suitable past time in the context which can play the role of TP point and then locating the predication as lying somewhere in the past of that time.[16] Thus the status of the TP point is presuppositional. The subscript "TPpt" in the entry (11.b) stands for special constraints to which the choice of TP points is subject.

(11) b.

⟨ *past perfect*, VP-predicator ⟩

⟨ { tp | tp < n } TPpt, λP. t | P(t), t < tp ⟩

To simplify matters further we also treat the temporal adverbial *the previous day* as if it were an indivisible lexical unit, instead of analysing it as a complex expression with internal semantic structure (i.e. as a definite description with *day* as its nominal head, modified by the prenominal adjective *previous*).[17] One important feature of the semantics of this phrase is the implicit argument of *previous*: In order to interpret an occurrence of this adverbial it is necessary to recover a day *d* from the context so that the referent of *the previous day* can be identified as the day immediately preceding t. Thus the semantics of *the previous day* also involves a presuppositional component, connected with this implicit argument.

(11) c.

⟨ *the previous day*, VP-operator ⟩

λt. ⟨ { t′ | day(t′) } impl. arg. *previous*, λP. t″ | day(t″), t″ ≻ t′, t ⊆ t″, P(t) ⟩

Our last entry is that for the verb *leave*. We only specify a semantics for the intransitive use of *leave* which is found in (9). (A general entry would contain this use as one among several alternatives, with varying argument frames

[16] This formulation fails to account for complications having to do with "sequence of tense"; since these are irrelevant to this paper, they are being ignored.

[17] For a discussion of temporal definite descriptions like this one, which takes the compositional aspect of such phrases seriously, see Kamp & Schiehlen (2001).

and varying meanings.) This 'intransitive' use comes with two presuppositions: (i) the prestate presupposition which this use of *leave* shares with its other uses (and, speaking more generally, with other change-of-state verbs) and (ii) the presupposition connected with its implicit argument.

(11) d.

⟨ *leave* (intr.),			V ⟩
	Nom	∅	
e_{lv}	x	l′	

$\lambda\{t,x\}.\langle\{$ [l̲ : loc(l)] impl. arg. *leave* , [$s_{lv}{}^0$: IN(x,l)] prest. *leave* $\},$

[e_{lv} $s_{lv}{}^1$ l′ : $e_{lv} \subseteq t$; $s_{lv}{}^0 \prec e_{lv} \prec s_{lv}{}^1$; $l' \neq l$; e_{lv}: leave(x,l); $s_{lv}{}^1$: IN(x,l′)] ⟩

(12) presents the semantic representations of the complex nodes of the IP of (10). We proceed bottom-up, from the representation for the lower VP node in (12.a) to that of the IP node in (12.d).

(12) a. lower VP node:

$\lambda\{t,x\}.\langle\{$ [l̲ : loc(l)] impl. arg. *leave* , [$s_{lv}{}^0$: IN(x,l)] prest. *leave* $\},$

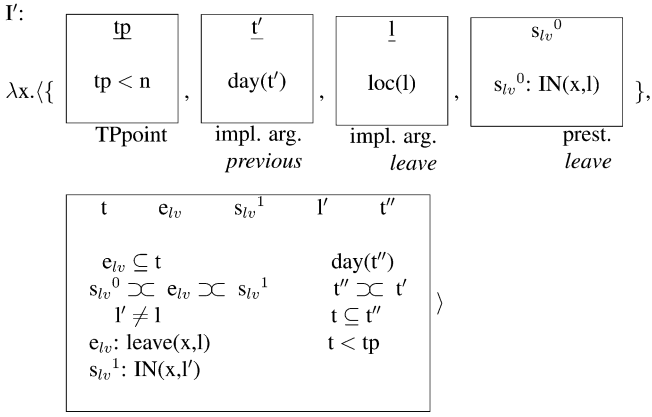
[e_{lv} $s_{lv}{}^1$ l′ : $e_{lv} \subseteq t$; $s_{lv}{}^0 \prec e_{lv} \prec s_{lv}{}^1$; $l' \neq l$; e_{lv}: leave(x,l); $s_{lv}{}^1$: IN(x,l′)] ⟩

(12) b. upper VP node:

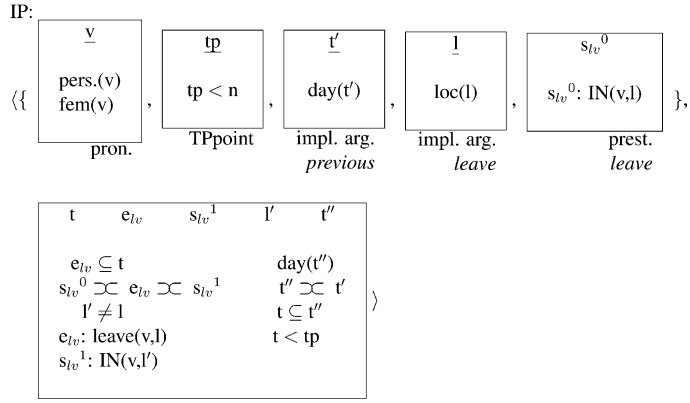
$\lambda\{t,x\}.\langle\{$ [t′̲ : day(t′)] impl. arg. *previous* , [l̲ : loc(l)] impl. arg. *leave* , [$s_{lv}{}^0$: IN(x,l)] prest. *leave* $\},$

[e_{lv} $s_{lv}{}^1$ l′ t″ : $e_{lv} \subseteq t$; $s_{lv}{}^0 \prec e_{lv} \prec s_{lv}{}^1$; $l' \neq l$; e_{lv}: leave(x,l); $s_{lv}{}^1$: IN(x,l′); day(t″); $t'' \prec t'$; $t \subseteq t''$] ⟩

(12) c. I′:



(12) d. IP:



I have refrained from stating the composition principles according to which these representations are obtained from those of their daughter nodes and refer the reader to relevant literature.[18]

(12.d) will be used as representation of the complement clause of (1.b) when we construct its preliminary representation in the next section.

4.3. *Preliminary Representation of (1.b) and Its Resolution*

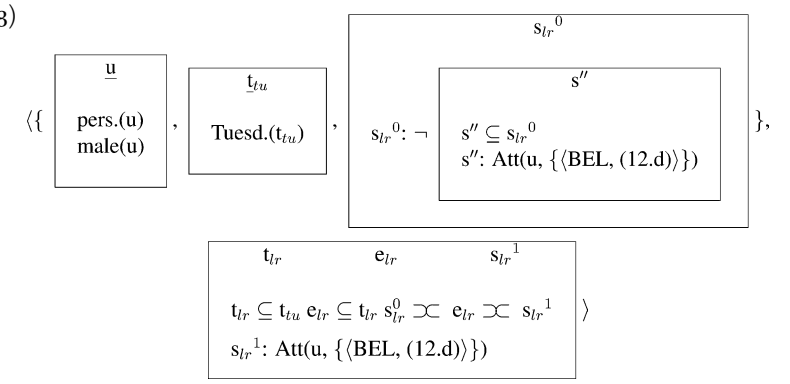
The preliminary representation for (1.b) has the same general form as that for (1.a). Once again we are dealing with a sentence which reports a change in Bill's attitudinal state. So once again the representation of the *that*-clause,

[18] There is to my knowledge no place in the existing literature where these principles are given in exactly the form in which they are needed here. Discussions of similar principles can be found in Kamp (2001a).

i.e. (12.d), enters into the representation at two places—into the representation of the belief that is part of Bill's resulting attitudinal state and into the representation of the prestate of *learn*, which consists in Bill not holding that belief.

I will forego the syntax-semantics interface details involved in obtaining this preliminary representation. The remaining steps which lead from the representation (12.d) of the embedded IP to that of the full (1.b) involve little that has not yet been illustrated in the construction of (12.d). There is one exception, viz. the binding of the embedded tense by the tense of the matrix verb. But as announced in Section 2, this is an issue where I am relying entirely on existing work of others and a proper exposition of how the embedded tense is bound by the matrix tense according to any of the existing proposals would involve us in much extra detail that would detract from our real concerns. So we have decided to let this matter rest, and we proceed straight to the preliminary representation of (1.a).

(13)



(13) has explicit representations of a total of eight presuppositions, the three displayed in (13) itself and the five of the embedded (12.d).[19,20] It is important

[19] Note well that we are dealing with 8 presuppositions, not 13. The fact that (12.d) occurs in (13) in two different places does not mean that it introduces 2×5 presuppositions into (13). Since the two occurrences of (12.d) in (13) stem from the same constituent of (1.b), each of (12.d)'s presuppositions counts as only one, which demands a single resolution. In particular, there must be single resolutions for the anaphoric discourse referents in these presuppositions, with antecedents which provide a justification of the presupposition in question that are valid for both positions in which it occurs. In fact, the matter is more complicated than this description suggests. For more on this see Kamp (2001a).

[20] It could be argued that there is one presupposition of (1.b) which (13) fails to display. The verb *learn*, it might be said, is factive in the sense that the new information which *learn* describes its subject as getting is presented as true (or 'as fact'). Factivity of stative

to note that the presuppositions are found in two different attachment sites. For justification of the three displayed in (13) the only contextual information available is the primary context provided by (8). In contrast, the justification of the five presuppositions of (12.d) may resort to 'secondary' context information as well. However, as noted in Section 1, the question how (8) can provide the needed secondary context information is one of those which embody the interest which (1) holds for us and which we will have to look into carefully.

Six of the eight presuppositions of (13) are anaphoric, with the anaphoric discourse referents u, t_{tu}, v, tp, t' and l. For two of these, u and v, the context (8) offers only one possible resolution:

(i) u := b; (ii) v := m;

The story about the other four is a little more complicated, and it is a somewhat different story for each. First t_{tu}: Adverbial uses of names of days of the week, such as that of *Sunday* in (1.a) and of *Tuesday* in (1.b), are anaphoric in roughly the following way. The interpretation of such a name occurrence involves the identification of some other day *d* which can serve as the 'origin' from which the reference of the given occurrence can be computed. The day referred to by the name is then either the first day after *d* that fits the name (e.g. the first Tuesday after *d*) or the last one before d. Which of these two is the intended referent depends on further factors, from which it is possible to tell whether the state or event described by the clause in which the name occurs is located after *d* or before it. Often the origin is the day of the utterance time. But it can also be some other contextually salient day. In the case of the *Tuesday* of (1.b) the natural choice for *d* is

---

attitude verbs such as *know* or *regret* is widely regarded as presuppositional, and much the same considerations that have been adduced in support of that view equally support a presuppositional treatment of the factivity of the 'inchoative' epistemic verb *learn*. Personally I am inclined to think that *learn* is factive, but have nevertheless chosen to ignore this aspect of its semantics here.

The factivity of *learn* is connected with another aspect of its meaning. "x learns that p" seems to entail that the information that p did not just become available to x, but that x accepted this information, adopting the belief that p is true. In my analysis of (1) I have assumed this acceptance aspect as part of the meaning of *learn*. It could be argued that the acceptance aspect of *learn* and its factivity go hand in hand. If that is so, the policy I have adopted here—of taking the one aspect on board but not the other, is of course not quite right. But it is a decision which has no serious consequences for the issues that concerns us.

The factivity issue might have been raised also in connection with the verb *hear* as it occurs in (1.a). According to my own intuitions *hear* (in the sense of 'come to be informed about' in which it is used in (1.a)), is less clearly factive than *learn*, but quite possibly it has a factive use as well.

the Sunday referred to by *Sunday* in (1.a).[21] (What day that is, is yet another question. It is one which we passed over when discussing the interpretation of (1.a), and it is a matter which cannot really be settled, the interpreter of (1) must simply accommodate the referent by assuming that whoever produced (1) must have been talking about some particular Sunday.) I am assuming further that the event described in (1.b) follows the one described in (1.a).[22] On these assumptions the referent of *Tuesday* is the Tuesday immediately following the day referred to by *Sunday* in (1.a), and that is what we will take it to refer to throughout the remainder of the paper.

Next the location discourse referent l. It might easily appear as if its resolution is as predetermined as those of u and v; for these cases are similar in that for each of these three discourse referents there is exactly one antecedent in the discourse context. In the case of l this is p, the only discourse referent in the discourse context representation (8) which satisfies the constraint that what it represents must be a location. Indeed, p seems a natural resolution for l—in other words, the leaving spoken of in (1.b) is a leaving from Paris—and that is the resolution which I will assume in this section. However, in Section 4.4 we will see that this is not the only possible interpretation for l.

Two anaphoric discourse referents remain, tp and t′. For these resolution is not unequivocal either. t′ could be resolved either to t_{tu} or to t_{su}. And in case t′ is resolved to t_{su}, tp could also be resolved to either t_{tu} or t_{su} (although this latter choice makes no difference to the resulting sentence interpretation). In this section we focus on the option of resolving t′ to t_{tu}. This means that *the previous day* is interpreted as referring to the day immediately before the day represented by t_{tu}—in other words, to the Monday between that Tuesday and the Sunday represented by t_{su}. In this case the only coherent resolution for tp is that which identifies it also with t_{tu}.

With these additional decisions our resolution set is extended to:

(iii) t_{tu} := "the Tuesday immediately following t_{su}";
(iv) l := p;
(v) t′ := t_{tu};
(vi) tp := t_{tu}.

---

[21] For some more discussion on this see Kamp & Schiehlen (2001).

[22] This interpretation is strongly suggested by the tenses of the matrix verbs in (1): the succession of the two simple past tenses *heard* and *learned* of (1) carries a presumption that the event of the second sentence followed that of the first. But many discussions in the literature show how delicate these matters can be. See Lascarides & Asher (1993); Asher & Lascarides (2003).

4.4. *Other Possible Interpretations of* (1)

Before we enter more deeply into the reconstruction of the interpretation determined by the resolution equations (i)–(vi), we will quickly review what other interpretations are possible for (1). In doing that I will not question the interpretation of (1.a) given in (8), but only look at alternative interpretations of (1.b) relative to this interpretation of (1.a). About these alternative interpretations I will have nothing more to say in the sequel. So the reader who isn't particularly interested in this matter and more concerned not to lose the thread of the main argument of this paper, can skip this section and move straight to 4.5.

We noted that for some of the anaphoric discourse referents of the preliminary representation of (1.b) the context that is provided by (1.a) offers only one possible resolution. For *he* it is 'Bill', for *she* it is 'Mary'. The matter of *Tuesday* is more complicated. I do not think that interpretations other than the one we have been considering can be excluded absolutely. It seems just possible to interpret *Tuesday* as referring to the last Tuesday before the utterance time in cases where that is a different Tuesday from the first one after the Sunday that was mentioned in the first sentence. (Note that when *Sunday* in the first sentence is itself interpreted as the last Sunday before the utterance time, and the utterance time is later in that week—e.g. on Friday or Saturday—then this second interpretation of *Tuesday* will *de facto* coincide with the one assumed in Section 4.3.) A second possibility would be for Tuesday to refer to the Tuesday immediately before the Sunday of the first sentence. As we noted in 4.3, this interpretation is not a very likely given the tenses of the two matrix clauses. But again, it is difficult to exclude it categorically. Since these interpretations seem so marginal, however, I won't consider these possibilities further.

I already mentioned in 4.3 an alternative interpretation for *the previous day*. On this interpretation the phrase would refer not to the day before Tuesday but to the day before Sunday. In order that this option lead to an intelligible interpretation we need a different resolution of the implicit location argument l of *leave*: When the leaving took place the day before Mary was in Paris, then presumably it was not a leaving from Paris but a leaving for Paris, and thus the place that Mary left must have been a place that is different from Paris. Since no place other than Paris has been mentioned, such a resolution of l requires accommodation. It should be stressed in this connection that there are not just speakers for whom this second interpretation of (1) is possible, but even some for whom it is the more salient one. One important implication of this is that the principle, to my knowledge first stated

explicitly in van der Sandt (1992), that resolutions of anaphoric discourse referents which do not involve accommodation—cases of *binding* in van der Sandt's terminology—have preference over resolutions which require it cannot be universally valid in this simple form. (Cf. Beaver 1995, 2005 for earlier indications that there is more to this principle than the simple formulation suggests.) There is no doubt that something like this principle must be true, but evidently a more refined formulation is needed. Such a refined formulation will have to pay attention among other things to the types of different anaphoric presupposition triggers. In particular, a distinction will almost certainly have to be made between overt pronouns and implicit arguments such as that of intransitive *leave*. (See also the remarks in Section 3.2 on trigger-dependent rules of presupposition justification.) In my estimation this exhausts the range of interpretations for (1.b).[23]

4.5. *Assumptions about the Persistence of Other Agents' Beliefs*

I will take it for granted that the resolutions (i)–(vi) yield full justifications for the presuppositions that contain the discourse referents for which they provide anaphoric discourse referents, i.e. that (8) entails each of the conditions that are contained in these presuppositions when the anaphoric discourse referents are replaced by their chosen antecedents. This leaves us with two remaining justifications, for the non-anaphoric pre-state presuppositions of *learn* and *leave*. It is the justification of the second of these presuppositions, the prestate presupposition of *leave*, which brings into focus those issues which motivated the choice of (1), and it is to this justification that the remainder of this section and all of Section 4.6 will be devoted. Once the justification of this presupposition is in place, the justification of the prestate presupposition of *learn* will be unproblematic.

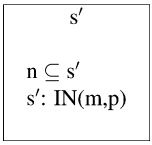
Given the resolution of v to m and of l to p, the prestate presupposition of *leave* says that Mary was in Paris immediately before the occurrence of the leaving event described in (1.b). We should keep in mind, however, that neither the event nor its presupposition are presented as mere facts but as contents of Bill's belief state. Contextual information which justifies

[23] Claims of this sort are notoriously prone to error—one always has to be wary of special contexts, which would open the way to other interpretations, but which are so unusual that a normal interpreter would not assume them unless the context was explicitly presented to him. Actually demonstrating that a given set of readings exhausts the interpretational possibilities for a given sentence or discourse on the basis of theoretical assumptions is almost always very hard, and well beyond our current capacities. So, as things are, such claims have to be handled with the greatest circumspection.

the prestate presupposition must therefore also be information about Bill's belief state, and it must be information about his belief state just before the time when on Tuesday he learned about Maria's leaving. But as things stand (8) provides no direct information about Bill's beliefs at that particular time. The attitude which (8) does ascribe to Bill is a belief which he holds as a result of the event of his hearing on Sunday about Mary being in Paris. Since it was the result of that event, Bill must have had it for at least some time on Sunday. But how long Bill stuck to this belief is something about which the interpreter has no direct information. However, beliefs, like many other propositional attitudes, have a tendency to persist, and so it is a natural assumption that Bill is still in this belief state when the new information reaches him on the following Tuesday. In fact, this assumption is so plausible that an interpreter of (1) would hardly be aware that he was making it. We will see presently that it doesn't really solve the justification problem posed by *leave*'s prestate presupposition, but it is a natural assumption nevertheless and it appears that a discourse like (1) exerts a strong pressure on the interpreter to make it.

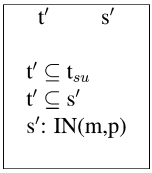
The assumption that Bill's belief about Mary's presence in Paris persists until Tuesday may seem innocent enough at first sight; but what really does this assumption amount to? What *is* the belief about Mary that is supposed to have persisted until the time when Bill got the new information on Tuesday? There are two different ways in which the persistence of Bill's belief about Mary being in Paris can be construed. One possibility is that Bill retains the belief that Mary was in Paris at the time when he was informed about this, i.e. the belief that Mary was in Paris on Sunday. But it is also possible for the belief to have persisted as the belief of Mary being *currently* in Paris. The difference between these two forms of persistence is as important as it is evident. From a truth-conditional perspective persistence of the second kind isn't persistence in the strict sense of the word. As noted in Section 2.1, the truth-conditional content of a belief about what is currently the case evolves with the time at which the belief is entertained. For instance, on Sunday the content of Bill's belief that Mary is currently in Paris is the proposition of Mary being in Paris on Sunday, whereas the content of the "same" belief on Tuesday is that of Mary being in Paris on Tuesday. Note, however, that in the representational format we have been using this evolution of truth conditions goes hand in hand with invariance of representational form. For it is the same content representation, repeated in (14), which carries these different truth conditions depending on whether it represents a belief entertained on Sunday or a belief entertained on Tuesday.

(14)



As opposed to this second kind of persistence, the first kind of persistence we mentioned is one which succeeds in preserving truth-conditional content of the belief that Mary was in Paris on Sunday. But this stability of truth conditions can, in the given case, only be achieved at the cost of a change in representational form: If we assume that Bill initially represented the belief he acquired on Sunday in the form in which it is given in (8) and repeated above as (14), its representation must have changed in the meantime if it is still to represent, when entertained at the later time on Tuesday, the proposition that Mary was in Paris on Sunday. In the abbreviated notation we first used in (8) and have been using since, the new, truth condition preserving representation can be presented in a pleasingly compact form:

(15)



In this particular instance, however, we do well to remind ourselves of the abbreviatory nature of this notation. As the representation of a propositional attitude, (15) conceals the internal and external anchors that would be explicitly represented in the earlier notation. A representation in that notation would have displayed among other things an internal anchor for a discourse referent t'_{su} together with the external anchor $<t'_{su},t_{su}>$. So there is more to the transition from the belief represented in (14) to the one represented in (15) than meets the eye in a superficial comparison of (15) and (14).

Yet we must be careful not to overstate this point. It is likely that Bill has an anchored representation for the mentioned Sunday, and quite possibly also one for the particular time on that Sunday when he got the information about Mary being in Paris. In that case most of what is needed to realise the transition from (14) to (15) is already in place and all that remains is to change the indexical link of the state s' to the time n by the non-indexical link to the anchored representation of the given Sunday (or of the particular time on Sunday when Bill's first belief change occurred).

For the interpreter of (1) there is no conclusive way of telling which of the two persistence assumptions about Bill's belief is the (more) appropriate one. In the next section we will spell out the implications of either option. I will use the remainder of the present section for two rather speculative remarks on the reasons why and the ways in which beliefs about the current state of the world are transformed into beliefs about what the world was like at some earlier time. First, an observation that goes back more than two decades and probably more:[24] The time it takes for the belief that a certain condition currently obtains to be transformed into the belief that this condition obtained at some earlier time (where this earlier time may be fixed with more or less accuracy) will vary as a function of what the condition is. When I observe that there is a dead blackbird lying on the roof and form the corresponding belief that this is currently the case, then this belief is likely to survive for longer than when I observe that there is a living blackbird sitting on the roof. We humans are equipped with elaborate networks of expectations as to how long different conditions are likely to last (in various circumstances) and these expectations guide us in our doxastic behaviour. And the longer the 'life expectation' we associate with a given condition, the longer it will take for us to abandon the belief that the condition currently obtains after we found out about its obtaining. And what is true for our own beliefs about such conditions also applies to our expectations about the doxastic behaviour of others. For instance, had the first sentence of (1) spoken of Mary being in the kitchen rather than in Paris, we would be much less inclined to make the assumption that Bill maintained his belief that she was still there until Tuesday.[25]

The second observation concerns the notion of doxastic strength. Some beliefs are held more firmly than others—we are more resistant to giving them up in the light of contrary evidence, we attach higher probabilities to their propositional contents, we are prepared to place higher stakes on outcomes which depend on their being true, etc. (I do not mean to imply that all these criteria must necessarily coincide in what they tell us about the

[24] Early work on this topic of which I am aware is that of D. McDermott. But I suspect that in AI and/or in the Philosophy of Mind an awareness of the importance of this factor may go back to an even earlier date.

