

# States in Flux: Logics of Change, Dynamic Semantics, and Dialogue

Oliver Lemon

Doctor of Philosophy  
University of Edinburgh  
1995



## Declaration

I declare that this thesis has been composed by myself and that the research reported here has been conducted by myself unless otherwise indicated.

Oliver Lemon

Edinburgh, 1995

## Abstract

Verbal communication is a complex process of negotiation towards a common aim, maximizing each participant's information about the world. Conversations typically involve dispute, correction, and agreement between their participants. How could a formal theory begin to describe such phenomena? In particular, how might a mathematical theory of utterance interpretation (a "formal semantics") handle the idea that language use systematically *changes* the beliefs of communicating agents?

The contributions of this thesis, in the fields of logic and formal semantics, are largely constructive. The research is carried out within two main "traditions", the research programmes of "Belief Revision", or Theory Change, and Dynamic Semantics (and to a lesser extent, Paraconsistent Logic). These theories are integrated, and applied in a formal model of dialogue. Thus, the objective of the thesis is to examine a marriage between "logics" of Theory Change (which describe rational modifications to theories in the light of new information) and systems of Dynamic Semantics (which describe an agent's processing of incoming linguistic information). This venture involves providing a semantics for Theory Change, extending Theory Change systems to the first-order case (i.e. where theories are expressed in predicate logic), and interfacing the result with existing systems of dynamic semantics. The project is placed in the larger context of formal semantics for dialogues; a theory of *changing* information states in communication, and a theory of the rational agency of communicants.

Philosophically speaking (in part I) the notions of information and communication come under scrutiny, as do some foundational issues for a formal semantics of communication. Particular definitions of information and communication are argued for; definitions which concentrate on transformations over epistemic states as the primary subject matter of formal semantics. A view of model-theoretic semantics is developed, which also concludes in the importance of an epistemic approach.

In terms of formal systems (part II), the central project is the development both of a semantics (called "Revision Semantics") for existing (propositional) systems of theory change (Gärdenfors 1988), and an extension of rational theory change systems to the first-order case, also with semantics. This approach results in an account of utterance interpretation which allows for arbitrary *change* in the information state of the processor (or agent), rather than simple information growth, or update (as in Update Semantics for example). A variety of simple completeness results are established, linking different Theory Change systems with *revisable* systems of Dynamic Semantics. The systems are *robust* in their handling of apparent contradictions in dialogue.

In part III, systems allowing revision are applied in the analysis of dialogue contributions known as "repairs" or "corrections". A treatment of constituent negation clauses and other related phenomena is provided in terms of revisable dynamic semantics. A model of communication is proposed, by which agents can dispute and repair information conveyed earlier in a dialogue (either by themselves or by another agent).

The potential for a fruitful combination of paraconsistent logics (logics which are *robust* in that they do not "explode" under contradiction) with revisable information states is also considered, so that gradual revision of inconsistencies becomes possible, along with an account of rational agent autonomy.

In addition, the notion of *change* in information is discussed with regard to the 'objects' referred to during communication. The "dialogue referents" emerging from revisable dynamic systems are epistemic "partial objects" whose properties can change.

In the body of the thesis, links between theory change systems and intuitionistic logics are also considered, as well a variety of approaches to the epistemic modals.

The main contributions of the thesis are the provision of revisable systems of dynamic semantics, the development of first-order theory change systems (with dynamic semantics), and an analysis of communication in terms of changing information states of rational agents.

## Foreword and Acknowledgements:

"God keep me for ever completing anything. This whole book is but a draught of a draught. Oh, Time, Strength, Cash, and Patience!"  
(from Herman Melville's "Moby Dick")

The motivations for this project sprang from 3 major sources; undergraduate courses in non-classical logics run by John Chidgey of the Philosophy Department at the University of Manchester, Jeff Paris's mathematics course in Theory Change (there known as "Belief Revision"), and from the community of logicians, philosophers, and semanticists at Edinburgh's Centre for Cognitive Science (CCS) and Human Communication Research Centre, and the Departments of Philosophy, Computer Science, and Artificial Intelligence. During my time at the CCS weekly discussions in the non-standard and applied logics ("NSAPLOG") reading group were invaluable. Tutorial discussions with students have also been irreplaceable.

Amongst the many individuals "without whom", I must mention Andreas Schöter, Lawrence Cavedon, Jonathan Ginzburg, Paul Dekker, Peter Ruhrberg, Tsutomu Fujinami, Darren Brierton, Greg Restall, Andrew Gillies, Ulrike Hahn, and especially Carl Vogel and Rodger Kibble. Thank you all for influencing this research in various complex and subtle ways!

I owe an enormous debt to the various supervisory talents of Paul Schweizer, Robin Cooper, and Peter Milne, each of whom helped to shape this thesis. I have also had the good fortune to be able to discuss this work with Ian Pratt, Greg Restall, and participants at conferences in the Czech Republic and Georgia.

Much of the content of Chapters 3 and 4 was delivered as a paper at the Logica International Symposium, Liblice Castle, in the Czech Republic, June 1995. A version of that material appears in the proceedings of LOGICA'95 [72].

Thoughts from chapter 6 were delivered at the Prague International Colloquium on "Objects, Structures and Meanings" (in memory of Pavel Tichy), September 1995.

In addition, some material from chapter 5 was presented at the Symposium on Logic Language and Computation, held in Gudauri, Georgia, in October 1995. That paper [71] is to appear in the collection "The Tbilisi Symposium on Logic, Language, and Computation", published by the Center for the Study of Language and Information (CSLI), Stanford.

This research was supported by a Science Research Studentship from the Department of Education of Northern Ireland (DENI).

I am especially grateful for the support and tolerance of my family, Maisie, Sue Bain, John Hicks, Masja Kempen, the Catchers, and everyone at the Welcome Nursery.

# Contents

<b>Part I</b>	<b>2</b>
<b>1 A Formal Semantics of Communication</b>	<b>2</b>
1.1 The significance of change . . . . .	2
1.1.1 Origins . . . . .	4
1.1.2 Communication and Change . . . . .	5
1.2 Definitions and Distinctions . . . . .	7
1.2.1 Methodology and Empirical goals of the thesis . . . . .	9
1.3 Change in Dynamic Semantics . . . . .	10
1.4 Communication and Information . . . . .	13
1.4.1 Communication . . . . .	13
1.4.2 The problem of communication . . . . .	15
1.5 Information . . . . .	18
1.5.1 The Informational Metaphor . . . . .	20
1.5.2 Information change: Eliminativism versus Constructivism . . . . .	22
1.6 Mathematical Models in Formal Semantics . . . . .	23
1.6.1 An epistemic/doxastic semantics . . . . .	28
1.7 A Revision Semantics for communication . . . . .	29
1.7.1 Conclusion . . . . .	31
<b>2 Formal Preliminaries in Dynamic Semantics and Theory Change</b>	<b>32</b>
2.1 Outline . . . . .	32
2.2 Information in communication . . . . .	32
2.2.1 What is information growth? . . . . .	33
2.2.2 Information: Syntax and Semantics . . . . .	35
2.3 Properties of a Dynamic Semantics . . . . .	36
2.4 Two dynamic traditions: DS and TC . . . . .	39
2.4.1 Integrating the Accounts . . . . .	43
2.5 Problems for standard dynamic semantics . . . . .	43
2.6 Review of the AGM results . . . . .	45

2.6.1	Expansion . . . . .	47
2.6.2	Revision . . . . .	47
2.6.3	Contraction . . . . .	48
2.6.4	Identities . . . . .	49
2.7	Constructing Contractions . . . . .	50
2.7.1	Multiple (or “General”) Contraction . . . . .	51
2.7.2	Constructing General Contractions . . . . .	53
2.7.3	Structuring theories . . . . .	54
2.8	Modelling Theory Change . . . . .	55
2.8.1	Intuitionistic Semantics for extension . . . . .	56
2.8.2	Grove’s Spheres . . . . .	56
2.8.3	“Supermodels” . . . . .	57
2.9	Logics of Information . . . . .	58
<b>Part II</b>		<b>63</b>
<b>3</b>	<b>A Dynamic Semantics of Theory Change</b>	<b>63</b>
3.1	Introductory Remarks . . . . .	64
3.2	Related research . . . . .	65
3.3	Definitions . . . . .	68
3.3.1	Non-standard definitions . . . . .	70
3.3.2	Partial Worlds versus sets of total worlds . . . . .	71
3.4	Theory Extension and Model Elimination . . . . .	73
3.5	Theory Contraction and “Downdating” . . . . .	74
3.5.1	Desiderata for a genuine “downdate semantics” . . . . .	75
3.6	Theory change as a model of dialogue . . . . .	77
3.7	Updating and AGM theory extension . . . . .	79
3.7.1	Downdate Semantics rationality postulates . . . . .	82
3.8	Systems of Theory Contraction and Downdate Semantics . . . . .	83
3.8.1	Supermodel Semantics . . . . .	84
3.8.2	Full Join Downdate Semantics (FJDS) . . . . .	87
3.8.3	Partial Join Downdate Semantics (PJDS) . . . . .	88
3.8.4	Defining Revision . . . . .	90
3.8.5	Reprise and critique . . . . .	90
3.8.6	A “merely” formal semantics? . . . . .	91
3.9	Application: Revisable Dynamic semantics . . . . .	92
3.10	Conclusion . . . . .	93
<b>4</b>	<b>Revision Semantics</b>	<b>95</b>

4.1	Two dynamic problems: implosion and chivalry . . . . .	96
4.2	The “Downdating” Approach . . . . .	101
4.2.1	Revision Semantics . . . . .	104
4.2.2	Propositional Theory Change . . . . .	107
4.2.3	Epistemic Modals in Revision Semantics . . . . .	108
4.2.4	Properties of Revision Semantics . . . . .	112
4.2.5	Ups and Downs in Discourse and Dialogue . . . . .	113
4.2.6	A Logic for Conditionals . . . . .	115
4.3	Paraconsistency versus Chivalry . . . . .	116
4.4	Constructivity in Dynamic Semantics . . . . .	117
4.4.1	A constructive account of epistemic modality . . . . .	119
4.5	Conclusion . . . . .	120
<b>5</b>	<b>First-Order Theory Change Systems with Semantics</b>	<b>122</b>
5.1	First-Order Extension and Intuitionistic Semantics . . . . .	123
5.1.1	Intuitionistic Downdates . . . . .	128
5.2	Levels, Systems, and Rationality Postulates . . . . .	129
5.2.1	Intuitions behind the Rationality Postulates . . . . .	132
5.3	First-Order Theory Change . . . . .	137
5.3.1	FOTC; syntax . . . . .	139
5.3.2	A Semantics of FOTC . . . . .	141
5.3.3	Revisable Modal Predicate Logic . . . . .	143
5.3.4	Completeness of FOTC wrt RMPL . . . . .	145
5.3.5	Comparisons . . . . .	148
5.4	Adding Dynamics of Anaphora . . . . .	148
5.4.1	First-Order theory extension with Anaphora . . . . .	150
5.4.2	Semantics of DPL . . . . .	151
5.4.3	Revisable DPL . . . . .	153
5.4.4	Completeness of $FOTC_I$ wrt RDPL . . . . .	157
5.5	Dynamic Modal PL, FUL, and $FOTE_{III}$ . . . . .	158
5.5.1	First-order Update Logic (FUL) . . . . .	161
5.5.2	Adding Downdates and Revision . . . . .	163
5.5.3	Revisable DMPL . . . . .	163
5.5.4	Adding some modals . . . . .	165
5.5.5	First-Order Revision Logic . . . . .	167
5.6	Prospects and Conclusions . . . . .	168
5.6.1	FOTC with growing domains . . . . .	168
5.6.2	Free Logic for Existential contractions . . . . .	169



5.6.3	Summary and Conclusions . . . . .	172
<b>Part III</b>		<b>174</b>
<b>6</b>	<b>Linguistic Data: analysing theory changes in communication</b>	<b>175</b>
6.1	Overview . . . . .	175
6.2	Theory manipulation in dialogue . . . . .	177
6.2.1	General features of the model . . . . .	177
6.2.2	Game-theoretical Dialogue Semantics . . . . .	179
6.3	Repairs and Corrections in Dialogue . . . . .	179
6.3.1	Self-correction . . . . .	182
6.3.2	Hearer Correction . . . . .	183
6.4	Analysis . . . . .	184
6.5	Mechanisms of Correction: Linguistic Cues for Theory Revision . . . . .	186
6.5.1	Constituent Negation as Revision Cue . . . . .	186
6.5.2	Intonation and “Editing Terms” as revision cues . . . . .	191
6.6	Other Theory Contractions . . . . .	192
6.6.1	Contraction “for the sake of argument” . . . . .	192
6.6.2	Negative Definitions of Concepts, and definition by revision . . . . .	193
6.6.3	Universal Claims with Exceptions . . . . .	193
6.6.4	Counterfactual and Future Conditionals . . . . .	194
6.6.5	Removal of sub-theories . . . . .	194
6.7	Revision Semantics in a formal theory of communication . . . . .	195
6.7.1	Axioms for assertion (update) . . . . .	196
6.7.2	Axioms for retraction (downdate) . . . . .	197
6.7.3	Axioms for modals . . . . .	197
6.7.4	Axioms for (teacher/pupil) repair . . . . .	198
6.8	Conclusion . . . . .	199
<b>7</b>	<b>Paraconsistency, Autonomy, and Change</b>	<b>200</b>
7.1	Motivation: the case for equal rights . . . . .	201
7.2	Rationality and Paraconsistency . . . . .	203
7.2.1	Paraconsistency and Dynamic Semantics . . . . .	203
7.2.2	Paraconsistency and Theory Change . . . . .	204
7.2.3	Autonomy and Paraconsistency . . . . .	207
7.3	TC and DS with inconsistencies . . . . .	207
7.3.1	Paraconsistent Update Semantics (PUS) . . . . .	209
7.3.2	Paraconsistent Revision Semantics (PRS) . . . . .	210
7.3.3	Paraconsistent Predicate Logic (PPL) . . . . .	212

7.4	Information Quality Measures . . . . .	214
7.4.1	Axioms for resolving disagreement . . . . .	216
7.4.2	Information Quality in DS states . . . . .	217
7.5	Summary . . . . .	219
	<b>Bibilography</b>	<b>220</b>

**Part I**  
**The General Framework**

# Chapter 1

## A Formal Semantics of Communication

“The bud disappears when the blossom breaks through, and we might say that the former is refuted by the latter; in the same way when the fruit comes, the blossom may be explained to be a false form of the plant’s existence, for the fruit appears in its true nature in place of the blossom. These stages are not merely differentiated; they supplant one another as being incompatible with one another. But the ceaseless activity of their own inherent nature makes them at the same time moments of an organic unity, where they not merely do not contradict one another, but where one is as necessary as the other; and this equal necessity of all moments constitutes alone and thereby the life of the whole.”

(G. W. F Hegel, [52])

### 1.1 The significance of change

Logicians have interpreted the world in various ways — the problem is that it changes. The quotation given above is relevant not only as a description of the flux of information which occurs in communication, but as a reminder that the world itself changes. New aspects of the world, and of our information about it, continuously override the old; not just in any arbitrary way, but so as to constitute coherent unities; ‘reality’ and ‘theory’. The world changes in accordance with natural regularities, and our acceptance of

changing information about it is likewise constrained by our standards of rationality<sup>1</sup>. Our information about the world comes from two major sources; language and observation. Both these sources are liable to radical change and even inconsistency. Agents often receive conflicting linguistic information, and the world throws up ‘contradictions’<sup>2</sup> via temporal change: what is the case today may not be so tomorrow.

Moreover, the power of language lies in its ability both to *express* and *change* the information that people have about the world; how it is, was, may be, or may become. Indeed in some cases the language used by somebody else will actually *force* agents (if they are rational) to change their beliefs and to hold new ones. The research programme known as *Dynamic Semantics* assumes this notion of *changing* individual information states as the basis for a theory of meaning (although it shall be disputed that such a theory could give an exhaustive account of *meaning*).

As has been mentioned, the significance of change lies not only in the philosophy of language. *Theory Change* or *Belief Revision* systems have been motivated by the notions of change in databases, legal systems, and epistemic states of (either real or artificial) agents (see eg: William James [58]), as well as in scientific theories<sup>3</sup>. Amongst other philosophers, Quine has pointed out the importance of the provisional nature of all theories (the “Quine-Duhem thesis” see eg: [80]). Figures in the philosophy of science (Popper, Kuhn, Lakatos) have stressed that all theories encounter (sometimes radical) revision in the face of empirical refutation. Carl Hempel (see [54]) attempted to systematise scientific theory change, providing part of the motivation for “belief revision” systems. Nevertheless, in contrast to these historical precedents, the focus of this thesis is the application of “logics” of theory change as a central part of a dynamic semantics of dialogue. An account is sought of how the processing of linguistic structures is seen to construct and modify agent-relative theories and information states.

---

<sup>1</sup>These rationality constraints apply to coherence of discourse, narrative, and dialogue, as well as scientific theory change.

<sup>2</sup>In the sense that, unless new information *supplants* prior information, rather than simply adding to it, contradictions ensue.

<sup>3</sup>In addition, for a discussion of the importance of change in the Philosophy of History, see [70].

### 1.1.1 Origins

One way of approaching this thesis is via the argument that agent-relative theories, or their contents (individual information states), play a crucial role in communication. By a “theory” I mean the syntactic characterization of an agent-relative epistemic state (or information state); the sort of thing which has previously been dubbed a “belief state,” or “belief system”. However, the “belief states” which I consider have little connection with the usual philosophical concerns about propositional attitudes. Indeed, in order to avoid any misleading entanglement with the literature on the propositional attitudes, I prefer (along with most current researchers) to do away with the label of “Belief Revision” in favour of “Theory Change.”

An integration of formal semantics with an account of communication is long overdue<sup>4</sup>, and certainly, some theory-like construct is central to achieving that goal. For not only are theories indispensable in our everyday theory of persons, but they are also the best starting point from which to approach the semantics of dialogue and a theory of communication in general. Moreover, some philosophers argue that an approach centered around (what I have called) theories is the best starting point for a scientific analysis of cognition in general (see eg: Dennett [24]). In addition, most research in AI and formal semantics assumes constructs similar to agent-relative mini-theories (eg: “background knowledge”, “conceptual schemes”, or “frames”) as an essential foundation. In general, then, formal treatment of such pragmatic and contextual factors is central to an account of communication

Against this, some philosophers might urge that the notion of a “theory” is too vague to be of any special interest whatever. Such comments are correct insofar as they point out the lack of precision in the commonsense notion of “belief systems”, but also fail to recognise that such “theories” can be refined and formalized to a useful extent. In part, this thesis embodies an attempt to further that project.

In particular, I take operations of *theory change* (TC) or “belief revision” (transformations over theories) to be of central importance in an adequate explanation of communicative linguistic behaviour. Indeed, communication is explicated as a series of (potentially) theory-changing speech-actions, as a process of iterated “belief revision”.

---

<sup>4</sup>For example, see [44]

Crucially, the stance taken here on the central importance of the notion of a theory (or “information state”) does not imply any ontological commitment to those theories above the level of that of theoretical entities<sup>5</sup>. Theories do not exist like houses and trees, but are syntactic classifications of *epistemic states*. It is in this sense that I go on to advocate semantics as a branch of mathematical epistemology.

Ontologically speaking, I construe a theory essentially as a rational agent’s epistemic approximation of a state-of-affairs, (or a “set-up”, or situation<sup>6</sup>). Structures distinctly like theories are central to an account of linguistic understanding, and are to be eliminated only at that expense.

### 1.1.2 Communication and Change

The vast majority of semantical research in the philosophical tradition concentrates on language as a relation between propositions and the world, rather than as a tool of communicants. Obviously this concern with language-world correspondences (referential truth-conditional semantics) overlooks much of the mechanics of everyday language use. However, worse than that, this traditional emphasis produces a somewhat blinkered vision of some central theoretical points concerning language use in general. Communicative behaviour relies more often on speaker-hearer correspondences, as it involves negotiation for the purpose of co-ordinating action. Some recent formal linguistic theories (eg: DRT [61], DPL [46]) have concentrated quite convincingly on the modelling of a linguistic agent’s interpretation of utterances. However, this hearer-centered perspective<sup>7</sup> also produces a philosophically misleading approach to semantics. Both the old and new traditions in formal linguistics and philosophy of language fail, in a damaging way, to do justice to simple communicative considerations. Dialogues are often used to resolve disagreements, repair misconceptions, and negotiate theories. Such dynamic processes require the interpretation of utterances not only to add to the information of dialogue participants<sup>8</sup>, but also, potentially, to *change* their information.