[25] It is worth noting that this variant of (1) has a peculiar ring to it. Apparently the form of (1) impels the interpreter to assume that the current belief state acquired on Sunday persists until the time of the belief change on Tuesday. For the just mentioned variant the persistence of this belief over a period of the given duration is inherently less plausible than it is for (1) itself and this seems to be the reason why the variant sounds odd in a way that (1) does not. See also Section 4.7.

degrees of confidence of which they are the supposed manifestations, but they are all important.) Once we are prepared to distinguish between different degrees of doxastic strength, we are in a position to distinguish between instantaneous transitions form beliefs about current states to beliefs about past states and transitions which are gradual. In particular, it is now possible to conceive of Bill's belief that Mary is currently in Paris as involving, over the period from Sunday till Tuesday, a gradual decay rather than a sudden switch from belief to non-belief. In fact, the more plausible picture would now be that Bill's belief that Mary is currently in Paris coexists with his belief that she was in Paris at the time when he was informed about this on Sunday and that the strength of the former slowly decays while the latter continues in full force.

Taking degrees of doxastic strength into account would arguably lead to a deeper and more natural analysis of the interpretation of (1) than I am able to give in this paper. But degrees of belief cannot be represented in the formalism we are using and to modify the formalism so that variations of doxastic strength can be adequately represented is a non-trivial matter. There is no question of developing such a modification here and all I can do is leave this as a task for the future.[26]

I should add that I am not persuaded that matters of doxastic strength are essential to the *semantics* of discourses like (1), even if it is relevant to the extra-linguistic reflections that their semantic representations are likely to set in motion. The reasons for these doubts will become visible in section 4.7.

4.6. *Implications of the Two Persistence Assumptions*

Let us assume that both persistence hypotheses we discussed in the last section—the hypothesis that Bill's belief at the relevant time on Tuesday has the content given by (14) and the hypothesis that it has the content given in

[26] One possibility which comes to mind is that of extending the formalism with an open-ended set of doxastic indicators '[BEL, μ]' where the μ are terms denoting measures of the doxastic strength. (The measures might be real numbers, but they could also belong to some other type of confidence scale.) The further question then is what kinds of terms should be admitted in the extension. In order to state general hypotheses about the decline of belief strength over time it would seem desirable to admit variables ranging over the confidence scale. A further question concerns the types of complex terms that should be allowed, and the operations that may appear in those terms. But whatever the details, it is clear that such a formalism would constitute a very substantial change from the one we are using. And so long as no decisions have been made about its exact properties, there isn't very much that we can say about its impact on semantics, or, for that matter, about its implications for the analysis of (1).

(15)—are possible accommodations which an interpreter might entertain in his attempt to make sense of (1.b). For all that has been said so far, either one is a candidate. But which one the interpreter adopts makes a good deal of difference to how he will understand what happens when Bill learns about Mary's departure.

First consider the assumption that Bill retained the belief he formed on Sunday in the form given in (14). Then the new information he received on Tuesday was in contradiction with what he believed at that time. For he still believed that Mary was in Paris, which is obviously inconsistent with the discovery that she left Paris the day before. So the interpreter who assumes Bill's belief to have persisted in this form must also assume that the new information which Bill got on Tuesday led him to a belief *revision*: Apparently Mary left earlier than Bill thought. So that thought had to be given up and replaced by one which is compatible with the new information.

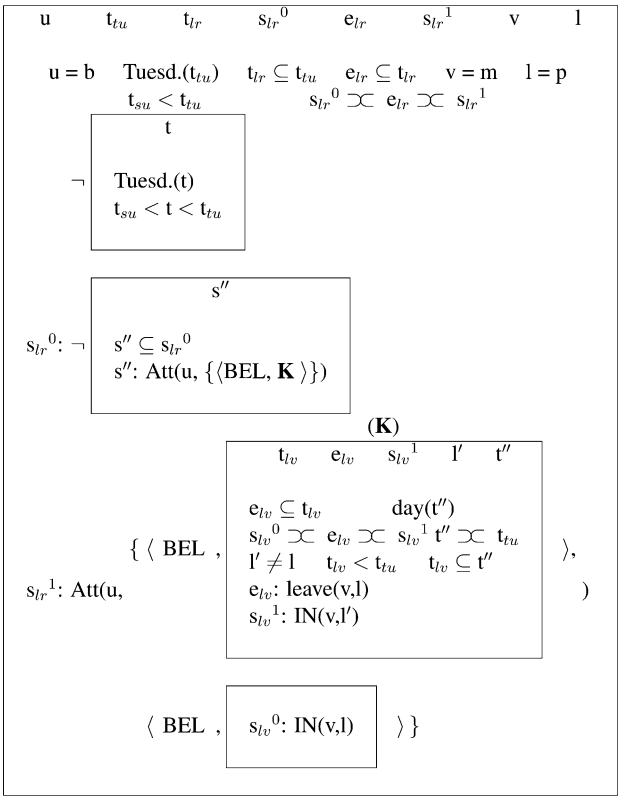
Note well, though, that the new belief state which results from this revision still involves persistence of the condition represented in (14) in a more limited form. If Mary left Paris on Monday then she must have stayed in Paris until some time on Monday. Thus (14) remained true until some time on Monday, rather than becoming false already on Sunday, which would have been possible too.

It should be emphasised that the belief revision of which we have spoken is a revision of Bill's beliefs, not of the assumptions made by the interpreter. It is Bill who is forced on Tuesday to abandon the belief of the form (14) which he held until then and to replace it by the belief that Mary remained in Paris only until some time on Monday. Only if the interpreter himself were to assume, on the strength of his interpretation of (1.a), that Mary was still in Paris on Tuesday would he have had to backtrack and correct his own perception of the facts. But there is no obvious reason why even an interpreter who adopts the persistence hypothesis under discussion would make this additional assumption as well.

The second persistence hypothesis attributes to Bill, at the time on Tuesday when he received the information that Mary had left, the belief that Mary was in Paris on Sunday. This belief was not contradicted by the new information which reached him and so there was no cause for belief revision. It should be observed, however, that this hypothesis too isn't quite what justification of the prestate presupposition of *leave* requires either. This time the belief which it attributes to Bill just before he gets the new information is too weak—it is just the belief that Mary was in Paris on Sunday, not that she remained there until Monday. So once again Bill must accommodate.

However, all that is required of him on this hypothesis is an addition to his beliefs, without the need to discard any he had.

(16)



The differences in the interpretations which result from the two hypotheses are significant, but they are limited nonetheless. The interpretation processes that are guided by the two hypotheses converge in that they end up attributing to Bill the same final belief state (the one in which he is said to have been as the result of his second belief change on Tuesday). They differ only with regard to what they entail about his beliefs immediately before that. (16) represents what is common between the two interpretations of (1.b)—Bill's final belief state, together with the condition that up to the time when he received the new information he was not in that state.[27,28]

[27] (16) is dependent on the representation (8) of (1.a) which serves as context for the justification of the presuppositions of (13). I have not repeated (8) here for reasons of space.

[28] Some may doubt whether prestate presuppositions are presuppositions at all. Or they

In (16) the accommodated belief that Mary remained in Paris till she left on Monday has been integrated into the final belief state ascribed to Bill. **K**, the content representation of the belief specified in the first Att-condition of (16), is the same DRS which also represents the content of the belief that appears in the second Att-condition. The new conditions involving t_{tu} state that the time represented by t_{tu} is the Tuesday immediately following the Sunday represented by t_{su}.

It should be clear how (16) could be extended to a fuller representation of the interpretations that result from the two persistence hypotheses. In either case a further Att-condition would have to be added which describes the relevant part of Bill's attitudinal state just before he got the new information on Tuesday. As nothing of interest is revealed by these full representations, I see no point in displaying them.

4.7. *Are the Persistence Assumptions Really Involved?*

Do the persistence hypotheses involved in the two interpretations just described play in the interpretation of (1) the role that I have been attributing to them? There appear to be good reasons for doubting this. For one thing, isn't (16), which captures the common denominator of the two interpretations described in the last section, all that an interpretation of (1) should yield? And if (16) is all that interpretation of (1.b) should yield, could this representation not be obtained by simpler means? For instance, couldn't we simply assume that interpretation of (1) proceeds like this:

(17) (1.b) says that as a result of what Bill learned on Tuesday he adopted the belief that Mary left Paris on the preceding Monday. This entails the belief that she was in Paris at the point on Monday when she left (since leaving Paris on Monday evidently entails being in Paris at the time on Monday when one leaves, and so you cannot believe the first without believing the second.)

---

may—a more moderate disagreement with the position I have assumed here—take it that they are conditions which function as normal entailments of positive declarative occurrences of the change-of-state verbs which trigger them and behave as presuppositions only when the trigger occurs within the scope of some operator, such as negation or a question. (For an early defence of a position similar to this see Wilson (1975).) Someone who holds this view of prestate conditions might either want to maintain that the persistence hypotheses we are discussing are irrelevant to the interpretation of (1) or he would want to describe their role in different terms. I do not think however, that this issue matters. In particular I do not believe that it affects the question whether the hypotheses play any role at all in the interpretation of sentence sequences like (1). More on this in the next section.

The problem with (17) is that it fails to do justice to the role that the first sentence of (1) plays in the interpretation of the second. And that isn't quite right. It is essential to the interpretation of (1) that the prestate of the event of Mary's leaving Paris which the second sentence speaks of is the same state that is spoken of in the first sentence—or, more precisely, that Bill must end up believing that they are one and the same state. This is something that the strategy described in the last section does capture, but (17) does not.

Persistence of the state of Mary being in Paris does play a role in the interpretation of (1). But as we will see presently, it almost certainly doesn't play the role that it was assumed to play in the account described in Section 4.6.

Let us, as a first step towards this, begin by considering the variant of (1) that is given in (18):

(18) On Sunday Bill heard that Mary was in Paris.
On Tuesday he learned that on the previous day she had visited the Swiss Embassy.

There is much that (18) and (1) have in common. The *that*-clause of the second sentence of (18) requires, just like the *that*-clause of the second sentence of (1), the context that is provided by the *that*-clause of the first sentence. And like in the case of (1) the context is not enough as it stands, but needs to be amplified by the assumption that the state of Mary's being in Paris persists till the relevant time on Monday. In the case of (18) this assumption is needed in particular to identify the intended referent of the NP *the Swiss Embassy*.[29]

So much for the similarities between (18) and (1). But there is also an important difference. In (18) the content of the second *that*-clause is *consistent* with the amplified context. (In fact, it is not just consistent with it, but entails it.) Because of this the final belief which (18) attributes to Bill is that on Monday Mary was still in Paris and that, while still in Paris, she went to the Swiss Embassy there.[30]

---

[29] There are countless Swiss embassies across the globe, so the interpreter needs further information in order to determine which one all those embassies this occurrence of *the Swiss Embassy* refers to. One kind of information that will serve this purpose is information about where the given embassy is located. And that is precisely the information that emerges when persistence is applied to the context condition that is provided by the *that*-clause of the first sentence: If Mary was still in Paris on Monday, then it was the Swiss Embassy in Paris which she visited.

[30] So even if we were to assume that Bill's belief about Mary being currently in Paris persists till Tuesday, the new information described in (18) would not require him to revise that belief.

As a next step notice the similarity between the interpretation of (18) and that of the discourse (19), in which the contents of the *that*-clauses of (18) are presented as facts.[31]

(19) On Sunday Mary arrived in Paris. On Monday she visited the Swiss Embassy.

In (19) too the first sentence provides the context for the interpretation of the second sentence, and once again, in order that this context can deliver what the interpretation of the second sentence demands of it, it has to be amplified by a persistence assumption. For sequences of factual sentences like (19) the role of persistence in interpretation is well-known and has been closely studied. These studies have shown that it is a general property of narrative texts (i.e. texts which describe episodes that develop in time) that states which have been introduced at one point in the text should be understood as continuing to hold until the discourse gives notice that this is no longer the case. More fully, if the state s has been introduced by sentence S_n of the discourse and has been asserted by S_n to hold at t, and the sentences S_{n+1}, ... following S_n describe what is the case at times later than t, then s will be understood as still holding at these later times, unless and until it follows from one of these sentences that s has come to an end at or before one of those later times.[32]

The interpretation of (19) involves a straightforward application of this persistence principle. The state of Mary being in Paris that is introduced by S_1 as holding on Sunday is assumed to persist at and beyond the later time described by S_2 unless it follows from S_2 that the state has terminated before or during that time. Since visiting the Swiss Embassy in Paris is compatible with this state, the prediction is that the state still holds at and continues to hold beyond the time of this visit on Monday.

The third point to note is that just as (19) is a close factual counterpart to (18), (20) is a close factual counterpart to (1):

---

[31] I have changed the *was* of the *that*-complement of the first sentence of (18)—(*that*) *Mary was in Paris*—into *arrived*. The discourse *On Sunday Mary was in Paris. On Monday she visited the Louvre*. is strange because we are inclined to perceive an opposition between *Sunday* in the first sentence and *Monday* in the second—as if Mary was no longer in Paris when she went to the Louvre. This is a rhetorical effect which has nothing to do with what we are after and which we do not get in (1) or (18), where *Sunday* modifies the matrix verb and not the verb of the *that*-clause. The first sentence of (19) matches that of (1) and (18) in this respect, because it too carries no implications about the duration of Mary's stay in Paris.

[32] In the discourse literature this persistence property of narrative discourse is also known as *monotonicity*. Analyses of monotonicity and formulations of the "monotonicity principles" which are responsible for this kind of persistence can be found for instance in Caenepeel & Sandström (1993) and Reyle & Roßdeutscher (2001).

(20) On Sunday Mary arrived in Paris. On Monday she left.

Here too the persistence principle makes the right prediction: The state of Mary being in Paris is assumed to persist until the time at which the discourse indicates its termination. In the present case this means persistence at least until the time of the event which the second sentence describes. This assumption makes it possible to interpret the second sentence as describing an event of Mary leaving Paris. But in this case the event description entails that the state ends there. So the interpretation of (20) which results is that the state of Mary being in Paris, which starts at some time on Sunday, continues until some time on Monday, when it is terminated by the event of her leaving Paris.

The parallels between the interpretations of (18) and (1) on the one hand and those of (19) and (20) on the other are striking—so striking that it is almost inescapable to see them as involving the same principles. In particular, interpretation of the belief attributing discourses seems to make use of the same persistence principle as interpretation of the factual discourses, the only difference being that in the case of (19) and (20) persistence is applied to a primary context while in (18) and (1) it is applied to a secondary context. What this suggests is a theory of interpretation in which the complement clauses of attitude attributions can be treated as if they combine to form bits of narrative discourse—more precisely: which allows sequences of such embedded clauses to be incrementally interpreted according to the same principles which existing theories of discourse semantics assume for bits of narrative discourse consisting of non-embedded sentences.

I will not attempt to formulate such a theory here, but only make a few informal observations about its implications for the interpretation of discourses like (18) and (1).[33] The first implication is this: When interpreting a sequence of attitude reports like those in (1) and (18), the interpreter builds, as he goes along, not only a primary discourse context, but also a secondary one. Both contexts grow as discourse interpretation progresses and are updated according to the same principles of language interpretation, and these principles include principles of persistence or monotonicity: Both

[33] I believe that formalisation should not be all that difficult. This should be so at least for a restricted version, in which there is besides the primary context just one secondary context. In general the context structure can be much more complex, with different secondary contexts for different agents and/or different times. Even within a 'secondary' context which corresponds to a given agent's attitudinal state at a given time there may be a further hierarchical structure of 'subcontexts', e.g. reflecting the different modes of the attitudes of which it is composed. (Recall the remarks about complex attitudinal states in Section 3.1.)

types of context come with a default assumption that stative conditions which are initially said to hold at certain times persist until there is clear evidence to the contrary.[34]

But of course, even if primary and secondary contexts come about in much the same way, their roles are very different. Primary contexts are what they have been made out to be ever since Stalnaker laid the foundations of dynamic interpretation: They capture the content of what has been interpreted already and provide the context for the utterances that follow. The role of a secondary context is different. What that role is depends, first and foremost, on the status of the sentence or clause that has been responsible for its latest update. For instance, when the clause is the *that*-clause of an attitude attribution, the updated context will serve to characterise the content of the attributed attitude; and if it is the *that*-clause of an attribution of belief change, as in the examples we have been looking at, then the updated context typically serves as specification of the content of the belief which results from the attributed change.

In this theory persistence has regained its place in the interpretation of discourses like (1) which (17) denies it. Superficially this might look like a revindication of the interpretation strategy described in Section 4.6, but even a cursory reflection suffices to see that it really is nothing of the kind. In the new theory persistence is made to play a very different part from the one it was made to play in 4.6. The persistence assumptions that are postulated by the present proposal are *not* assumptions about the persistence of the subject's beliefs. Their status is a more abstract one. They are just what we have described them to be: interpretation principles which are operative not only at the level of primary but also of secondary contexts. Precisely what that amounts to can be assessed only in the light of the role which the secondary context ultimately comes to play in the interpretation of the given discourse. But whatever that may tell us about the persistence principles that are used in the updating of various secondary contexts, it is clear that the applications of these principles to the interpretation of (1) or (18) have nothing directly to do with the beliefs which the subject Bill may have entertained during the time between his first belief change on Sunday and his second one on Tuesday. From this point of view the present theory is anything but a reaffirmation of the strategy of Section 4.6.

---

[34] As far as this is concerned, the present approach seems to be much in the spirit of what Stalnaker must have had in mind when he introduced the concept 'secondary context' and coined this term for it.

For some readers it may have been plain from the start that the theory we have just outlined gives the better reconstruction of how people process and understand a discourse like (1). But for someone who approaches the problem presented by (1) from the perspective from which I—for better or worse—came to this problem, the point of view we have now reached is not self-evident. Let me, to conclude this section, briefly recapitulate this perspective and what it implies for the interpretation of a discourse like (1).

In a few words, the perspective was this: To correctly interpret the content of an attitude attribution, the interpreter often needs antecedent information about the attitudinal state of the attributee. When the given attribution is one of a sequence, then this information can usually be obtained, in part or in whole, from the attributions which precede it. Therefore the interpreter must, as he proceeds with his interpretation of the successive sentences of the sequence, build a representation of the attitudinal state of the attributee (or attributees) in question, and use, each time he tackles a new attribution, the representation of the attributee's attitudinal state that he has already obtained to make sense of what it is this new attribution contributes.

From this perspective it seems a natural assumption that in a case like (1), where the new sentence attributes an attitude at a later time than the state of which the interpreter has already obtained a representation, the interpreter should try to infer, on the basis of what he knows about the attributee at the earlier time, what his state should be like at the later time that is concerned in the new attribution. But natural though this extrapolation from the case of several attributions at a single time to different attributions at different times may seem at first sight, it doesn't seem—this is the central moral of this paper—to be what is actually going on. The persistence involved in the interpretation of such attribution sequences is *not* the persistence of any attitudes on the part of the attributee.

This is not to deny that persistence plays no part in the mental life of the attributees which such sequences describe. Surely, if they are at all like you and me, persistence will be as much a part of how their mental states evolve as it is of ours. But whatever assumptions we, as interpreters, may be inclined to make about the persistence of the attitudes that are attributed to them by the sentences of which we are the recipients, it appears that these assumptions are not part of the interpretation process as such. Or at any rate, if they are part of the process at all, they do not seem to be part of it in the manner suggested in Section 4.6.

5. Summary and Conclusions

This paper has focused on the analysis of one example—a brief and fairly simple discourse, consisting of two attitude attributions to a single subject. Like other examples which have been discussed in the literature the two attributions are understood as related to each other in content. The example differs from those other examples in that (i) its attributions are not attributions of attitudes as such, but of attitudinal changes; and (ii) these changes are asserted to have taken place at different times. Because the attitude change attributed by the second sentence occurs some time after the change described by the first sentence, a complication arises for the use that can be made of the first sentence as context for the interpretation of the second. This complication has been the paper's central focus.

In my attempt to deal with this problem I have outlined two interpretation strategies. The first of these is based on the idea that interpretation of a sentence which claims that a subject was in a certain attitudinal state at a certain time t, or undergoes a change of attitudinal state at t often requires antecedent information about the subject's state at or just before t; so, when the antecedent discourse only provides information about the subject's attitudes at other times, then it will be necessary to extrapolate from that information to the subject's attitudes at the time to which the new attribution pertains.

The main conclusion of this paper is that plausible as this idea may seem initially, it is almost certainly not correct for discourses like (1). Rather, in successions of attitude attributions the clauses which describe the contents of the attributed attitudes form a kind of subtext, whose dynamic interpretation is subject to the same rules that govern dynamic interpretation of discourse in which no attitude attributions are involved. The difference between this second perspective and the first one doesn't manifest itself when all attributions speak of the subject's attitudinal state at a single time. It becomes visible only when the attributions concern different times (or, but that is a theme for another paper, different agents). It is this which makes sentence sequences of the type of (1) a genuine challenge for the semantics of attitude attribution.

The account of the interpretation of (1) which emerged from this study as the more likely one is still awaiting formalisation. I don't expect that a formalisation would encounter serious obstacles, but we won't be sure until the work has actually been done. Since the work hasn't been done yet, the last part of the story this paper has told had to be kept at an informal level. And because of the informal character of this last part, the much more

formal presentation of the earlier parts may in retrospect look like a luxury one could well have done without. But this impression is widely justified at best. For without the attention that we have paid to formal detail in these earlier sections, the issue with which we have grappled in the final section wouldn't even have come into proper view.

References

Abusch, D., 1997. Sequence of tense and temporal *de re*. *Linguistics and Philosophy* 20, 1–50.

Asher, N., Lascarides, A., 2003. *Logics of Conversation*. The MIT Press, Cambridge, MA.

Beaver, D., 1995. *Presupposition and Assertion in Dynamic Semantics*. Ph.D. dissertation, University of Edinburgh. Revised version published with CSLI Publications, Stanford, CA, 2001.

Beaver, D., 2005. Accommodating topics. In: H. Kamp, B.H. Partee (eds.), *Context-Dependence in the Analysis of Linguistic Meaning*. Elsevier, Amsterdam, 73–86.

Caenepeel, M., Sandström, G., 1993. A discourse level approach to the past perfect. In: M. Aurnague et al. (eds.), *Semantics of Time, Space and Movement—Proceedings of the 4th TMS International Workshop 1992*. IRIT Université Paul Sabatier, Toulouse, 167–182.

Dekker, P., 1993. Existential disclosure. *Linguistics and Philosophy* 16, 561–587.

Geurts, B., 1995. *Presupposing*. Ph.D. dissertation, University of Stuttgart.

Geurts, B., 1999. *Presuppositions and Pronouns*. Elsevier, Amsterdam.

Heim, I., 1983. On the projection problem for presuppositions. In: M. Barlow, D. Flickinger (eds.), *Proceeding of the Second West Coast Conference on Formal Linguistics*. Stanford University Press, Stanford, CA, 114–125.