<sup>5</sup>see Quine, “On Mental Entities” p.223 in [80]

<sup>6</sup>These being notions derived from Russell, the relevant logicians, and situation semantics respectively

<sup>7</sup>Paul Dekker calls this a “recipient oriented” semantics in [21].

<sup>8</sup>As is the case in Update Semantics and Dynamic Predicate Logic, and their progeny.

If the explanatory scope of Dynamic Semantics is to be extended to cover communicative contexts, so that actual *dialogue* contributions can be formally treated, an incorporation of processes of information *change* is essential.

I take the task of a formal semantics of dialogues to be the construction of a mathematical theory of the *interpretation* and production by agents of utterances during communication. This project is to be sharply distinguished from the usual philosophical task of semantics as uncovering the *meanings* of utterances. This “agent oriented” approach is an epistemological, rather than metaphysical, venture.

### Outline of the body of the thesis

Firstly, the remainder of this chapter is organized as follows:

I define some basic terms and make distinctions. Then the assumptions underlying the thesis are laid bare and given some initial argument. The exploration revolves around a discussion of the issue of mathematical modelling issues in semantics. The precise area of application that concerns me here is the mathematical modelling of linguistic communication in terms of changing “information states.” The main worries will be:

(1) How could a mathematical theory *explain* communicative phenomena?

and, (2) What notion of “information state” is needed in a formal semantics of communication?

The concern of the first chapter is to establish and argue for the philosophical underpinnings of the approach, which I am only able to hint at later on. In particular I focus here on the notions of information and communication, and foundational problems with the venture of formal semantics in general.

Having established some basic foundations in the first two chapters (part I), the thesis becomes increasingly formal in part II, until the applications and discussion of part III. Once the groundwork of part I is established, the following chapters concentrate on formal problems in Theory Change, Dynamic Semantics, and dialogue modelling, and logical tools with which to solve them. It is to this end that the theory change and dynamic semantical systems introduced in chapter 2 are combined.

One clear point which motivates the shift towards first-order theories (in chapter



5) is that information exchange in communication cannot, in general, be described at a coarse propositional level. In particular, a restriction to the propositional calculus fails to account for the “grain” of correction exploited by constituent negation clauses, and anaphor resolution. Analysis of constituent negation clauses (in chapter 7) reveals that a predicate calculus style of formalization is required to account for the fine-grained incrementality of the interpretation of corrections. This point counts against many AI type belief models, which take propositions as the unit of change (eg: AGM belief revision). Of course, the more fine-grained systems still have to satisfy the propositional-level constraints (AGM rationality postulates).

Initially (in chapters 3 to 7), the performance of an ideal and gullible rational agent in communication is investigated. Later the issue of agent autonomy is explored (Chapter 8), introducing more sceptical agents. An *autonomous* rational agent is one who will revise their current information if the result of that revision coheres better with their previously established theory, and generally forms a more informative<sup>9</sup> system. Thus a form of limited agent autonomy (limited only by their rationality) is preserved, relative to each individual’s operational theory.

In addition, the claim that some notion of partial or abstract, arbitrary or conceptual object is essential in an informational account of language understanding is explored in Chapter 6. I explore the idea that such objects, or “dialogue referents” can change in various ways.

The central concepts of the thesis are now considered.

## 1.2 Definitions and Distinctions

Some basic notions for the ensuing enquiry are defined below. Many of these definitions will be given a formal characterization in the main body of the thesis.

A **theory** is the syntactic form of an agent’s epistemic approximation of some portion of reality. A theory is a partial description of the world as perceived by an agent.

By **agent** I mean an idealized rational information processor, with the capacity for

---

<sup>9</sup>A way of measuring the relative merits of competing theories shall be explored.

perception and action, including linguistic action.

By **information** conveyed by an utterance I mean the change in the content of an agent's *theory* induced by their *interpretation* of that utterance.

By **communication** between agents I mean *information exchange* for the purpose of changing the contents of theories belonging to those agents.

By **exchange** of information I mean the construction and revision of agent-relative theories (rather than the actual transfer of "pieces of information" from one agent to another.)

An agent's **interpretation** of an utterance is taken to be its construction of an epistemic approximation (a partial theory) to the state-of-affairs supposedly described by that utterance, relative to the agent's background beliefs and context set by prior dialogue.

Notice some consequences of the above definitions. Information states, as I define them, are partial, cannot be separated from interpretation, and are thus agent relative. This definition contrasts with some current approaches (eg: Barwise [10]) to information, as shall shortly be explored. *Informational* factors are to be distinguished from *logical* ones. The class of logical possibilities is much wider than the class of informational options, for information must obey *rational* epistemic constraints as well as logical ones. (See the "multiple extensions problem" later in this chapter.)

The notions of *rationality* and *logicality* are therefore understood to be distinct. Rational options are always a subset of the logical options, but not vice versa.

- The *communicative effect*<sup>10</sup> of an utterance is the actual change that it induces in the epistemic state of its hearer.
- An utterance is *informative* (or has communicative effect) only in the case that it *changes* (either expands or contracts) the range of epistemic possibilities of the hearer. (A consequence of this is that non-eliminative statements are genuinely informative.)

---

<sup>10</sup>Dynamic Semanticists usually take this to be the *meaning* of an utterance

- To *understand* an utterance, upon hearing it, is to recognize the change it would make for your epistemic state (beliefs) if you were to incorporate it as sound information. This is a *hearer's interpretation*. Thus hearers can understand utterances which they don't actually believe.

### 1.2.1 Methodology and Empirical goals of the thesis

The overall explanatory aim of the thesis is to draw out the principles and structures underlying the notion of changing information states as they are manipulated in communication. The empirical work to be achieved is the extension of the data coverage of systems of Dynamic Semantics (see chapter 7). Currently, there are many dialogue contributions<sup>11</sup> which cannot be formally analysed because they have to do with the correction, repair, and recovery strategies of the participants. These dialogues involve contraction of established information, and, more often than not, revision of contradictory information. Most commonly, such dialogue contributions are manifested as *constituent negation* clauses. For example:

**Example 1** (*van Leusen 1994*)

A: *They gave Tim a blue kite.*

B: *No, they gave Tim a YELLOW kite.*

The semantics developed here, it is hoped, will advance the formal treatment of such utterances and other dialogue contributions in dynamic semantics (again, see chapter 7).

Furthermore, the systems developed will go some way to explaining the manner in which communicants manage to agree on referents, and share information about "objects". As with all scientific endeavours, some degree of predictive power is expected of the model developed. Predictions will take the form of rationality constraints upon options for the next utterance in a dialogue.

---

<sup>11</sup>See, for example, corrections in the the Map Task [15] dialogues.

The data to be covered will be drawn from the linguistics literature on corrections and repairs in discourse and dialogue (van Leusen [99], Carletta, Caley, and Isard [15], Levelt [73]). Cases will also be presented in the form of “talk experiments”; invented “mini-dialogues” which illustrate certain phenomena which revisions might handle. Like “thought experiments”, “talk experiments” are supposed to uncover conceptual constraints on rational language use. Their purpose is to explore possible dialogues, and uncover dialogue strategies. In particular, the use of epistemic modals, constituent negation, and question forms will be significant. The data draws out the kinds of information structures which can feature in revisions, and the types of expressions used to effect such changes.

### 1.3 Change in Dynamic Semantics

If the spirit of dynamic semantics (which I argue is that “*communicative effect* is the change induced in the hearer’s information state”) is really to be taken seriously, then dynamic systems do not only require a notion of update (simply adding a sentence and its consequences), but also ways of revising and contracting information (handling contradictions, and retracting previously established information.) The repair and recovery strategies of agents in dialogue have implications for a theory of information in communication.

Formal semantics needs to be integrated with an account of communication in general. The setting for nearly all dynamic theories of semantics has been *discourse*, whereas it is clear that *dialogue* is a basic form of language use<sup>12</sup>. A primary purpose of language use is to change the epistemic (or, more precisely doxastic) states of other agents. Formal semantics has commonly abstracted away from communicative issues, relegating them to the “dustbin of pragmatics.” Empirical adequacy demands, however, that communication be seen as the correct setting for semantic theories in general (otherwise many dialogue contributions cannot be analysed.) Indeed, one might regard discourse as a restricted case of dialogue. The upshot for formal semantics is that individualistic considerations become more important; different individuals have different semantic options,

---

<sup>12</sup>See eg: [44]

and the agent-relativity of information states becomes a driving force behind communicative language use.

In contrast to the stance sketched above, the more philosophical notions of “Meaning”, “Truth”, and “Reference” are commonly understood to transcend the epistemic limitations of agents. These notions are, therefore, peripheral in the account of communication sought here. However, they still play an important role at the boundary of that account. Meaning, truth, and reference, as they relate to the world, perform the function of being the regulative ideals of information exchange in communication. The account of informational factors in dialogue sought here is to rely upon (a formalization of) the epistemic states of agents, and largely eschews the discussion of meaning and truth. Information is understood to be an essentially epistemic notion (contra Channel Theory, Situation Theory), which is a facet of an agent’s interpretation of communicative signals and perceptions, as opposed to an independent substance inherent in reality. Indeed, I offer a definition of information for formal semantics, in terms of change.

**Definition 1** *A signal “carries information” if it has the capacity to change an agent’s commitment to a set of epistemic possibilities.*

This definition does not restrict information to being true, and relieves it of any metaphysical burden. It also classifies signals which might have previously been thought of as absurdities as informative. Utterances which contradict the hearer’s information (resulting in changing it, rather than monotonically increasing it), are informative. Information so defined is amenable to formal treatment, and is quite adequate for the purposes of formal semantics.

Note that (epistemic) possibilities are connected to the world. They are ways that the world might be, so that information is about the world, but not in it.

### **Defining communication; the primacy of semantics**

Communication is construed as an attempt by agents to converge on a shared (or mutually acceptable) *theory* about the world. In other words, communication is modelled

as iterated theory change between agents. An advantage of this approach is that it (at least) allows a putative theory of communication, and is amenable to formal treatment and analysis. The notions of theory building and revision become central to an account of communication. Theories are syntactic structures, their contents are semantical (information states). Information *processing* is handled syntactically while the *content* of an information state or theory is given by way of model-theoretic semantics (see eg: Wansing [106]). For practical purposes, the computational tractability of theory change systems (symbol processing accounts) supercedes that of a corresponding intensional semantics (which operates over sets of possibilities). In that case, one might ask, why bother with the semantics at all? Some have argued that if a semantics is merely an alternative characterization of syntactic results (ie: is a “merely” formal semantics), then it ought to be dispensed with. What is the point, some might ask, of looking at theory change through semantical goggles? There are at least 4 replies to this critique:

(1) Semantics provides a notion of content.

(2) Syntactical systems are based on semantical intuition in the first place. Semantical notions like truth and consistency preservation provide constraints on admissible processing algorithms. Without semantical ideas as a foundation and guide, there would be no constraints on admissible syntactic systems.

(3) A semantics allows the exploration of correctness and completeness for a formal system. Semantics functions so as to define a notion of correctness for information processing accounts.

(4) Semantical exploration, intuition, and insight, can lead to the invention of new processing algorithms.

These considerations (as well as the desired “marriage” with Dynamic Semantics) motivate the provision of a semantics for Theory Change (in the next chapter), and a

primarily semantical model of communication (as opposed to a symbol processing account). Of course, the first-order theory change systems of chapter 5 provide a symbol processing approach to the phenomena modelled by their corresponding systems of dynamic semantics. It shall also be demonstrated that the semantics for theory change is not “merely” formal, but preserves intuitions about the changing commitments of agents to “inhabit” various “realms” of possibility.

## 1.4 Communication and Information

The methodology of this thesis is the application of logic in order to reveal features of information structures relevant to natural language interpretation in communication. This venture requires a clear understanding both of what counts as communication, and what an explanation in terms of “information states” actually is (or should be). Such an understanding requires a rigorous conception of information, in both its syntactic and semantical guises, as well as criteria which establish what is to count as successful communication. I have already given basic definitions of these concepts as I think they should be employed. They are now analysed in some detail. Aspects of both these foundational conceptions are noticeably lacking in current theories of formal semantics, which employ a variety of notions of “information state” and are explicitly *hearer-centered*.

### 1.4.1 Communication

The first issue which must be addressed is; What is to count as successful communication?

The proposal embodied in the following chapters is that successful communication between two agents should be defined as their increasing *convergence upon a shared theory* about the world.

This proposal begs the following questions:

- (1) What notion of *theory* is needed to support an account of communication?
- (2) What notion of *sharing* is needed for communication?
- (3) What notion of *convergence* can ensure that communication is so often successful?

To answer these in brief: a theory is taken to be a syntactic structure, simply a set of sentences of some formal language. The *content* of the theory is a semantical object; the set of possibilities which support that theory (the set of possibilities to which an agent who espouses that theory is committed). A theory becomes *shared* in a very weak sense; only in that the agents find it *mutually acceptable*. This agreement on a “shared” theory can be established through dialogue, when the *contents* of each agent’s theories are changed so as to be isomorphic. This establishing of a shared theory *structure* between individuals is taken to be the starting point for a “structural realism” about the entities referred to during communication. One cannot require that shared theories are syntactically identical, for this would be an impossibly strong requirement, but agents can hope to establish a strong structural correlation between their respective theories. This structural similarity amounts to “dialogue referents” or theoretical objects playing the same role in each agent’s theory. Of course, there is much to be said about what the notion of “the same role” amounts to, especially under change. I shall return to this issue later in the thesis.

As for *convergence* criteria, theories ought to approach each other’s version of the world as much as possible, as well as seeking consistency and truth. Ultimately, the quest for consistency and its preservation drives theory change processes, and the notion of truth is the regulative ideal to which all good theories aspire. In this regard, it is also crucial to acknowledge that (without physical confirmation) no agent can be certain that they have communicated successfully, whilst admitting that in practice they manage it all the time, and that agents can detect and repair communication *failure*. Indeed, the notions of falsification and subsequent repair of theories are central to an account of communication as iterated belief revision.

Communicative failures are grounded out through *action*. With only language to go on, agents can never have knowledge that a communicative utterance has been successful (“the problem of communication”). Action is unambiguous in a way that words cannot be. Actual physical copresence (or deixis)<sup>13</sup> is often needed to verify shared reference, whereas mere “linguistic copresence” is always open to possible falsification and sub-

---

<sup>13</sup>see Clark and Marshall [18]



sequent revision. If one takes for granted that communicants can, in the last resort, actually leave aside the resources of the spoken word and finally just point to relevant objects<sup>14</sup>, then there comes a point where action in the physical world eventually grounds out reference and predication.

This point applies to the issue of the degree of sharing which a theory of communication needs to allow its agents. An adequate account of communication needs at least to admit that agents share (a part of) the world, or more explicitly, an environment that they perceive similarly.

### 1.4.2 The problem of communication

Important issues can only be addressed here once some criteria are in place for what counts as sharing reference and communicating successfully about the same thing. A first attempt might be to say that, as a default assumption, people assume that they share referents. However, it is clear that communication often brings agents to realise that they have in fact failed to share referents (eg: talking at cross purposes, and so on), and they are then able to remedy the mistake. The fact that this type of problem actually arises needs to be accounted for by any theory of communication. There is a vital distinction between an objective “third-party” view of a conversation, and the perspective available to, and required by, the dialogue participants. The latter is to be the subject of a model of communication, being the “information states” of the individual agents.

This brings about the “problem of communication”; that of a complete lack of any guarantee that communication is occurring within the framework of a shared conceptualization, or classificatory and referential scheme. An important consequence is that “shared” theories are in fact not, as discussed above, strictly speaking available. Agents can have approximations as to what they might assume their dialogue partner’s beliefs to be. (These nested epistemic states are built up primarily by use of questions and the “test” epistemic modals, e.g. “John might be in”, “Presumably you’re on your way.”, “That must be John.”)

I am concerned with an account of information “exchange”, or rather, reconstruc-

---

<sup>14</sup>Leaving aside the Wittgensteinian objections (from [109]) that this too is a conventional linguistic act, and that verification of physical copresence is impossible for many parts of language use .

tion, in dialogue. Here, information is conceived of as any stimulus altering the range of epistemic states of a cognitive agent<sup>15</sup>. These epistemic states *can* be about the world, so that some information is about the world, or they can relate only to fictional objects. In opposition to this, some researchers (eg: Dretske [27], Barwise [10]) take information to be essentially *true*. Of course, there is no *real-world* information in fictional statements, but I seek a wider conception of information, which is not to be restricted simply to the analysis of factual declarative sentences. There are informative statements which are about non-actual states of affairs. There are even informative statements (utterances which change the information state of their processor) which are false.

Now, examine a brief statement of the problem of communication.

Dynamic semantics (or cognitive semantics generally) often supposes that although the referent that each participant has *in the (physical) world* is the same object, they may have different conceptualizations of, or ways of referring to, that object in their private (mental) worlds, or epistemic states.

So one can say that agents share the same (dialogue) referent if and only if their conceptualizations of it in that context are fixed and identical. The question now is: how can agents know if their conceptualizations are the same without physically pointing to or otherwise indicating an object? The point being that we may need to rely after all on the physical world in order to define the identity relation even if we are interested in only the one in conceptualizations<sup>16</sup>.

Take an example:

**Example 2** A: - "*That's a good book.*"

B: "*Yes, Donleavy is very witty.*"

A: "*Oh! No. I meant Peake's book!*"

Could one say here that A and B now definitely share the same (dialogue internal) referent? Agent A may be talking about the book S written by Peake whilst agent B may have in mind a different book T, also written by Peake. For them to resolve the dialogue referent, they might be finally forced to point to the actual book. Perhaps even pointing

---

<sup>15</sup>Of course, the stimulus, or signal, is *interpreted*, which is the point of doing the semantics.

<sup>16</sup>These points were put to me by Tsutomu Fujinami.

to that book may be a part of a conceptualization? Therefore, some theorists are tempted to say that dialogue partners must share the same book physically.

The question here, put forward by its critics, is whether dynamic semantics might be possible without reference to the physical world. Ultimately of course, the answer must be negative. I have already argued that reality acts as the ideal to which agent's epistemic states approximate. But the issue demands further clarification.

The problem is one of sharing reference in communication, and crucially, *knowledge* that reference is shared. So, identity criteria on referents ("talking about the same thing") seem to be decisive.

On a "cognitive" semanticist's explanation/description of events, reference relations are not independent of humans, needing to be "grasped" correctly in communication (the Fregean "magical" theory of reference), but are constructed and refined according to task and context. Thus there is not just one correct reference relation, but a family of such admissible relations. The idea is that we can explain "talking about the same thing" simply by assuming that people share the same epistemic strategies, that is that they assume that reference is shared, to begin with, and then refine this belief during the discourse (using the kinds of operations which I provide in "Revision Semantics"). They do have to pay real attention to what their perception of the real world is, of course, so that some kind of causal/direct reference account is also a crucial part of the story. Dynamic Semantics really is not denying that the world has a crucial role to play in utterance interpretation. The problem was that, in order to share reference, at some stage agents are forced to actually go and point to objects in the world. However, experience dictates that only sometimes is this necessary, even for physical objects. In addition, this argument cannot accommodate cases where agents discuss "abstract" and "non-existent" objects, for example "democracy" and "the proposed Nuclear Tests". Surely there are true statements to be made about democracy and Nuclear Tests, which are not (and cannot be) determined by going and pointing to them.

On the issue of truth and falsity, these notions are of course considerably less central to a dynamic account. The business of DS revolves around the assertions that agents are prepared to make on the basis of their (partial and changing) information about the world. From the point of view of formal semantics (rather than philosophy) "truth on the basis

of available evidence" is more of interest than any more abstract notion. Of course the world ("the way things really are") is the final arbiter of disputes; but evidence drawn from the world is used to revise an agent's conceptualizations of it. Therefore, Dynamic Semantics does not ignore the world at all, it simply gives agents' epistemic states related to it, their "theories" concerning it, considerably more weight in a semantic theory. All in all, when the focus is on communication, the epistemic states of agents are of central importance, and reality is peripheral.