Kamp, H., 1990. Prolegomena to a structural theory of belief and other attitudes. In: C.A. Anderson, J. Owens (eds.), *Propositional Attitudes*. CSLI Publications, Stanford, CA, 27–90. Reprinted in: K. von Heusinger, A. ter Meulen (eds.), *Meaning and the Dynamics of Interpretation. Selected Papers of Hans Kamp*. Brill, Leiden, 2013, 513–583.

Kamp, H., 2001a. The importance of presupposition. In: C. Rohrer, A. Roßdeutscher, H. Kamp (eds.), *Linguistic Form and its Computation*. CSLI Publications, Stanford, CA, 207–254. Reprinted in: K. von Heusinger, A. ter Meulen (eds.), *Meaning and the Dynamics of Interpretation. Selected Papers of Hans Kamp*. Brill, Leiden, 2013, 455–501.

Kamp, H., 2001b. Presupposition computation and presupposition justification. In: M. Bras, L. Vieu (eds.), *Semantic and Pragmatic Issues in Discourse and Dialogue: Experimenting with Current Dynamic Theories*. Elsevier, Amsterdam, 57–84. Reprinted in: K. von Heusinger, A. ter Meulen (eds.), *Meaning and the Dynamics of Interpretation. Selected Papers of Hans Kamp*. Brill, Leiden, 2013, 425–454.

Kamp, H., 2003. Einstellungszustände und -zuschreibungen in der DRT. In: U. Haas-Spohn (ed.), *Intentionalität zwischen Subjektivität und Weltbezug*. Mentis, Paderborn, 209–289.

Kamp, H., van Genabith, J., Reyle, U., 2011. Discourse Representation Theory. In: D. Gabbay, F. Guenthner (eds.), *Handbook of Philosophical Logic*, 2nd Edition. Vol. 15. Springer, Berlin, 123–394.

Kamp, H., Reyle, U., 1993. *From Discourse to Logic. Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer, Dordrecht.

Kamp, H., Roßdeutscher, A., 1994. Remarks on lexical structure and DRS construction. *Theoretical Linguistics* 20 (2/3), 97–164.

Kamp, H., Schiehlen, M., 2001. Temporal location in natural languages. In: H. Kamp, U. Reyle (eds.), *How we say WHEN it happens. Contributions to the Theory of Temporal Reference in Natural Language*. Niemeyer, Tübingen, 181–232.

Kratzer, A., 1998. More structural analogies between pronouns and tenses. In: D. Strolovitch, A. Lawson (eds.), *Proceedings of the 8th Semantics and Linguistic Theory Conference*. Cornell University Press, Ithaca, NY, 92–110.

Lascarides, A., Asher, N., 1993. Temporal interpretations, discourse relations and commonsense entailment. *Linguistics and Philosophy* 16, 437–493.

Ogihara, T., 1994. Adverbs of quantification and sequence-of-tense phenomena. In: M. Harvey, L. Santelman (eds.), *Proceedings of the 4th Semantics and Linguistic Theory Conference*. Cornell University, Ithaca, NY, 251–267.

Ogihara, T., 1996. *Tense, Attitudes and Scope*. Kluwer, Dordrecht.

Reichenbach, H., 1947. *Elements of Symbolic Logic*. Macmillan Co, New York.

Reyle, U., Roßdeutscher, A., 2001. Temporal underspecification in discourse. In: C. Rohrer, A. Roßdeutscher, H. Kamp (eds.), *Linguistic Form and its Computation*. CSLI Publications, Stanford, 255–283.

van der Sandt, R., 1992. Presupposition projection as anaphora resolution. *Journal of Semantics* 9, 333–377.

Stalnaker, R., 1988. Belief attribution and context. In: R. Grimm, D. Merrill (eds.), *Contents of Thought*. University of Arizona Press, Tucson, AZ, 140–156.

von Stechow, A., 2003. Binding by verbs: Tense, person and mood under attitudes. In: M. Kadowaki, S. Kawahara (eds.), *Proceedings of the 33rd Annual Meeting of the North East Linguistics Society*. GSLA Publications, Amherst, MA, 379–403.

Wilson, D., 1975. *Presupposition and Non-Truth-Conditional Semantics*. Academic Press, London.

INTERVIEW WITH HANS KAMP*

Alice ter Meulen and Klaus von Heusinger

AtM: First on your background: you grew up in the Netherlands, where you got your bachelors degree in physics in Leiden and studied logic with Evert Beth in Amsterdam.

HK: When I went to Leiden University in 1958, I wanted to study theoretical physics. As the completion of my undergraduate studies got nearer, I felt an urge to do something different for a year before disappearing behind the doors of the Institute for Theoretical Physics. Dick de Jongh, who was also studying mathematics and physics in Leiden, had told me about Professor Evert Beth who had founded his Institute for Formal Logic and the Foundations of Science at the University of Amsterdam just a couple of years before. Dick himself had decided to go there to do a Masters degree and then, eventually, a Ph.D. Dick whetted my appetite so much that we ended up going to Amsterdam together. As far as I was concerned, this was just a kind of time-out to explore a world of science I might never belong to in earnest. But, as the year at Beth's Institute progressed, I was offered a student assistantship and became increasingly integrated into the research group Beth had established. So the temptation to stay and make formal logic and its applications the subject of my further academic work became harder and harder to resist. In the end I gave in and decided to stay.

During those years in Amsterdam I developed a strong interest in the relation between logic and language, one of Beth's many interests. Open as he was to new developments, he organized, during my second year, a seminar on Noam Chomsky's book ***Syntactic Structures***, which had just appeared, but was already widely hailed as heralding a completely different way of looking at grammatical structure.[1] Young linguists also participated, among them Albert Kraak, Wim Klooster, Pieter Seuren, Hugo Brand Corstius, some

* This text derives from a prolonged, in-depth interview with Hans Kamp that took place in the Azores, Stuttgart and Köln. The transcription of the audiotapes was edited, extensively rewritten and often expanded in various ways. We'd like to acknowledge our gratitude for help with its content and form from Barbara Partee, Antje Roßdeutscher, Dana Scott, Blanca and Emily Kamp, Malte Willer, Daniel Altshuler and Tillmann Pross.

[1] Chomsky, Noam, 1957, ***Syntactic Structures***. Mouton, The Hague.

© ALICE TER MEULEN AND KLAUS VON HEUSINGER, 2013 | DOI:10.1163/9789004252882_023
This is an open access chapter distributed under the terms of the CC BY-NC-ND 4.0 license.

of whom had been told by their more established colleagues not even to consider Chomsky's new work (or else ...), so we were all sworn to some kind of secrecy as to their presence. For me this was an early opportunity, rare in Europe in those days, to become familiar with the basic ideas of generative grammar. Beth did not only have an abiding interest in the relation between formal logic and natural language, but also in its relation to computation and thought. These interests came together in the project for which Beth had been able to find a sponsor in Euratom, the European joint project for the development of peaceful uses of nuclear energy. The project's aim was to explore the possibilities of using Beth's Semantic Tableau Method—then still quite new—as the basis for the development of automated systems capable of effective and psychologically plausible logical reasoning. The computer was to find proofs for deductively valid inferences using the tableau-based deduction algorithms that the group was to design and make available in machine-readable format. Being part of that group was an important factor that kept me in Amsterdam, although I don't think I ever contributed anything to this project. But I didn't work any less hard than the others. The project provided a wonderful opportunity to be part of a larger unit with a clear scientific purpose and a strongly felt, joint commitment. Besides, it provided a certain material comfort.

During my second year in Amsterdam I met Richard Montague, who had chosen to spend part of his sabbatical year there. (The other part was spent with Andrzej Mostowski in Warsaw.) The reason why I got to know Montague better than I otherwise would have had little to do with logic. Besides a superb logician, Montague was also a very accomplished organist. Someone at my father's office happened to be a relative of the verger of the St. Laurens church in Alkmaar, the site of one of Holland's most famous organs, at that time already used for widely hailed recordings by highly regarded organists. Since it was within the verger's discretion to decide who would be admitted to the organ outside regular services, it was possible to arrange for Montague to come along to Alkmaar and try it out. The event was a very big thing for me and I think it also meant something to him. In any case, he did remember me when I wrote to him a couple of years later, which led to my being accepted eventually as a Ph.D. student in the Philosophy Department of UCLA, where Montague was a professor.

I got my Masters in Philosophy and Logic from Beth's Institute about a year and a half after Beth had died—he died during my third year in Amsterdam. The topic of my master's thesis was a paper by Alfred Tarski on the calculus of formal, deductively closed systems, ***The Calculus of Deductive Systems***, in which he develops a type of algebra whose points are deductive

theories from some formal language and whose operations are complement, union and intersection of deductive theories.[2] These algebras have many of the properties of Boolean algebras, but only some of them are Boolean. More precisely, they are Boolean if and only if the formal language that provides the points of the algebra is sufficiently simple, for instance, a language of classical propositional logic with a finite number of atomic sentences. One of the interesting challenges for this approach to the deductive structure of formal languages is to determine how the properties of the languages generating the algebras manifest themselves as properties of the algebras they generate. Tarski wrote this paper in the nineteen thirties in his native Poland, but he did not include any proofs. He had a hard time getting his work published—he once told me that in the twenties and thirties, when Poland made stellar contributions to mathematical and philosophical logic, the prevalent anti-Semitism made it hard for Jewish scholars to become successful academics and that publishing in top journals was one of the difficulties someone like him had to cope with. In the case of this paper he was eventually given just enough space to state the definitions and theorems, but had to suppress all the proofs. So what I did, partly following a suggestion from my advisor and partly because I really wanted to understand the paper, was to figure out these proofs. My masters thesis was really nothing more than a list of proofs for all the theorems of the paper with a little bit of commentary. It certainly wasn't a very original piece of work; but doing it was very useful for me. And it confirmed Tarski as my hero-of-logic Number 1. (He shared first place with Gödel, who, I still regret to this day, I never met in any way or even set eyes on.)

In my second year in Amsterdam Beth asked me who I wanted to be supervised by during my studies in Amsterdam. I chose Karel de Bouvère, because I wanted to work mostly on model-theory and that was what de

---

[2] Tarski, Alfred, 1983, ***Logic, Semantics, Metamathematics***, 2nd edn., edited by John Corcoran, Hackett Publishing Company, Indianapolis, IN.

Tarski, Alfred, 1930, Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften I, ***Monatshefte für Mathematik und Physik***, 37, 361–404. English translation under the title 'Fundamental concepts of the methodology of the deductive sciences' in: Tarski 1983, 60–109.

Tarski, Alfred, 1935, Grundzüge der Systemenkalküls. Erster Teil, ***Fundamenta Mathematicae***, 25, 503–526. English translation by J.H. Woodger under the title 'Foundations of the calculus of systems' in: Tarski 1983, 342–383.

Tarski, Alfred, 1936, Grundzüge der Systemenkalküls. Zweiter Teil, ***Fundamenta Mathematicae***, 26, 283–301. English translation by J.H. Woodger under the title: 'Foundations of the calculus of systems' in: Tarski 1983, 342–383.

Bouvère did, whereas model-theory had become less prominent in Beth's own work. When my masters degree came in sight and I had to make a decision whether to go on to do a Ph.D., I decided to look for a place as Ph.D. student in one of the California strongholds of mathematical logic. My reason was, apart from a sense of adventure, that I had made up my mind to work on the interface of model-theory and proof-theory to see if Beth's method of Semantic Tableaux could be used to define new decidable fragments of predicate logic. I felt that, after Beth's death, nobody in Amsterdam was an optimal Ph.D. supervisor on that topic.

My time at UCLA (1965–1968) was, both by the standards of today and by those at the time, quite pampered. I received a University Fellowship, which liberated me from any TA work, so that I could devote myself entirely to my dissertation. However, I did have to take courses, some of which later proved invaluable to my further career and development. Crucial input that I received in this way, which had an immediate effect on my work, came from Arthur Prior, who was at UCLA as Visiting Professor in the Philosophy Department during my very first semester. Prior was offering an upper division course on Tense Logic, covering most of the material that appeared not long afterwards in his book ***Past, Present and Future***.[3] Priorean Tense Logic—the system obtained by extending classical propositional logic with the two tense operators ***P*** ('it was the case that') and ***F*** ('it will be the case that')—and the many variants of it that were formulated by Prior himself and by others subsequently, is purely 'topological' in that it can only express non-metrical temporal relations such as temporal precedence or the relation of temporal overlap. Prior raised the question whether it was possible to use such topological systems also as a basis for the expression of metrical relations, e.g. relations having to do with *how long p* was the case before *q* was the case. This problem became the starting point for what was to become my dissertation. I decided at that point, some time in the Fall of 1965, to shelve the project with which I had come to UCLA—that of using Semantic Tableaux to try and discover new decidable fragments of classical predicate logic, which at that point had not led to any significant results—a topic that, to my regret, I have never quite managed to get back to. The main result of my dissertation was that, if you want to express some form of temporal metric within a topological calculus like Prior's (***P***, ***F***)-calculus, you need a logical system with more powerful operators. The 2-place propositional operators ***S***(***ince***) and ***U***(***ntil***) not only made it possible to express

[3] Prior, Arthur, 1967, ***Past, Present and Future***, Oxford University Press, Oxford.

metric relations, it also proved to be in an important sense 'optimal' in that—provided the temporal order is a linear, order-complete relation—they make available all that a topological tense logic could possibly give you: Any tense operator that can be defined semantically can also be expressed by a formula of the ($\boldsymbol{S}$, $\boldsymbol{U}$)-calculus. The ($\boldsymbol{S}$, $\boldsymbol{U}$)-calculus eventually found a place within Computer Science, first as an ingredient in certain methods of program verification and, more recently, as a building block in specification languages in chip design. But honesty obliges me to stress that these applications were entirely unforeseen and that they are utterly different from what motivated the development of the calculus at the time. Prior thought that the tenses of the verb should formally be analyzed as modal operators. He and others took their inspiration from the formal work in modal logic that, in the wake of Saul Kripke's achievements of the late fifties and early sixties, had reached a remarkable degree of popularity and sophistication. The philosophical drive behind this work was the widespread perception that, even if modality might have something to do with truth in possible worlds, the modal expressions that are found in natural languages should not be construed as involving reference to possible worlds. If possible worlds were to play any part in the analysis of such expressions, they should do so only indirectly and implicitly, not as part of the logical forms of those expressions, but only in the meta-language, in which their semantics can be given a model-theoretic explication. Prior and others not only held such a view with respect to modality, but also with respect to temporal reference. The tenses, they argued, involve no more explicit reference to times than modal adverbs like *necessarily* or *possibly* involve explicit reference to worlds. Hence there should be a way of explicating the role time plays in our understanding of those expressions without construing them as explicitly referring to or quantifying over times. And, moreover—and a good deal less plausibly—this point of view was then extended to also cover other kinds of temporal expressions, such as calendar terms and temporal adverbials. Priorean Tense Logic was meant to accomplish this. What matters most for the present story is that Prior thought of Tense Logic as relevant to the analysis of natural language and that I saw the work of my dissertation in that same spirit—as part of a general project of providing logical analyses of linguistic meaning. That temporal expressions of natural languages do not refer to times any more than *necessary* and *possible* refer to possible worlds was a central motive behind my dissertation no less than I took it to be a central motive for Prior. And the fact that words like *since* and *until*, with their apparent operator-like syntax, are as expressively powerful as the formal results of the thesis seemed to establish carried a great temptation: I was only too eager to

interpret it as evidence that Tense Logic gave us the right means to explain how natural languages generally handle time. It took me much of the following decade to persuade myself that this simply wasn't right: there is no justification for such a nominalistic attitude towards temporal reference in natural language. For me the most obvious reason for this was that from a linguistic point of view expressions like (*at*) *ten minutes to five* can lay as much claim to referring to times as noun phrases like *Arthur Prior*, *The University of California at Los Angeles* or *the father of Tense Logic* can lay claim to referring to people and institutions. But there are a good many other reasons as well.

AtM: Are there other people you met during those years at UCLA who you feel had an important influence on your work or intellectual development?

HK: One of them was Alfred Tarski. Tarski came to UCLA as a visiting professor for (I believe) the Spring Semester of 1967. That was the first time I met him. I got to know him personally then, up to a certain point, partly through his Ph.D. student Judith Ng, who came with him during his semester at UCLA. But I had already come to see Tarski as one of the fathers of modern logic during my years in Amsterdam and that he has always remained. I still remember that in the class he was offering that semester at UCLA he also discussed the Calculus of Deductive Systems. That gave me a sense that the work I had been doing on my masters thesis in Amsterdam hadn't been a complete waste. I did mention to him once that I had spent some time with that paper in Amsterdam and it was then that he told me about the difficulties he had encountered back home before he escaped to the US.

But there were many other outstanding logicians at UCLA at the time, in particular C.C. Chang and Yanis Moschovakis in mathematics and David Kaplan and David Lewis in philosophy. Kaplan was away for most of the time I was at UCLA as a graduate student and I got to know him better only later. Another person, for whom I had already formed a great admiration before I got to UCLA, was Dana Scott. At that time Scott taught at Stanford, but he would occasionally come down to UCLA, partly in order to work with Montague on a monograph on set theory by Montague, Scott and Tarski, that had been in the making for some time, and was nearing completion then.

AtM: But your Ph.D. advisor was Montague?

HK: Yes, Montague was my advisor. As a point of fact, I wrote the dissertation essentially on my own. I had even had some difficulty in getting Montague to sit down to read my results before I got the dissertation typed and submitted.

But if that suggests I was one of those graduate students who are ignored or neglected by their advisors, nothing could be further from the truth. As a matter of fact, I was immensely lucky and privileged in the kind of tutelage I did get from Montague. The form it took had to do with his often wanting somebody around when he did his own work, as a kind of sounding board or sparring partner. During most of the three years I was at UCLA I happened to be the only person whom he considered suitable for this role. In particular, during 1966 and 1967 I was present for quite a bit of the time in which he developed the ideas that can be found in some of the papers that were collected in ***Formal Philosophy*** after his death.[4] That gave me a unique opportunity to see from up close how new ideas emerge, how they develop and mature, and what can and should be done by way of moulding them into a form that makes them not only useful to oneself, but also capable of being communicated to others. To witness how some of these papers took shape—I am thinking here in particular of: *On the Nature of Certain Philosophical Entities*, *English as a Formal Language* and *Pragmatics and Intensional Logic*—and to be able to watch Montague's truly extraordinary ability to combine penetrating intuitions with an uncanny gift for formalization was the best graduate education anyone could ever have had. I am immensely grateful to this day that my thesis supervision took this particular form.

AtM: Alonzo Church was there at UCLA. How did the development of lambda calculus influence Montague's work?

HK: Church came to UCLA in the fall of 1966, upon his retirement from Princeton. UCLA had offered him an attractive deal, to set up an infrastructure that would enable him to continue his work as editor of the review section of the ***Journal of Symbolic Logic*** (JSL). It made it possible for him to accommodate his archives as well as much of his own personal papers. In those days, when there wasn't yet any kind of electronic support for archiving and data collecting, that was a *sine qua non* for the continuation of the review section work. Without this, I doubt very much that Church would have come to UCLA. He had devoted a good part of his life to the JSL's review section, work he considered extremely important to a new field, which Mathematical Logic certainly was when he first took the review

[4] Montague, Richard, 1974, ***Formal Philosophy: Selected Papers by Richard Montague***, edited and with an introduction by Richmond Thomason, Yale University Press, New Haven, CT.

section under his wings and certainly still was in the sixties, especially when compared with other, well-established branches of mathematics like number theory, analysis, geometry, topology or abstract algebra.

But, as far as I remember, there was little, if any, connection between Church's coming to UCLA and the influence of the lambda calculus on Montague's work. Of course, the presence of the lambda calculus in Montague's work is self-evident. Montague's Higher-order Intensional Logic (HOIL) is an intensional system of the typed lambda calculus. But Montague was closely familiar with Church's work on the lambda calculus well before HOIL was given its final form and used in the linguistic and philosophical applications that we know from the papers in ***Formal Philosophy***. When I came to UCLA, the lambda calculus was a prominent focus for Montague in two quite distinct contexts. On the one hand, there was the question of the consistency of the untyped lambda calculus, a problem that was solved by Scott in 1969. This was a problem in which Montague too had taken an acute interest, as it was closely related to his own work on set theory and on at least superficially similar problems there, such as the consistency of Quine's system known as ***New Foundations***.[5]

On the other hand, in the mid-sixties Montague was in search of a formal system that could serve as a general formal framework for addressing philosophical questions, including the formulation of model-theoretic semantics for fragments of natural languages. Here it was important to build on the typed lambda calculus—the possibility to refer to higher-order and not just to first-order entities was an essential desideratum for such a general framework. But another desideratum that Montague had come to see as essential was that the system admit of forms of non-extensionality—in this regard Montague was part of the modal logic revolution within philosophy that is now mostly identified with Kripke's work, but to which Montague himself had also been one of the early contributors.

This constituted perhaps the most important opposition between the general perspective represented by him and his colleagues in Southern California, on the one hand, and the persisting extensional commitments of Quine and his followers, on the other hand. Montague saw HOIL as an optimal answer to the question what a formally precise, consistent and practically viable logic of intensionality ought to be like. The importance he

[5] Quine, Willard van Orman, 1937, New foundations for mathematical logic, ***American Mathematical Monthly***, 44, 70–80. Reprinted in: Willard van Orman Quine, ***From a Logical Point of View***, 2nd rev. edn. Harvard University Press, 1980, Cambridge, 80–101.

attached to HOIL is illustrated by the following remark I still vividly remember him making to me at one time in 1966 or 1967: "I always thought until now that the right framework for doing philosophy properly was Zermelo-Fraenkel set theory—I now realize it is higher-order intensional logic."

No one intending to develop a non-extensional logic could have ignored Church's contributions to this, his work on the ***Logic of Sense and Denotation***, in which he tried to formalize the ideas that Frege had formulated about sense and reference and the relations between them in ***Über Sinn und Bedeutung*** and other writings.[6,7] All versions of Church's ***Logic of Sense and Denotation*** are very complex. A measure of the complexity of Church's original formulation of the system is that even someone of his extraordinary intellectual powers and his proverbial carefulness could have overlooked an essential flaw in it. It collapses into a purely extensional system, which is, of course, precisely what it was meant not to do.

Montague had played an active role as member of the doctoral committee of David Kaplan, whose dissertation was about the ***Logic of Sense and Denotation***. I suspect that some of the decisive ideas about HOIL may have taken hold at that time, as a result of the intensive interactions between him and Kaplan. But, as far as I remember, it wasn't the interaction with Church himself that was immediately relevant to HOIL's final formulation. Of direct importance to that formulation was a lively correspondence in 1965/66 between Montague, Kaplan, Scott and John Lemmon, mostly over the handling of constants and variables (both first- and higher-order) within such a non-extensional system. I very much doubt that Montague consulted Church, in correspondence or in person, over the exact formulation of HOIL. In retrospect this lack of contact between two members of the same department, who were aware of the closeness of their interests, may seem odd, especially if they respected each other as much as I know Montague respected Church. I suppose Church must have had a high respect for Montague as well, though this is just an assumption. But the explanation might well be that Church was a very private person. He would not be one to go to someone as much his junior as Montague for a chat or for some informal

---

[6] Church, Alonzo, 1951, A formulation of the logic of sense and denotation, in: Paul Henle, Horace M. Kallen and Susanne K. Langer (eds.), ***Structure, Method, and Meaning: Essays in Honor of Henry M. Sheffer***, Liberal Arts Press, New York, 3–24.