## 1.5 Information

Contemporary cognitive science is keen to explain the mind as an "information processing system", and in this way commits itself both to a computational and informational model of scientific explanation. Typically, however, explanations in the established sciences take universal laws and initial conditions *causally* to force the occurrence of the event to be explained<sup>17</sup>. In that case, what explanatory force do the notions of "information" and computation have? Whilst computation in the abstract is very well understood, information does not have this virtue (Information Theory<sup>18</sup> as a branch of mathematics deals with *quantities of information* in signals subject to interference, as opposed to the *content and structure* required in NL interpretation.) The menagerie of different structures masquerading as "information states" is too populated to explore fully here, and for now it is enough to ask what a formal semantics requires of the notion. In particular, where is information to reside on the following scale of distinctions? (In order of increasingly "abstract" ontological status):

- Interpretation - theory change - dynamic semantics (Beliefs "in flux")
- Theories - static belief systems - doxastics (Justified Belief)

---

<sup>17</sup>Hempel. C. [54] "The Deductive-Nomological model of scientific explanation".

<sup>18</sup>As established by Shannon and Weaver [91]

- Knowledge - epistemics (Justified True Belief)
- Meaning - traditional semantics (Reference)
- Truth - metaphysics

From my earlier comments, it should be clear that I take “information” to reside at the first point of the above scale. The second and third points on the scale correspond to “static” information, or information that has been accepted (or believed) by an agent (their “background knowledge”). The following everyday uses of the term “information” can help to support this view. For example;

“I got some good information from Harry today”

“There was very little information in today’s news,”

and, “I’m writing to inform you of my change of address”

“You are not very well informed.”

The basic features to be extracted here are:-

1. Information is something that agents *have*.
2. Agents *generate* information via perception and interpretation.
3. Agents use language to *express* and *change* information in narrative and communication.
4. Agents typically gain more information than that which is explicitly communicated by a sender. *Context* fills in “gaps”, and *inference* creates more detailed information. Information is thus *underdetermined* by a signal.

In order to accommodate these features, the broad approach taken in this thesis agrees with much of the spirit of current work in dynamic semantics, where the epistemic state of an agent plays a crucial role in their interpretation of linguistic expressions. However, I argue (contrary to the literature on Dynamic Semantics) that *meanings*, whatever they are, cannot be identified with transformations over epistemic states.

### 1.5.1 The Informational Metaphor

The current “informational trend” in semantics, philosophy, and logic can be traced back perhaps to Dretske’s [27] “Knowledge and the Flow of Information” (1981). Since then, most active researchers have paid more than lip-service to the movement. The problem seems to be that information has been tied too strongly (for the purposes of semantics anyway) to truth and knowledge. Semanticists ought to be interested in information as it relates to belief — a much weaker notion. Frege’s concern with the “cognitive significance” of propositions (in [33]) (in particular the informativeness of identity statements) is at the root of the issues here, and ought to be taken as a starting point for a theory of information as it relates to semantics.

Along with writers such as Asher and Landman, I take information to be truth-value neutral, thus trivially permitting false information, as something close to propositional content. However, for researchers such as Perry, Israel, Dretske, Fodor, Barwise, and Seligman, if a signal carries the information that  $p$ , then  $p$  is *true*.

Crucially (and this counts against a Realist or Platonist conception of information), in addition to correct information agents can have the *wrong* information too; that is, a partial model of the world which reality won’t suffer gladly. However, one ought not to look to reality as the source of mis-information, but to the fallible inferential and representational capacities of epistemically limited agents.

There is a multitude of sub-doxastic processes going on in the brain, amongst them the processing of speech. On this account, the sub-doxastic level of processing is where information emerges. The more substantial and stable doxastic level, of interpretation and production of language, is where information emerges in concrete form. Information, then, is not solely linguistic, and is not in the world, but emerges from the relation between a cognitive agent and the world. It is a mistake to think of information as “stuff” either in the head or in the world. We speak both of people “having information” and physical objects (newspapers, signals) “containing/ conveying information” but these metaphors obscure a crucial distinction. Only agents have information, whereas certain relations between agents and objects give rise to this information. This

distinction is illustrated by the fact that 2 individuals may extract quite different information from the same physical object (e.g. a text), so that objects have the potential to inform (due to their causal relations in the world), but it is agents who *choose* either to believe or disbelieve (to adopt or to reject) the information which they generate from objects. Also, note here the crucial role of inference in determining the information conveyed by a signal. Conventions of interpretation license various inferences from certain kinds of structures.

However, it is enough for a formal semantics of communication to note that the notion of an information state as a set of possibilities has some important virtues. The first is that it is ontologically neutral. Information states as sets of possibilities, or functions mapping variables onto entities in a model, are silent about the modes of existence peculiar to the variables in its domain, or entities in its range. Intuitively, however, such a notion seems a little bizarre if one persists in confusing the model for a description of actual psychological events. It is difficult to conceive of communication as simply the mutual exchange of sets of possibilities. Indeed, something stronger is required in a suitable conception of information for communication. As was drawn out earlier, some stable definition of a *subject* is essential for agents who believe that they are talking about the same thing. This point places a requirement on a theory of communication to provide an account of persistent “informational entities” as the subjects of dialogue. A notion of “dialogue referents” is required around which disputes can revolve and be resolved. So, although the dynamic or epistemic approach to information is to be applauded, it needs strengthening by way of a solid understanding of the informational entities (pegs, discourse markers, proto-objects etc.) which are constructed and manipulated during communication. This is partly my objective in chapter 6.

To conclude this discussion; formal semantics does not require a realist and truth-laden notion of information. In fact, if all information is true, as the philosophers require, then most of the “information” dealt with in formal semantics is not information at all.

### 1.5.2 Information change: Eliminativism versus Constructivism

Another point of departure in a debate concerning a suitable notion of information for communication concerns the conditions under which we are to say that information has increased or grown. The approach of [101], [46], and [21] (amongst others) has been that information increases at the expense of possibilities. That is, as Sherlock Holmes would have argued, if we eliminate all possibilities except one, then that possibility, however remote, must be the truth. To restrict a theory of information to this insight is a serious mistake, for such an approach entirely neglects the sense of information growth in which possibilities are constructed, and avenues for investigation opened up<sup>19</sup>. Dekker's Eliminative Dynamic Predicate Logic<sup>20</sup> allows a very minimal "constructive" aspect of information growth; the introduction of new variables. I advocate a significant extension of this approach, with an emphasis on revisions. In addition to the elimination of possibilities a more important role must be allowed for the construction of new avenues of possibility, or the re-investigation of previously discarded ones. This is more than just the introduction of new discourse referents, and demands an entire theory of information revision.

The appropriate formalisms construe information states as combinations of sets of variable assignments and sets of possible worlds in a first-order model-theoretic semantics. Given the above discussion, such a formal characterization might at once appear (perhaps like most work in formal semantics) both technical and crude, but by the end of this chapter it should be clear that such an approach is useful in important respects, especially in comparison with the stances taken by other theories. In particular, a metaphysical conception of information as having agent independent existence (eg: the infons of Situation Semantics [10]) should be avoided, while on the other hand, a construal of information in terms of an agent's mental representations (in any strong sense of "representation") must be dismissed. While the discussion above has, I hope, sketched the horns of this dilemma, the following discussion points to an acceptable escape route.

---

<sup>19</sup>If Holmes has never even heard of Dr. Moriarty, he cannot eliminate him from his enquiries.

<sup>20</sup>This is somewhat of a misnomer, as the next chapter reveals.



## 1.6 Mathematical Models in Formal Semantics

Semantics as part of a “Cognitive Science” is often caught precariously between the concerns of logic and philosophy on the one hand, and the demands of an empirical psychological account of linguistic behaviour on the other. Psychological evidence repeatedly supports the claim that humans do not behave logically, while at the same time cognitive science seeks logic-based information processing mechanisms behind reasoning and linguistic activity. In order to avoid this seeming conflict, informational and logical factors must be distinguished. I provide just such a distinction in the following chapter, using the Multiple Extensions Problem (from the literature on Theory Change) to motivate rationality constraints on change as informational constraints on interpretation. I eschew descriptive psychological considerations largely in favour of the normative ideal of an ideal rational agent. The main body of the thesis, then, employs mathematical modelling as its primary methodology. But how can a mathematical model of interpretation, a “formal semantics”, be seen as scientifically explanatory?

In part, the following is a sketch of the explanatory framework of a Formal Semantics which properly pays its dues to communication. The sketch builds on the same foundations as most current work; the model-theory of language presented in Wittgenstein’s *Tractatus* [108]. However, it modifies them in a way that takes seriously epistemic states<sup>21</sup> emerging from communicative language use, as the subject matter of a scientific semantics.

Wittgenstein’s conception in the *Tractatus* was (in part) to provide a mathematical model of language, along the lines of the mathematics of mechanics. For instance, as Hertz wrote of puzzles about the “nature” of force in mechanics:

“Why is it that people never in this way ask what is the nature of gold, or what is the nature of velocity? Is the nature of gold better known to us than that of force? Can we by our conceptions, by our words, completely represent the nature of any thing? Certainly not. I fancy the difference must lie in this. With the terms “velocity” and “gold” we connect a large number of relations to other terms; and between all these relations we find no contradictions which offend us. We are therefore satisfied and ask no further questions. But we have accumulated around the terms “force” and “electricity” more relations than can

---

<sup>21</sup>A terminological problem here is that, strictly speaking, these states are not epistemic (since knowledge carries the presumption of truth), but doxastic.

be completely reconciled amongst themselves. We have an obscure feeling of this and want to have things cleared up. Our confused wish finds expression in the confused questions as to the nature of force and electricity. But the answer which we want is not really an answer to this question. It is not by finding out more and fresh relations and connections that it can be answered; but by removing the contradictions existing between those already known, and thus perhaps by reducing their number. When these painful contradictions are removed, the question as to the nature of force will not have been answered; but our minds, no longer vexed, will cease to ask illegitimate questions.”  
(Hertz, [55], quoted in [59])

The mathematical models in mechanics are completely determined by their formal properties, and as such their limits of applicability are specified, as it were, from inside the system itself. The parallel idea of explaining from within a mathematical system the limits of a language thus eschewed cultural/psychological origins of linguistic “concepts”, concentrating instead on the formal structure of relations between propositions. Just as spurious questions concerning the reality of “lines of force” in mechanics were to be dissolved by regarding them as relations in a model which is practically applicable, so confusions surrounding linguistic items could be bypassed through consideration of the formal function of their corresponding model-theoretic counterparts. The logical formalisms of Frege and Russell were then to be used to construct a mathematically precise modelling tool formally capable of *showing* how internal structures of linguistic expressions *represent* corresponding structures of objects in the world, and their modes of combination into facts.

This was then, broadly, the “picture theory” of language, the intention of which was to elucidate the workings of a language by way of modelling the structure of its terms as they are used. The theory was to provide an understanding of the nature of language from within a mathematically rigorous system, therefore avoiding the circularity of stating explicitly a theory *about* language. Crucially, properties of a language are to be understood via a mathematical model, the structure of which demonstrates, rather than states, the nature of that language. Terms are to be understood by way of their functions in models.

But how does a model “show” us structure, and what is its relation to the language in question, and to the world? Methodological issues concerning the interpretation of model structures become central to the Tractarian position, and indeed the enterprise of

formal semantics as a whole. Wittgenstein, of course, held that the language-world connection could not be stated in the model-theory, only “shown” by it. Thus the *Tractatus* left the language-world connection to itself, as something ineffable, leaving only models as a kind of metaphor for language. Logical relations hold only within a symbolism, so that formal models can never be anchored (as a matter of logic) to the world that they describe, but only by convention.

A model exhibits its structure purely by virtue of the formally specified relations holding over its elements. This allows theorists to see, say, “lines of force” not as entities in the model, but simply as instances of a certain kind of relation over those entities. “Lines of force” then become a mathematical, functional, or structural property of model-theoretic entities, and similarly for the logical connectives of a language. The model, by its formal structure, is supposed to show relations between entities, so that beyond describing configurations of entities, the model can assert nothing about them. But now a number of other questions arise.

- (1) The manner in which the model is to forge a connection to objects in the world (how the model-theoretic entities relate to the physical ones. i.e. The ontological status of entities and relations in the model)
- (2) How the terms of the language in question are to be interpreted in the mathematical formalism of the model-theory.

The model-world connection (1) employs the, notoriously difficult, notion of “representation”, along with a correspondence theory of truth. As Wittgenstein used it, the notion of representation had a distinctly *constructive* aspect, so that agents consciously *build up* logical structures which represent the world. “Representations,” then, are not re-runs of sensory experiences, but something quite different. The notion, from [61], of a Discourse Representation Structure (via the “construction algorithm”) appears to have precisely this feature. On the other hand many interpretations of Wittgenstein invite a decidedly pictorial mental image or snapshot conception of “representation.” The partial models of Situation Semantics seem to follow this line more closely. In any case, models are to be judged true or false depending on their accordance with reality. If the structure of entities in the model is mirrored by a corresponding configuration of objects

in the world then the model is declared true. This correspondence theory of truth thus demands an isomorphism between language and the world, and the concern is then how such a strong connection between model and world is to be guaranteed. Not only must there be a method which connects model-theoretic entities with their real-world counterparts, but there is also to be a connection between mathematical relations over the symbols in the model and relations between the real objects. To establish the truth of a certain model one is to determine whether or not the real world objects stand in the modelled relations to each other. How then does this stance claim to avoid such issues as the reality of lines of force? If they are relations in a model and nothing more, then so be it, but once one imposes a correspondence test for model-theoretic truth, one is forced to look for such relations in the world, thus putting the initial question once more. At this point, it seems that a route away from this isomorphism test for truth is needed if the Tractarian picture is to be salvaged. To this end, I suggest the adoption of a "structural realism" regarding formal semantics. The issue revolves around the interpretation of model-theoretic relations. As argued above, the requirement that they point through the model to relations in the world begs precisely the question that the Tractarian model-theory of language set out to avoid.

What can, and should, be argued here, is that relations constructed in the model-theory of language are not interpreted as being in the world itself, but in the categorisations and explanations which agents construct in order to deal with perceived correspondences in the world (ie: relations are in an agent's information). Thus, models should be interpreted as models of epistemic or doxastic structures, rather than parts of the world. This argument certainly needs some elaboration. A shift in interpretation of the relations in a model, and their truth conditions, is required here. Recall that the intention behind Wittgenstein's notion of representation was that models are built by agents as they interpret language. Propositions show certain relations between named objects. Once these stated relations are regarded as statements about an agent's perceptions and interpretations of the world, rather than as directly corresponding to reality, the models which agents construct are not directly related to the world, but of the epistemic situations presented to them by other agents and their own senses. Thus a relation of "force" communicated by one individual to another vis a vis two objects is, first and foremost, a de-

scription of the first individual's epistemic state, concerning their perceived correspondences and regularities in the world. Thus, once theorists move into the communicative stance on language use (rather than the classical world-language centered position) they free model-theoretic relations from the metaphysical burden placed upon them by a correspondence theory of truth. But what theory of truth remains? Linguistic acts partly serve to articulate an agent's epistemic "possibility space", based on their perception of correlations. What is it then to judge a statement as true? Presumably, agents label a statement "true" if the logical-epistemic space that it prescribes for the objects that they take it to refer to are compatible with their own. *Judgements* of truth, truth on the basis of available evidence, rather than apprehension of pure metaphysical truth, are central to language seen primarily as a vehicle for communication about perceived regularities. The epistemic analysis of truth is more important to formal semantics than the traditional metaphysical stance. The truth conditions (or rather, judgement acceptance conditions) for a proposition now become a matter of compatibility between models built up by interpreting the proposition and the world-model of the interpreter. This is the semantical version of what has been called "Belief Revision". This revised Tractarian position regards the important part of language as laying models against other models for comparison, rather than, as Wittgenstein had it, a model "laid against reality like a measure."<sup>22</sup> In their own ways, both DRT and Situation Semantics fall into this Tractarian trap (DRT via embedding conditions and Situation Theory via a realist approach to relations). The methodological lesson should be that entities and relations in a model need distinguishing in an important ontological sense. Whilst the entities in a model correspond to informational approximations to individuated objects in the world, the relations stated between them are a feature of the communicant's judgements, beliefs, and perceptual capabilities. Thus entities have a (slightly) stronger ontological position than do relations. An approach, such as Situation Semantics, which fails to appreciate this distinction entirely (by regarding the world as composed of relations and objects) does not do justice to language use as observed in communication.

The final question here concerns the language-model connection. As with the world-model isomorphism, we are entitled to ask what guarantee we have regarding the math-

---

<sup>22</sup>Tractatus 2.1512

ematical model's applicability to our actual descriptive language and practice. Montague Semantics and DRT provide explicit translation rules from syntactic structures to model structures. In DRT, the assembly of DRSs even seems to appease the constructivist aspect of Wittgenstein's "representations." Ideally, what is sought is an incremental procedure which takes each word in a linguistic expression to construct or modify a logical model of the entities and relations between them posited by that expression. The functional view of "meaning" adopted in modern-day information-based semantics takes this view of sentence and discourse processing, but approaches are divided on the issues of representation and compositionality. Given the discussion above, the "epistemic" approach taken in Dynamic Semantics is clearly to be advocated.

### 1.6.1 An epistemic/doxastic semantics

Semantics, as part of a cognitive *science*, is presumably an explanatory venture, affording the theorist some degree of predictive power regarding the utterances of agents (in a given context). Cognitive science is often described as the study of the mind as an information processing system. If this is to be taken at all seriously, paradigms are required which really do attempt to give explanatory information processing accounts which are noticeably lacking in the semantics of dialogue. I have attempted to argue that concentration on dynamics and epistemics, or more precisely, doxastics, is the most promising way to approach these issues.

Other approaches, such as Channel Theory, are forays in mathematical metaphysics, because there information and its flow are considered to be agent-independent parts of reality. Instead, the project of formal semantics is a mathematical epistemology, because information is generated by cognitive agents. I have said earlier that information is any signal which (potentially) expands or contracts an agent's set of epistemic possibilities (their "commitment slate"). Meanings, by way of contrast, are tied to reference and truth. So "Meanings just ain't in the head, and information just ain't in the world." This has the further consequence that all the modalities in which are of interest in a semantics of communication are simply the epistemic modalities.

### A note on the modals

The standard position currently adopted in formal semantics is the acceptance of the Kripkean distinction between metaphysical and epistemic necessity. Instead, I should like to follow Quine in his statement:

“Hume was right, I hold, in discrediting metaphysical necessity. . . . *Sub specie aeternitatis* there is no necessity and no contingency; all truth is on a par.”

(Quine, [81], p.140)

The real question here is whether or not the Kripkean distinction is of any use in an account of communication as iterated theory change, or state revision. A standard example in dynamic semantics of a “metaphysical modal” might be along the lines of, “He survived. He might have died.”

The dynamic account of the meaning of “might”, derived from Veltman [100], cannot handle such a discourse, leading to the adoption of the epistemic/metaphysical modal distinction. Groenendijk, Stokhof, and Veltman [47] also claim that information states can be enhanced with a set of “metaphysically possible worlds” for such metaphysical modals to range over.

However, unless these “metaphysical possibilities” are still relative to an agent’s epistemic state<sup>23</sup> this account does not square at all with the position I take throughout the thesis; that information states are *epistemic* objects, and that utterances express the epistemic state of a speaker and request hearers to revise their own theories. My claim is that utterances such as “He survived. He might have died,” are still *epistemic* claims by a speaker, albeit of a non-eliminative nature. When I speak of “counterfactual” or “counterdoxastic” modals I shall in fact be referring to those modals whose operation is not *eliminative* on epistemic/doxastic states, but “constructive”.

## 1.7 A Revision Semantics for communication

Given the above discussion, it seems clear that the integration of Theory Change systems with Dynamic Semantics is a viable project, which is both technically and theoret-

<sup>23</sup>In which case they are hardly metaphysical.

ically interesting. If one revises the dynamic account of meaning, adopts a structural realism concerning individuals in model-theoretic semantics, recognises the importance of change in communication, and accepts the doxastic conception of information and the primacy of semantics, then one is led to the project which the rest of the thesis attempts to carry out.

From the perspective of Artificial Intelligence research, the dynamic semanticist's notion of an update is useful in modelling the change in a knowledge-base due to a change in the world that it describes. On the other hand, dynamic semantics requires agents which can process conflicting information about the world.

Updates in communication are triggered by simple assertions which do not conflict with an agent's current information, whereas *revisions* are the result of processing and incorporating fresh and possibly inconsistent information.

In some of the literature in the field of Dynamic Semantics there is an explicit recognition that the notion of an update alone is not sufficient for a full semantical theory, and that some notion of "belief revision" will have to be worked out. But there, so far, is where the issue has been left.

"If the speaker attempts to exchange the information that  $\phi$ , and the hearer has information to the contrary, then the exchange is simply taken to come to a halt. For the exchange to proceed in such a situation, a higher order discussion may be required . . . as well as some method of belief revision. Since . . . belief revision fall[s] beyond the scope of the present undertaking, we just have to settle for expelling the occurrence of inconsistency of information."

(Paul Dekker, [21], pages 211-2.)

An integration of revisions and updates is therefore to be desired both for the AI and Formal Semantics communities. As far as dynamic semantics is concerned, the issue is that while updates are always concerned with information growth, a dual notion of revision, or repair of information is required. Theory change deals with rational agents which are required to dispense with certain current information in order to accommodate new information. This process can even occur just hypothetically, during a dialogue, where an agent "suspends disbelief" for the sake of argument. Such revisions have not yet been taken seriously enough in the formal semantics of dialogue, but there is a large and technically impressive "belief revision" literature which, although neither semantically nor linguistically oriented, is available for integration into such a new ac-



count.

### **1.7.1 Conclusion**

The chapter raises foundational questions for the current “informational” vogue in formal semantics, and cognitive science in general, and similar issues for a theory of communication in that vein. In particular, an important problem for any theory of communication was raised (the issue of “common ground” or “shared” information states), whilst a rigorous conception of “information” as an explanatory device in formal semantics was developed. A formal definition of communication was also provided. Definitions of information and communication (in terms of doxastic states and their transformations) were given, and argued to be suitable for a formal semantics regarded as a scientific, as opposed to a purely philosophical, venture. I argued for a revision of the dynamic account of meaning, and a “structural realism” about model-theoretic semantics. I argued that the integration of (ultimately First-Order) Theory Change methods with systems of Dynamic Semantics could lead to an empirically adequate semantics of dialogues, and a putative theory of communication.

The remainder of the thesis carries out some of the formal work that this chapter has, in part, motivated.

## Chapter 2

# Formal Preliminaries in Dynamic Semantics and Theory Change

### 2.1 Outline

My intention in this chapter is to give at once a brief introduction to technicalities in the fields of Dynamic Semantics and recent work in Theory Change (or “Belief Revision”), and to develop some issues arising from their application in a putative theory of communication. The most important concepts of the thesis (information and change) are formalized. In addition, along with the recognition that information is often incomplete and inconsistent, criteria for any satisfactory ‘informational logic’ are developed, and some non-classical logical ideas that will be useful in the applications of later chapters are discussed.

### 2.2 Information in communication

Specific conceptions of information and communication were argued for in the preceding chapter. Given that discussion, now simply note that the ideas about ‘information growth’ which are currently employed in Dynamic Semantics (DS) are unsatisfactory, insofar as they fail to properly capture many kinds of informative statements made during communicative dialogues (these linguistic constructions are properly explored in

chapter 6). Specifically, an account for *repair* during communication breakdown is offered, and systems of *Revision Semantics* (RS), which deal with corrections in dialogue, are developed. A particular concern is the extension of the coverage of formal semantic theories to an analysis of *dialogue* rather than discourse. The techniques of Revision Semantics enable the analysis of dialogue contributions for which discourse-based theories can offer no explanation. It is in this setting, then, that a Revision Semantics is seen to be central to an account of information states in communication.

### 2.2.1 What is information growth?

Researchers in formal semantics have embarked upon an account of the information gained and lost by an agent through the interpretation of utterances. Acquired information is not just the explicit surface content of an utterance, but is also given in the inferences licensed by believing that utterance<sup>1</sup>. Thus the information gained through utterance interpretation is highly context-sensitive; being relative to the inferential capacities of the agent (its “background knowledge”), and to the information provided in the foregoing dialogue or discourse. Earlier it was argued that formerly, philosophers have been confused about a theoretically useful notion of information (making it overly objective) either because they fail to properly take into account the element of subjectivity required in linguistic interpretation, or because they are concerned with a different problem altogether. Information has been tied to truth by philosophers taking a ‘God’s eye’ view of the content of discourse. I define information content via *change* in epistemic state for an agent. Important consequences of this are:

- (1) that an utterance can be informative for an agent even if it is false, and
- (2) that utterance interpretation can be informative *without* elimination of possibilities for an agent.

Ever since Kripke’s semantics for Intuitionistic Logic [66] there have been increasingly more elaborate attempts to capture formally the intuitive idea of increasing information. The technical device for capturing information states is most commonly an un-

---

<sup>1</sup>cf: Montague’s goal of capturing all valid entailments.

structured collection (set) of possibilities. Every time information is gained it precludes some of these possibilities. Information produces (provisional) commitment where there once was ambivalence.

The *eliminative stance* on knowledge growth exploits the idea that information grows via the elimination of possibilities. These sets of possibilities grow smaller under information growth, ultimately until only one possibility under consideration (“the real world”) is left. Information growth thus proceeds by a “Sherlock Holmes” method of elimination of possible models.

*Intuitionistic logics* describe the process of gaining information (in this case mathematical knowledge) as movement outward along the branches of a tree-like structure (a partially ordered set of points), where each branch point is associated with a certain amount of established (and thence irrefutable) information. Points further along branches “hold” more information than lower points in the partial order. Gaining information is thus a process of travelling ever outward along (infinitely long?) branches of this “tree of knowledge.”

In *Relevant Logics* the idea is very similar to the intuitionistic case. Points in the frame semantics for relevant logics are referred to as “information states, theories, or possible situations”<sup>2</sup> which have the same natural ordering as intuitionistic frames.

In the following chapters it is shown how the intuitionistic (and therefore also the relevant) way of thinking about information states leads to an overly naive approach to information change; in particular the approaches fail to address the Multiple Extensions Problem for information loss. Of course, such ‘logics’ were never intended to tackle phenomena which require non-monotonicity.

In addition, there is an alternative which needs to be explored here, for it is closer to intuitions about the information carried in statements in dialogue: the *constructive stance*. The intuitions behind Discourse Representation Theory (Kamp 1981 [61]) seem to be similar. Each utterance interpretation is taken to construct an entity or relation in

---

<sup>2</sup>see Greg Restall in his review of Entailment vol. 2

one model, until a detailed “picture” of some situation is achieved<sup>3</sup>.

However, it is worth asking whether these two conceptions of information growth are really in conflict, as they are commonly thought to be, and indeed, whether these are the only ways in which information can be said to grow. One view might be that the eliminative and constructive construals of information growth are really just two sides of the same coin. Eliminative considerations become paramount when we consider information semantically, while a syntactic focus (as in AGM belief revision and DRT) results in the constructive view. But such a view would be to mistake the force of the argument for non-eliminativity as information growth. For our purposes it is information *change* rather than growth which should be the focus of a theory of communication. In particular there is an important sense in which a dialogue partner’s insistence that he in fact knows nothing about X actually results in a form of information growth on the hearer’s part; in at least two possible ways. Firstly, the hearer’s information about the speaker grows; he knows that she is ignorant about X. Secondly the hearer may be informed just in terms of the subject matter of X; he may never have heard anything at all about X before.

### 2.2.2 Information: Syntax and Semantics

I shall write about “theories” (syntactical objects; sets of sentences) as a way of expressing doxastic states and “information states” (semantical objects; sets of models and assignments) as an alternative name for the things so expressed (contents of theories). In harmony with this syntax/semantics or processing/content distinction<sup>4</sup> I explore rational systems of *Theory Change* and of *Dynamic Semantics*. Theory change (TC) systems deal with constraints upon the modifications that theories should go through in the light of new, and potentially conflicting, information about the world. Systems of Dynamic Semantics (DS) attempt to describe the *meanings* of expressions in Natural Languages as changes in the information states of their users.

---

<sup>3</sup>Jaap van der Does clearly has this in mind in his “Dynamics of Sophisticated Laziness” [95]. Also, Wittgenstein’s notion of “representations” in [108] is distinctly constructive.

<sup>4</sup>See Wansing [106] for more on this point.

Theory Change systems, since they deal with finitely describable structures (closed sets of sentences), are computationally much easier to deal with than the infinities of “possibilities” dealt with in Dynamic Semantics. Despite this computational advantage however, it is the semantics that is primary in this account, in the sense that it provides the intuitions from which different Theory Change systems develop. A central goal of the thesis is DYANA’s<sup>5</sup> first objective;

“to determine those aspects of the structure of information states which are relevant to Natural Language interpretation.”

In the following chapters the structure of such information states is explored with special reference to dialogue. Specific attention is given to cases where utterance interpretations conflict either with the hearer’s current beliefs, or with information previously conveyed by the speaker. In such cases, information states must undergo repair and correction.

### 2.3 Properties of a Dynamic Semantics

Common to all systems of Dynamic Semantics (Discourse Representation Theory<sup>6</sup>, Update Semantics<sup>7</sup>, Dynamic Predicate Logic<sup>8</sup> and their progeny) is the following philosophical claim, which serves to unite these otherwise often disparate approaches;

“The meaning of a sentence is the change that it induces in its hearer’s information state.”

The philosophical considerations of the preceding chapter demand some modification of this slogan, but for now it serves to introduce the cornerstones of the dynamic approach: that agents ‘have’ *information states* (sets of possibilities  $s \in S$ , where  $S$  is the powerset of the set of all possibilities) and that these states *change* under utterance inter-

<sup>5</sup>The ESPRIT basic research project 6852

<sup>6</sup>DRT, Kamp [61]

<sup>7</sup>US, Veltman, [101]

<sup>8</sup>DPL, Groenendijk and Stokhof,[46]

pretation. Information states are generally taken to be *sets of possibilities*<sup>9</sup> which an agent has under consideration at any point in the interpretation of a dialogue or discourse. Following Veltman, the basic mode of change in an information state is the *update*, written  $[\cdot]$ , which is to correspond to the *addition* of information to a state. Write  $s[\phi]$  to denote the state resulting from the update of  $s$  by a formula  $\phi$ .

One objective of this thesis is to provide truly dynamic systems, in the sense that the usual notion of update does not capture the notion of *change* in information state sufficiently for the purposes of an empirically accurate formal semantics of Natural Languages. Writers such as Dekker<sup>10</sup>, Asher and Veltman acknowledge that such an account is required, but do not tackle the issue in their published works. To remedy this deficiency a new “downdate” operator  $]\cdot]$  is supplied, the function of which is to *remove* specific information from a state. I shall describe as a *Revision Semantics* any dynamic system which accommodates loss of information (“downdates”), as well as its growth under updating, and the consistency-preserving update operation of revision.

The following properties of dynamic systems are useful in classifying the different approaches<sup>11</sup>. Here, they are surveyed with reference to a Revision Semantics.

1. ELIMINATIVITY is the property of a dynamic semantics whereby interpretation can *only* eliminate possibilities. The result of interpreting a formula in a state  $s$  is always a subset of  $s$ .

$[\phi]$  is eliminative iff  $s[\phi] \subseteq s$

Interpretation thus guarantees update of information in the sense that possibilities have been narrowed down. Purely eliminative updates are *monotonic*, in that they reflect the dynamics of simple addition of information. Obviously with revisions in mind, a strictly eliminative semantics is undesirable, for some expressions retract pre-

---

<sup>9</sup>Possible worlds, situations, possible variable assignments, possible ‘individuals’, or combinations of these, depending on the particular system.

<sup>10</sup>See for example [21] page 157.

<sup>11</sup>For a full account here, I refer the reader to [46], and [21].

viously established information, and actually open up epistemic possibilities. Thus, while retaining many of the eliminative operations familiar from Update Semantics, a dynamic semantics with revisions will also employ some non-eliminative processes.

2. DISTRIBUTIVITY is the property whereby (when computing the update of a state  $s$  with a formula  $\phi$ ), only properties of individual elements in  $s$  are relevant, and not global properties of the state as a whole.

$\llbracket \phi \rrbracket$  is distributive iff  $s \llbracket \phi \rrbracket = \bigcup_{i \in s} \{i\} \llbracket \phi \rrbracket$ .

Update semantics is non-distributive because its epistemic modal “might” tests global properties of an information state (‘maybe  $p$ ’ tests the entire state  $s$  for non-absurdity under update by  $p$ , which need not hold of singleton subsets of  $s$  to be successful<sup>12</sup>). Retaining this account of the epistemic modals also would require non-distributivity for a Revision Semantics<sup>13</sup>.

3. NON-COMMUTATIVE CONJUNCTION;

$\phi \wedge \psi \neq \psi \wedge \phi$

This property of conjunction is due in particular to the dynamic treatment of the epistemic modals, specifically the modal “might” (or “maybe”), denoted  $\diamond$ . So that: “It might be John. It’s Mary.” does not have the same meaning as “It’s Mary. It might be John.” The latter is nonsensical. Formally then;  $\diamond p \wedge \neg p \not\equiv \neg p \wedge \diamond p$

Note that a text such as :

“It’s John. It might have been Mary.”

is not absurd. Past tensed modals are often argued to be ‘metaphysical’, rather than epistemic, since they reflect the way the world may or may not have been, rather than properties of an agent’s epistemic state. However, this Kripkean distinction between epistemic and metaphysical modalities, ought not to be imported into formal semantics (see eg:[47]). Counterfactual modals still express epistemic properties, and no set of ‘metaphysically possible worlds’ is needed in order to treat such expressions. A Revision Semantics allows a treatment of such counterfactual modals as testing entailments of a state under revision.

<sup>12</sup>See [21] p. 155 for a full account of the non-distributivity of US.

<sup>13</sup>See Jan van Eijck and G. Cepparello [97], who argue that non-distributivity is the essence of a good analysis of “maybe.”



## 4. NON-ELIMINATIVE MODALS

One requirement common to all systems is that interpretation of epistemic modals should never actually diminish the set of possibilities under consideration. Note that, in an eliminative semantics, this requirement means that interpretation of modals is entirely uninformative, because under update by modals information states either remain static or become absurd. A Revision Semantics ought not to tamper with this property of US.

## 5. DYNAMIC ENTAILMENT.

The treatment of anaphora in first-order dynamic semantics is exemplified by the following theorem of DPL:

“A man was mad.” “So, he was mad.”

$$\exists x Mx \models Mx$$

Syntactically free variables (standing for pronouns) are semantically bound.

Of the above properties, the following are adopted as desiderata for a Revision Semantics: non-commutative conjunction, dynamic entailment, and the non-eliminativity of the modals. Distributivity would restrict a treatment of modals, and eliminativity would make a treatment of contraction and revision impossible (see the impossibility theorem of chapter 3).

## 2.4 Two dynamic traditions: DS and TC

There follows a brief history of the development of the research programme of Dynamic Semantics. Then the next section is devoted to a similar exposition regarding dynamics over *syntactic* structures; research in Theory Change or “Belief Revision.”

### DRT, US, DPL, and EDPL

The seminal insights of the dynamic approach to semantics were provided by Discourse Representation Theory (DRT) [61] and Irene Heim’s “File Change Semantics” (FCS) [53], which introduced “updating” in its earliest form and the idea of *discourse referents* or “indefinite objects” [63] which enabled a treatment of pronominal anaphora. These projects

formulated something close to the current dynamic account of meaning. Kamp's DRT [61] made explicit use of the notion of "mental representation"; insofar as the Discourse Representation Structures generated in DRT were taken to be mental structures representing the content of utterances via an indispensable level of "discourse representation". The theory made empirical progress in its treatment of the classic 'donkey' sentences (noted by Geach<sup>14</sup>), but required a level of discourse representation to account for anaphoric relationships.

The next major development in the dynamic tradition was Veltman's Update Semantics (US, [101]), which deals with information growth via elimination of the possible worlds constituting information states, and provides a novel treatment of the epistemic modal operator "*might*". Update Semantics is a propositional dynamic system, which does not accommodate DRT's account of anaphora. US is also non-representational, in the sense that its information structures are about the world, rather than an indispensable level of mental entities (DRT's "discourse referents"). Indeed, at this point it is important to distinguish the two varieties of information employed in the dynamic tradition. Whilst US interpretation has to do with an agent learning about the world, other systems (eg: DPL) focus only on information about possible anaphoric relationships within a discourse. Call the former *discourse external* information, and the latter *discourse internal information*.

Groenendijk and Stokhof's Dynamic Predicate Logic (DPL 1991)[46], in which information states were considered to be sets of variable assignments to be employed in the treatment of intersentential anaphora, neglects the issue of discourse external information. DPL accounted for the same data as Kamp's DRT, but was couched in "*non-representational*" terms; it demonstrated how anaphoric relations could be treated without a separate level of discourse representation. However, an argument in the preceding chapter concluded that the debate over the representational commitments of various semantical theories has often confused a technical issue for a more substantive philosophical one.

---

<sup>14</sup>eg: "If a farmer owns a donkey he beats it."

Both US and DPL contain a non-commutative conjunction. Both possess a deduction theorem. However, DPL is non-eliminative, whilst US is non-distributive. Both allow a notion of updating information in discourse, but while US interpretation is concerned with updating information about the external world, DPL interpretation focuses on the passing along of possible variable assignments for use in the interpretation of subsequent anaphoric expressions. DPL also failed to be faithful to its ancestors (Heim's [53] and Kamp's [61]) by *losing the idea of a genuine domain of discourse referents*<sup>15</sup>.

Paul Dekker (in "Transsentential Meditations") [21] set about integrating these two formal approaches to discourse interpretation, in the system dubbed "Eliminative Dynamic Predicate Logic" (EDPL), which was to remove DPL's non-eliminativity, and adapt its logic to that underlying US (thus modelling both discourse-internal and discourse-external information update). Information in EDPL is about the possible values of **growing** sets of variables, and is encoded by way of **partial** variable assignments, so that there are two kinds of dynamics within information states, and so two kinds of update. An EDPL information state is a domain of variables together with sets of variable assignments to them, so information is gained either by the reduction of partiality (further specifying possible variable assignments), or by extension of the domain of those partial variable assignments (finding new objects introduced in the discourse.) And so EDPL was born. However, its notion of update, as Dekker (quoted below) explicitly mentions, involves only the *addition* of information in discourse, and not its *revision* in dialogue.

"The system does not allow a speaker to convey information which conflicts with information the hearer has. In order to solve such disagreement, some kind of belief revision seems to be required and revision falls beyond the scope of the present undertaking." ([21] p. 210)

In addition to this omission, it is difficult to square the motivations behind Dekker's eliminativity property with the fact that the introduction of new discourse referents in EDPL in fact opens up new epistemic possibilities for an agent. In other words, EDPL is not, strictly speaking, eliminative. For interpretation of existential statements can extend an information state.

Later on in the thesis a revisable DPL-like system is provided; one where DPL informa-

---

<sup>15</sup>See Dekker in [22] on the "dumping" of subjects.

tion states can be revised. Later still (also in chapter 5), systems which deal both with changing discourse internal and external information are supplied.

### The Changing Theory of Theory Change

As mentioned earlier, philosophical insights concerning an account of theory change have come from many areas: Pragmatics (William James [58]), Philosophy of Science (Hempel [54]), and Philosophy of Law (Makinson). Belief Revision has also been used in an account of human communication (Galliers [39]), and is part of a large movement in “non-monotonic logics”.

Formal work in Theory Change started out as “Belief Revision” research, with the following general question. How ought an agent to rationally change its belief system while accommodating new (possibly contradictory) information? In any answer to this question the main delicacy to be dealt with is the **Multiple Extensions Problem** (MEP)<sup>16</sup>. For a simple instance of the MEP consider the following theory:

$$T = \{p \rightarrow q, p, q\}.$$

Imagine that an agent needs to remove  $q$  from this theory (perhaps because the original evidence for  $q$  is now in doubt). The agent cannot simply remove  $q$  from  $T$ , for under logical closure it returns (via Modus Ponens). So either one of  $p$  and  $p \rightarrow q$  must also be removed. But how is a rational agent to choose which of these sentences to dispense with? Alchourron, Gärdenfors and Makinson (AGM)[1] developed a series of formal answers to this problem in the case of propositional logic. Various assumptions of the AGM approach have been disputed, and led to the investigation of different types of changes (eg: Fuhrmann [35]) over different structures (eg: Ryan [88]). In the psychological literature, work has been carried out on the actual behaviour of individuals undergoing belief change (Elio and Pelletier [28], Harman [51]). In Artificial Intelligence, TC has been part of a whole movement. Systems of non-monotonic inference, defeasible inference, truth maintenance, and default reasoning, each closely related to TC, have been of much interest to researchers in AI.