[7] Frege, Gottlob, 1892, Über Sinn und Bedeutung, ***Zeitschrift für Philosophie und philosophische Kritik***, NF 100, 25–50. Translated by M. Black under the title 'On sense and reference' in: Peter Geach and Max Black (eds.), ***Translations from the Philosophical Writings of Gottlob Frege***, 2nd edn., Oxford, 1960, 56–78.

advice. And on the other hand, but by much the same token, he wasn't either the sort of person that one would feel comfortable approaching about something, unless one had worked out what one wanted his advice on in pretty much all relevant detail.

I mentioned that Montague saw HOIL as a new formal framework for addressing philosophical questions and that questions of natural language semantics were just one item on his much longer agenda. Even so it is probably right to say that it has been through its applications to natural language semantics—the application that became known as Montague Grammar—that HOIL has become most widely known and used. Since Montague Grammar was, almost from the time it became established through Montague's seminal contributions to natural language semantics, the concern of linguists and only marginally of philosophers, HOIL (and some of its variants like Gallin's Ty2), are much better known in linguistics than they are in philosophy.[8] This is one reason why Montague Grammar stands out among the possible and actual applications of HOIL and it is a reason for dwelling on this one of Montague's accomplishments a little longer.

AtM: A sideline: the richer notion of an individual concept was already in Carnap's ***Logische Aufbau***, right?[9] How did that work tie into Montague's HOIL?

HK: The notion also played a role in Carnap's ***Meaning and Necessity***.[10] Carnap's work was a very important input to what was going on at the time when Montague and Carnap were both on Kaplan's dissertation committee and later on, when HOIL was honed into final shape. I do not know how much interaction there was between Montague and Carnap at that time. Carnap was the official advisor of Kaplan—but Montague once told me that he did most of the actual advising. How much actual personal interaction there was between Carnap and Montague in the early sixties or before I do not know. My impression is that there wasn't much, probably due to similar barriers that I suggested earlier may have limited direct interactions between Montague and Church. By the time I came to UCLA Carnap was retired and lived a quite secluded life, partly because of his failing eyesight.

---

[8] Gallin, Daniel, 1975, ***Intensional and Higher-Order Modal Logic***, North-Holland Publishing Company, Amsterdam.

[9] Carnap, Rudolf, 1928, ***Der logische Aufbau der Welt***, Weltkreis-Verlag, Berlin-Schlachtensee, Neuaufl. Meiner, Hamburg, 1998.

[10] Carnap, Rudolf, 1947, ***Meaning and Necessity: A Study in Semantics and Modal Logic***, The University of Chicago Press, Chicago, IL.

As regards individual concepts, yes, these did play a crucial part in the discussions of what would be the best way to define a non-extensional type-theoretical system. One feature of HOIL that is anything but self-evident against the background of extensional systems like ordinary predicate logic or the extensional typed lambda calculus is the different treatment of variables and constants. This is a difference that in HOIL holds at all levels, but it is easiest to explain in relation to individual constants and variables. Individual constants behave in such a way that they can be used as designators of (non-rigid) individual concepts such as *the US president*. They can be used that way because the semantic value of a constant in an intensional model can vary from world to world. Variables, on the other hand, are assigned values independently of worlds; that is, variable assignments are defined in such a way that they assign individuals to individual variables, and not individual concepts. To handle variables and concepts in this way constitutes one from among several options that present themselves when we pass from an extensional to a non-extensional system. It is now common place that the option Montague chose is a viable one, which among other things is compatible with a uniform assignment of types to argument slots. For instance, a 1-place predicate P of individuals has the type $\langle e, t\rangle$; both '$P(v)$', where v is an individual variable, and '$P(c)$' where c is an individual constant, are well-formed formulas and their semantics works out as intended. But at the time this wasn't obvious—not until one had seen things spelled out and been able to verify that it was actually possible to proceed in this way. Whether the choice made by HOIL is the conceptually right one is, of course, another question. That question is connected with another, truly fundamental question: Is intensionality the right kind of non-extensionality, and, more specifically, is it the kind of non-extensionality we find in the languages we speak? In a way the answer to that question was already known at the time when HOIL was given its definitive characterization: it is 'No, not in general'—a point made at the time in a review *Is Necessity the Mother of Intension?* of some of Montague's work by Fred M. Katz and Jerrold J. Katz.[11] I thought then, as I do now, that this review was unsympathetic beyond the call of duty. But its title was spot on. The difficulty it points at is that intensionality isn't the right instrument for dealing with the propositional attitudes, because its inability to deal with the problems of logical omniscience—an inadequacy of Intensional Logic that

[11] Katz, Fred and Jerrold Katz, 1977, Is necessity the mother of intension?, ***The Philosophical Review***, 86(1), 70–96.

strictly speaking vitiates all its applications to the attitudes and the semantics of attitude reports. The merit of Church's ***Logic of Sense and Denotation***—in all its consistent versions—is that it contains safeguards against this intensional oversimplification. But the price, as we noted earlier, is that the system is very difficult to use in applications. Because of this, HOIL is at a significant advantage even there where it is strictly speaking wrong to apply it. Here too natural language semantics has adopted a policy of closing an eye to the snakes in the grass; for otherwise it would be very hard to discern the insights that an intensional analysis *can* give, even in cases where the grass isn't snake-free.

In the Fall of 1966 Montague gave a seminar on the application of logic to the semantics of natural language. From the start the perspective that informed what he presented in the seminar was the one epitomized in his often quoted remark that 'there are no fundamental differences between natural languages and the artificial languages of formal logic'. After nearly half a century of intensive work on the syntax and semantics of natural language much of which would not have been possible without his contributions, it is easy enough for us today to come up with a whole catalogue of differences between natural languages and the formal languages that Montague must have been thinking of then—enough, it would seem, to hold Montague's slogan up to ridicule. What we should not forget is how bold and fruitful this position was in those days, and to make good on that position in the way that Montague did. The common attitude among logicians and logically minded philosophers was that natural languages are ever so many Heath Robinson machines, which we somehow manage to make do the things that we need them for, but which only hang together through endless numbers of *ad hoc* devices, without rhyme or reason and that look so fragile when we observe them through the glasses of a formal logician that it is a lasting mystery why they do not break down more often than they do. I mention the seminar because it was in its course that what was to become Montague's own form of Montague Grammar took on more specific contours. A crucial occasion that I still remember clearly was when one of the seminar participants—Bob Mattison, a graduate student from Berkeley who had come to UCLA to work with Montague and who had become closely integrated in the group of devotees that attended the graduate seminars Montague offered—had unearthed one of the papers on categorial grammar by Adjukiewicz, quite hard to get hold of then, and brought it to the attention of the seminar. To my knowledge this is how Montague became aware of this work on categorial grammar. After that the seminar took a clear turn, in the direction of Montague Grammar as we know it,

in which semantic values of complex linguistic expressions are described as compositionally determined by the semantic values of their smallest meaningful parts (that is, of the words and the semantically contributing morphemes from which they are made up) via a series of immediate dependencies that are made visible by the expression's syntactic tree. Montague recognized right away the value of Adjukiewicz-type grammars for the project of giving a compositional account of the form-meaning relation—because the syntactic analyses they assign to complex expressions make the compositional relation between syntax and semantics especially transparent.

AtM: In the formal work in natural language in the fifties and sixties, the two figures that stand out are Chomsky and Montague. In certain respects their aims can be seen as quite similar. Yet, there seems to have been little contact between the two and, in fact, not much by way of mutual appreciation. Why do you think this was so?

HK: The lack of mutual interest between Chomsky and Montague—to the extent that there was any perception of the other's achievements, it seems to have been largely negative—may seem strange in one way, since they were motivated by the same deep insight and concern: that natural languages, for reasons that we now find commonplace, but that weren't at the time, were subject to far greater systematicity than had been assumed until then or at least than anyone had been prepared to openly assert. The relation between form and meaning, they both saw clearly and posited emphatically, had to be essentially lawlike, and for natural languages this had to be in some important sense by 'design'—just as for formal languages, even if the designing and the designers were different.

But for Montague, coming from formal logic and, even more importantly, from the model-theoretic school of Tarski, where it is the received view that the conceptually transparent way of defining correct inference and entailment should be given in model-theoretic terms—a theory of form alone could not possibly suffice by itself. Logical form, whether close to or distant from surface form, is only (more) form, and therefore cannot provide an explanation of the form-meaning relation, unless it is backed up by an account of the semantics of the logical form language in model-theoretic terms.

Closely connected with this was a firm anti-psychologistic methodology. The only trustworthy evidence for or against a theory of the form-meaning relation for natural languages was supposed to be given by speakers' judgments about entailment relations between sentences. Exactly how a theory

should account for the positive and negative entailment pairs that can be ascertained in this way for a given language on the basis of speakers' judgments is pretty much a matter of what works best. And Montague would not have wanted to draw any conclusions from the details of a theory that describes a language successfully according to this criterion to what is actually going on in the minds of speakers that produce or interpret the expressions of that language and come up with their judgements of what follows from what. He did allow for the relevance of syntactic simplicity or its faithfulness to basic syntactic intuitions. For instance, he thought the syntactic structures that such a theory assigns to the sentence of a given fragment of English should be close to our central intuitions about English sentence structure. But as far as I can recall, Montague wanted to see those judgments as the only points of contact between his theoretical descriptions of the semantics of natural language fragments and the reality of how those languages are actually used and understood.

For Chomsky such considerations about giving a separate account of the semantics and/or logic of 'logical forms'—i.e. those forms that the theory presents as the direct harbingers of meaning—never seem to have had any purpose. This is not because he saw intuitions about what follows from what as irrelevant in general, but because he thought they could not tell us much about that specific aspect of human cognition that enables us to acquire and use a human language. Too many other cognitive factors enter into logical and semantic judgments and into the cognitive processes that are the substance of such judgments to allow any conclusions vis-à-vis the language capacity as such, as distinct from all the other cognitive modules that are presumably involved.

None of this explains why Montague should not have adopted Chomsky's proposals for the theory of grammar for his purposes, i.e. as the syntactic basis for a model-theoretic account of semantics. But Montague felt very strongly that the theory of grammar that Chomsky proposed at the time—the Extended Standard Theory of ***Aspects of the Theory of Syntax***—lacked the formal rigor he needed as syntactic basis for his model-theoretic definitions of semantic values.[12] I have sometimes speculated whether Montague's attitudes might have changed had he lived longer. But I am not sure. By the time he died there already seems to have been a kind of entrenchment; quite a bit would have had to change somewhere along the line.

[12] Chomsky, Noam, 1965, ***Aspects of the Theory of Syntax***, MIT Press, Cambridge, MA.

AtM: Was the fragment strategy suggested by the need to avoid circularity, having no truth predicate and hence no threat of paradoxes?

HK: No, I don't think that was the real reason. The basic reason is very simple. If you want to describe a language in the rigorous way that Montague wanted there is little else you can do, short of giving a rigorous treatment of the syntax and semantics of the language as a whole, with all its different constructions and its entire vocabulary. Though I guess that too would count as a fragment—an 'improper' fragment, just as a set is an improper subset of itself. But it is one that is just too big to handle to start with, or for that matter, whenever you want to explain how some part of the language really works. Giving such a complete account for a language in its entirety is unattainable even today. It was totally out of the question at the time when the enterprise was just getting under way, and as a way of showing how such a theory could be made to work, it would probably have been the wrong thing to try, even had it been a realistic possibility. And besides, it isn't even clear that the notion of a human language in its entirety is well-defined.

It is true that most work today, both in syntax and in formal semantics, looks like it has abandoned the principle that all that we can do is to describe the syntax, or the syntax and semantics, of certain fragments of the languages we study. What such studies typically present are analyses of individual words, morphemes or constructions, which take the form of particular syntactic or syntactic + semantic principles. The implicit assumption behind such proposals is that those principles proposed in such individual studies can be integrated into a single comprehensive syntax, or a single comprehensive syntax-cum-semantics, for the language as a whole. The worry of those who complain about the failure of such studies to present explicitly stated grammars, consisting just of a syntax or of a syntax and semantics, for fragments that include the particular phenomena to which the study is devoted is that this implicit assumption may be illusory. Putting together a grammar which integrates all such principles, for a fragment that includes all these constructions, morphemes and words, is anything but a straightforward exercise. The ways in which the different principles interact can be just as intricate as the principles themselves and these interactions have as a rule not been considered by those who have stated and studied the principles involved. How tricky it can be to integrate rules that were designed to deal with particular constructions into a single system is something that is all too familiar to grammar developers in Computational Linguistics. I have heard and seen a little bit of this in connection with the LFG grammar and parser for German that have been developed over a

number of years by a group of my colleague Christian Rohrer. One problem with such parsers and the grammars they implement is that unforeseen interactions create the possibility of assigning certain strings syntactic structures that they simply cannot have.

You also raise a question about the notion of truth: is it, or is it not, one of the tasks of a grammar for a natural language to specify under what conditions any declarative sentence of the language is true? This is an important and delicate issue—one whose importance and difficulty we have been aware of since the work by Gödel and Tarski in the nineteen thirties. Their results have taught us that you cannot have a consistent semantics, of the sort that nearly all current formal approaches to natural language semantics are trying to put into place, for a language that has certain expressive powers, and these powers are present almost before you know it—not only in human languages as they exist, but also in most reasonably comprehensive fragments of them. It is curious that this point is so rarely discussed in semanticist circles. It is part of our overall semantic project that the accounts we propose for certain parts of the language should eventually be put together into a semantic account for the language as a whole. But if that is the ultimate goal, then at some point some much more drastic adjustments will have to be made than many of us ever bother to articulate even to ourselves. Since natural languages do have the means to provide structural descriptions of their own expressions, and since they do have overt truth predicates, such as *is true, is a truth* and so on, they are among those languages for which a syntax and semantics of the sort we are trying to provide for them are impossible.

Interestingly it was Montague himself who, in his justly famous paper ***Syntactical Treatments of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability***, demonstrated that this problem goes deeper and affects many more expressions than just those that refer to truth.[13] All these results—Montague's as well as Tarski's and the many other semantic paradoxes that have been discussed since the early days of the 20-th century—go to show that in the syntactic and semantic descriptions of most and probably all human languages there comes a point where safeguards have to be put in place against the semantic paradoxes. There are

[13] Montague, Richard, 1963. Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability, ***Acta Philosophica Fennica*** 16, 153–167. Reprinted in: Richard Montague, ***Formal Philosophy: Selected Papers by Richard Montague***, edited and with an introduction by Richmond Thomason, Yale University Press, New Haven, CT, 1974, 286–302.

various proposals for how this might be achieved, for instance, via Property Theories of the sort proposed by Açzel, Turner, Chierchia and Fox, or by partial definitions of semantic notions, in the spirit of Kripke's ***Outline of a Theory of Truth*** or of the stabilization approach of Herzberger, Gupta and Belnap.[14]

AtM: Did Barbara Partee take part in Montague's seminar at UCLA at that time?

HK: No, she did not. During the years I spent as a graduate student at UCLA I met her only once, in my last year there. She was known as this very bright young linguist, from the first crop of graduate students who got their degree with Chomsky at MIT. Barbara had been invited to give a talk in the philosophy department; if I am right, that was in 1968 and the topic was negation. I believe that was the only time I saw her before I got my Ph.D. from UCLA and left. The collaboration between Montague and her started after that, somewhere between 1969 and when Montague died in 1971. So it was sadly short-lived. Joint advising was I believe an important part of it. I can think of only one case right now, viz. the advising of Michael Bennett, but for all I know there may have been others. But, of course, such cooperations could not have been enough by themselves to turn Barbara into the champion of Montague's method that she soon became. For Barbara the discovery of the formal approach to natural language semantics Montague had just put on the map must have been a pivotal experience, which proved to be decisive for her whole further career. Unless I am mistaken the term 'Montague Grammar' is due to her. In any case, if it hadn't been for her, it is quite certain that Montague would not have had the impact within linguistics he did have; and I am not sure whether he would have had more than a fleeting impact. The fact that Barbara was already established as one of the new leading generative syntacticians, and the persuasive arguments she was able to

---

[14] Açzel, Peter, 1988, ***Non Well-founded Sets***, CSLI Publications, Stanford, CA.

Turner, Ray, 1987, A theory of properties, ***The Journal of Symbolic Logic***, 52(2), 455–472.

Chierchia, Gennaro, 2006, Broaden your views. Implicatures of domain widening and the "logicality" of language, ***Linguistic Inquiry***, 37(4), 535–590.

Fox, Danny, 2007, Free choice disjunction and the theory of scalar implicatures, in: Uli Sauerland and Penka Stateva (eds.), ***Presupposition and Implicature in Compositional Semantics***, New York, Palgrave Macmillan, 71–120.

Kripke, Saul, 1975, Outline of a theory of truth, ***Journal of Philosophy***, 72(19), 690–716.

Herzberger, Hans, 1982, Notes on naïve semantics, ***Journal of Philosophical Logic***, 11(1), 61–102

Gupta, Anil and Nuel Belnap, 1993, ***The Revision Theory of Truth***, MIT Press, Cambridge, MA.

give for why formal semantics was needed, were essential to its acceptance within a community where acquiring an operative mastery, or even just a passive mastery, of the formal tools used in MG was a non-trivial matter—then even more so than today.

AtM: How did Rich Thomason get involved with this?

HK: David Kaplan had asked him to take charge of the posthumous edition of the papers of Montague: ***Formal Philosophy***. I am not sure how much of an interest Thomason had at that point in Montague's work. He was more known as a logician with among other things an interest in modal logic and there is the well-known joint work with Robert Stalnaker on adverbs, which appeared at roughly the same time as ***Formal Philosophy***. The choice of Rich as editor was evidently a very good one. For one thing he got the collection out in a remarkably short period of time, especially if one keeps in mind that he cannot have been familiar with more than a few of the papers in it, when he was put in charge.

AtM: Linguists, and perhaps also many philosophers nowadays, tend to think of Montague as the creator of Montague Grammar. But does that image do justice to his own perception of the contributions he had made?

HK: You are right to ask that. In fact, one almost certainly should answer the question with 'no'. For most of his professional life Richard was a logician, with important contributions to his name in proof theory, set theory, the paradoxes (really an offshoot of his work on proof theory and on Gödel), recursion theory and model theory. His formulation of HOIL, another important contribution, came comparatively late in his short career, and it was only this that enabled him to do the work on natural language semantics. But even within the narrower confines of this part of his total oeuvre he himself seems to have thought of his work in natural language semantics as just one of the different possible purposes to which HOIL could be put. As things have turned out, there is only one paper from his hand in which HOIL is applied to questions that do not directly belong to natural language semantics, *On the Nature of Certain Philosophical Entities*. Whether more such papers would have followed had he lived longer can only be a matter of speculation.

AtM: That paper never had much of an impact neither in the philosophy community, nor in linguistic community, right? Since Montague's paper *English as a Formal Language* with its direct model-theoretic interpretation was too hard to read for many people, the paper *On the Proper Treatment of*

Quantification in Ordinary English became like the gold standard for natural language semantics.

HK: I think that one reason why *On the Nature of Certain Philosophical Entities* didn't have more of an impact on the philosophical community than Montague's papers on natural language semantics is that the paper remained an isolated contribution. There never were any follow-ups with more concrete applications of the ontological and logical proposals the paper makes to problems that preoccupy philosophers working in the respective areas to which the paper intends to make a contribution. Montague didn't have an opportunity to do so himself, and nobody in any of the relevant branches of philosophy felt compelled to pick up the torch and show the value of these proposals for that branch of philosophy. Some of the contributions made by Montague's papers on natural language are remarkably salient. Apart from the fact that these papers show that something can be done for significant parts of natural languages that many had doubted could be done at all, they contain at least one specific proposal that was an instant hit. I am thinking of Montague's insight that you can treat all noun phrases—both the referential and the quantificational ones—as denoting properties of properties. That idea has made us rethink some fundamental questions about reference, predication and truth. It opened our eyes to a new range of ways in which the syntax of a language can map onto its semantics and it resolved a puzzle that had been with us ever since Frege had pointed out that the grammatical similarity of sentences like *Socrates is mortal.* and *Every man is mortal.* should not mislead us about the fundamental differences in their logical forms. There is no question but that to perceive that difference and to integrate it in the design of a formal theory of reference, predication and quantification in the way Frege did has been one of the signal achievements of our civilization. But it left (or should have left) a sense that something was still missing: if the logical forms of these two sentences are as radically different as predicate logic seems to tell us, then why should natural language grammar be so radically misleading—how can it get away with misleading those whom the language, with its overtly recognizable grammatical structure is after all designed to serve? I think it is fair to say Montague has given us most of the answer to that question.

AtM: The history of generalized quantifiers started in mathematical logic. Was it an independent development at that time?

HK: As far as I can tell, the answer is 'yes'. For instance, the work on generalized quantifiers of Mostowski goes back to well before Montague's work on

Montague Grammar.[15] And some of the signal results on generalized quantifiers in mathematical logic, such as Keisler's theorem that adding *there are uncountably many* to first-order logic preserves its axiomatizability, do not seem to have anything to do with natural language.[16] In fact, the overlap between results about generalized quantifiers that might interest a mathematical logician and those that can be seen as relevant to natural language semantics isn't all that large. Barwise's result about non-axiomatizability of first-order logic + *most* with its standard semantics might qualify as an example, as might Johan van Benthem's results about the semantics of generalized quantifiers with finite domains, or the results of Keenan, Westerståhl and others about polyadic quantifiers, assuming we count these as generalized quantifiers too.[17] But as far as the impact of Montague's ideas about noun-phrase semantics is concerned, the seminal contribution is no doubt the paper *Generalized Quantifiers and Natural Language* by Barwise and Cooper.[18] In fact, this paper signals a historical rapprochement between natural language semantics and mathematical logic. Barwise and Cooper were colleagues at the University of Wisconsin. Barwise, an outstanding mathematical logician with an acute interest in foundational questions, had become convinced that fundamentally new directions were needed to make progress with certain problems in the foundations of mathematics. In search of alternative directions he decided to attend one of Cooper's seminars on natural language semantics. Cooper, who had done his Ph.D. with Barbara Partee, was one the world's experts on Montague's work, and his own work at the time can be roughly described as embracing some variant of Montague Grammar. The methods he taught and exemplified in the seminar must have persuaded Barwise that here was a way of describing natural language meaning that was capable of revealing the ways in which we express things in natural language faithfully enough and that at the same time

---

[15] Mostowski, Andrzej, 1979, ***Foundational Studies. Selected Works***, I–II, North-Holland Publishing Company, Amsterdam.

[16] Keisler, Jerome, 1970, Logic with the quantifier "there exist uncountably many", ***Annals of Mathematical Logic***, 1, 1–93.

[17] Barwise, Jon, 1979, On branching quantifiers in English, ***Journal of Philosophical Logic***, 8(1), 47–80.

van Benthem, Johan, 1986, ***Essays in Logical Semantics***, Kluwer, Dordrecht.

Keenan, Edward and Dag Westerståhl, 2010, Generalized quantifiers in linguistics and logic, in: Johan van Benthem and Alice ter Meulen (eds.), 2010, ***Handbook of Logic and Linguistics***, 2nd rev. edn., Springer, Berlin, 859–910.