For the purposes of this thesis, it is important to recognise that TC systems have, un-

---

<sup>16</sup>First noted by Rescher in 1964, [82]

til now, only been supplied for *classically closed* and *consistent* theories in *propositional* logic. In chapter 5, ideas from the theory of “multiple contraction” (Fuhrmann [35]) are employed in an account of *first-order* theory change, and elsewhere, ideas from paraconsistent logic (see Restall and Slaney, [84]) are used to relax the assumption of consistency. Taken together, these moves enable an account of *autonomous* theory change; that is, where agents choose (rationally) whether or not to accept new information, rather than automatically accommodating it (as the AGM approach describes.)

### 2.4.1 Integrating the Accounts

This remainder of the thesis is an attempt to marry the insights of formal Belief Revision/Theory Change systems with the dynamic semantics of US, DPL, and Modal DPL, and explore the outcome. Theoretically interesting results concern the extension of the AGM approach to theories defined over more complex languages than propositional logic and the invention of a semantical approach to theory change, leading to a variety of completeness results. Such systems are useful for a number of reasons, the most pressing one being the development of an account of information exchange in a *formal theory of communication*.

One of the main issues to be dealt with will be how contradictions ought to be handled in a robust information preserving predicate logic. Obviously, insights from paraconsistent logics will be useful in this regard.

The systems developed will address anaphor binding revision and information state revisions for theories expressed in both propositional and first-order logics. There follows a preview of the problems that such an account will treat.

## 2.5 Problems for standard dynamic semantics

In this section a few specific problems for standard dynamic semantics (by which I mean non-revisable DS) are pointed out. The addition of revisions to standard DS offers a treatment of these problems. If DS is to extend its data coverage (increase its empirical adequacy), then these issues need to be addressed.

**Data coverage: ‘dirty dialogues’**

In general, a class of problems emerge from the fact that the data coverage of standard systems is restricted to a set of “clean” dialogues. “Clean” dialogues are those which do not involve conflict, correction, or repair on the part of either speaker or hearer. Repairs happen when a speaker wishes to retract or modify information gained by interpreting a preceding utterance. In fact, the vast majority of real dialogues contain such contributions, which include constituent negation clauses such as:

“No, I meant the house not the lake.”

Evidence for this claim can be gleaned from even a cursory glance at any real dialogue corpus; for example the *Map Task* corpus [2] at Edinburgh’s Human Communication Research Centre<sup>17</sup>. Chapter 7 treats these phenomena in some detail; for the moment they are mentioned in respect of the empirical adequacy of standard systems.

**The Implosion problem**

Information states in standard systems of DS are *implosive under contradiction* (theories, on the other hand, *explode* under inconsistency.) All information is lost when an inconsistency arises.

$$s \llbracket Px \wedge \neg Px \rrbracket = \emptyset$$

All information built up by the preceding dialogue is lost once an inconsistency arises with regard to any variable or proposition in that dialogue. The problem also occurs for contradictions in Update Semantics. This shortcoming violates any rational idea of informational economy, for agents should at least be able to retain information that is not “connected” to the contradiction. In addition to avoiding implosion, it is desirable that a semantics of dialogue cover cases of resolution of contradiction. Agents commonly use dialogue to enter into disputes, and resolve them.

---

<sup>17</sup>See [15] for self-repair data

### The “downdate problem” for pronouns

The “downdate problem” in DPL (and its extensions) arises when all information about possible anaphor bindings is lost once an anaphor is resolved. If this resolution is later to be changed, say because of a further dialogue contribution, then the analysis collapses. DS requires a more robust notion of subjects from which information can be recovered.

**Example 3** *“The man with red hair was carrying a boy up the stairs. He was whistling.”*

Our instinctive binding of the pronoun “he” to “man” standardly means (ie: in the treatments offered in standard DS) that this assumption cannot later be revised in order to recover a whistling boy.

First-order theory change systems, with dynamic semantics, (which I develop in chapter 5) remedy this problem.

## 2.6 Review of the AGM results

The central problem dealt with in the AGM literature (see eg: [1], [42], [40], [41]) is that of how to consistently add new sentences (which might conflict with a current theory) to a deductively closed set of sentences (or “theory”). Rather than truth preservation during inference, which is the aim of the deductive rules of classical logic, Theory Change takes consistency preservation under change as its ultimate goal. Such systems work on two levels, with two ‘logics’. The *theory-internal* logic provides a notion of deductive closure for the theories which evolve by way of rational theory change; the *theory-external* ‘logic’ governing rational change. The base (theory-internal) logic assumed in AGM theory change is Classical Propositional Logic (CPL), so that “theories” are taken to contain all the sentences that classical logic dictates as derivable from the explicitly stated sentences of the theory. The adoption of CPL produces a number of, perhaps unwelcome, consequences, as well as a number of interesting questions, depending on the particular application to be made of AGM theory change results. In particular, classically closed theories are infinite, consistent, and they contain irrelevancies (the relevant logician’s “Paradoxes of Relevance”). I later explore criteria that an “Informational Logic” ought to meet, and argue that, for the purposes of Dynamic Semantics, CPL is not the correct

base logic to employ. Still later I show that AGM results can indeed be preserved under the adoption of a variety of non-classical logics (cf: Restall and Slaney [84].)

Growth of knowledge is understood in the AGM tradition as a syntactic operation of repeated set union and logical closure; the prime directive being that of maintaining consistency under change. Where new information conflicts with that already present in the current theory a process of *revision* must take place (on pain of theory explosion). Note that there is no question of whether or not new information should be accepted; the question is rather how to achieve this whilst at all costs preserving consistency. Later I shall develop a system which allows agents rationally to *judge* whether or not to accept new information. Call this property **Rational Autonomy**. The ideas developed here have obvious application in the (Formal) Philosophy of Science, but I argue that they are crucial too for formal linguistics, as tools for the fixing of reference and agreement between dialogue participants in general.

We speak of sets of sentences  $T$  or  $K$  as belief states, belief systems, or theories defined over some formal language  $\mathcal{L}$ . There are three basic operations over these sets in AGM theory change; Expansion, Contraction, and Revision. The functions correspond to (consistency 'blind') information growth, information loss (where the matter of consistency preservation scarcely arises), and consistency preserving information growth respectively.

Returning now to the Multiple Extensions Problem (MEP), the initial question was how to determine a suitable contracted theory from a set of subtheories. The force of the MEP was that logic alone does not provide us with any criteria by which to make such a choice. This problem illustrates the divide between logical and non-logical (informational) factors in theory change. The distinction between logic and information is a crucial one, which demands extra-logical criteria to be applied to the problems of theory change. It is precisely the business of rational theory change systems to supply such criteria.

We now briefly review each of the AGM operations in turn.



### 2.6.1 Expansion

A new proposition  $\phi$ , which does not contradict any sentence in the existing theory  $T$ , is added to  $T$ , and the resulting set of sentences is closed under logical consequence. Expansion is denoted  $T + \phi$ , and defined as follows:

Where  $\vdash$  is the deducibility relation of the logic under consideration (classical propositional logic in the AGM case), and similarly,  $Cl$  denotes closure under logical consequence:-

$$T + \phi = \{\psi \mid T \cup \{\phi\} \vdash \psi\}$$

or equivalently:

$$T + \phi = Cl(T \cup \{\phi\})$$

Expansion, then, is a straightforward matter. The situation is not so simple for revision and contraction, because here there is no unique way to specify what should happen as a matter of logic<sup>18</sup>. Rather, revision and contraction operations must satisfy certain set of **rationality postulates**, which are important extra-logical, or informational, criteria.

The notion of expansion in theory change captures the notion of a (non-anaphoric) update in Dynamic Semantics precisely. To anticipate later chapters somewhat, in general;

$$T + \phi \text{ corresponds to } s_T \llbracket \phi \rrbracket$$

where  $s_T$  is the information state which is the content of theory  $T$  and  $\llbracket . \rrbracket$  is an eliminative update function, taking states to (increasingly specified) states.

### 2.6.2 Revision

A new proposition  $\phi$  (which may be inconsistent with the current theory  $K$ ) is added, and some sentences from  $K$  are removed in order to maintain consistency. Thus revision is a *consistency-preserving update*. Revision of  $K$  by  $\phi$  is denoted  $K \dot{+} \phi$ .

---

<sup>18</sup>By the Multiple Extensions Problem

$K_{\perp}$  denotes the inconsistent (or overdefined) theory, which is just the set of all sentences of the logical language (denoted  $\mathcal{L}$ ) in which the theory is expressed. Thus  $K_{\perp} = \mathcal{L}$ .

On the other hand,  $\emptyset$  denotes the empty theory, which contains no sentences save the theorems of  $\mathcal{L}$ .

(K+1) For any sentence  $\phi$  and any belief set  $K$ ,  $K\dot{+}\phi$  is a belief set (**Closure**)

(K+2)  $\phi \in K\dot{+}\phi$  (**Success**)

(K+3)  $K\dot{+}\phi \subseteq K + \phi$  (**Inclusion**)

(K+4) If  $\neg\phi \notin K$ , then  $K + \phi \subseteq K\dot{+}\phi$  (**Vacuity**)

(K+5)  $K\dot{+}\phi = K_{\perp}$  iff  $\vdash \neg\phi$

(K+6) If  $\psi \leftrightarrow \phi$ , then  $K\dot{+}\phi = K\dot{+}\psi$  (**Congruence**)

(K+7)  $K\dot{+}(\phi \wedge \psi) \subseteq (K\dot{+}\phi)\dot{+}\psi$  (**Composition I**)

(K+8) If  $\neg\psi \notin K\dot{+}\phi$  then  $(K\dot{+}\phi)\dot{+}\psi \subseteq K\dot{+}(\phi \wedge \psi)$  (**Composition II**)

The effect of (K+5) is to ensure that  $K\dot{+}\phi$  is consistent unless  $\phi$  is a contradiction.

The 6th postulate ensures that content, rather than surface syntactic form, determines revision. The notion of identity of content is captured here by identity of truth-conditions.

The effect of postulates 7 and 8 taken together is that  $T\dot{+}(p \wedge q)$  is equivalent to  $(T\dot{+}p)\dot{+}q$  so long as  $q$  does not contradict the sentences in  $T\dot{+}p$ .

### 2.6.3 Contraction

Contraction is the most important of the AGM operations. Expansion does not present any real challenges, since it merely employs logical closure, and it turns out that revision is definable in terms of contraction and expansion.

A sentence  $\phi$  is to be removed from a theory  $T$ , or belief system  $K$ . The main problem to be solved is that some other sentences from  $K$  may have to be removed in order that the resulting system does not contain  $\phi$  under closure. Working within CPL, attention

is focussed on the retraction of contingent sentences. Contradictions are never in (non-absurd)  $K$  in the first place and so never need to be removed<sup>19</sup>, and removal of a logical theorem, or tautology, is plainly absurd. Contraction of  $K$  by  $\phi$  is denoted  $K \dot{-} \phi$ .

(K $\dot{-}$ 1) For any sentence  $\phi$  and any belief set  $K$ ,  $K \dot{-} \phi$  is a belief set (**Closure**)

(K $\dot{-}$ 2)  $K \dot{-} \phi \subseteq K$  (**Inclusion**)

(K $\dot{-}$ 3) If  $\phi \notin K$  then  $K \dot{-} \phi = K$  (**Vacuity**)

(K $\dot{-}$ 4) If  $\not\vdash \phi$  then  $\phi \notin K \dot{-} \phi$  (**Success**)

(K $\dot{-}$ 5) If  $\phi \in K$ , then,  $K \subseteq (K \dot{-} \phi) + \phi$  (**Recovery**)

(K $\dot{-}$ 6) If  $\psi \leftrightarrow \phi$ , then  $K \dot{-} \phi = K \dot{-} \psi$  (**Congruence**)

(K $\dot{-}$ 7)  $K \dot{-} \phi \cap K \dot{-} \psi \subseteq K \dot{-} (\phi \wedge \psi)$  (**Composition I**)

(K $\dot{-}$ 8) If  $\psi \notin K \dot{-} (\psi \wedge \phi)$  then  $K \dot{-} (\psi \wedge \phi) \subseteq K \dot{-} \phi$  (**Composition II**)

#### 2.6.4 Identities

The following AGM results demonstrate that revision and contraction functions are interdefinable.

If a contraction function satisfies (K $\dot{-}$ 1) to (K $\dot{-}$ 4) and (K $\dot{-}$ 6), then a revision function obtained via the Levi identity satisfies (K $\dot{+}$ 1) to (K $\dot{+}$ 6).

$$\text{LEVI IDENTITY: } K \dot{+} \phi = (K \dot{-} \neg \phi) + \phi.$$

If a revision function satisfies (K $\dot{+}$ 1) to (K $\dot{+}$ 6), then a contraction function obtained via the Harper identity satisfies (K $\dot{-}$ 1) to (K $\dot{-}$ 6).

$$\text{HARPER IDENTITY: } K \dot{-} \phi = K \cap (K \dot{+} \neg \phi)$$

---

<sup>19</sup>I shall later argue that this really goes against the spirit of theory change, and advocate a paraconsistent approach.

Note that these crucial results allow (in the propositional case) a reduction of the problem of specifying revision operations to the (simpler) one of producing a suitable contraction function.

An interesting question arises in the first-order case: can similar identities be found in the first-order case (where theories are sets of sentences of predicate logic)? First-order postulates for revision and contraction involve constraints on names, identities, and the quantifiers, which have the potential to block a first-order Levi-identity. This question is postponed until chapter 5.

## 2.7 Constructing Contractions

The AGM literature provides several contraction functions which meet the constraints imposed by the rationality postulates.

Each of these methods relies on the idea of choosing some appropriate subset of the contracted theory, which fails to imply the offending sentence. Thus, define  $T \perp p$  as the set of all *maximal* subsets of theory  $T$  which fail to imply  $p$ . Following Fuhrmann [35], this set will sometimes be called the set of *remainders* of  $T$ .

Again, the force of the MEP is that there are in general many suitable remainders from the point of view of pure logic. AGM research explores a variety of ways of selecting appropriate elements from the set of remainders, as an illustrative example I examine “partial meet contraction.”

**Partial meet contraction** relies on the assumption that there is simply some way of choosing the best elements of  $T \perp p$ , via a selection function  $\alpha$ . So,  $\alpha$  chooses a set of preferred remainders of  $T$ . Then define  $T \dot{-} p$  as the set of sentences which such preferred remainders (maximal subtheories) of  $T$  have in common. So that:

$$T \dot{-} p = \bigcap \alpha(T \perp p)$$

It turns out that the rationality postulates for contraction are obeyed by partial meet contractions, and that any contraction satisfying the postulates is a partial meet contrac-

tion. So, the contractions constructed by way of the partial meet contraction function meet the rationality criteria laid out in the postulates (see [42] for the proof.).

### 2.7.1 Multiple (or “General”) Contraction

Fuhrmann (1994) extends the notion of Partial Meet Contraction to the case where contractions remove a *set* of sentences from a theory, rather than a single sentence. This generalization of the AGM approach provides the notion of a package remainder; the subtheories ( $T \perp B$ ) left over after removal of an entire set of sentences (B).

#### Definition 2 Package Remainders

for sets of sentences  $X, T, B$ ,

$X \in T \perp B$  iff

- $X \subseteq T$
- $X \not\vdash B$ , and
- $\forall Y : X \subset Y \subseteq T \Rightarrow Y \vdash B$

#### Fuhrmann’s General Contractions

Andre Fuhrmann’s “Essay on Contraction” [35] contains an elegant generalisation of the AGM approach which will be particularly useful in the construction of systems of First-Order theory contraction. A limitation of the AGM approach is that it considers only the retraction of one sentence from a set of sentences. Fuhrmann (and Hansson) relax this constraint in their work on *multiple contraction*: the removal of *sets* of sentences from belief sets.

There are two notions of contraction here which are of particular interest: *Choice Contraction* (denoted  $\exists$ ): the removal of at least one of a set of sentences from a theory, and *Package Contraction* (denoted  $\forall$ ): the removal of an entire set of sentences from a theory.) Chapter 5 produces evidence that such contractions occur in everyday linguistic practice.



General contraction is a generalisation of the AGM approach, incorporating the results covered above (Choice and Package contractions coincide when the set to be removed is a singleton). Fuhrmann proposes the following sets of rationality postulates:

Where, for sets of sentences  $A$  and  $B$ ,  $A \vdash B$  iff  $B \cap Cn(A) \neq \emptyset$

(" $\vdash$ " is "Scott consequence": some of  $B$  are amongst the consequences of  $A$ .)

### Package Contraction

(PC1)  $B =_A C \Rightarrow A \forall B = A \forall C$  (**Congruence**)

(PC2)  $A \forall B \vdash B \Rightarrow \vdash B$  (**Success**)

(PC3)  $A \forall B \subseteq A$  (**Inclusion**)

(PC4)  $\alpha \in A \setminus A \forall B \Rightarrow \exists A'$  such that  $A \forall B \subseteq A' \subseteq A$  and  $A' \not\vdash B$  and  $A', \alpha \vdash B$   
(**Relevance**)

### Choice Contraction

(CC1)  $Cn(B) = Cn(C) \Rightarrow A \exists B = A \exists C$  (**Congruence**)

(CC2)  $A \exists B \Vdash B \Rightarrow \Vdash B$  (**Success**)

(CC3)  $A \exists B \subseteq A$  (**Inclusion**)

(CC4)  $\alpha \in A \setminus A \exists B \Rightarrow \exists A'$  such that  $A \exists B \subseteq A' \subseteq A$  and  $A' \not\vdash B$  and  $A', \alpha \Vdash B$   
(**Relevance**)

Where, for sets of sentences  $A$  and  $B$ ,  $A \Vdash B$  iff  $B \subseteq Cn(A)$

(“ $\perp$ ” is “Bolzano consequence”: all of  $B$  are contained in the consequences of  $A$ .)

The Relevance postulates replace AGM Recovery, which also makes AGM Vacuity redundant. The major omission here is Closure, which Fuhrmann omits because General Contractions are not restricted to theories (logically closed sets of sentences). However, this virtue need not worry us, for Inclusion and Relevance together imply that whenever a theory is given as input, theories are returned<sup>20</sup>.

## 2.7.2 Constructing General Contractions

The partial-meet contraction method of AGM needs to be altered only slightly in order to cover package contraction. I refer the interested reader to Fuhrmann [35], where full details and proofs are to be found.

Recall that  $B$  is now (potentially) a *set* of sentences. Then define partial meet package contraction via the above generalisation of AGM’s “maximal subtheories”; package remainders:

$$A \perp B = \bigcap \alpha(A \perp B)$$

The contractions so defined meet all of PC1 - PC4.

Choice contractions are characterized in a similar fashion, employing the notion of *choice remainder*: the set of all maximal subsets of  $A$  that do not *contain* set  $B$  is written  $A \angle B$ . The difference between choice and package contractions can easily be traced back to their employment of Bolzano and Scott consequence respectively.

### Definition 3 Choice Remainders

$X \in A \angle B$  iff

- $X \subseteq A$

---

<sup>20</sup>See page 51 of Fuhrmann’s “Essay on Contraction”

- $X \not\models B$ , and
- $\forall Y : X \subset Y \subseteq A \Rightarrow Y \models B$

Then put  $A \exists B = \bigcap \alpha(A \angle B)$

Such functions satisfy CC1-4.

### 2.7.3 Structuring theories

Rational theory revisions are not (and cannot) be made on logical grounds alone. There are many cases where pure logic cannot dictate the revision that takes place, for logic does not pronounce upon the pedigree of facts. There are various ways of giving structure to theories, which allows revision to be computed in a less arbitrary way than by using selection functions.