[18] Barwise, Jon and Robin Cooper, 1981, Generalized quantifiers and natural language, ***Linguistics and Philosophy***, 4, 159–219.

satisfied the standards of rigor that a mathematician considers a *sine qua non* if anything tangible is to be achieved. In this way the basis was laid for a remarkably fruitful cooperation, which lasted until Barwise's death and of which *Generalized Quantifiers and Natural Language* was the first major result.

AtM: Is there any connection you can see between *Generalized Quantifiers and Natural Language* and the development of Situation Semantics (SS) that conquered the world soon after and that was part of Barwise's central focus till his untimely death in 2000 and has also been very prominent in Cooper's work since the early eighties?

HK: The first paper that was truly about Situation Semantics that I recall was *Scenes and other Situations*.[19] When I read this paper—if I remember correctly, that was shortly before it appeared in print—I was very impressed by it. Because of its situation-theoretic approach it is of course quite different from *Generalized Quantifiers and Natural Language*, but, on the other hand, it shares with that paper many methodological and technical assumptions that are distinctive of Montague Grammar—so much so that it could not possibly have been written by someone who wasn't thoroughly familiar with Montague's work. In fact, the continuity between Montague Grammar and Situation Semantics should not be overrated any more than it should be underrated.

Situation Semantics soon became the joint project of Barwise and John Perry. In fact, as far as I know, one of the reasons why Barwise came to Stanford on a permanent basis was for him and John Perry to be able to work together in that way which is only really possible when you work and live in the same place. But at that point SS had already taken shape to a considerable extent. That was already so in the academic year 1981–1982, when Barwise and Perry were both part of a group of people, convened by Stanley Peters, who spent a year at the Center for Advanced Study in the Behavioral Sciences. The other members were Stan himself, Lauri Karttunen, Robin Cooper, Manfred Bierwisch, Johan van Benthem and some younger people, some of them brought along by Stan and Lauri from the University of Texas, where they were both working at the time, of whom I remember in particular Edit Doron, Elisabet Engdahl and Hans Uszkoreit. During that year at

[19] Barwise, Jon, 1981, Scenes and other situations, ***Journal of Philosophy***, 78, 369–397. Reprinted in: Jon Barwise, ***The Situation in Logic***, CSLI Lecture Notes 17, CSLI Publications, Stanford, CA, 1989, 5–36.

the Behavioral Sciences Center, SS was already widely discussed throughout the US and beyond.

AtM: What do you see as the fundamental step from PTQ and what I call 'plain vanilla' Montague Grammar, on the one hand, towards a more dynamic semantics on the other? And what was the role of SS in this transition?

HK: Whether SS is a form of dynamic semantics, or a step in that direction, depends on how you look at it. The SS of *Scenes and Other Situations* is no more dynamic than classical MG. The difference is just that in SS the semantic realities of certain expressions—including in particular naked infinitival complements of perception verbs—are situations rather than any of the entities made available by the ontology of HOIL. In SS properly speaking this assumption is integrated into a theory whose central semantic notion is that of a situation supporting a formula, instead of that of a formula being true at a possible world, or 'index' in the technical sense adopted in possible worlds semantics. However, you can also think of SS as a theory of how utterances can incrementally build the situations that support them, and from this perspective it can be seen as dynamic in a way that MG is not. But I am not sure to this day whether this second perspective does justice to the way in which Barwise and Perry themselves saw SS, whether at the time when SS got under way in the early eighties or later.

AtM: Speaking of dynamic semantic theories, you yourself have spent a lot of your time developing Discourse Representation Theory (DRT). Was DRT the result of your year at the Stanford Center?

HK: No, DRT already existed. In fact, it goes back to a problem that Christian Rohrer once put to me in Stuttgart about the difference between the French tenses 'imparfait' and 'passé simple'. Christian had become interested in the use of formal methods when working for his Ph.D. under Coseriu in Tübingen, first in the context of machine translation. After a couple of years in the US and another two to write his 'Habilitation' in Germany he had become Professor for Romance Linguistics at the University of Stuttgart (at the age of 29!) and obtained a grant from the German Science Foundation for the application of formal logic to the semantics of natural language. By the end of the seventies he and his group had begun to concentrate on problems having to do with the use of the tenses in French and it was in that context that in 1978 he draw my attention to one of the classical problems in this domain: How do French speakers and authors choose between PS and Imparfait and what are the theoretical principles that govern such choices? French philologists had come up with various largely informal

accounts of the difference between PS and Imparfait, among other things that the PS conveys punctuality whereas the Imparfait conveys durativity of the events they are used to describe. Christian at that point thought that the model-theoretic approach of Montague Grammar might be the right framework for capturing this and other distinctions that had been proposed in the literature. But when you start thinking more carefully about what this punctuality-durativity contrast could actually amount to, you come to see that the standard methods of mathematical model-theory are not really the right tools for capturing the insights that the more traditional philologists who had been thinking about tenses in Romance had undeniably had. For instance, what could it be for an event described with the help of the PS to be 'punctual'? Surely not that such events last for only one 'point of time'—one instant of the time we use in physics to describe physical processes such as atomic or astronomical motion. The right perspective, it seemed, which could make coherent sense of the claim that PS events are punctual and Imparfait events are not, is one according to which discourse, with its choice of particular time forms, creates its own ontology, and, as part of this, its own 'discourse time'. It is in the sense of this discourse time that the distinction between the 'punctual' PS events and the 'non-punctual' Imparfait events can be given what seemed to me a plausible explication. In particular, it is possible for an event to be punctual in the sense of discourse time and yet cover an extended stretch of real time.

AtM: That sounds like it is more aligned with Montague's program to model philosophical entities than with the Davidsonian theory of events as arguments of verbs.

HK: It is really quite different from what either Montague or Davidson were trying to do. Distinctive of an account of tensed discourse along the lines I was indicating is that it postulates a level of semantic representation of discourse distinct from the syntactic representations of the sentences from which the discourse is made up, on the one hand, and from the reality that the discourse is describing on the other. It is the difference between the discourse representations of this intermediate level and the reality described by the discourses represented by them that is responsible in particular for the difference between the discourse time generated by a discourse representation and the 'real time' of the physical world that is being described. And the relation between discourse time and real time is such that an event from the discourse representation may be punctual in the sense of discourse time, while the real event that it represents is extended in the sense of the real time of the described world. This was the conceptual nucleus which

after a couple of years led to DRT as a general method for doing natural language semantics, in which discourse representations (DRT's Discourse Representation Structures) act as intermediaries between natural language discourse and the real or fictional worlds described by it. Because of the central role played by its DRSs DRT can be characterized as a logical form theory, much like earlier, usually not very systematic attempts to capture the meanings of natural language sentences by specifying their logical forms in predicate logic. In both these types of logical form account, the logical forms proposed have independently grounded meanings (the grounding is either model-theoretic or proof-theoretic or both). It is the logical form which specifies the meaning of that to which it is assigned as logical form. But compared with that kind of logical form approach DRT is a logical form theory with an extra twist, in which the representations are in general representations of larger units than single sentences. In fact, it was the use of DRSs as logical forms, and thus as intermediaries between natural language and reality, that seems to have provoked quite a strong resistance in the early years of DRT.

KvH: What were these problems that people had with accepting this additional representation layer?

HK: Actually, DRT soon came under attack from different quarters. One of these was the point of view of those that became the core of what is now often considered the 'true' Dynamic Semantics, starting with the work from the late eighties of Groenendijk and Stokhof (G&S). They also advocated a form of incremental interpretation to deal with donkey pronouns, but criticized DRT for making what they saw as the unwarranted claim that such an account had to involve a distinct level of semantic representation. And it seems that since then quite a large part of the community sees their alternative proposal and its many subsequent variations as Dynamic Semantics in the true sense of the word. It is also often said in conjunction with this that DRT is a kind of pseudo-dynamic theory, a doubly unfortunate combination of unwarranted representationalism on the one hand and a failure to distill the central dynamic concepts on the other. It should be stressed in this connection that Groenendijk & Stokhof's critique was directed at the theory presented in the paper *A Theory of Truth and Semantic Representation* (TTSR).[20] That paper focuses on the phenomenon of donkey pronouns—anaphoric pronouns whose anaphoric antecedents

[20] The paper is included in this book in Part Four.

are indefinite noun phrases in the same or an earlier sentence. G&S were right, or almost right in their claim that, when it comes to accounting for the donkey pronoun phenomena that TTSR discusses, a DRS interface between natural language syntax and described world truth conditions isn't needed. What we can do instead is to complicate the semantic values that are assigned to natural language expressions, such as information states and Context Change Potentials in lieu of propositions. Arguably, though, that isn't quite enough. One also needs a record of what variables have already been used and at which embedding levels; usually that information can be recovered from the information states. When a variable has not yet been used at a given level, the assignment functions that are part of the corresponding information state impose no restrictions on the possible values to that variable, but if the variable has been used, some values will have been excluded. But there is no *a priori* reason why used variables should always come with such a restriction on their possible values. It could be a feature of particular natural languages or fragments, and thus something one could rely on when dealing just with those. But it isn't something that can be taken for granted and, if true, that should be explicitly and carefully argued for.

AtM: A complaint one sometimes hears from people who have adopted Dynamic Semantics in the spirit of Groenendijk, Stokhof, Veltman and others is that DRT isn't really a dynamic theory of semantics at all—that it is just the combination of a purely static model-theoretic semantics for the logical form language whose formulas are the DRSs and some translation algorithm for converting sentences and texts from the natural language into the logical form language.

HK: Yes, that is of course a way you could describe DRT, and it is one that doesn't make it sound very dynamic. But what makes DRT a dynamic approach in my own eyes even so has to do with two things that this description leaves unsaid. First, the algorithm for forming sentences and discourses into DRSs is incremental. This is most obvious for the case of multi-sentence bits of discourse. Here the algorithm starts with constructing a DRS for the first sentence, then uses this DRS as a guideline (or 'discourse context') for the construction of a DRS for the next sentence, which gets integrated into the first DRS to form a DRS that represents the first two sentences together; this last DRS then acts as discourse context in dealing with the third sentence and so on. The role that DRSs play in the applications of the translation algorithm is thus a double one: on the one hand, a DRS represents the content of what has been translated already—it identifies

that content via the truth conditions that it has by virtue of the model-theory of the DRS language; on the other, the DRS serves as discourse context for what it is to be interpreted next. DRSs thus emerge as a kind of unification of content and context. They function as links between what has already been processed and what is still to come. And the construction algorithms that make use of them are 'dynamic' in that they build DRSs incrementally, sentence by sentence.

At the same time the double role that DRSs play in DRT may be taken as an argument for an approach that makes use of them as constituting a separate level of semantic representation. The fact that they can at the same time capture the content of the bits of language from which they have been derived and play an effective part in guiding further interpretation can be seen as in indication that they capture some of the reality of incremental processing that goes on when human speakers interpret a discourse. Seen from this angle DRT takes on the status of a theory of mental processing of language and of the mental representations that are produced and manipulated in the course of those processes. That is, of course, no refutation of the argument G&S launched, but it is making a case for the logical form approach of DRT that is largely independent from their critique.

AtM: When one thinks of DRT as a logical form theory which transforms sentences and sentence sequences into 'logical forms' belonging to some logical formalism, one is reminded of Game-theoretical Semantics (GS) as proposed by Jaakko Hintikka in the nineteen seventies.[21] Was GS a source of inspiration for you?

HK: I had looked at GS quite closely when it became popular, partly through Hintikka's own publications and partly through those of his student and collaborator Esa Saarinen. So I was certainly aware of GS when I was working on the first explicit DRT construction algorithms, that of TTSR and that for a small fragment of French which focuses on temporal matters and deals specifically with the Passé Simple and Imparfait. (This second paper appeared, in French, in the French journal ***Langages*** and doesn't seem to have had much visibility.[22]) But what struck me as unsatisfactory about GS from the start, and still does so today, is that it was quite casual and

[21] Hintikka, Jaakko and Gabriel Sandu, 2010. Game-theoretical semantics, in: Johan van Benthem and Alice ter Meulen (eds.), ***Handbook of Logic and Language***, 2nd rev. edn., Elsevier, Amsterdam, 361–410.

[22] Kamp, Hans, 1981, Référence Temporelle et Représentation du Discours, ***Languages***, 36–64.

inexplicit about the syntax of the natural language sentences for which game-theoretic analyses were being proposed. That makes it very difficult to see exactly what the theory predicts, especially in cases where the crucial questions have to do with the interactions between pronouns and their quantificational antecedents. One way in which you become aware of this difficulty is when you want to prove for certain fragments of natural language that a given game-theoretical semantics proposed for them is first-order in that it assigns all sentences of the fragment first-order truth conditions. In GS such proofs are often desirable even for fragments where it seems intuitively clear or plausible that the game-theoretical treatment proposed for them is first-order. This is because the truth conditions directly assigned in GS are always second-order on the face of it: they always involve quantification over game strategies, which are functions from finite sequences of individuals to individuals. But in order to carry out such a proof —that a certain game-theoretical treatment of a certain natural language fragment *is* first-order—it is necessary that the treatment provide you with a fully explicit definition of the truth conditions it assigns to each of the sentences of the fragment. And that presupposes, among other things, that the syntactic structures of the sentences to which the game-theoretic algorithm assigns these truth conditions must have been made fully explicit. So, for me this was not an example to follow. If I got anything from GS that was directly relevant to the development of DRT it was a sharpened awareness of how important it is to be fully explicit about syntactic structure when you do natural language semantics. But, of course, that lesson is also plain from Montague's work, with which I had been imbued as a graduate student.

AtM: What you say about the second-order form of game-theoretically defined truth conditions reminds me of other work of Hintikka's from the sixties and seventies which addresses the question which parts of natural language are first-order and which irreducibly second-order. I am thinking in particular of what Hintikka has said about branching quantifiers.[23]

HK: Yes, that is a point on which Hintikka has been proved right many times over. At the time many people did not find his own examples of irreducibly second-order sentences—examples involving branching quantifiers that he claimed arise in sentences with more than one complex noun phrase of the form *a relative of every villager and a relative of each townsman*—wholly convincing. But at least in my own case that problem seemed no longer decisive,

[23] Hintikka, Jaakko, 1973, Quantifiers vs. quantification theory. ***Dialectica***, 27(34), 329–358.

once Barwise had come up with quite different examples involving quantifying determiners like *most* and *few*. In fact, today we are aware of another category of natural language sentences that are not first-order, because they involve plural indefinites and plural pronouns anaphoric to them. A classical example of this is Kaplan's *Some critics only admire each other*, for which he showed quite early on that its truth conditions are essentially second-order. That natural languages like English are not first-order, and for reasons other than overt quantification over properties, sets or predicates, is a point of the first methodological importance. But the moral isn't necessarily that in order to get things right we have to move to GS.

KvH: There were claims that when you formulated DRT you were influenced by AI and knowledge representation?

HK: No, I don't think that is true. What is true is that I participated in some of the meetings that the Sloane Foundation organized in the late seventies and early eighties as part of its effort to establish Cognitive Science as a distinct scientific discipline. And it is also true that in Cognitive Science the existence of mental representations isn't really questioned. The issue here is what representations are like and how they are used. I think that that probably had an influence and that it kept in check whatever sense I might otherwise have felt that I was out in left field. But at that point DRT was already in place. I do not think that contact with cognitive scientists had much, if anything, to do with its conception.

But one doesn't really have to look towards AI or Cognitive Science to find support for the idea of a representational approach to semantics. Something that has puzzled me for many years about the resistance against representation within formal semantics is that most of the fiercest opponents that I have come across have been people with a background in formal logic. This seems odd to me insofar as formal logic has traditionally been concerned as much with logical form as with anything else. Perhaps the most important and fruitful single insight in the history of logic, due to Aristotle, that from premises that tell you something about a given domain or situation you can arrive at further conclusions about that domain or situation through the application of purely formal operations, in which the content can be completely ignored. Logical form—how the given contents are represented rather than what content is represented—has therefore been an essential ingredient of formal logic since antiquity.

So why should people with a background in formal logic be suspicious of a logical form approach such as DRT to the semantics of natural languages? One reason, no doubt, is the intuition that appears to have moved

Montague: the 'logical forms' of natural language sentences ought to be the very structures that are ascribed to them by the syntax for the natural language itself. This is part of Montague's claim that there is no significant difference between natural languages and the artificial languages of formal logic. And we do know quite a bit about what that syntax is like on independent grounds, to some extent from intuitions that can be elicited from native speakers, from traditional work on grammar and arguably also from the more sophisticated methods of modern generative syntactic theory. In fact, this would be an attractive position and potentially an argument against the logical form approach of DRT, if only the 'logical forms' provided by the natural language syntax could pass muster as logical forms in the traditional sense attached to the term in formal logic, viz. as forms that can serve as premises and conclusions in formal accounts of logical deduction. But unfortunately no one has ever succeeded in showing convincingly that such grammatically motivated 'logical forms' can be made to serve in natural systems of formal logical deduction.

A related factor in the resistance to DRT as a logical form theory may have been the rise of model theory as a complement or alternative to proof theory and as providing the conceptually correct and universally applicable way of defining the basic concepts of inference, first and foremost that of logically valid consequence. The conviction that natural languages are not first-order may have further contributed to the conviction that the model-theoretic method is clearly to be preferred. Of course, that is not a good argument against the logical form approach, for there is nothing wrong with higher-order logical forms. In fact, none of the considerations I have mentioned seem to be good reasons against representational alias 'logical form' approaches as such. When it comes to what could have motivated the aversion against logical form approaches that seemed to be abroad at the time, I really feel at a loss.

AtM: As far as early criticism of DRT is concerned: it didn't come just from Dynamic Montague Grammar, but also from those who reject Dynamic Semantics in any form, and advocate some form of E-type analysis of donkey pronouns, right?[24]

---

[24] Evans, Gareth, 1980, Pronouns, ***Linguistic Inquiry***, 11, 337–362.

Cooper, Robin, 1979, The interpretation of pronouns, in: Frank Heny and Helmut S. Schnelle (eds.), ***Syntax and Semantics 10: Selections from the 3rd Groningen Round Table***, Academic Press, New York, 61–92.

HK: The E-type analysis in its original form—specifically the versions proposed by Gareth Evans and by Robin Cooper—antedates DRT by a few years. A legitimate reproach that could be made to the early presentations of DRT is that they did too little to compare the proposal they put forward with these alternatives. Let me try to say very briefly what I take to be the essence of the E-type account and the original motivation for it. The E-type account of donkey pronouns analyzes them as a kind of shorthand for definite descriptions. Evans' original motivation, I believe, for trying to deal with donkey pronoun phenomena in this way was the then widely held view that a satisfactory account of definite descriptions was already in place, in the form of Bertrand Russell's Theory of Descriptions.[25] In his view replacing donkey pronouns by certain definite descriptions provided an account of them, since it reduces sentences containing them to sentences for which a good account was already in place.

There are two empirical assumptions that come with such an account. The first is that donkey pronouns, in spite of the fact that they are anaphorically linked to an indefinite antecedent, nevertheless convey the kind of unique identifiability that is part of the meaning of Russellian descriptions. The second is that there is a systematic way to recover the definite descriptions that are to be substituted for the donkey pronouns to which the analysis is being applied. Evans was clearly aware of both these assumptions. He argues extensively that the first assumption is correct. And he shows his awareness of the difficulties with the second assumption: it isn't always possible to extract the substituting description by means of a strictly linguistic algorithm from the surrounding text, and in such cases an appeal to the context is necessary. Such cases are also a problem for Cooper's version, which offers a text-based algorithm for the construction of the substituting description. Even with this concession, however, the E-type theory makes a definite empirical commitment through its first assumption. Some of the predictions it makes in view of this assumption are different from those made by DRT and other, non-representational dynamic accounts of donkey phenomena. One such difference relates to discourses like *John owns a donkey. He keeps it in the stable next to his farmhouse*. Evans' E-type account makes the prediction that this discourse is felicitous only if the speaker has some specific donkey in mind in connection with her use of *a donkey*, viz. that donkey to which the definite description that is to be substituted for *it* picks out.

---

[25] Russell, Bertrand, 1905, On Denoting, ***Mind***, 14(4), 479–493.

In its original form, DRT predicts no such constraint: it assigns to a two-sentence discourse like the one mentioned simply the truth conditions of an existential quantification, contributed by the indefinite *a donkey* with the predications from both sentences within its scope. For sentence sequences of this kind, that are used and understood as describing situations in the real world, the E-type prediction seems right, so here dynamic theories are in need of correction. But the matter seems different for donkey sentences like Peter Geach's *If a farmer owns a donkey, then he beats it* or Irene Heim's telling example *If you are buying a sage plant here, you have to buy eight others with it*, said by one person to another who is contemplating buying a sage plant from a market stand, where they only sell sage plants in squares of nine. In these examples the kind of specificity that the pronoun in a donkey discourse seems to be imposing on its indefinite antecedent appears to be absent, and the absence of a prediction of such an effect is therefore desirable.

Subsequent refined versions of the E-type theory, initially by Irene Heim in her 1990 paper on donkey anaphora and by Stephen Berman in his dissertation, were designed to do justice to both these intuitions.[26] They combine the distinguishing features of the E-type approach—donkey pronouns are analyzed through reduction to non-pronominal terms and indefinites, including the indefinite antecedents of donkey pronouns analyzed as existential quantifiers—with a Situation Semantics of the kind developed by Angelika Kratzer.[27] To give a satisfactory account of all cases, however, especially of the recalcitrant conditionals with two similar indefinites in the antecedent, such as Jan van Eijck's *If a man lives with another man, he shares the housework with him*, or the bishop sentence *If a bishop meets a bishop, he blesses him*, even more complex situation-based analyses are needed, which bring substantial numbers of distinct situations into play. One problem with those analyses is that the situations they appeal to are in part very abstract. This has become more pronounced as the situation-based account has developed further, especially in the work of Paul Elbourne.[28] In these versions it becomes increasingly difficult to see exactly what differences remain

---

[26] Heim, Irene, 1990, E-type pronouns and donkey anaphora, ***Linguistics and Philosophy***, 13, 137–177.

Berman, Stephen, 1991, ***On the Semantics of Wh-clauses***, Ph.D. dissertation, University of Massachusetts, Amherst, MA. Published with Garland, New York, 1994.

[27] Kratzer, Angelika, 2002, Facts: Particulars or information units? ***Linguistics and Philosophy***, 25, 655–670.

[28] Elbourne, Paul, 2005. ***Situations and Individuals***. MIT Press, Cambridge, MA.

with DRT-based theories, which capture the relevant semantic properties in terms of discourse referents and DRS structure. A further important consideration is the role of event semantics: if, as increasing numbers of semanticists are taking for an established fact, events are an essential part of the ontology of natural language, then a situation theorist who accepts this fact, has to ask himself what the relation is between events and situations, and in particular, whether events, needed in any case, could not shoulder some or all of the burden of accounting for the local donkey phenomena we find in conditionals.

KvH: Returning to event semantics, any anecdote you recall about the work you did with Christian Rohrer?

HK: Nothing particularly amusing. One thing I should say first. The one who brought me to Stuttgart in 1972 was Franz Günthner. Franz was one of the members of Rohrer's group at the University of Stuttgart at that time. But it wasn't until 1978 that I spent more time in Stuttgart, long enough to do some real work. That is when the foundations were laid for the work that Christian and I did together in the eighties.

One thing I do recall is that in 1972 Rohrer's group was housed in the Schlossstraße in Stuttgart, in the attic of some rather gloomy building that, as one of only a few in that part of town, had survived the war. It is there that I developed the ideas about the role of discourse time in the processing of tenses and temporal adverbs as the basis for a linguistically meaningful distinction between punctuality and durativity. In particular, the manuscript of what subsequently appeared as the paper *Events, Instants and Temporal Reference* was written there, quite quickly by my standards, in the course of a weekend with much writing and not much sleep.[29] I still quite vividly recall that Rohrer's group shared the attic floor with the son of the caretaker of the building, who must at that time have been in his early twenties, or very late teens. He was a rather gloomy and taciturn character, whose impression on all of us was all the more sinister because he never said anything and looked like he thought the world would soon come to an end, and that when the time came, he personally would have nothing against helping it over the edge. Spending all or most of the night with this character next door and nobody else in the building was a little disconcerting, but fortunately I was too busy to give such thoughts much room to evolve. In any case, nothing ever happened and the man was probably perfectly harmless. Perhaps he

[29] This paper is included in Part One of this book.

would even have proved to be quite nice, had we made more of an effort to be nice to him.

I remember that summer, as well as the next one, when I returned for a slightly shorter stay, with great fondness. Christian had succeeded in getting a number of people to come to Stuttgart for short periods in the summer, including Dov Gabbay and Lennart Åquist. Together with the permanent members of the group we had regular meetings in which we took turns telling the others about our recent work and sometimes it was very recent, meaning from the week or the day before. All in all these summers were remarkably productive and a lot of fun.

KvH: You mentioned Heim's work. Were the two of you working independently? When did you first hear from her?

HK: The first time Irene and I met was at a conference at the University of Konstanz in 1978, where I presented material from the paper *Events, Instants and Temporal Reference* already mentioned. Irene was at that point a graduate student in linguistics at the University of Massachusetts, but her reputation had already reached beyond the places where she had been a student and I remember that I knew of her before we set eyes on each other at the meeting. Then, during the academic year 1981/82 that I spent at the Stanford Behavioral Sciences Center, Irene was at Stanford as holder of one of three fellowships that had been given to Julius Moravçsik. The other two fellows were Mürvet Enç and Susan Stucky. At that point the foundations of Irene's File Change Semantics (FCS) and DRT were already firmly in place. As I said, the first version of FCS goes back to 1979 and thus antedates the first explicit version of DRT by about a year. During the year at Stanford Irene and I talked a fair number of times, though not on a regular basis.

KvH: So your ideas developed in different places, but they were based on the same sources?

HK: That may depend on what you understand by 'the same source'. Irene was a Ph.D. student of Barbara Partee at the time, and I had talked to Barbara in the spring of 1980, a couple of days before that year's Amsterdam Colloquium (or 'Montague Colloquium', as it was then still called), which on that occasion took place some time in March. The talk I gave at the Colloquium discussed some of the ideas of DRT, but without the formal definitions, which came together only a few months later—some time in May, if I remember correctly—during a research stay at the Cognitive Science Center of the University of Texas. On the day before the meeting Barbara gave me her famous ball-example: (i) *One of the ten balls is not in the bag. It is*

under the sofa. (ii) *# Nine of the ten balls are in the bag. It is under the sofa.* In (i) *it* can be straightforwardly understood as referring to the missing ball, but in (ii) it cannot. The moral of this can be summarized as follows: (i) and (ii) have the same second sentence, and, while their first sentences are different, they express the same proposition, i.e. are true in the same possible worlds. But the interpretation of the second sentence that is possible in (i) is not possible in (ii). So the difference between (i) and (ii) cannot be explained in terms of the propositional contents of their first sentences, since they express the same proposition. Nor can it be explained in terms of the content or form of the sentence containing the pronoun, since that sentence is the same in (i) and (ii). Therefore the decisive difference must be the forms of the first sentences, as distinct from their contents. It is the form of the first sentence, and not just its propositional content, which provides the context that enables the pronoun to refer to the missing ball. This ball-example came to play an important role in the general argumentation that something is needed beyond standard Montague Grammar to give an adequate account of donkey phenomena. Barbara's instantaneous reaction, when she gave this example in response to an alternative illustration of largely the same point that I put to her showed to me that she had been thinking about similar problems, and probably more deeply than I had. In retrospect it seems very likely that it was connected with her advising of Irene's dissertation. So she may have been an important 'common source'.

AtM: Another possible common source that comes to mind, though, is David Lewis's account of unselective binding from the first half of the seventies. Did you see the connection at the time when FCS and DRT were developed between unselective binding and what you were after?

HK: There is no question but that I was aware of Lewis' *Adverbs of Quantification*.[30] I was present at its first presentation (at a conference on natural language semantics organized by Ed Keenan in Cambridge in 1974) and must have read the paper repeatedly between then and the time when I started thinking about incremental discourse interpretation. But here too there is an important difference in motivation. Lewis was concerned with the type of expression mentioned in the title of his paper, viz. quantificational adverbs. His paper is just about the semantics and logical forms of sentences that contain such adverbs, not about multi-sentence discourse.

[30] Lewis, David, 1975, Adverbs of quantification, in: Edward L. Keenan (ed.), ***Formal Semantics of Natural Language***, Cambridge University Press, Cambridge, 3–15.

It was only the work of Craige Roberts about 'modal subordination' in her 1987 dissertation and related work by Peter Sells from around the same time that showed how Lewis' observations can be extended to multi-sentence discourse. But that work makes quite explicit use of DRT.[31] The common ground between Lewis' *Adverbs of Quantification* and my *A Theory of Truth and Semantic Representation* is what the two papers have to say about conditionals in which the quantificational adverb, representing some kind of universal or generic quantifier, is tacit. For such sentences they propose what are in essence the same logical forms, and so are the motivations underlying them. Lewis treats the *if*-clauses of his adverbially quantified sentences as restrictors of his unselective quantifiers, which describe types of 'cases'. In the treatment of TTSR the same *if*-clauses are seen as specifying hypothetical states of affairs and the logical form for the sentence as a whole is obtained by interpreting the main clause in the context provided by this hypothetical state of affairs description. In Lewis' terms: the main clause is interpreted as making a statement about the 'cases' described by the *if*-clause. But the overall motivations were nevertheless quite different.

In TTSR, the emphasis was on the interpretation of donkey pronouns in discourse no less than in sentences. The aim was to show that the mechanism responsible for the behavior of donkey pronouns in donkey sentences was the same as that governing their interpretation in 'donkey discourses' (such as Partee's ball example). The only difference is that in donkey sentences the state of affairs description that serves as interpretation context for the pronoun is hypothetical. Overt quantificational adverbs were not dealt with in the first papers on DRT. And the phenomenon of unselective binding, which stands out as the most salient similarity between Lewis' treatment of adverbs of quantification and early DRT, emerges in the two theories in quite different ways. For Lewis unselective quantification is part of the nature of adverbial quantifiers; in DRT it is a kind of side effect, due to the general architecture of logical forms (the DRSs) and the particular way in which it analyzes indefinite noun phrases. In this respect, by the way, Heim's File Change Semantics is much like DRT; at least this is so for the account presented in Ch. 2 of her dissertation.[32]

---

[31] Roberts, Craige, 1990, ***Modal Subordination, Anaphora, and Distributivity***. Garland, New York.

Sells, Peter, 1984, ***Syntax and Semantics of Resumptive Pronouns***, Ph.D. dissertation, University of Massachusetts, Amherst, MA.

[32] Heim, Irene, 1982, ***The Semantics of Definite and Indefinite Noun Phrases***, Ph.D. dissertation, University of Massachusetts, Amherst, MA. Published with Garland, New York, 1988.

AtM: From what you said earlier about what led you to DRT, it would seem that your motivation and Irene's were quite different. The real issue for her was novelty/familiarity, as the feature that distinguishes definites from indefinites: the referents of definite descriptions are supposed to be known, the referents of indefinites are supposed to be new. That distinction is so very different from MG, and much of her dissertation is concerned with it.

HK: That is true. But it is important to keep in mind in what ways the familiarity-novelty distinction is different from Montague Grammar. It is not just that familiarity and novelty are notions that do not have a natural home in MG (not at any rate as it was understood and practiced at the time). But it was also important because of the way in which it repositioned the category of indefinite noun phrases within the grammar: definites and indefinites are now seen as forming one category of referential noun phrases, distinct from the other main category, the quantifying noun phrases. This repositioning is part of a quite new perspective on questions of reference, quantification and variable binding. It differs from the traditional view, which distinguishes between referential and quantifying phrases, but treats indefinites (and sometimes also all or some definite descriptions) as quantifying, and from Montague's perspective according to which all noun phrases, even proper names, are generalized quantifiers.

KvH: You were saying that your thinking about the Passé Simple and Imparfait in Stuttgart in 1978 set you on the course that led to DRT, that your talk in the Montague Colloquium of 1980 was a crucial next step, but that the decisive step was still missing and that this missing link came to you only during the months you spent a little later at the University of Texas. What was this missing link?

HK: In the form in which DRT was then formulated, it was the insight that the model-theoretic truth definition for the basic DRS-language—the formalism of which the DRSs are the formulas—should be given in terms of partial assignments that are extensions of other assignments. The point here is that, although truth definitions of formalisms like the predicate calculus are usually thought of as 'bottom-up' in that they define the semantic values of complex expressions from the semantic values of their constituents, those definitions are in an important sense, and for the very same reason, also 'top-down'. When you are interested in what the definition has to say about some particular complex expression, you typically start with a clause which states that the expression has a certain value in a given model M at a given index i and under a given variable assignment g and then use the

definition to reduce this statement to statements about smaller and smaller constituents, until you have reached the atomic constituents. In the course of such a reduction you move from the initial variable assignment *g* to different assignments; and when you use partial assignments (which is also possible when you give the truth definition for classical predicate logic, although for some reason this is not usually done), you find that with every reduction that involves a change in the assignments under consideration this change takes the form of *extending* the assignments involved thus far. DRT exploits this feature of growing assignments in a new way. Assignments now play the part of contexts of interpretation in a setting in which the complex expressions are DRT's logical forms, i.e. its DRSs. The assignments that enter into the evaluation of such an expression are made to play the role of contexts for the evaluation of constituents (i.e. sub-DRSs and DRS-conditions) of the given DRS. This means that many of the clauses of the recursive truth definition for DRSs and DRS-conditions have to refer to two assignments rather than one—the assignment that serves as context and the one with respect to which the constituent is said to be true, and which is an extension of the former. To get a truth definition of this form to do just the right thing must appear as something close to a triviality these days. But at the time it took some careful consideration; in any case it took me some time to see how actually to do this.

KvH: When you were at Stanford discussing DRT with the other people in the group, and beyond, how did they react to it?

HK: One of the aims of the group was—I hope I am speaking here also in the spirit of all or most of the other members, and am not just projecting my own self-centered perspective—to bring SS and DRT together, to see what was shared, what was different and how one might make a choice between them or achieve a synthesis of them.

This goal took a more concrete form in the course of that year; we were going to write a book together in which SS and DRT would each offer their own treatment of a certain fragment of English for which two members from the group were designated to provide a syntactic description. Unfortunately, this syntactic description—basically just a necessary preliminary for the semantic parts that would build on this foundation—never made it beyond the pre-final version and eventually that also meant the end of the project as a whole. Perhaps that was just as well, for as I came to see things, partly as the result of my effort to fulfill my side of the commitment, the project was based on what was probably a misconception of the relationship between SS and DRT and of what might be seen as their respective strengths and

weaknesses. What I came to perceive as a strength of SS is its emphasis of the fact that, when we think or speak about the way the world is, we typically focus on certain small parts of it, certain 'situations'. It is in relation to the particular situation or situations that we are talking about that the words we are using are to be interpreted and evaluated. SS endeavors to flesh this insight out by developing a detailed theory of the structure of situations and using this structure in analyzing the semantics of language. My own attempts at situation-semantic analyses of certain linguistic constructions left me with a sense that the range of options SS offers for analysis in natural language semantics is almost too rich for its own good. Often the range of possible choices for the theorist is so large that the framework gives no guidance as to how one should proceed—one is faced with a true *embarras du choix*. In this way the strength I mentioned easily turns into what a linguist might feel to be a weakness.

DRT is, both in its aims and in its form, a very different enterprise. It is more directly focused on the question how language represents content. What content is, and in particular, whether situations play any part in it, are less prominent questions than in SS. Central to its conception is the finiteness and partiality of the information provided by linguistic descriptions, below, at and above the sentence boundary (i.e. by sentence constituents, by complete sentences and by multi-sentence discourse). Of course, when the idea is put like this, there is nothing really new to it. DRT makes special use of this partiality; but it is compatible with the adoption of a conventional model-theoretic semantics, which acknowledges no ontological categories that cannot be found in HOIL. This more specifically linguistic perspective might have been combined with SS by formulating a situation-based model-theoretic semantics for some suitable DRS-language. I still think that such combination of SS and DRT might have been illuminating for our understanding about how the two are related and that such a combination might have been fruitful. But the two weren't put together then, and it hasn't been done since. And in any case, the usefulness of such a combination would perhaps be less now than it might have been at the time, since eventualities play a so much more prominent and accepted part in semantics.

AtM: SS was marketed as semantic realism—and Barwise at the time rejected any form of representationalism, as DRT is. You mentioned you had formed the opinion that SS offers far more options than linguistic analysis can use, and that this renders it unhelpful as a tool for linguistic analysis. That is in some ways reminiscent of what Chomsky and other generative syntacticians have seen as a desideratum for grammar frameworks. They

should be subject to the constraints built into the human 'language engine', which select for only a small number of the totality of logically possible languages; only the languages belonging to this small subset are humanly possible languages. Your complaint also reminds me of one that has been leveled at certain categorial grammars, which provide for often large numbers of alternative syntactic analyses of intuitively unambiguous sentences, such as e.g. *Mary kisses John*.

HK: There are a number of different points you are raising, too many probably to sort out properly in an answer that isn't going to be excessively long. What I was thinking of in particular just now when I said that SS isn't giving enough guidance to the natural language semanticist was work that Robin Cooper and I embarked on in the early nineties, which explored the different options in Situation Theory for the semantic definition of the logical constants of classical predicate logic: negation, disjunction, conditionals, conjunction and the universal and existential quantifiers. We started with negation, but we never got beyond it. The main reason was that the situation-theoretic framework left us with just too many options to explore without seeming to provide us with any natural handle on how to choose between them and select those that would be relevant to the analysis of negation in natural language (or, for that matter, for any other philosophical or logical applications).

The kind of spurious ambiguity problem, like that of getting multiple syntactic analyses for *Mary kisses John*, is a somewhat different issue. In order for this problem to arise for SS we would need explicit situation-theoretic descriptions of fragments of natural languages, of the sort that has been presented in work Cooper and by Peters and Gawron.[33] But as far as I know the syntactic descriptions proposed there do not lead to the structural ambiguities that are distinctive of Categorial Grammar.

A problem of overgeneration of structural descriptions, by the way, also seems to lurk off stage in Elbourne's proposal to deal with housemate- and bishop-sentences in his non-dynamic situation-based theory of donkey phenomena. What is it about the form of such English sentences that

---

[33] Barwise, Jon and Robin Cooper, 1993, Extended Kamp Notation: A graphical notation for situation theory, in: Peter Açzel et al. (eds.), ***Situation Theory and its Applications***, CSLI Publications, Stanford, CA, 29–53.

Cooper, Robin, 1993, Generalized quantifiers and resource situations, in: Peter Açzel et al. (eds.), ***Situation Theory and its Applications***, CSLI Publications, Stanford, CA, 191–211.

Gawron, Marc, and Stanley Peters, 1990, ***Quantification and Anaphora in Situation Semantics***, CSLI Publications, Stanford, CA.

determines the particular logical forms Elbourne proposes? Or does the syntactic form of those sentences underdetermine those logical forms, allowing for a great many, with at least one among them that fits Elbourne's proposals?

AtM: You have reminded us of your view of DRT as a more cognitively real way of doing semantics. But applications of DRT are normally formulated as involving algorithms that convert given syntactic structures incrementally into semantic discourse representations. What about the psychological reality of those syntactic structures? What remains of the cognitive claims of DRT, if there is no psychological reality to the syntactic structures which existing versions of the theory take as their point of departure?

HK: That is a serious problem and it has also been a rather frustrating one. All applications of DRT to natural language—and that is of course what DRT is primarily about—have been suboptimal or problematic for this reason. In the most ambitious single attempt of this kind that I have thus far been a party to—the book ***From Discourse to Logic*** (FDL) with Uwe Reyle—we tried to finesse the problem by assuming a syntax of which we said emphatically that it was just a stand-in for more serious syntactic theories which we hoped others would at some point put in its place.[34] We hoped that others would come with the syntactic theory of their choice, motivated by serious syntactic and/or psychological considerations, and define on the basis of it a DRS construction algorithm employing essentially the same construction principles that we were presenting in the book. Of course, from the perspective of actual sentence processing such a replacement would still be a serious idealization in that it describes syntactic parsing and semantic processing as sequential. Actual parsing by human interpreters is clearly a process in which syntactic and semantic structures are identified in tandem and 'on line', as the successive parts of each sentence become available. But that is the kind of idealization that the existing formulations of DRT share with pretty much all other work within theoretical linguistics that concerns the syntax-semantics. So we were well aware, when we wrote the book, that the way we proceeded was far from optimal. But we thought that it would enable us to show enough of the essential properties of the algorithms in question to make it possible for people with more specific syntactic

[34] Kamp, Hans and Uwe Reyle, 1993, ***From Discourse to Logic. Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory***. Kluwer, Dordrecht.

commitments or predilections to see how to adapt our proposals to the syntax of their choice. Later on I have often worried whether that had been the right strategy. On the one hand, we did get a certain amount of flack about the naïveté of our syntax; on those detractors our explicit disclaimer—that this was just a make-do syntax and that it did not in any way represent a deeper syntactic commitment on our part—was apparently wasted. On the other hand, the adaptation of our construction algorithm to different syntactic frameworks that we were hoping for never really took off.

But I'd like to think that times have changed and that a new perspective on syntax is gradually emerging, which makes it possible to distinguish between (a) what syntactic structures should be assigned to particular grammatical expressions, and (b) what general syntactic principles should be invoked to explain why it is that those syntactic structures should be assigned to the expressions of a given language and not others. Not that all controversies about the syntactic structure of particular expressions have been resolved. There are many disputes remaining, both within individual syntax theories and between them. But especially the extensive work on syntactic annotation, beginning with the Penn Tree Bank, has led to a considerable degree of convergence on the substance of syntactic structure, even where the debate over underlying principles and the overall architecture of a syntactic theory predicting these structures continues.

Restating DRS construction algorithms as operating on syntactic structures used in large annotation enterprises is now becoming an option that did not exist in this form in the first half of the nineties. On the other hand, a lot has since been learned about the internal structure of words. It is now widely recognized that many words are complex syntactic and semantic structures, built from 'roots' and 'functional elements'. This development also poses a new challenge for DRT: construction of semantic representations should go, compositionally, all the way from the building blocks of the words up to the level of the full sentence (and beyond to that of multi-sentence discourse). Antje Roßdeutscher and I are currently working on such 'root-based' DRS construction methods.

AtM: Returning to what you said about the year at the Stanford Behavioral Sciences Center and the interaction with Situation Semantics: In Situation Semantics and Situation Theory there was an increasing focus on information and information flow. How would you now characterize information structure? What should it do for you or what primitives should it have?

HK: 'Information' is one of those words that are under lots of stress these days. On the one hand, everybody and his grandmother are talking about

information and more often than not, one has the impression that few expressions could be less informative. On the other hand, there is also a formally precise and well-understood cluster of information-related notions in the tradition of Shannon.[35] The theory of information in this tradition is of immense importance and the foundation for most of what makes our electronically shaped world tick the way it does. But it is still very difficult to establish convincing and practically useful links with information as it is conveyed by language and captured in thought. In fact, there is some well-known information-related work on natural language that is closer to mathematical information theory. This is the work by Paul Zipf, and further research in the tradition he established. Zipf's Law is a good example of the way in which the information density of individual words of a natural language is distributed over its lexicon.[36] But these results are very far removed from what we normally think of when we speak of the information that one obtains from a verbal message.

Recently, however, these two worlds of information have finally been converging—dangerously so, I am almost tempted to say. In the branch of Computational Linguistics that has come to be known under the name of Distributive Semantics the semantics of a word is captured in terms of a vector that encodes its co-occurrence frequencies with other words in some corpus. This information is useful in many meaning-related computational tasks. But the really hard question is how this kind of information about word meaning can be related to the ways in which the semantic contributions of words to the truth conditions of sentences, or to logical forms such as DRSs. I do not know how far anyone has got with finding answers to this question. But the pressure to build a bridge between these two approaches to meaning is high and growing. So it may not be unreasonable to expect that, if there are any ways of building such a bridge at all, they will be built before long.

But all this only pertains to the notion of information itself and not to the notion of information flow. Here too, the two traditions I have been hinting at are very different. Information theory in the Shannon tradition sees information flow as involving coding, code transmission and decoding. Many of its results concern coding complexity and its optimalizations, transmission noise and channel capacity. The problems about information in linguistics

---

[35] Shannon, Claude E. and Warren Weaver, 1949, ***The Mathematical Theory of Communication***, University of Illinois Press, Urbana, IL.

[36] Zipf, George K., 1949. ***Human Behavior and the Principle of Least Effort***. Addison-Wesley, Cambridge, MA.

are different because, first, the code is given to begin with—the code is just the natural language as available to its users, with its rules of phonology, morphology, syntax, semantics and use. So the problem of natural language semantics is to a large extent that of finding perspicuous ways of representing or identifying the information carried by different bits of this code. There is, of course, also the question how the information carried by bits of natural language code succeeds in getting transmitted from one agent to another. Questions of noise are already prominent in phonetics and phonology, but it is obvious that the redundancies in linguistic messages make it possible for them to get across as smoothly as they do most of the time. Even when strictly phonetic-phonological recognition is impaired, there is redundancy at the levels of syntax and semantics. Therefore the problem of speech recognition is really a problem of language recognition that involves many different levels of linguistic representation, including also levels of semantic representation which enable the decoder to assess decoding hypotheses for their consistency and coherence with the current context. Translated into the terminology I am using right now, information somehow has to get across—coded, transferred, decoded—at a number of distinct but interacting levels.