Advocates of “epistemic entrenchment” have a particular way of prioritizing or producing a hierarchy of beliefs. For constructive propositional belief revision, postulate a relation  $\leq$  on sentences reflecting their degree of epistemic warrant, ordered via the consequence relation. Then, define contraction of sentence  $\phi$  from knowledge base  $K$  as follows.

$$\theta \in K \dot{-} \phi \text{ iff } \theta \in K \text{ and } \phi \leq \theta \text{ or } \vdash \phi$$

An obvious problem with this approach is that, while it avoids the initial ‘arbitrariness’ of the use of the selection function in choosing suitable theory remainders, it merely pushes the problem of selection into the theory itself. Theories are now structured sets of sentences, but how is it determined where sentences fit in that structure? The selection of entrenchment position of a sentence in a theory is just as arbitrary as use of the selection function. Mark Ryan faces this same problem in his study [88] of “Ordered Theory Presentations”; the problem of selection of remainders is replaced by the problem of choosing an ordering on sentences in a theory.

What is needed, I suggest, is a more principled way of selecting appropriate remain-



ders, which does not rely on some previously imposed ordering. Theories must carry their structure with them, especially if, as Ryan observes, they are to undergo *iterated revision*. The model of dialogue proposed here sees communication as a process of iterated change.

The IQ measures (“Information Quality”, presented in chapter 8) for first-order systems, are intended as (the beginnings of) a principled way of replacing epistemic entrenchment hierarchies and informationally guiding autonomous rational revisions.

## 2.8 Modelling Theory Change

The systems and results presented above for theory change are entirely syntactical; they concentrate on the particular *form* in which a theory is expressed, and as such they provide perfectly good algorithms for theory change. So why might one need to go about *modelling* theory change? Modelling allows important notions such as *content* and *completeness* to be formalized and verified (I said more in defence of semantics in the preceding chapter). A semantics provides a standard of correctness against which methods of information processing (the syntactic manipulations) can be compared.

Semantically speaking, a modelling for a theory gives us a notion of the *content* underlying a particular set of sentences. Contents of theories will be thought of as sets of “possible worlds”<sup>21</sup>; the more informative a theory is, the fewer “worlds” it admits as possible. A complete theory prescribes just one possible world as the real one. (The obvious link here is with the concept of a total information state in Dynamic Semantics.) Once a semantics for contraction or revision is in place there is a basis for checking soundness and *completeness* of various systems of theory change.

There are a number of ways in which one might go about modelling theory change, but surprisingly little work has been done on the subject. Gärdenfors (1986) gives a modelling for theory *extension* in terms of the semantics of intuitionistic logic. However, Grove’s “Spheres” [48] have been the only direct attempt to provide a modelling for theory *revision*. As Fuhrmann [35] points out, Grove’s Spheres are closely related to Lewis’ [74] semantics for counterfactuals. My semantics for revision, developed in

---

<sup>21</sup>In the first-order case, sets of variable assignments are also involved.

chapter 3, is given using the concept of “*supermodels*”, and is directly applicable to the *dynamic semantics* of natural language. I now review each modelling strategy in turn.

### 2.8.1 Intuitionistic Semantics for extension

In “Knowledge in Flux”<sup>22</sup> [42], and a number of earlier papers, Peter Gärdenfors proves that propositional theory extension is complete with respect to the semantics of propositional intuitionistic logic. This is a result which I extend to the first-order case in chapter 5. The semantics of Intuitionistic Logic (IL) is based on the idea that models encode growing states of knowledge. Once a proposition is held to be true in IL it remains true (Persistence). Thus the connection with theory extension is intuitively obvious (to anticipate once more):

**Theorem 1**  $T + p \vdash \alpha$  iff  $s_T \llbracket p \rrbracket \models_{IL} \alpha$

Alpha is in a theory under extension by a proposition  $p$  if and only if it is supported by the intuitionistic model of the theory updated by that proposition.

### 2.8.2 Grove’s Spheres

Grove’s modelling [48] is directed specifically towards AGM theory *revision*. Groves spheres also make clear the link between systems of theory change and the semantics of counterfactuals. Systems of nested spheres  $\{\Sigma_P : P \subseteq W\}$  are families of subsets of  $W$  (the set of all possible worlds), centered upon a particular subset  $P \subseteq W$ , corresponding to a proposition  $P$ .

Write  $[\alpha] = \{w \in W \mid w \models \alpha\}$  for the content of a proposition  $\alpha$ .

Similarly, let  $[T] = \{w \in W \mid \forall \alpha \in T, w \models \alpha\}$  be the content of a theory  $T$ . In addition  $\mu_{\Sigma_P}[\alpha]$  is the set of smallest spheres in  $\Sigma_P$  which non-emptily intersect  $[P]$ . Systems of nested spheres satisfy four conditions:

- (S1) Minimality:  $P$  is the smallest sphere in  $\Sigma_P$

---

<sup>22</sup>Pages 138 - 142

- (S2) Maximality:  $W$  is the largest sphere in  $\Sigma_P$
- (S3) Finite Nesting:  $\forall \alpha, \forall \Sigma_P$ , if  $[\alpha] \neq \emptyset$  then  $\mu_{\Sigma_P}[\alpha] \neq \emptyset$ <sup>23</sup>
- (S4) Connectedness: if  $P, Q \in \Sigma_P$  then  $P \subseteq Q$  or  $Q \subseteq P$

(S3) and (S4) ensure that for every consistent  $\alpha$  and every family of spheres, there is always a unique smallest sphere which non-emptily intersects  $[\alpha]$ .

Lewis's 1973 semantics [74] for counterfactual conditionals differs only in the respect that there spheres are centered on single worlds (total information states), rather than sets of worlds (propositions.)

Although Grove states his semantics for revision, it can easily be use to generate contractions too. To find  $T \dot{-} p$  consider the set  $A$  of the closest worlds<sup>24</sup> to  $[T]$  at which  $\neg p$  is true.  $T \dot{-} p$  is just the set of propositions true in all of the worlds  $[K] \cup A$ .

### 2.8.3 "Supermodels"

The modelling which I develop in the next chapter is specifically tailored to interface with the systems of dynamic semantics presented above, although it is obviously formally equivalent to contractions based on Grove's spheres. Supermodel semantics could be described as "spheres for dynamic logicians", although their focus is on contraction rather than revision, and I give them in terms of models, rather than possible worlds (though this is not really a distinction of any substance.) Models are eliminated under update (semantical theory extension), and sets of "supermodels" are constructed during "downdating" - the semantical analogue of theory contraction. The result is a family of dynamic systems which are dynamic in two directions; information growth (model elimination) and information loss (supermodel construction).

---

<sup>23</sup>Rules out infinite chains of inclusion; like Lewis's limit assumption in the semantics for counterfactuals.

<sup>24</sup>Some work must be done to ensure that there is such a set.

## 2.9 Logics of Information

A Formal Semantics of communication relies, it is argued, upon the coordination of information states belonging to communicating agents. The question quickly arises, then, as to the particular logical principles governing deducibility over information states (the “inner logic” of information states). The base logic under which theories are to be closed ought, then, to be an “informational logic.” Here, some basic constraints to be met by any adequate logic based on the notion of information states<sup>25</sup> are surveyed. An “informational logic” is one in which the deducibility relation is thought of as inference based on the available information. Various other names have been given to such logics (“evidential inference”, “logics of inquiry”, and doxastic and epistemic logics in general.) The intuition to be captured here is that  $\Gamma \vdash \phi$  corresponds to the evidence  $\Gamma$  making the inference  $\phi$  available. This informational perspective forces us to reject certain of the standard rules of classical logic. In the work to be developed in the main body of the thesis, classical logic is retained as the base logic. I return to non-classical themes at the close of the thesis.

### 1. Paraconsistency

The first quibble with classical logic regards its approach to inconsistency. Inconsistent input is common, especially in dialogue, so Dynamic Semantics needs a logic which treats contradiction robustly. Theory change systems work so as to preserve consistency, so that contradictions would never be met by the underlying inner logic of the theories themselves. However, such an idealization is too strong for a proper treatment of the data. Thus “Ex falsum sequitur quodlibet” (EFSQ), the classical law “ $\perp \vdash \phi$ ”, is informationally invalid. Having contradictory information does not license an agent to infer just anything. I require theories to be *non-explosive* under contradiction, and information states to be *non-implosive*. AGM theory works with a classical base logic and avoids explosion by way of the revision function. I explore a combination of AGM theory with paraconsistency in chapter 8, the point being to allow contradictions to be explicitly dealt with, rather than ruled out of the picture altogether.

---

<sup>25</sup>Compare with Wansing’s “Logics of Information Structures”[106], Andreas Schöter’s [89].

## 2. Partiality

The classical law of excluded middle,  $\vdash p \vee \neg p$ , is invalid from the informational perspective. For information states, or theories, must be able to be *incomplete*; agents can be agnostic about certain propositions. However, this can be modelled without the use of a partial base logic. Theories are partial in that they fail to contain information about certain propositions. Model this partiality via *sets* of possible models or worlds. Within each theory (ie: in terms of the internal logic of the theory) it is true that  $\vdash p \vee \neg p$  is the case, but this does not commit a theory to holding either  $p$  or its negation (disjunctions are not witnessing, one might say). Partiality is accommodated by way of the external logic of theories, rather than their internal deductive systems. In this respect, AGM systems are partial, and Update Semantics is similarly partial.

## 3. Relevance

The crucial insight of relevant logic [3] is that entailments hold between formulae in information states if and only if there is an “informational” link from antecedent to the consequent (as opposed to a mere connection of truth-values). One might say that “A (informationally) implies B” only if the content of A makes the content of B available. Thus the usual ‘paradoxes’ of relevance (eg:  $A \rightarrow (B \rightarrow A)$ ) ought to be ruled out in the logic underlying communication. Both AGM theory and systems of Dynamic Semantics are “irrelevant.” I return to the issue of conditionals in a later chapter.

## 4. Non-monotonicity

Non-monotonicity is otherwise known as defeasibility, revisability, or *non-persistence*, this is really a meta-logical requirement. Normal “monotonic” systems accumulate inferences and information. Theory extension is monotonic, for example. However, the notion of change requires that previous information states may be overridden by the reception of new information. Prior commitments may have to be lost via the acquisition of new information. This is the whole point of AGM theory change, and one which is widely accepted in the AI and logic communities. “Truths” (other than the logical laws) in an information state may later be discovered to be “false” or just of undecided status, and vice versa. The eliminativity of interpretation in standard dynamic systems makes them monotonic. The next chapter explores how DS informa-

tion change can be made non-monotonic, using a modelling of AGM theory change.

### 5. Constructive negation

Logics of information processing ought to give equal weight to both positive and negative information. Negative information is classically taken to be supported by lack of negative information. Thus if a state fails to support  $p$ , it (classically speaking) supports  $\neg p$ . However, negative information ought to have an independent status to that of positive information. The point is that one cannot infer  $\neg p$  just from the fact that there is no positive information about  $p$ .

### Base Contraction versus Paraconsistency

Closure of theories under classical consequence might, given the above discussion, be too problematic a requirement in many applications of theory change. For one thing, it leads to the familiar problem of logical omniscience, if theories are taken to be information states of an agent. A related issue is that classically closed theories are infinitely large sets, which hardly fits in with a view of agents as finite. A pressing problem is the fact that classically closed theories “explode to absurdity” under inconsistency. The theory of “base contraction” was motivated by some of these considerations. The basic idea behind “base contraction” is that contractions occur only over *open* sets of sentences; the “theory base”. The theory base is later closed, after contraction, in order to generate the full theory.

However, the approach taken in base contraction, wedded as it is to classical logic, overlooks an alternative route away from these problems. As Restall and Slaney [84] point out, a “theory” that does not make use of logical closure is hardly a theory at all. For this reason, and the supporting fact that AGM results can be preserved under (some) modifications of the base logic (or logic internal to theories), I prefer to explore ways of changing the logic underlying theory change in order to escape some of the problems of classically closed theories (again see chapter 8).

I turn to the formal specification of an informational base logic for a dynamic semantics of communication in a later chapter. The following chapter provides a semantics for theory contraction, in the propositional case, which interfaces with systems of dynamic

semantics.

# **Part II**

# **Formal Development**



## Chapter 3

# A Dynamic Semantics of Theory Change

Theory change systems (AGM et al.) have, as yet (and in the literature as a whole), only been stated at the propositional level. They take theories to be sets of sentences (syntactic structures) closed under classical propositional consequence. Systems of dynamic semantics, on the other hand, assume that information states are *semantical* structures, often of *first-order* complexity.

One of the lessons of the first chapter was that “downdates” of information states need to be taken seriously if dynamic systems are to fulfil their empirical aims. Currently, dynamic systems deal only with information growth in the sense of decreasing partiality, but dynamic semantics needs to process *changing* information both from a changing world and from differently informed linguistic agents. Information exchange in communication is to be modelled by way of theory construction, contraction, and revision, within the frameworks provided by two allied perspectives; dynamic syntax (theory change or “belief revision”), and dynamic semantics. As explained in the preceding chapter, the syntactic characterization of theory change derives from the AGM literature on “Belief Revision”. Dynamic semantics, on the other hand, demands systems which see theory change as affecting semantic structures, or “information states”. Some well-explored AGM syntactical theory contraction algorithms are used to construct new systems of propositional *downdate semantics*. That such research is desirable,

note Hans Rott's request [86] for;

"... the invention of a semantics for belief contraction."

The resulting "downdate semantics" forms the core of new extensions of current dynamic semantical systems which are capable of processing *changing* information.

### 3.1 Introductory Remarks

The discussion of chapter 1 resulted in the assertion that communication can be modelled as agents' convergence upon a mutually acceptable theory about the world. This account involves the construction and revision of agent-relative *theories*, rather than a common or shared "information state". Theories here ought not to be construed solely as sets of sentences to which individuals assent under ideal conditions, for the *content* of a theory is also of crucial importance when one considers how it should be changed<sup>1</sup>.

Thus the syntactic operations of theory change go hand in hand with an account of the semantical objects and processes to which they relate. The content of a theory will be given in terms of structures of *possibilities* admitting of transformations which mirror syntactical theory expansion, contraction, and revision. The syntactical operations determine how information in a theory is *processed*, while the semantics describes the information *structures* which are being transformed. Furthermore, the operations of theory contraction and revision are taken to be an integral part of the dynamics of a serious Dynamic Semantics of dialogues. Such a semantics not only needs to posit information structures which are revisable, but ought to provide an account of how the appropriate revisions are to be carried out.

The general problem for a genuine Downdate Semantics is that, whereas in the case of theory extension or semantic update a unique extension or update can be determined on logical grounds alone, for contracting theories and semantic downdates there is (in general) no purely logical way of determining the required contraction or downdate<sup>2</sup>.

---

<sup>1</sup>This is even tacitly acknowledged in the AGM case, where the congruence postulates attempt to ensure that propositions with the same 'content' have the same dynamic effect.

<sup>2</sup>Jaspars [60] seems not to acknowledge this important point. The multiple extensions problem shows that constraints provided by logical considerations are not strong enough to provide a computable con-

Thus extra-logical, or informational, factors must be brought to bear on the problem. While there exist syntactical systems developed in the AGM literature (Alchourron, Gärdenfors, & Makinson, [1]), these are only stated at the propositional level. Thus, going on to consider the dynamics of first-order theories (in chapter 5), not even the syntactical work is in place yet.

### 3.2 Related research

Gärdenfors ([40]) proved the completeness of AGM theory *extension* with respect to Intuitionistic Logic (IL). This was relatively unsurprising given the Kripke semantics for IL in terms of growing states of knowledge, and the very similar motivations behind the AGM expansion postulates. An interesting extension of IL is Nelson's logic of constructible falsity [77], the system N. Here negative information is taken to have a semantic status on a par with that of positive information, whereas IL formulates negation as intensional. Pearce and Rautenberg [78] showed that an extension of AGM theory expansion where falsity or rejection is seen as an independent concept is complete with respect to N.

Note, however, that these results apply only to monotonic systems; those where information grows and persists. Constructive negation represents the rejection of information rather than its retraction or revision, so that the above completeness results describe only the semantics of theory growth. This begs the question of a set of completeness results for a semantics of theory contraction and revision with respect to the AGM postulates. These I supply in this chapter.

#### Jaspars' "ud"

An obvious way to start a semantics of theory contraction would be to put the Kripke semantics for IL as it were "in reverse," so that retraction of a proposition is seen as a relation in the opposite direction to the usual information extension order for Kripke

---

traction algorithm. See Rescher [82], Gärdenfors [42], Fagin, Ullman, & Vardi [29], Galliers [38]

models. Jaspars [60] takes this approach in the Kripke semantics for N (“Nelson models”), introducing “up- and down- dates” as modal operators.

A system dubbed “**ud**” for derivations in an up-and-downdate formulation of the semantics of N is given in Jaspars [60]. “**ud**” is sound with respect to an up-down language over Nelson models (p. 97), completeness is also shown (p. 163). The system sees updates and downdates as actions denoted by modal operators. approach results in the following (where  $\downarrow$  is a ‘downdate’ function):

“ $\downarrow p[q]$ ” means that  $q$  always results from the retraction of  $p$  from some information state  $s$  in model  $M$ .

Thus: “retraction of  $p$  means  $q$ ” (where  $\leq$  is a partial order on  $S$ ):

(1)  $M, s \models \downarrow p[q] \Leftrightarrow \forall s' \leq s : M, s' \not\models p \Rightarrow M, s' \models q$  Thus, if  $p$  is not supported but  $q$  is, at all points lower in the information growth order than  $s$ , then  $s$  supports the assertion that “retraction of  $p$  means  $q$ .” This seems a rather strange way of thinking about down-dating. Removal of  $p$  from a theory does not “mean” anything much; it certainly does not *entail* anything (in the sense of ‘support’), unless one wants to remove  $p$ ’s consequences as well, so that removing  $p$  might “mean” forcing the removal of  $q$ . This would be to stretch the idea of entailment too far though, for the action of theory contraction should not be thought of as supporting any evidence at all; but quite the opposite. What Jaspars has defined is the idea that “ $q$  is in the residue of  $s$  after contraction by  $p$ .” This notion can easily be defined in the systems of the preceding chapter, both syntactically ( $q \in T \dot{-} p$ ) and semantically ( $(s_T] p \models q$ ), but it is not the focus of the research presented here.

AGM theory change is constructed so as to tackle the multiple extensions problem (MEP); the fact that there is (generally) no unique logical way of contracting a theory. The approach embodied in “**ud**” singles out just one theory contraction approach, and a fairly naive one at that. In this way, the prescription that information contraction is understood by way of (1) ignores the MEP. The prescription is that the correct way of contracting a theory is given by the semantical insights afforded by Nelson models. But this is surely too restrictive a view.

In this chapter a range of semantical approaches complete with respect to AGM contraction methods is developed, so that a generic system for semantical theory change

is available, as opposed to a single prescriptive system. One of the great merits of the AGM tradition is its broad approach to theory change within a treatment of the important logical problem posed by the MEP. Any semantical approach to theory change needs to respect this logical point. Whilst, computationally speaking, unique revision and contraction functions are desirable, the burden of uniqueness cannot be placed on the logical language underlying the contraction process, but on the nature and structure of the particular information states that a revision system processes. Jaspars' approach makes the burden of choice of contraction a matter of logic rather than of informational, or rational, considerations.

### **Removing information in Dynamic Semantics**

The issue of downdating in dynamic semantics has been acknowledged by researchers in the field (Dekker [21], Asher [6], Vermeulen [104], Jaspars [60]), but few have been tempted to start tackling it. Dekker ([21], p. 205) introduces the "common ground" as the "strongest common downdate" of two information states (all the information that the two states agree upon, via set intersection.) However, this is not the kind of downdate that I shall be concerned with. I explore a function on information states of two arguments; the state to be downdated and the expression whose semantical contribution is to be removed from it.

Vermeulen [104] adds to dynamic predicate logic expressions of the form  $x\exists$  which are interpreted as instructions to forget the value of a variable  $x$ . Whilst, later, I shall be concerned with this kind of downdate in first-order systems one should also be careful not to "jump the gun" in quite this way. Forgetting the values of variables is not the be all and end all of downdating, for there are many other syntactic units by which to downdate in dialogue (see chapter 7), especially involving dialogue-external information. Before issues in the dynamics of anaphora and modality are addressed, the propositional work must be firmly in place.