Related to this is an issue of general importance for the methodology of natural language semantics. Formal semantics has been operating on an assumption that can also be traced back to some of the deep convictions of Montague, that fits in with what many semanticists with a background in formal logic had come to see as the true and only possible way forward. This is the conviction that languages can be studied as self-contained systems, with a syntax and a semantics, which make certain forms with their associated meanings available for those who use them. On such a view the study of natural language should take the form of (a) studying the syntax and semantics of a language as a self-contained system, in abstraction from what uses are or can be made of it, and (b), building on the results of this part of the investigation, investigations of how speakers use this independently existing system. This is the classical semiotic hierarchy most often connected with the name of Morris, with the study of syntax presupposed by the study of semantics and the study of pragmatics (i.e. of the use of language) presupposing the study of both syntax and semantics. I have become increasingly convinced over the last few years that this methodology has its limits and that there is an urgent need for a new approach to natural language semantics, where the use of an expression by the speaker and its interpretation by the recipient are distinct, but coupled processes. Superficially this is more like the picture of information transmission in the

Shannon-Weaver tradition, though the differences remain more important than the similarities. Such an approach to linguistic meaning is forced to stick its neck out in ways that go well beyond the commitments involved in the currently received way of doing formal semantics, in which semantic values or logical forms are assigned to expressions in abstraction from their role in the coding and decoding of verbal messages. In particular, we must not only be prepared to make assumptions about the form of the *results* of interpretation, but also about the forms in which thoughts are available to the speaker at the point when she encodes them in words. DRT, as a theory that tells us something about the form of the semantic representations that interpreters extract from the verbal inputs they receive, offers some specific proposals about the final results of the communication process, those established at the receiver's end. But we also need concrete proposals for the semantic representations of thoughts that stand at the beginning of the communication chain—those ones the speaker turns into the words that she thinks best serve her communicative purpose. One way to deal with these different aspects of verbal communication is to assume that the thought representations that serve as input to language production are the same as those that result from interpretation. That is no doubt a hazardous assumption, and one that won't be true in general without various qualifications. But for the analysis of a range of problems, especially those having to do with reference—of specific indefinites, definite descriptions (in connection with the referential-attributive distinction), proper names and the various uses of demonstratives—distinguishing between the production and the interpretation of such expressions and then looking at how they are coordinated in actual communication appear to me now as essential to further progress.

There is one aspect to information and information flow in verbal communication that has acquired a status of its own. This is what in current linguistic semantics and pragmatics is usually referred to as 'information structure'. 'Information structure' too is a term with almost as many meanings as users, and its use varies in particular with regard to what it is taken to include. What I take to be the common core to all uses in the current formal semantics and pragmatics literature is the division of information into focus and background. Here the basic agenda was set by Mats Rooth's 1985 dissertation.[37] Our understanding of the semantics and pragmatics of focus,

[37] Rooth, Mats, 1985, ***Association with Focus***. Ph.D. dissertation, University of Massachusetts, Amherst, MA.

and of its prosodic, syntactic and morphological realizations in different human languages, have made steady progress over the past 25 years, though there still is much that remains to be done. Usually information structure is meant to include more than just the focus-background distinction.

Another notion, or set of notions, that have to be distinguished from the focus-background division, is that between topic and comment; but there are different notions of topic, and with that, corresponding notions of topic-comment division. Topic notions differ from each other along several dimensions. One of these dimensions is the size of the unit that a topic is the topic of. In particular, there is a notion of sentence topic that interacts with focus in important ways, which were first mapped out in the work of Daniel Büring.[38] His work has taught us that systematic investigations of the semantics and pragmatics of focus soon reach a point where the interaction with the topic-comment distinction has to be taken into account as well. But apart from sentence topics there are also more global notions of topic—topics in the sense of 'topics of discussion', where there is a close connection with topics as issues or questions under discussion. The study of this notion of topic is part of the study of the structuring of various types of discourse and texts and has quite different ramifications, which reach into different areas of pragmatics.

Finally, the theory of linguistic information structure is sometimes understood as including the distinction between presupposed and asserted information. Presupposition has played a central part in the development of DRT, largely through the crucial contribution made by Rob van der Sandt and Bart Geurts around 1990.[39] Van der Sandt argued that anaphora (including donkey anaphora) and presupposition are just two sides of the same coin and suggested that DRT was a natural framework to deal in a uniform way with both. But with this proposal comes a fundamental revision of the overall architecture of DRT. DRSs are now constructed in two stages: (i) a first stage, in which a 'preliminary' representation is constructed for the current sentence; this preliminary representation contains explicit representations of all presuppositions triggered by presuppositional elements in

---

[38] Büring, Daniel, 2007, Semantics, intonation and information structure, in: Gillian Ramchand and Charles Reiss (eds.), ***The Oxford Handbook of Linguistic Interfaces***, Oxford University Press, Oxford, 445–473.

[39] van der Sandt, Rob, 1992, Presupposition projection as anaphora resolution, ***Journal of Semantics***, 9, 333–377.

Geurts, Bart, 1996, Local satisfaction guaranteed: A presupposition theory and its problems. ***Linguistics and Philosophy***, 19, 211–257.

the sentence, including the 'referential' presuppositions triggered by pronouns and other definite noun phrases; and (ii) a second stage, in which the presuppositions of the preliminary representations are resolved, and the representation that results from the resolution is merged with the context representation. Furthermore, when anaphora is treated as a form of presupposition the original motivation in early DRT for constructing DRSs top down—the online processing of anaphoric pronouns—disappears. Largely because of this, the top-down method has in practice been almost universally replaced by bottom-up algorithms. This has been true in particular for me and for those with whom I have cooperated on DRT-related projects during the past twenty-odd years. Within the setting of DRT, the presuppositional/non-presuppositional opposition therefore has a different status when compared with that of other aspects of information structure: it is an integral part of the architecture of DRT as it is now most commonly used. To represent other information structural notions, such as the focus-background opposition or the various notions of topic, additional representational provisions have to be introduced into the DRT formalism. Some of the needed extensions of the formalism are already in place, but there is quite a bit that remains to be done.

AtM: The concern in SS with information flow led eventually to a separate development of Situation Theory, viz. Channel Theory. I myself don't know whether Channel Theory has proven useful in computer science or mathematical logic and model-theory. But I have the impression that it has not had much of an impact on work in linguistics. Linguists like to have handy toolkits that they can apply to new problems. But SS never gave us much of such a practical toolkit, honed for application to linguistic problems. This also relates to more general questions of reception. As far as visibility was concerned, there seems to have been quite a difference between SS and DRT in the early days. Whereas SS was widely hailed, especially within philosophical circles, DRT had less visibility at that point. In fact, there seems to have been considerable resistance to it. Do you have an explanation for this difference?

HK: I have kept up neither with the more recent developments of Channel Theory nor with its reception within the linguistic community. So this is something I do not dare comment on right now. But as far as the early reception of SS and DRT is concerned, I think there are a number of things that I can say. First, there is a difference in the scientific communities that the two approaches address and/or to which they have something useful to offer, or to which they would have liked to have something to offer. The

fundamental issues that SS raised have a much broader appeal. At least they were seen as such at the time, and, I think, rightly. SS was concerned with questions about the central ontological categories of thought and language and thus with the content and nature of the propositional attitudes, and it presented these issues in a form that had immediate appeal to the philosophical community. And there were a number of points SS was making that seemed important and right and that could be appreciated without detailed technical knowledge about earlier work such as Montague Grammar.

In my own recollection there was in particular the paper I mentioned already, which Barwise had written before he joined the Stanford faculty and which I thought was marvelous both in its concrete accomplishment and in the way it was pointing ahead. This was his *Scenes and Other Situations*, in which he shows that situations, pieces of the actual world, must be invoked as the semantic values of the infinitival complements of naked-infinitive perception verb constructions, as in *John saw Mary kiss Fred*. This paper made a strong case for a radical revamping of semantics in general, in which situations play a central role, and a double one: on the one hand, as semantic values of such expressions as the infinitival complements of perception verbs and, on the other hand, as indices, taking the place of possible worlds. This was widely seen as nothing short of a revolution, welcomed among other things, because it held the promise of doing away with a notion that many philosophers regarded with much suspicion, viz. possible worlds, and replace it with entities, viz. situations, that for the most part were much more narrowly circumscribed, and therefore much more tangible and tractable. Thus SS promised a healthy realism to take the place of what many saw as the murky metaphysics of possible worlds. I think it is right to say that in the end quite a few of the initial supporters of SS came to see it as failing to fulfill this promise. There are a number of reasons for this disappointment that come to mind. The first is that one of SS's original cornerstones, viz. Barwise's analysis of naked infinitive perception sentences, came to be seen as less uncontroversial than it had appeared initially. This was because of the alternative proposal first made by Jim Higginbotham, according to which the contributions of infinitival complements to the truth conditions of naked-infinitive perception sentences are the *events* that these complements describe.[40] As events became increasingly important in semantics and the need to acknowledge them more widespread, this alternative has gained in popularity, also because it provides a natural

[40] Higginbotham, James, 1983, The logic of perceptual reports: An extensional alternative to Situation Semantics, ***Journal of Philosophy***, 80, 100–127.

explanation of the fact that the infinitival complements of perception sentences must be descriptions of events and cannot be descriptions of states. The second factor was that the use of situations to replace possible worlds as indices of evaluation turned out to be something less than the unequivocal step forward that it had looked like at first. One problem here is a matter of detail. What is needed is a formally explicit theory of situations. Barwise and Perry and some of their collaborators saw this need clearly and took great pains to deal with it. Situation Theory was soon distinguished from Situation Semantics and its first task was to provide a formal axiomatic basis of what situations are. More often than not, when things get more formal and technical, popularity wanes. That is an old story, which has made many scientific theories disappear from general view just because they became too sophisticated to be understood and appreciated by anyone unwilling to take the trouble and time to work through the formal details.

But there was also another problem with the use of situations as indices. This problem has to do with the fact that the space of situations is much richer and more complex than the space of possible worlds. On the one hand, this was something that the situation semanticists wanted—for one thing, because it promised a finer individuation of propositions and therefore a better account of propositional attitudes. On the other hand, too much richness can be a mixed blessing. To repeat what I said earlier, too much richness in a framework for doing semantics means that the formalism may fail to give you any useful guidance to what the linguistically important semantic constructs are and what form the semantic analysis of linguistic forms should take.

AtM: How does all this relate to your infamous, but never published paper *A Scenic Tour Through the Land of Naked Infinitives*? Was that an attempt to get closer to SS?

HK: Yes and no. It was, on the one hand, an attempt to become clear about what was involved in a formal situation-based semantics for some tractable formal language. The paper develops, for a modest extension of what is syntactically speaking just first-order predicate logic, a model-theoretic account of what it is for a formula of the formalism to be supported by a situation. This gives one, for each sentence *j* and each situation-based model *M*, as the *proposition expressed by j in M*, the set of those situations of *M* that support *j*. Since the logical consequence relation of Situation Semantics is weaker than that of classical logic—this was a central point for Barwise and Perry and it is confirmed by the formalization that '*Scenic Tour*' presents—the notion of propositional identity produced by such a semantics is finer

than that generated by classical logic, i.e. it is easier for two sentences to express distinct propositions. Barwise and Perry had pointed to this fact as yet another point in favor of SS as compared with the possible worlds semantics of Montague and others: because propositions are more finely individuated, SS might offer a way out of the logical omniscience problem that besets intensional analyses of the propositional attitudes. At this point my paper turns into an argument against SS. The logic generated by the formalization of the support relation proves to be that of the strong Kleene semantics, which is, though weaker than classical logic, still much too strong to give a satisfactory solution to the logical omniscience problem. For instance, take the case of belief sentences with *that*-complements. According to the standard intensional treatment of such sentences the contribution that the *that*-complement to the truth conditions of the sentence as a whole is the classical proposition it (the *that*-complement) expresses. That gives rise to the omniscience problem because there are many cases of sentences S_1 and S_2 that are logically equivalent and hence express the same classical proposition, but whose equivalence is hard to see. It is possible for an agent *a* to sincerely assent to the one, but not to the other. In such cases the classical treatment entails that either both *a believes that* S_1 and *a believes that* S_2 are true, or else neither of them. That seems unmotivated and the problem is obviously that classical propositions 'collapse' too many distinctions. Since SS offers a finer individuation than classical semantics and logic, it may at one point have seemed natural to expect that this finer individuation could also block the unwanted collapse of the truth conditions of belief sentences. But, alas, it succeeds in this only to a quite limited degree: there remain many pairs of sentences S_1 and S_2 that are not only classically, but also situation-semantically equivalent, but whose equivalence is hard to see or establish. For such pairs the omniscience problem remains.

AtM: We have touched in a few places on connections between your work and cognitive science. On the other hand, you also mentioned computational linguistics more than once. Of course, for the last twenty years you were part of a department of computational linguistics, the IMS (Institut für Maschinelle Sprachverarbeitung) of the University of Stuttgart. The relation between cognitive science and computational linguistics always struck me as quite complex, and also as a topic of highly diverse opinions. Is there anything you can say about this relation?

HK: I am surely not the best person to ask about this. But here are a few things that come to mind. First, it is important to realize that computational linguistics is and has been many things to many people. As is so often

the case with a new discipline that gets its inspiration from very different neighbors and tries to cope with their respective expectations, it has gone through different fashions, some of which have become long-lasting methods or perspectives, while others have more or less disappeared from the scene, perhaps lingering in obscurity and waiting for a revival when their time will have come again. One of the disputes that was high on the agenda of many computational linguists in the 70's, 80's and 90's and that also had its run at the IMS, was precisely this: What was, or could be, or should be, the relation between computational linguistics and cognitive science?

That dispute took more than one form: Should computational linguistics —as a science, rather than an engineering enterprise—be the study of how human beings carry out the computational tasks involved in language processing, or should it identify, from a strictly functional perspective, what computations are needed to carry out certain given language-related tasks and then design the most efficient algorithms to implement those tasks? But among those who defended the second position, there was a further point of debate: Would the most efficient algorithms be *de facto* the very same that are also used by human speakers, and would it therefore be useful to study human language-related cognition more closely, as probably the best way to discover what are the most efficient algorithms from a purely functional point of view? Or is the starting position of the computer, with all its resources—online word lists and thesauri, vast amounts of encyclopedic knowledge and so forth—so different from that of us humans that its best language processing algorithms would be unlikely to have much in common with the ones employed by us? In that case the study of human cognition would be useless even as a means of discovering useful clues of what the best language processing algorithms for computers might be like. These disputes never got resolved, even if they helped to clarify where everyone stood. But perhaps the most noteworthy aspect of all this was that at some point the issue simply went dormant. The reason for that was that, during the time when the debate was taken seriously and thought to be important, there had been a broadly shared understanding of computational linguistics, presupposed by all sides participating in it, that computational linguistics was about the computation of linguistic representations of the kind proposed and discussed in computational linguistics. With the rise of statistics in computational linguistics—which is now as prominent at the IMS as it is in all other major centers of computational linguistics—the debate has become marginalized, because the rough consensus about what representations are to be computed exists no longer. The debate between 'statistical CL' and 'logic-based CL' is an even more fundamental one: What counts

as a 'representation' of linguistically relevant information? Or is even a liberal interpretation of the term 'representation' misleading in this context? Is the information about linguistic expressions that is identified and used by statistical CL algorithms, and perhaps even by us, of a fundamentally different character than theoretical linguistics and 'logic-based CL' have long been assuming? In other words, the fundamental debate has become even more fundamental than it was. It is now about the nature of information in general, in a setting where information in the form of the representations traditionally assumed in 'logic-based CL' is only one among several options.

AtM: Has the IMS ever been involved in natural language systems for artificial agents, I mean robots?

HK: That is an interesting and most relevant question. Actually at the IMS not much of that kind of work is going on. There have been some speech recognition projects, dealing with the problem of filtering out background noise, crucial to applications of automated language processing in speech-based communication between drivers and their vehicles. Actually, we do not usually think of cars as robots. But many of the issues that make robotics so important for theoretical linguistics are already manifest in the exchanges between drivers and their vehicles. What distinguishes the problems that arise in connection with verbal communication with robots from problems in other areas of computational linguistics is that robots have other information channels, besides the ones used for the transmission of language. The point of being able to talk to your car as you are driving is, on the one hand, that the car has non-linguistic means of acquiring information—the car may be equipped with sensors that take in information about the road, and the conditions of travel, including what is delivered by its navigation system that at a minimum can identify its current location. The car's language facility should understand verbal requests for such information and answer on the basis of the information it has gathered, or look for that information first, in case it finds that it doesn't have it. On the other hand, the car will also have some of the features of an agent. It can perform certain actions, such as switch off the air conditioner or open the sunroof, and it should be possible to tell it in words that one wants it to do such things. Finally, in really sophisticated systems the car should be able to decide to initiate a verbal exchange, perhaps just by telling the driver something that it thinks he ought to know. But such an initiating utterance by the car could start an ensuing dialogue and the car should be prepared for the reactions such initiations are likely to provoke, and ideally also be able to deal with less expected reactions.

What such a robot must be able to do with language as an integral part of its cognitive system confronts us with issues that are very different from those that have so far been addressed within formal semantics. Let me give an example from a project, led by Alois Knoll from the Technical University in Munich, the task of which was to develop a natural language facility for a robot, who is to put together a wooden toy plane from various building pieces, including slats, screws and bolts. Such a robot must be able to connect verbal predications, as we find them in, for instance, a sentence such as *A slat with three holes is lying behind the green bolt*, with information that the robot takes in through its built-in camera. That is, image recognition, in the sense of converting what is registered by optical sensors into a verbal description of it, is essential to the language component of such a system. The same is true of other modules that process sensory information and that must make this information accessible to verbal realization. The converse of this, being able to picture a scene on the basis of verbal input involving such information, or updating such a picture on the basis of new verbal information of this sort, is equally important. This problem, of the interface between visual and verbal representations, is one that, if I am right, people have been aware of for a very long time. But it wasn't until the eighties that the first formally and computationally precise proposals were made about how the two forms of interpretation might differ and how they could nevertheless interact. I am thinking here of the work by Barwise and John Etchemendy, which they presented as an interactive program that could be used as a kind of logic tutor. Their system *Tarski's World* is one in which reasoning can make use of both geometrical and propositional representations, where propositional representations are basically just logical formulae built from elementary clauses of the form $red(a)$, $behind(s, b)$ and so forth.[41] A lexicon that enables the language processing system to link its atomic predicates to information accessible via other channels is just one of the many demands placed on the language modules of robots with sensors. It is paradigmatic for the kinds of problems that have to be solved, if we want to take the notion of equipping robots with language seriously.

AtM: Some people have stated the ambition to build a robot that people can interact with without recognizing that it is a robot—a robot that can pass the Turing Test. What can be said about that ambition in the light of the difficulties of which you have just spoken?

---

[41] Barwise, Jon and John Etchemendy, ***Tarski's World***. http://ggweb.stanford.edu/tarskisworld

HK: It is important to keep in mind that Turing's test is supposed to apply to computers, and not to robots: The person who is to decide whether he is interacting with a robot or with another person is communicating with his interlocutor only by typing in messages and seeing messages displayed on his console. There is no other interaction between the two—nothing remotely like when, say, a human and a robot are in the process of jointly performing a complex task and talking to each other in order to be able to perform the task better. Tricking a human into believing that he is dealing with a human interlocutor is at least a possibility. There have been decades of debate over how hard or easy it is for a computer to fool the one communicating with it via a keyboard and console and over what can be concluded, when the computer succeeds. But the point is that the only way in which the human partner and the computer interact in the classical Turing Test are the messages that the partners send to each other; they do not interact in any other way.

To conduct a meaningful conversation about anything in natural language, in which the partner shows the capacity of reacting adequately to things that we put to him and that could not have been predicted at that point of the conversation, so that the reactions cannot be simply canned, that is already a vastly more complicated task than Alan Turing seems to have thought. It is vastly more difficult in one part for the reasons mentioned in connection with robots. Producing and interpreting speech as relating to information that one has from some other source is a hard problem, even when that source is not sensory. Even building a natural language dialogue system that can sustain conversations about some topic, about which all the relevant information can be encoded without too much effort in digital form, such as chess, is much more than we can chew off at present, even now, close to sixty years after Turing conceived his test. As far as the applicability of Turing's test to robots is concerned, here we have the obvious problems that—so far at least—robots look so different from human beings: You know that you are dealing with a robot before the poor thing has had a chance to say or do anything. Trying to build a robot that can fool the judge to the point of there being nothing that distinguishes his partner in action and conversation from a human seems as pointless, as it feels creepy. For the kinds of robots we are likely to interact with in the future it would seem that the closest one can get to a positive Turing Test result is that the judge develops a personal attitude to it, treating the robot as another person and perhaps being unable to do otherwise. But that is a long shot from being convinced that one is dealing with a human person.

AtM: Information from linguistic input is used in reasoning by combining it with information from other sources. As you were implying, a robot must be able to reason with the information it gets from what you are saying to it in combination with information it has or can acquire through other channels. But isn't this ability to reason with information obtained from linguistic input something we expect from automated language processing systems generally?

HK: When your system uses formulas of predicate logic as logical forms for natural language sentences and discourses, of course it can use any kind of theorem prover for predicate logic: you convert your inputs into predicate logic formulas, let the theorem prover carry out its deduction and reconvert the conclusion into natural language. This is a very old recipe, as already in the eighties there were projects that tried to combine this idea with DRT in that natural language would be turned first into DRSs, these would then be translated into formulas of predicate logic, the theorem prover of one's choice then let loose on those formulas and the conclusion turned back into natural language with or without the use of DRT. A propos, you can also develop deduction systems that operate directly on DRSs. Uwe Reyle and I did this in the early nineties for the DRS language of Chapters 1 and 2 of FDTL. But, of course, since that DRS language is equivalent to first-order predicate logic, and there are very simple translation algorithms for turning DRSs of this language into first-order formulas and conversely, formulating such a DRS-based deduction algorithm can't be hailed as a particularly telling result.

But how useful can such deduction modules be, when the task is to provide a robot with reasoning facilities that are able to interact in the right way with its language processing modules? Much of the reasoning on which we humans depend is not deductively valid. It is non-monotonic in one way or another and often hard to distinguish from what might be described as our learning new information, expressed in the conclusion that we adopt. For the robots we have been talking about, the matter would be no different. One of the important and exciting lessons of Learning Theory that seems pertinent here is that it may well be that there is no principled distinction between learning and reasoning. Within the general setting of information acquisition and use, instances of traditional deductively valid reasoning reappear as marginal and in a sense degenerate special cases of something much more general. This is a development about which I know far too little to speak about it with confidence. If the perspective on reasoning it suggests is right, the importance of deduction systems for AI and Cognitive Science

is probably much smaller than some logicians, myself included, have long liked to believe.

There is also another point that should be mentioned in this context. The DRS languages we need for natural languages transcend the limits of first-order logic. The need for such higher-order DRS languages arises in particular, when plurals are taken into account as well as singulars, as argued in detail in Chapter 4 of FDTL and more recently in our DRT survey in the 2nd edition of the ***Handbook of Logic and Language***.[42] Since complete proof systems are impossible for second- and higher-order formalisms, the best we can hope for from a proof system for such a formalism is that it cover all that human reasoning will ever use. But how will we, as theoreticians, know when enough is enough? The classical picture of the role of deductive logic in information processing is hence being eroded from two sides. On the one hand, we cannot hope for systems capable of proving all that is deductively valid and, on the other, the reasoning systems that robots, like we, really need must be in a position to draw inferences that are not deductively valid, hence transcending the bounds of purely deductive logic.

AtM: DRT as I know it shares with other systems of Dynamic Semantics that it is strictly incremental: as discourse processing goes on, DRSs only grow and nothing ever gets erased. That seems unrealistic in general and especially so when we think of dialogue. When people talk to each other, they often argue, they make conflicting statements, and, hopefully, eventually one will persuade the other and the other will withdraw the claim he had made. Is there a way to model such conversational processes in DRT? And how much does that affect the central assumptions of the theory?

HK: There is a whole range of interesting questions here that this touches on and that haven't yet been discussed. To start with, yes, of course updating isn't always monotonic in dialogue, and the same is true, by the way, if less prominently, in spoken monologues and written texts. But it is important to distinguish between different ways in which this can happen. Probably the most common type of case, and the one you seem to be thinking of, is that of argumentative dialogues, in which the participants argue over the truth or tenability of a certain claim or position. The aim of this kind of verbal exchange is that the discussants come to an agreement on the point

[42] van Eijck, Jan and Hans Kamp, 2010, Discourse Representation in Context, in: Johan van Benthem and Alice ter Meulen (eds.), ***Handbook of Logic and Language***, 2nd rev. edn., Elsevier, Amsterdam, 181–252.

at issue, either because one convinces the other of his position or because the debate ends inconclusively, with each of the participants sticking to his position, or because a kind of Hegelian compromise or 'synthesis' is reached, to the effect that both are aware that what they were arguing over meant something different from what they thought, or that it really didn't mean anything at all, when you come down to it. Let us focus on the case where either participant A convinces participant B of his position, or vice versa, and assume—quite implausibly, but we will come back to this—that the dialogue consists entirely of assertions. How do we describe such a dialogue? A kind of minimal proposal is to restrict attention to that information which is publicly accessible from the overt execution of the dialogue—that is, from the utterances the participants actually perform. We can think of such information as appearing on a kind of Lewisian scoreboard, consisting of three parts: (i) a component containing the statements that have been made by A, (ii) a component containing the statements that have been made by B and (iii) a component containing the publicly accessible common ground, as it gets established in the course of the conversation. The task then is to describe how a new utterance by either A or B modifies the information on the scoreboard. The new scoreboard will have a representation of the last speaker's utterance added to his scoreboard component and, depending on how the other speaker reacts to the utterance, possibly also to the common ground component. One of the things a speaker can do in this dialogue model is to retract a claim previously made. For the sake of simplicity, let's assume that such retractions take the form of the speaker asserting the negation of the claim *S* that is withdrawn. The effect of this on the scoreboard will then be that the representation of *S* is removed from the speaker's component on the scoreboard and replaced by a representation of neg-*S*. And in case a representation of neg-*S* is also part of the other speaker's component, neg-*S* will also be added to the common ground component. Retractions of this kind need not involve removing material from the common ground component. But that kind of retraction is possible as well. For instance, a participant may retract a claim he made earlier on, that was accepted by his interlocutor at that point, so it was added to the common ground. The speaker's present disclaimer will have the effect that the earlier claim is now removed. Besides these two types of non-monotonic scoreboard evolution, there may be yet others that could also be described in this simple dialogue model. But the more important point is that the model oversimplifies the reality of actual dialogues in a number of ways. First, it is rare for conversations to consist exclusively of assertions. In fact, such dialogues hardly ever occur in real life. Conversations are almost always

made up of utterances of different speech act types and a proper record what the participants in a dialogue say to each other should record the speech act type of each utterance as well as its propositional content. This already makes the components of the score board more complex; but in addition, in a coherent dialogue there will be certain relations between the utterances of A and B, such as, for example, when A asks a question and B reacts to that by giving an answer to it. These relations are as a rule also part of the overt, accessible information pertaining to the discourse and should be recorded too. This entails that the order in which the participants make their respective contributions will have to be recorded as well.

A more ambitious project of dialogue description is one which includes in the description of the successive stages of a dialogue not only a record of the publicly accessible information, but also partial descriptions of the attitudinal states of the participants. These richer descriptions provide a basis for an analysis of how dialogue partners can draw conclusions about each other's beliefs, desires and intentions on the basis of what the other is saying. Much of what is involved in drawing such conclusions is or involves Gricean reasoning with implicatures, but there is more. For instance, the reasoning involved in the recognition of indirect speech acts or the interpretation of novel metaphors. Such inferences are for the most part defeasible, which may give rise to further cases of non-monotonicity. I mention this just to give some indication of how much is arguably involved in describing what goes on in the course of a conversation and what a description of it must include if it is to enable us to capture the Gricean forms of pragmatic reasoning that are widely seen as a central part of linguistic interpretation and *sine qua non* for the theory of meaning. In Stuttgart we have been working for many years on extending DRT so that it can supply the descriptions of the public and non-public components of dialogue stage descriptions I indicated above. Since the nineties we have been making use of an extension that permits detailed descriptions of complex mental states, consisting of propositional attitudes and entity representations. An extension in which one can represent the speech act types of utterances together with their content has come a long way too. This work has not yet been published in any form. But these are only some of the formal tools needed for the rich descriptions of dialogue and conversation stages we have just been talking about. A general theory of conversation—or, for a start, a theory of dialogue—which articulates how utterances determine the transitions between successive stages that result from them is yet another matter. If such a theory is to be formulated in a DRT-based framework, then the notational extensions I mentioned are indispensable. But that is only the beginning. Much more will be needed

in addition and by no means all of that is in place. You may feel that I have strayed quite far from the question you asked. And you wouldn't be wrong. But my reasons for saying all this is to convey my conviction that, if we want to take the kind of non-monotonicity you are talking about seriously, all these different aspects of the activity of conducting a conversation will have to be brought into the picture. At least in the context of natural language semantics non-monotonicity is, I think, always a matter of belief revision (or, more generally, a matter of changing attitude). That is why the problem of developing a satisfactory theory of conversation that can deal with its non-monotonic aspects is so very hard. In fact, it isn't just a problem that is hard to solve. The first hurdle is to find the right way of stating it.

AtM: These are all extensions of DRT that could be carried out by looking only at one natural language. But what about linguistic variability in semantics?

HK: So far the explicit use of DRT in cross-linguistic semantics has been quite limited. But there appears to be a growing number of linguists among those working on semantics in a cross-linguistic setting, who are becoming convinced that the dynamic phenomena DRT was designed to make sense of are quite prominent in many non-western languages, including, in particular, those in which sentence boundaries are less well-defined and therefore much less easily identified than is the case for languages like English, French or German. Some kind of dynamic framework is, for all we now know, absolutely indispensable to describe how form and context determine content in such languages. Arguably the most prominent representative of this new dynamic trend in the semantic description of broad spectra of different human languages is Maria Bittner. Bittner has developed a mode of semantic representation—involving her 'infotention states', which is importantly different from DRT, but which has nevertheless incorporated some of its central ideas.[43] She has also launched the idea of bringing out a collection of papers representing different forms of and perspectives on dynamic semantics, that should clarify how the different systems differ from each other and what options they offer the working semanticists who need a framework to describe the dynamic phenomena that they have encountered in the language or languages they are working on.

---

[43] Bittner, Maria, 2008, Aspectual universals of temporal anaphora, in: Susan Rothstein, (ed.), ***Theoretical and Crosslinguistic Approaches to the Semantics of Aspect***, Benjamins, Amsterdam, 349–385.

AtM: Quite a few of our colleagues think of you as the one who invented DRT and then did much to develop it further. As this selection of your papers shows, that is a rather limited view. But it does seem that over the last thirty years DRT has played a very large part in your work. Can you say something about how you see the future of DRT?

HK: Yes, it is definitely true that DRT has been of central importance to me over these last three decades. I guess the main reason why it has been so important is that for me DRT is the crystallisation of a quite radical change in my way of thinking about language, about its relation to human thought and about the ways in which it functions in human communication. And the changes that DRT has seen over the years are largely a reflection of my thinking about these questions.

The version of DRT that I believe is most widely known is no doubt the one presented in our book ***From Discourse to Logic***. But much has changed since the time, almost twenty years ago now, when FDTL was published; the versions of DRT that we at the University of Stuttgart have been using for at least a decade differ from the FDTL version in several quite fundamental respects. Not all changes have been motivated by the ideas I just alluded to. For instance, one important change is that in the current versions DRSs are constructed bottom up and that their construction involves two stages—first, construction of a preliminary DRS with explicit representations of all presuppositions and then resolution of those presuppositions in the context. These changes were made following van der Sandt and Geurts, who in the early nineties modified DRT to fit van der Sandt's treatment of presupposition justification as a form of anaphora resolution.[44] And another important difference of one of the DRT versions in current use is one in which DRS construction doesn't start with semantic entries for lexical items, but 'further down'—with the semantic specifications of lexical roots and the semantic contributions of sub-lexical functional heads. This change is motivated by work on the syntactic and semantic structure of the lexicon that has been undertaken by Roßdeutscher and the people in Stuttgart who are working with her.

---

[44] van der Sandt, Rob, 1992, Presupposition projection as anaphora resolution. ***Journal of Semantics***, 9, 333–377.

van der Sandt, Rob and Bart Geurts, 1991, Presupposition, anaphora, and lexical content, in: Otthein Herzog and Claus-Rainer Rollinger (eds.) ***Text Understanding in LILOG***. Springer, Berlin, 259–296.

But other important changes *are* reflections of the evolving ideas about the relation between language and thought and the ways in which language functions as an instrument in communication. One of these changes has been in place for quite some time now and descriptions of it can be found in the existing literature.[45] This change involves two steps: first, using DRSs as building blocks in a formalism for the description of complex mental states, consisting of a collection of propositional attitudes, as well as a collection of 'entity representations', which can serve as constituents in the propositional attitudes; and, second, an extension of the DRS formalism itself with conditions that attribute complex mental states described in this way to some agent at some time.

In more recent work, starting as joint research with Agnes Bende-Farkas on specific indefinites around 2000, the assumption that the DRSs constructed from linguistic input capture important features of the mental representations constructed during interpretation is extended to a similar assumption about the representations of thoughts that people turn into words when they express them in speech or writing.[46] I am as aware now, as we were then, of the dangers of such a step. The assumption that DRSs, or structures that contain DRSs as building blocks capture important aspects of mental representation is already one that can be met, and has been met, with considerable scepticism. But in assuming that these same structures also serve as inputs to language production we are surely sticking our necks a good deal further out. But hazardous as these assumptions may be, they have, I think, proved quite fruitful in dealing with a number of linguistic and philosophical issues, and we may as well hang on to them so long as no alternative, based on more solid evidence, is available to us.

Hypotheses about the form in which information is present to the mind both as input to language production and as the result of language interpretation are important insofar as they enable us to describe verbal communication as a process in which the recipient tries to build a representation from what he hears or reads which matches the intention of the speaker.

---

[45] Kamp, Hans, 2003, Einstellungszustände und -zuschreibungen in der DRT, in: Ulrike Haas-Spohn (ed.), ***Intentionalität zwischen Subjektivität und Weltbezug***, Mentis, Paderborn, 209–289.

Kamp, Hans, Josef van Genabith and Uwe Reyle, 2011, An updated survey of DRT, in: Dov M. Gabbay and Franz Guenthner (eds.), ***Handbook of Philosophical Logic***, Second Edition, Vol. 15, Springer, Berlin, 123–349.

[46] Kamp Hans and Agnes Bende-Farkas, 2001, ***Indefinites and Binding: From Specificity to Incorporation***. Lecture Notes for the 13th European Summer School for Logic, Language and Information (ESSLLI 2001) at the University of Helsinki, Finland, 13–24 August, 2001.

This way of describing the function of language is particularly useful in connection with a range of issues about reference, about how the different types of definite noun phrases of a language such as English succeed in transmitting reference in thought from speaker to addressee, and about what guides speakers in choosing one such reference device as opposed to another.

One would expect that such a 'bi-polar' approach to meaning transfer should not just be important in connection with reference. We can see it as a new setting for making formal sense of the Gricean perspective of linguistic meaning as 'non-natural meaning' and of the idea that success in communication is achieved when the addressee captures the communicative intentions of the speaker. It is a general feature of verbal communication that the speaker has to be careful when choosing the words she uses to get her message across and that she is guided in her choice by what she thinks is most likely to work given what she can expect her audience to know. Questions of this sort—whether the speaker's words will do what they are supposed to do and how she goes about choosing words that she thinks will do the job—arise for instance quite prominently in discussions of 'tropes' such as irony and metaphor; but so far these discussions have been necessarily informal, since there has been no satisfactory formal framework within which they could be made explicit.

Thinking about meaning and communication along these lines leads inevitably to the question what it is for two or more persons to share information, and how verbal communication can lead to information sharing. Accounting for this central aspect of communication—the *intersubjectivity* that it is capable of creating and on which it also relies—seems to me to be one of the great challenges for any theory of meaning and communication. Information can be shared without having to be objectively true. And this can happen in more than one way: in cases where those who share the information are subject to a shared illusion—each wrongly takes shared information to be true—and in cases where it is clear to all involved that what is shared isn't true objectively, as when we are engaged in fictional discourse.

AtM: Formal semantics originated in the sixties and became firmly established as a well-defined discipline in the seventies. What differences are most salient to you between natural language semantics as it was done then and how it is pursued today?

HK: The particular section of the philosophical and logical community by which I was formed and socialised in the second half of the sixties was what has since often been referred to as 'Southern California Semantics',

and as I indicated earlier in this interview, the two most important strands of that socialisation were (i) the model-theoretic approach to natural language semantics initiated by Montague, and (ii) the operator-oriented approach to semantic questions, which in my case was represented primarily by Prior's work on Tense Logic, but of which there were also many other instances (such as e.g. Hintikka's modal-logical approach to knowledge and belief). In both kinds of work we can, from the present perspective that has been shaped by developments over more than four decades, clearly discern the influence of ideas that came from formal logic and that nowadays are widely seen as alien to the way in which meaning is realised in the languages we speak. Montague's work contains ways of dealing with English syntax which could be seen as implementations of his dictum that 'syntax is the handmaiden of semantics', in the more provocative and more problematic interpretation of that phrase. And the attempts to capture aspects of natural language meaning in terms of Tense Logic and other variants of Modal Logic were inspired by the idea that many parts of speech in natural languages function as modal operators—that natural languages were much more like modal logics than we have come to think they are since then.

As I see things now, a significant part of the developments in natural language semantics during the seventies had to do with freeing ourselves from the sway of these alien conceptions. A more balanced and refined understanding of the relation between syntax and semantics was due largely and crucially to the work and personality of Barbara Partee. It is thanks to her that Montague's model-theoretic approach could have become the dominant method in formal natural language semantics that it is today. And she made that possible because her insight that the model-theoretic method which Montague had succeeded in applying to natural language was crucially important to semantics went hand in hand with that knowledge of generative syntax that enables us to see the syntax of natural languages as having a certain autonomy. Its structures are not only motivated by the meanings that they serve to support, but are partly determined by principles that are authentically syntactic, principles that have to do with questions of form and questions of form only. And then of course it was also essential that Barbara had already established for herself a reputation as an authority in English syntax.

In rethinking and redoing Montague's own work in ways that do justice to these insights about syntactic autonomy, and by gradually extending those results to cover an ever larger range of semantic phenomena in an ever larger number of languages, Montague Grammar has become what it is today. It constitutes a theory of the relationship between form and meaning that

acknowledges both form and meaning as realities in their own right, each subject to its own laws and principles which reveal themselves through mostly complex and partly independent webs of empirical evidence, but nevertheless are bound together in the tight and systematic manner without which human languages could be neither learned nor used. That connection is usually described as the 'compositionality' of natural language.

I cannot resist adding at this point an observation that has been with me ever since it was made clear to me that syntax is significantly autonomous. If syntax is the handmaiden of semantics, she comes to this task with properties of her own, of which it isn't self-evident that they are optimally suited to this task. The bulk of work in natural language semantics over the past decades has consisted in dealing with questions that have to do with the syntax-semantics interface; questions of the sort: 'How is it possible for this syntactic construction to carry this kind of meaning?' In fact, questions of this sort have been the main preoccupation of natural language semanticists, and they will undoubtedly continue to be a central concern for the foreseeable future.

But why should that be so? Why don't the syntactic forms of natural languages show the kind of semantic and logical transparency with which we are familiar from the languages of formal logic, such as the predicate calculus or the lambda-calculus? From this perspective—and thus from the vantage point of the formal logician—it easily looks as if the human race has made things unnecessarily difficult for itself, by developing languages in which meanings are expressed by syntactic means that seem to lack the kind of transparency that formal languages so clearly and admirably display. Why is it that humans have made things so difficult for themselves?

As a matter of fact I do not really endorse this last question. I am somehow convinced that the way of looking at language that prompts the asking of it is wrong. But why is it wrong and in precisely what ways? I suspect that coming up with true and persuasive answers to that question is one of the hardest challenges for a general theory of language; but I also expect that good answers may teach us more than almost anything else about the nature and likely origins of human language as a tool for communication and for the clarification of our own thoughts, and about the ways in which language is interwoven with non-linguistic aspects of human cognition.

My own emancipation struggle in the seventies was primarily against the assumption that so much of meaning in natural language is expressed in the form of modal operators. As I mentioned earlier, my doctoral dissertation had been on an extension of Priorean Tense Logic. The aim of the dissertation was to design a Prior-type tense logic that was optimally

expressive. Given that aim, the project of the dissertation could be described as a success. But in a way the very success of the project boded its own downfall. The main result of the dissertation was that the tense logic proposed in the dissertation is capable of expressing arbitrary temporal relations that are definable in the first-order theory of the relation of earlier and later. But when one looks more closely at *how* various such relations can be expressed in a language like English and compares those modes of expression with how the ways in which those relations can be expressed in the tense logic of the dissertation, then one cannot help being struck by the overwhelming discrepancies: the tense-logical paraphrases of the constructions that are used to express temporal relations in a natural language such as English bear virtually no resemblance to the constructions they paraphrase.

But the inadequacy of tense logic as a framework for dealing with the temporal dimension of natural languages goes further: A closer exploration of the various ways in which temporal relations are expressed in English and other human languages quickly reveals that there is virtually no linguistic justification for the idea that there is any role to play for tense operators in the semantic analysis of temporality in natural language. Many of the expressions we use in expressing temporal relations behave, by any reasonable criteria, as referring terms—terms that refer to times in the same way in which, say, as the name 'Barack Obama' refers to the man Obama. The most telling examples of such terms are dates—expressions such as *Wednesday*, *the first of October*, (*at*) *five to ten* and so on. But the referential function of temporal expressions is by no means restricted to them. If we take their referential appearance at face value, as I soon became convinced one should, then the semantics of temporal reference takes on a completely different character. Times are now part of the ontology of natural language, just as physical objects, persons and various other familiar and uncontroversial sorts of individuals.

Admission of times among the first class citizens of ontology wasn't the end of the story. The next step along the road of emancipation from the modal operator paradigm was the conviction that verbs and their various syntactic projections should be construed as describing and thereby introducing into the discourse events, processes and states. The arguments that led me to this second conclusion aren't exactly the same as those that led to the adoption of times as part of the ontology. But in spirit the arguments were quite similar; and the conclusions were similar too, in that they both led to a richer ontology than one would have had to acknowledge without them. For a model-theoretic approach of the kind that has become a kind of standard since Montague, these ontological extensions come with the

challenge to say precisely what models which reflect such a conception of ontology must be like: What are the different ontological sorts that make up the universes of such models, what is the internal structure of each of those sorts, and how are different sorts structurally related to each other?

This task still is a major problem for formal semantics as we have it today. In fact it is nothing but the model-theoretic guise of the problem that, in the apt phrase of Emmon Bach, is often referred to as the question of 'natural language metaphysics'.[47] So far, precise articulations of natural language ontology have been fragmentary at best whether as part of model-theoretic natural language semantics or in some other setting. Some parts of it are in reasonably good shape. In particular, it seems to me that our understanding of the structure of time and its formal articulation as part of model-theoretic semantics is now about as deep as it needs to be. But already when we turn to events and states the issues are much less clear, and explicit articulations of ontological structure much more tentative. On the whole, most of the work in this domain, to the extent that it exists at all, has been limited to such ontological domains as time, space and mereology. But there is a vast range of ontological questions that will have to be clarified and articulated formally as we proceed to more ambitious systems of model-theoretic semantics, which cover larger natural language fragments, and which contain better approximations of the semantics of ever larger parts of the lexicon. A notorious problem present the many different kinds of abstracta that we find among the denotations of words: properties and kinds, laws, obligations and contracts, virtues and vices and so forth.

There are various resources that anyone who wants to address these aspects of natural language ontology can and should look at—from world knowledge data banks such as Cyc to more linguistically motivated banks such as WordNet or FrameNet. But in my own limited experience what can be found in those sources almost invariably needs to be carefully rethought and recast before it can be incorporated into the particular model-theoretic semantics that one is trying to develop. I expect that in the years ahead of us work on the ontology of natural language will reassert itself as one of central preoccupations of our discipline.

This answer is a very long one already—but then, this is a very big question!—and so far the only issue it has addressed is the relation between semantics and ontology. But that, of course, is only one of the issues that have been prominent in formal semantics over these past forty years and

---

[47] Bach, Emmon, 1989, ***Informal Lectures on Formal Semantics***. SUNY Press, New York.

for which there nevertheless remains a great deal to be done. Some of the other big themes, both during these past decades and for the future, we have touched upon in this exchange as we went along: dynamics, hyperintensionality, generation as an (imperfect) mirror of interpretation, the internal syntactic and semantic structure of expressions that dictionaries treat as words. But other topics haven't even been mentioned. Among them are a number of difficult general problems at what is often described as the 'semantics-pragmatics interface'. The precise form of different patterns of Gricean reasoning that interpreters use to compute implicatures is one such issue. A second one is if and how utterances can be factored into a content component and component which captures the speech act type of the utterance: which elements of an utterance should be treated as constituents of the content it expresses and which as indicators of the type of speech act that is being performed? Another big and difficult problem is vagueness, to which linguists and philosophers keep returning, but which has proved remarkably recalcitrant.

Perhaps the biggest challenge of all is the still poorly charted field of non-literal meaning, including metonymy and metaphor. I think existing work on metaphor has made it quite clear that there are considerable differences between phenomena all of which have been described as metaphoric. What we need for a start is a finer, well-motivated classification of the phenomena—that is an indispensable foundation for any account of metaphor that is to be both comprehensive and formally precise. But to my knowledge such a classification is still outstanding, not to speak of a theory that it could serve as foundation.

A particularly irksome feature of metaphorical uses of language and of the creation and non-literal extension of lexical meaning is what might be called their 'semi-productivity': Sometimes it is possible to predict on the basis of fully general principles how the semantics of a word can be extended from its literal to its non-literal meaning, or how new words can be formed with a particular form and meaning. But in general such precise predictions are not possible. On the other hand, even where exact prediction is not possible, it is nevertheless as a rule quite clear that not everything goes. Sense extension, introduction of new words and other aspects of metaphoric language use are constrained by general principles. But these principles only determine outer bounds within which sense extension, creation of new words etc. are possible. These principles do not define, they only confine meaning. How the meanings of certain words are extended and what new words get introduced into a language, and when, are for the most part idiosyncratic.

For formal semantics, accounting for this kind of semi-productivity is a particular challenge. It is a challenge that in some ways is comparable to the challenge posed by vagueness: How can we be formally precise about what is, by its very nature, the very opposite of precise? For vagueness quite a bit of progress has been made with this challenge in the course of the past forty years or so. Here some creditable methods have been developed for describing the imprecise in precise terms. But no comparable frameworks have thus far been worked out to deal with the problems of sense extension, creation of new words and non-literal meaning more generally. If we had such a framework we would probably have a much better and deeper understanding of linguistic meaning, of its fluidity, flexibility and dynamics, than we do at present. But we don't; and developing such a framework may well prove to be the greatest challenge of all that semantics is facing today.