### 3.3 Definitions

This chapter explores logical and informationally rational ways of changing *consistent* and deductively closed partial theories<sup>3</sup>. Propositional systems are the focus of interest for the moment. Recall some standard **syntactical definitions** for logical theories:

A *theory*  $T$  is a set of sentences of some language  $\mathcal{L}$ . Here, theories are taken to be *partial* descriptions of the world<sup>4</sup>, since generally there will be propositions  $p$  such that neither  $p \in T$  nor  $\neg p \in T$ . A theory  $T$  is *complete* if  $\forall \phi \in \mathcal{L}$ , either  $\phi \in T$  or  $\neg \phi \in T$ . Theory  $T$  is *deductively closed* when  $T$  is the set  $\{\phi \in \mathcal{L} \mid T \vdash \phi\}$ , where  $\vdash$  is a deducibility relation over  $\mathcal{L}$ . We take  $\vdash$  to be the deducibility relation of classical propositional logic<sup>5</sup>. Write  $Cl(T)$  to mean the deductive closure of  $T$ . Write  $T_{\perp}$  for the inconsistent theory, so that  $T_{\perp} = \mathcal{L}$ . Write  $T_{\emptyset}$  for the theory deducible from the empty premise set (i.e. the set of tautologies.) Theory  $T$  is consistent if  $T \neq \mathcal{L}$ .

Following AGM notation,  $T \perp p$  is the set of all maximal subtheories of  $T$  which fail to imply proposition  $p$  under deductive closure (or, to follow Fuhrmann [35], the set of “remainders” of  $T$  after  $p$ ). This set is non-empty as long as  $p$  is not a logical truth. We also define a similar notion, the set of *all* theories (implicitly defined over a certain vocabulary) which fail to imply  $p$ :

**Definition 4**  $T - p = \{T' \mid T' \not\vdash p\}$

Semantically speaking, sets of sentences making up a theory are underwritten by information structures, or sets of models (or “possibilities”) for the language. A propositional model is a set of valuations (assignments of truth-values) to the propositional letters in  $\mathcal{L}$ , and can also be thought of selecting a subset of  $Atom\mathcal{L}$ , the set of atomic propositions, as those facts which hold in a certain *possible world*. Think of each individual model as a “possibility” or *total* description of how the world might be. A *set*

<sup>3</sup>I explore these issues in relation to *paraconsistent* theories in later chapters.

<sup>4</sup>Certainly the theories manipulated in linguistic communication concern only a limited subject matter.

<sup>5</sup>Although I shall later be concerned with non-classical systems.

of models is a set of such possibilities (considered disjunctively); a *partial* description of how things might be. As a theory gets larger (more complete), its corresponding model-set gets smaller, as more possibilities are eliminated. This is the general strategy of an *eliminative semantics*.

Write  $Mod(\mathcal{L})$  for the set of all models of the language  $\mathcal{L}$  (NB: not the model of the absurd theory.) This set corresponds to the empty theory, where all possibilities are available.  $Mod(\mathcal{L})$  supports no sentences, but will accept update by any non-contradictory sentence. In terms of Dynamic Semantics, the notation<sup>6</sup> is that  $Mod(\mathcal{L}) = S$ , and then individual states are the  $s \subseteq S$ .

With each partial theory  $T$  there is an associated set of models  $M_T \subseteq Mod(\mathcal{L})$ <sup>7</sup> Here, take the turnstile " $\models$ " to be the entailment relation of the semantics of classical propositional logic. (Entailment is defined as preservation of truth over the turnstile).

**Definition 5**  $M_T = \{m \in Mod(\mathcal{L}) \mid \forall \phi \in T, m \models \phi\}$

The absurd theory,  $T_{\perp} = \mathcal{L}$ , has no models, so that  $M_{T_{\perp}} = \emptyset$ .

Correspondingly, each set of models has an associated theory<sup>8</sup>;

**Definition 6**  $Th(M) = \{p \in \mathcal{L} \mid \forall m \in M, m \models p\}$

In the case of propositional languages there is a 1-1 correspondence between single models and complete theories;  $T$  is complete iff  $\exists m \in Mod(\mathcal{L})$  such that  $T = Th(m)$ . Thus a single model is the semantical correlate of a complete theory; a state of total information. (Of course, things become somewhat more complex in a first-order logic.)

To sum up here then, the syntax (theories) and semantics (models), link up in the following way:

---

<sup>6</sup>There is a potential confusion to be avoided here. Often, in systems of dynamic semantics, information states are written as  $s \in S$  where  $S$  is the *powerset* of the set of all possibilities (which I have called  $S$  here.)

<sup>7</sup>We shall later generalize this definition to cover sets of theories.

<sup>8</sup>This definition will also be generalized to cover sets of sets of theories.

$T_\emptyset$  corresponds to  $Mod(\mathcal{L}) = S$ , and

$\mathcal{L} = T_\perp$  corresponds to  $\emptyset$

Say that a set of models  $M$  *supports*  $p$  iff  $\forall m \in M, m \models p$

Say that a set of models  $M$  *rejects*  $p$  iff  $\forall m \in M, m \models \neg p$

Say that a set of models  $M$  is *agnostic* about  $p$  iff  $M$  neither supports nor rejects  $p$ .  $M \not\models p$ , or  $M$  *fails to support*  $p$ , iff  $M$  either rejects, or is agnostic about,  $p$ .

### 3.3.1 Non-standard definitions

**Definition 7**  $M \perp p$  is the set of all sets of models which fail to support  $p$ . A model-set  $N \in M \perp p$  fails to support  $p$  iff  $\exists m \in N$  such that  $m \not\models p$ .

Thus,  $M \perp p = \{N \mid N \not\models p\}$

It is from this set that specific *minimal* elements must be chosen for propositional down-dates. Note that (of course) some sets of models in  $M \perp p$ , further than being agnostic on  $p$ , may actually *reject*  $p$ . That is, there are  $Q \in M \perp p$  such that  $Q \models \neg p$

Now extend the function  $Th$ , taking sets of models to theories, to a function returning a set of theories from sets of sets of models, thus:

**Definition 8** Where  $\mathcal{M}$  is a set of sets of models

$Th(\mathcal{M}) = \{Th(M) \mid M \in \mathcal{M}\}$

$M_T$  can also be extended so that it takes sets of theories, denoted  $\mathcal{T}$ , into sets of sets of models;

**Definition 9**  $M_{\mathcal{T}} = \{M_T \mid T \in \mathcal{T}\}$

This allows the following theorem:



**Theorem 2 1.**  $Th(M \perp p) = T - p$

2.  $M_{T-p} = M \perp p$

So to elaborate, the set of theories associated with the set of sets of models failing to support  $p$  are those theories failing to imply  $p$ . The (set of sets of) models of the set of theories which fail to imply  $p$  are the sets of models which fail to support  $p$ .

Proof: first part

$$Th(M \perp p) = Th\{N \mid N \not\models p\} = \{Th(N) \mid N \not\models p\} = \{T' \mid T' \not\models p\} = T - p$$

Second part;

$$M_{T-p} = \{M_{T'} \mid T' \in T - p\} = \{M_{T'} \mid T' \not\models p\} = \{M_{T'} \mid M_{T'} \not\models p\} = M \perp p$$

**Example 4** (Where the notation for models is hopefully obvious, ie: for example "101" stands for "a true, b false, c true".)

$$\text{Take } \mathcal{L} = \{a, b, c\}$$

$$T - a = \{\emptyset, \{b\}, \{c\}, \{b, c\}\}$$

$$M \perp a = \{N \mid N \not\models a\}$$

$$\text{Then } Th(M \perp a) = \{Th(N) \mid N \not\models a\} = Th(\{000, 010, 001, 011\} \{ \emptyset, \{b\}, \{c\}, \{b, c\} \}) = T - a$$

### 3.3.2 Partial Worlds versus sets of total worlds

In the formalism presented above, *sets* of total worlds (or models) have been used to model the content of incomplete theories. Agnosticism with respect to  $p$  is modelled by a set of worlds, some of which support  $p$ , and some of which reject  $p$ . Why not, though, dispense with total worlds, and employ instead partial worlds in order to model partial theories, (as is advocated in Situation Semantics for example?) There are (at least) five reasons for sticking with total worlds;

- (1) preservation of classical results,

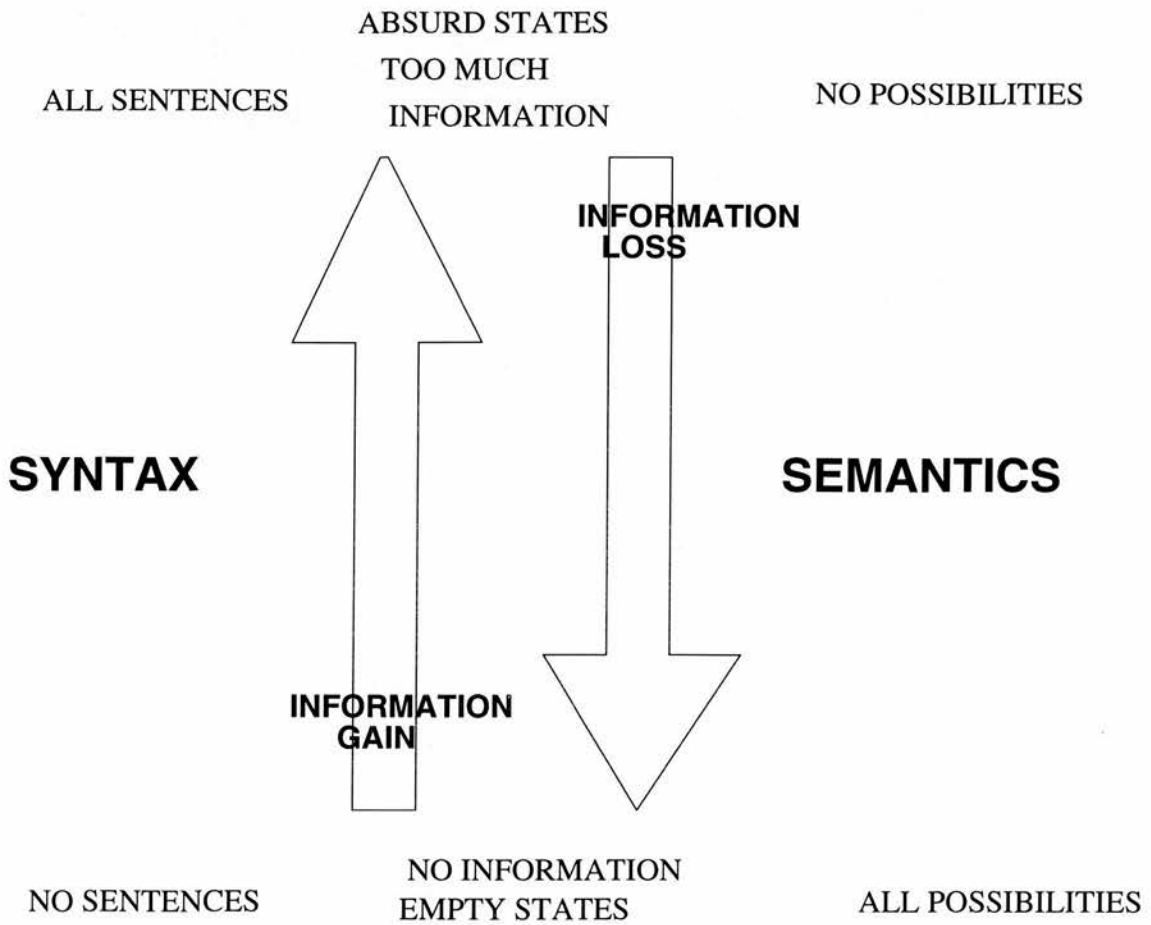


Figure 3.1: Theories versus States

- (2) interfacing with Update Semantics,
- (3) maintaining a clear distinction between theories and models,
- (4) maintaining eliminativity, and
- (5) formal equivalence between partial models and sets of total models.

Veltman's Update Semantics [101], which is one of the systems to be modified, is presented in terms of shrinking sets total models. There is just one partial model of an incomplete theory, so models will have to be extended (made less partial) rather than eliminated under theory extension. Conversely, partial models will have to be "shrunk" (made more partial) under theory contraction, and overwritten under theory revision. All these operations could be defined in partial model theory without much trouble, but, once this is carried out, the distinction between theories and their models seems to be largely lost. Thus the whole motivation behind a semantical approach is undermined by the use of partial models. For these reasons I shall operate on sets of total 'worlds', rather than one increasingly (and decreasingly) specified partial world. Besides, anything that can be expressed using a partial model can just as well be expressed using a set of total models; the approaches are formally equivalent in this sense, though their processing attributes will obviously differ.

### 3.4 Theory Extension and Model Elimination

Syntactically speaking, information growth is handled by way of adding new sentences to the theory, and performing deductive closure. Thus, for propositional logic, as before;  $T+p = Cl(T \cup \{p\})$ , where  $T+p$  is the extension of a theory  $T$  by a proposition  $p$ . This leads to the following simple result; that propositional theory extension is functional<sup>9</sup>.

**Lemma 1 1.**  $\forall T, p, \exists T'$  such that  $T' = T + p$

2. if  $T$  is closed and  $T + p = T'$  then  $T'$  is closed.
3. if  $T + p = T'$  and  $T + p = T''$  then  $T' = T''$

---

<sup>9</sup>see van Eijck [96] for the full details.

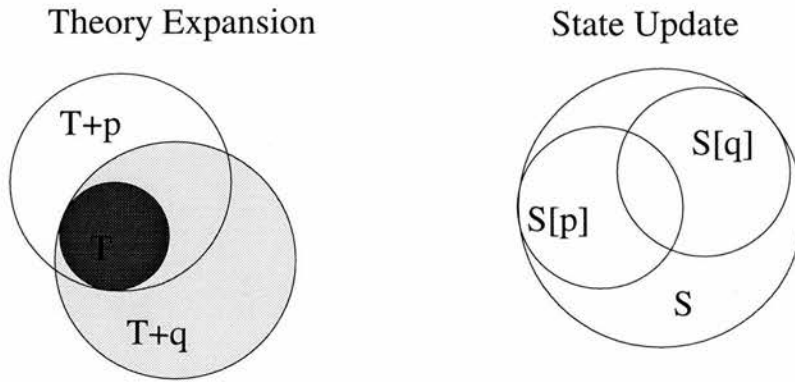


Figure 3.2: Expansion and Update

Semantically, the update process is described by model-elimination, so that, in general;

**Definition 10** (*Update*)

$M \llbracket p \rrbracket = \{i \in M \mid i \models p\}$  where  $M$  is set of models  $m \in M$ , and  $\llbracket . \rrbracket$  is the update function.

As each new proposition is interpreted in the context set up by models resulting from the processing of the foregoing propositions, the set of possible models for the theory becomes smaller and smaller. Eventually a state of total information (only one possible model) of the world remains.

The upshot is that, using model-elimination, a family of systems of Update Semantics can be produced, mirroring different theory expansion systems<sup>10</sup>.

### 3.5 Theory Contraction and “Downdating”

Van Eijck’s approach [96] to modelling systems of theory extension is now extended to theory *change* systems along the lines of Gärdenfors [42]. The rationality postulates for propositional belief revision form a set of syntactic constraints on expanding and contracting sets of propositions. These constraints can be met by a variety of (syntactically

<sup>10</sup>see van Eijck [96] for the results here.

defined) contraction functions<sup>11</sup>, which serve to preserve certain appropriate subsets of sentences from the original theory after the removal of some particular sentence. For example,

**Definition 11 Maxichoice Contraction**

when  $\phi$  is not a theorem of  $\mathcal{L}$ , the contraction  $T \dot{-} \phi$  can be defined as  $\gamma(T \perp \phi)$ , where  $T \perp \phi$  is the set of all maximal subsets of  $T$  which fail to imply  $\phi$ , and  $\gamma$  is a selection function choosing an element of this set. Thus, since  $T \perp p = \{T' \subseteq T \mid T' \not\vdash p\}$ , we have  $T \dot{-} \phi = \gamma(T \perp \phi)$

I associate a *downdate semantics* with this function, and prove that it is complete with respect to Maxichoice Contraction.

Note that whereas  $Cl(T \cup \{p\})$  (theory extension) forms a unique theory  $T + p$ ,  $T \perp p$ , which is a collection of suitable theories, has no such virtue. This is the source of much difficulty (indeed, the whole AGM system is a response to this problem). The way ahead now is to shift the syntactic rationality postulates for revision and contraction of theories into a semantical setting. Some useful insights for theory change systems emerge by way of the semantical perspective on belief revision. In addition, the provision of genuine downdates and revisions for semantic structures motivates a new family of systems of dynamic semantics<sup>12</sup>.

Ultimately the project involves finding new rationality constraints for revision at the predicate calculus level as well as for modal systems. Given a list of rationality postulates the task is then to find corresponding dynamic systems which meet them in terms of semantic operations of model elimination and model creation. This is the task of chapter 5.

### 3.5.1 Desiderata for a genuine “downdate semantics”

First of all, some definitions:

<sup>11</sup>eg: partial meet, full meet, maxichoice, transitively relational partial meet contraction.

<sup>12</sup>This is the subject of the next chapter.

**Definition 12 Implosive functions**

A function  $f$  is **implosive** on a set  $M$  iff  $f(M) = \emptyset$

**Definition 13 Eliminative functions**

A function  $f$  is **eliminative** on a set  $M$  iff  $f(M) \subseteq M$

Now impose the following basic restrictions on a downdate function for any system of downdate semantics. Where  $\Downarrow$  is a downdate function,  $\Uparrow$  an update function,  $M$  is a set of models, and  $p$  any proposition.

**Definition 14 Desiderata:**

(for any proposition  $p$  which is not a theorem of  $\mathcal{L}$ )

1.  $M \Downarrow p \not\equiv p$  (**Success**)
2.  $M \Downarrow p \not\equiv q$  iff  $M \models q$  (**Non-eliminative downdates**)<sup>13</sup>  
There ought not to be information growth after downdating.
3.  $M \Downarrow p \not\equiv M \Downarrow \neg p$  (**Integrity**)  
The requirement is for genuine downdates. In general, downdate by  $p \neq$  update by  $\text{neg}p$ .
4.  $M \Downarrow p \not\equiv \emptyset$  (**No Implosions**)  
Downdating ought not, in general, to result in absurdity.
5.  $M \Downarrow p \not\equiv S$  (**No Explosions**)  
Downdating ought, in general, not to result in the loss of all information.

These considerations bring about the following result.

**Theorem 3 Impossibility Result**

There are no eliminative and non-implosive downdate functions (for semantics with classical entailment).

**Proof:**

Assume that there is such a downdate function. Take a set of models  $M$  such that  $M \models$

<sup>13</sup>In fact I prove that there is no eliminative non-implosive downdate semantics (next theorem).

$p$ , so  $\forall N \subseteq M, N \models p$ . The eliminative downdate  $M \Downarrow p \llbracket$  is non-empty (non-implosive). Thus (by success)  $\exists m \in M \Downarrow p \llbracket$  such that  $m \not\models p$ . But since  $M \Downarrow p \llbracket \subseteq M$  (by eliminativity), it is the case that  $\forall m \in M \Downarrow p \llbracket, m \models p$ , so that  $m \models p$  and  $m \not\models p$ . So, unless the semantics is paraconsistent, there is a reductio.

In a classical semantics, without an account of theory change, the contraction of a theory with respect to a proposition  $p$  is tantamount to extension by  $\neg p$ . (Denial that  $p$  is assertion that not  $p$ .) Van Eijck's [96] second system for the extension of propositional theories works in just this way; he is able to provide a translation of that system into Veltman's Update Logic.

### 3.6 Theory change as a model of dialogue

Theory change systems are not just important for the philosophy of science, artificial intelligence (robotics), and computer science (databases); they are important as a model of linguistic communication too. A conversation can be seen as a co-operative attempt between agents (using "belief revisions") to agree upon acceptable theories about the world as it is, was, or may be. Theories are best thought of as an agent's epistemic approximations to reality. They make no representational claims.

I claim that people change their personal theories all the time during conversation; indeed that this evolution of information structures is the motivation behind communicative behaviour. Communication is not the only way that information structures get altered, perception of a changing world also brings about information change. Thus an empirically adequate dynamic semantics should account for information change (rather than monotonic update) due to both a changing environment and (potentially conflicting) input from (less than perfect) communicating agents.

I give an analysis of some relevant dialogue contributions in chapter 7. For the moment, in order to sketch the ideas here, consider a few "talk experiments":

**Example 5** John: *"The opera starts at 8pm, I think you said."*

Jackie: *"Mmmm, did I? No, I can't remember what time it is on at."*

Jackie's contribution in this dialogue is not a case of standard negation. She is not actually denying that the opera begins at 8 o' clock, but neither is she willing to affirm her earlier claim that it does. Rather, this example requires John to contract a proposition  $p$  (to the effect that the opera starts at 8pm) from his "mini-theory" about the immediate future. Theory contraction is thus applicable to a "logic of forgetting."

**Example 6** John has arranged to meet Jackie in the theatre later on.

Jackie: *"I'm not going out tonight."*

This example requires John to *revise* his theory about the coming evening (which supports a proposition  $p$  to the effect that Jackie is going out tonight), by  $\neg p$ . This task may reduce him to tears, but not to logical absurdity.

**Example 7** John: *"You can catch the 1pm train tomorrow."*

Jackie: *"Good. I was worried that there might be a strike."*

John: *"Ah. You're right. Now that you mention it - maybe you can't catch that train."*

This example is more complex, as it involves epistemic modality. John has made an assertion, and later wants to take it back, due to a modal expression uttered by Jackie. The modal has the effect of changing John's previous commitment.

**Example 8** John and Jackie have been discussing who might have stolen their amnesiac-of-the-year trophy.

John: *"I still think the suspect has to be someone with access to our front door keys; there is no sign of a forced entry"*



Jackie: “Whoever did it was very quiet.”

John: “Oh! Forget about this ‘suspect’. I remember now. I hid it - it’s under the bed.”

Here John is instructing Jackie to “remove” an individual (“the suspect”) from her current theory, because he doesn’t exist. The downdate here is by a discourse referent or variable (eg: see the earlier discussion of Vermeulen [104]). This type of contraction can only be adequately formalized in a first-order system. First-order contractions and revisions are the subject of chapter 5.

The above epistemic/linguistic examples illustrate two important points. Firstly, the update/downdate distinction in partial semantics allows us to capture the distinction between strong and weak denial in everyday dialogue. Weak denial can be modelled by way of theory contraction, leaving the epistemic status of a proposition indeterminate. Strong denial is modelled by updating a theory with a negation (this may necessitate a revision).

The second point to notice is that some of these examples would be much better handled in first-order logic, for we often need a robust notion of “dialogue referent” if certain epistemic problems are to be resolved. After contraction, for instance in the opera example, John is not just going to forget about the status of  $p$ ; he still has the discourse referent *the opera* to worry about. Similarly in the second example John needs to know of Jackie (a “subject”) just what its plans for the evening are.

This motivates the movement of the thesis, towards a first-order syntax and semantics for theory change (chapter 5), and a paraconsistent treatment of contradictions (chapter 8), where a robust and revisable notion of subject can be provided (see chapter 6). Such systems also allow dynamic anaphor binding, so that contraction and revision over sets of variable assignment functions can also be formalised.

### 3.7 Updating and AGM theory extension

Unlike the variety of systems explored by van Eijck [96], which provide clauses for theory extensions with respect to each of the logical connectives (including one for  $\diamond$ ), belief revision in the AGM tradition sees theory expansion as a very simple matter of adding

the new sentence (regardless of its logical form) and closing under consequence. I stick to a simple version of theory extension and make use of established results in the AGM framework.

### AGM Propositional Theory Extension (AGMTE)

Recall the standard definition of theory extension:

#### Definition 15 AGM Theory Extension

$$T + p = \{q \mid T \cup \{p\} \vdash q\} = Cl(T \cup \{p\})$$

From this one can make explicit the following clauses:-

$$T + (p \wedge q) = (T + p) + q$$

$$T + (p \vee q) = (T + p) \cap (T + q)$$

Note that in the case of negation I do *not*<sup>14</sup> stipulate that

$$T + \neg p = T' \text{ such that } T' + p = \mathcal{L}.$$

This would be identical to the intuitionistic account of negation, where  $\neg p$  is defined as  $p \rightarrow \perp$ . Rather, we simply add  $\neg p$  and close under consequence, as usual. In other words, we treat update by positive and negative information in the same way. Note that the AGM extension clause still leaves the way open for *genuine* downdating, where a downgrade is not seen simply as updating with a negation (as is a feature in van Eijck's syntactic correlates of Update Semantics). This openness is an essential virtue of the AGM system, leaving the way open for contractions and revisions. In this way the choice of an update semantics or theory extension system has implications for the further development of a downgrade system.

The result is that the above definition allows the development of a very simple Update Semantics.

---

<sup>14</sup>as opposed to van Eijck's systems

**AGM Update Semantics (AGMUS)**

The following is not quite an update semantics in the style of Veltman, for it lacks an account of epistemic modality (one can very easily be supplied). I consider modal extensions of the AGM systems in detail in the following chapter.

**Definition 16 AGM Update Semantics**

For all propositions  $p \in \mathcal{L}$  there is an eliminative update on sets of models  $M$  :-

$$M \llbracket p \rrbracket = \{m \in M \mid m \models p\}$$

And note the following derived clauses:

$$M \llbracket p \wedge q \rrbracket = M \llbracket p \rrbracket \llbracket q \rrbracket$$

$$M \llbracket p \vee q \rrbracket = M \llbracket p \rrbracket \cup M \llbracket q \rrbracket$$

The connection between AGMTE and AGMUS is given by the following;

**Theorem 4 (Completeness)**

For all propositions  $p$ ;

1.  $Th(M) + p = Th(M \llbracket p \rrbracket)$
2.  $T + p = T'$  iff  $M_T \llbracket p \rrbracket = M_{T'}$

This just states formally that AGM theory extension is accomplished semantically by model (or world) elimination.

**Proof:** for the first item by induction on complexity of  $p$ <sup>15</sup>.

*Base step:*

$$Th(M) + p = Cl(Th(M) \cup \{p\}) = Cl(Th(\{m \in M \mid m \models p\})) = Th(M \llbracket p \rrbracket).$$

Then by Induction Hypothesis

$$Th(M) + p = Th(M \llbracket p \rrbracket), \text{ and } , Th(M \llbracket p \rrbracket) + q = Th(M \llbracket p \rrbracket \llbracket q \rrbracket).$$

From this the rest of the induction is trivial.

Proof of the second part follows from the first. Since  $T$  and  $T'$  are deductively closed;

---

<sup>15</sup>See van Eijck [96] for a similar proof.

$T + p = T'$  iff  $Th(M_T) + p = Th(M_{T'})$

and so from the first part of the theorem

$Th(M_T) + p = Th(M_T \llbracket p \rrbracket)$

so that by Lemma 1 (third clause),

$Th(M_T \llbracket p \rrbracket) = Th(M_{T'})$ , and it follows that

$M_T \llbracket p \rrbracket = M_{T'}$

### 3.7.1 Downdate Semantics rationality postulates

Here I list a series of constraints on the expanding sets of models which model theory contractions. I adopt these postulates in addition to the desiderata on a genuine downdate semantics presented earlier in the chapter. The constraints can be thought of as rationality postulates for downdating information states. These are basically semantical versions of the familiar AGM postulates. We label them DS 1-8, and RS 1-8. Only propositional constraints are listed here. Recall that  $\llbracket \cdot \rrbracket$  is the downdate function on information states  $M_T \subseteq Mod(\mathcal{L})$  (or alternatively  $s \in S$ ), and  $\llbracket \cdot \rrbracket$  is the revision function on information states

#### Rationality Postulates for DOWNDATING

(DS1) For any sentence  $\phi$  and any set of models  $M_T, M_{T'} \llbracket \phi \rrbracket$  exists (**Closure**)

(DS2)  $M_T \subseteq M_{T'} \llbracket \phi \rrbracket$  (**Containment / Non-eliminativity**)

(DS3) If  $M_T \not\models \phi$  then  $M_{T'} \llbracket \phi \rrbracket = M_T$  (**Vacuity**)

(DS4) If  $\not\models \phi$  then  $M_{T'} \llbracket \phi \rrbracket \not\models \phi$  (**Success**)<sup>16</sup>

(DS5) If  $M_T \models \phi$  then  $M_T \subseteq (M_{T'} \llbracket \phi \rrbracket) \llbracket \phi \rrbracket$  (**Recovery**)

(DS6) If  $\psi \leftrightarrow \phi$ , then  $M_{T'} \llbracket \phi \rrbracket = M_{T'} \llbracket \psi \rrbracket$  (**Congruence**)

(DS7)  $M_{T'} \llbracket \phi \rrbracket \llbracket \psi \rrbracket \supseteq M_{T'} \llbracket \phi \wedge \psi \rrbracket$  (**Composition I**)

(DS8) If  $M_{T'} \llbracket \psi \wedge \phi \rrbracket \not\models \phi$  then  $M_{T'} \llbracket \psi \wedge \phi \rrbracket \supseteq M_{T'} \llbracket \psi \rrbracket$  (**Composition II**)

<sup>16</sup>The downdated sentence is not entailed by the resulting information state, unless it is a theorem.

### Rationality Postulates for Revision Semantics

- (RS1) For any sentence  $\phi$  and any “information state”  $M_T, M_T \llbracket \phi \rrbracket$  exists (**Closure**)
- (RS2)  $M_T \llbracket \phi \rrbracket \models \phi$  (**Success**)
- (RS3)  $M_T \llbracket \phi \rrbracket \subseteq M_T \llbracket \phi \rrbracket$  (**Containment**)
- (RS4) If  $\neg\phi \notin T$ , then  $M_T \llbracket \phi \rrbracket \supseteq M_T \llbracket \phi \rrbracket$ .<sup>17</sup> (**Vacuity**)
- (RS5)  $M_T \llbracket \phi \rrbracket = \emptyset$  iff  $\models \neg\phi$ . (**Implosion**)<sup>18</sup>
- (RS6) If  $\psi \leftrightarrow \phi$ , then  $M_T \llbracket \phi \rrbracket = M_T \llbracket \psi \rrbracket$  (**Congruence**)
- (RS7)  $M_T \llbracket \phi \wedge \psi \rrbracket \subseteq (M_T \llbracket \phi \rrbracket) \llbracket \psi \rrbracket$
- (RS8) If  $M_T \llbracket \phi \rrbracket \neq \phi$  then  $(M_T \llbracket \phi \rrbracket) \llbracket \psi \rrbracket \subseteq M_T \llbracket \phi \wedge \psi \rrbracket$

## 3.8 Systems of Theory Contraction and Downdate Semantics

Here I prove connections between various AGM methods for theory contractions and new systems of downdate semantics. The new downdate semantics given will be very simple, analogously to the AGM update system. All propositions participate in the same downdate procedure.

The following notion is useful in the construction of *minimal downdates* or maximal contractions, which are desirable for reasons of informational economy.

**Definition 17 (Maximal subtheory of  $T$  which fails to imply  $p$ )**<sup>19</sup>

$T' \in T \perp p$ , ( $T'$  is a maximal subtheory failing to imply  $p$ ) iff

1.  $T' \subseteq T$

---

<sup>17</sup>Revision is update when  $\neg\phi \notin T$

<sup>18</sup> $M_T \llbracket \phi \rrbracket$  is consistent unless  $\phi$  is logically impossible. In paraconsistent systems we do not make this requirement. Instead a coherence measure which respects informativeness of the theory as a whole is to be used in judging whether or not to revise, and there can be non-trivial information states containing contradictions.

<sup>19</sup>Or Remainder of  $T$  after  $p$

2.  $p \notin Cl(T')$
3. for any  $T''$  such that  $T' \subset T'' \subseteq T, p \in Cl(T'')$ .

Note that  $T \perp p \subseteq T - p$

This definition allows a theory contraction algorithm which makes use of a choice function<sup>20</sup>.

### Definition 18 Maxichoice Theory Contraction (MTC)

When  $p$  is not a logical truth, choose one of the maximal subtheories failing to imply  $p$ . Use a choice function  $\gamma$ , and define contraction thus:

$$T \dot{-} p = \gamma(T \perp p)$$

$$\text{if } \vdash p \text{ then } T \dot{-} p = T$$

Maxichoice contraction satisfies the AGM postulates for theory contraction.

In order to give the *semantics of maxichoice contraction* a semantical parallel of maximal subtheories is required.

### 3.8.1 Supermodel Semantics

Here I define a possible-worlds style downgrade semantics and prove that it is the semantical correlate of the AGM maxichoice contraction function. In order to achieve this, I introduce some new formal notions.

### Definition 19 Ignorant Supermodels

Notation;  $M_{T'} \in M_{T \perp p}$

$M_{T'}$  is a minimal supermodel set of  $M_T$  which fails to support  $p$  (or simply "a supermodel set of  $M_T$  ignorant of  $p$ ") iff.

---

<sup>20</sup>There are plenty of things wrong with this function (see Gärdenfors et al. [1], and Ryan [88])- I merely use it as a simple starting point

1.  $M_T \subseteq M_{T'}$  (supermodels)
2.  $M_{T'} \not\models p$  (ignorance)
3. for any  $M_{T''}$  such that  $M_T \subseteq M_{T''} \subset M_{T'}$ ,  $M_{T''} \models p$  (minimality)

Thus the set of minimal supermodel sets of  $T$  ignoring  $p$ , written  $M_{T \perp p}$ , is the minimally larger set of models (than  $M_T$ ) which support as much of  $T$  as they can without also supporting  $p$ . Note that  $M_{T \perp p} \subseteq M \perp p$  and that  $M_{T \perp p}$  never actually rejects  $p$ . Of course, for each  $T$  and  $p$  there are, in general, many minimal supermodel sets. The problem for a downdate semantics is now which one of these to choose. A simple illustration may prove useful at this point:-

**Example 9** Take  $\mathcal{L} = \{p, q, r\}$  and  $T = \{p, q\}$ .

Then  $M_T$  is the set consisting of the two models  $m$  and  $m'$  where  $Val(p, m) = Val(q, m) = Val(p, m') = Val(q, m') = 1$ , and  $Val(r, m) = 0$  but  $Val(r, m') = 1$ .

This theory supports/contains  $p$  and  $q$ , but is agnostic as to the status of  $r$ . Now suppose  $q$  is to be contracted from the theory (and this is not to say that  $q$  is false, remember.) Syntactically the theory simply reduces to  $T' = \{p\}$ , but semantically things are more fine-grained. There are two minimal supermodel sets which ignore  $q$ . One is the set  $M_{T'} = \{m, m', m''\}$  where  $p, q$ , and  $r$  take the values 1, 0, 0 respectively in  $m''$  (and  $m$  and  $m'$  remain as above). The other is identical but for a model  $m'''$  replacing  $m''$  which is identical to  $m''$  but for  $r$  taking the value 1. Thus  $T' = T \dot{-} p$  has two minimal supermodel sets ignoring  $q$ . Just one of them ought to be selected as the downdate  $M_T \parallel q$ . But how?

The answer provided by minichoice contraction is simply to choose one of the minimal supermodel sets which make up  $M_{T \perp p}$  by using a selection function  $\delta$  (in the same way as Maxichoice Theory contraction uses a selection function  $\gamma$  to choose a maximal subset of  $T \perp p$ ). Once constraints are put on  $\delta$  such that it matches the work done in the syntax by  $\gamma$ , there is a tight connection between Maxichoice Theory Contraction, and Minichoice Downdate Semantics. This is the idea behind **minichoice** downdates

**Definition 20 Minichoice Downdate Semantics (MDS)**

$$M_T \parallel p \models \delta(M_{T \perp p})$$

where  $\delta$  is a selection function on sets of sets of models  $M_{\mathcal{T}}$  (returning a set of models), mirroring  $\gamma$ , the selection fn on sets of theories, such that:

if  $t' = \gamma(T)$  where  $T$  is a set of theories, and  $t'$  is a theory, then

$$Th(\delta(M_{\mathcal{T}})) = Th(M_{t'}) = \gamma(T)$$

In other words,  $\delta$  chooses just that set of models which support the maximal sub-theory selected by  $\gamma$ .

Maxichoice theory contraction and MDS can be connected in the following way:

**Theorem 5 (Completeness of MTC with respect to MDS)**

$$1. Th(M) \dot{-} p = Th(M \parallel p \parallel)$$

and,

$$2. T \dot{-} p = T' \text{ iff } M_T \parallel p \parallel = M_{T'}$$

**Proof:** First part,

$$Th(M) \dot{-} p = \gamma(Th(M) \perp p) = \gamma\{T' \subseteq Th(M) \mid T' \not\vdash p\} = Th\delta\{N \mid N \not\vdash p \text{ and } M \subseteq N\} = Th\delta(M_{T \perp p}) = Th(M_T \parallel p \parallel).$$

Second part;

since  $T$  and  $T'$  are **deductively closed**<sup>21</sup>  $T \dot{-} p = T' \text{ iff } Th(M_T) \dot{-} p = Th(M_{T'})$

Then by part one,

$$Th(M_T) \dot{-} p = Th(M_T \parallel p \parallel)$$

Therefore, by lemma 1.3,  $Th(M_{T'}) = Th(M_T \parallel p \parallel)$

and,  $Th(M_{T'}) = Th(M_T \parallel p \parallel)$

Thus,  $M_{T'} = (M_T \parallel p \parallel)$ , as required.

**Claim:** MDS satisfies both the rationality postulates for downdating and the desiderata on downdate semantics.

<sup>21</sup>Proposals for revisions over non-closed theories (due to limited inferential competence of the agent), as in the literature on “belief bases”, would have to be furnished with a different proof here.



**Justification:** By definition, MDS is non-eliminative, non-implosive, successful, non-explosive, and has integrity. Therefore MDS satisfies the desiderata. Also, since MDS is the semantical analogue of MTC, and MTC satisfies the AGM postulates, then MDS satisfies the semantical versions of those postulates.

However, MDS does not supply a perfect downdating system.

**Problem:**

The set of models remaining after a minichoice downdate is sometimes too small. MDS can remove the partiality of an information state<sup>22</sup>, producing complete theories and unwanted information *growth* after downdate.

Alternatively, then, define a downdate to be the supermodel set which is the union of all the minimal supermodel sets (in the above example this is just the supermodel set  $\{m, m', m'', m'''\}$ ), and this produces a **full join downdate semantics**.

### 3.8.2 Full Join Downdate Semantics (FJDS)

So, in order to escape the problem with MDS, define downdate to be the union of all the supermodel sets which fail to support  $p$ .

**Definition 21 (FJDS)**

$$M_T \Downarrow p \equiv \bigcup (M_{T \perp p})$$

**Problem:** Although FJDS escapes MDS's problem with unwarranted information growth, the system in fact goes too far the other way, producing the largest possible downdates (in fact they are often too large) and thus supporting the smallest remaining theories (in fact they are often too small). In general, information which ought rationally to be conserved may be lost.

**Claim:** FJDS satisfies the rationality postulates and desiderata.

The obvious syntactic connection here is with full meet theory contraction (FMTC), where  $T \dot{-} p = \bigcap (T \perp p)$ .

---

<sup>22</sup>This is simply a reflection of the problems with maxichoice contraction (see the AGM literature) in the semantics. The problem is that  $p \in T \Rightarrow p \vee q \in T \dot{-} p$  or  $p \vee \neg q \in T \dot{-} p$

With  $\perp$  given by FJDS as above, we have (unsurprisingly) the following completeness result:

**Theorem 6** (Completeness of FMTC with respect to FJDS)

1.  $Th(M) \dot{-} p = Th(M \perp p)$
2.  $T \dot{-} p = T'$  iff  $M_T \perp p \models M_{T'}$

**Proof:**

First part;  $Th(M) \dot{-} p = \bigcap (Th(M) \perp p) = \bigcap \{T' \subseteq Th(M) \mid T' \not\vdash p\} = Th \cup \{N \mid N \not\vdash p \text{ and } M \subseteq N\} = Th \cup (M_T \perp p) = Th(M_T \perp p)$ .

Second part; as for previous theorem.

To get a more detailed idea of what's actually going on here, consider the following very simple example.

**Example 10** Take  $\mathcal{L} = \{p, q\}$  and  $T = \{p \rightarrow q, p\}$  (this is a complete theory, which contains  $q$  under closure.)

Then  $M_T$  is the set consisting solely of the model  $m$  where  $Val(p, m) = Val(q, m) = 1$ . This theory supports/contains both  $p$  and  $q$ . Now consider the full-meet contraction of  $q$  from the theory. Maximal subtheories of  $T$  failing to imply  $q$  are  $\{p \rightarrow q\}$  and  $\{p\}$ . Their meet is  $\emptyset$ , the null theory (so that FMTC would be too extreme).

Semantically, minimal supermodel sets of  $M_T = \{m\}$  are  $M$ , containing  $m$  and the model  $m'$  where  $Val(p) = 1, Val(q) = 0$ , and the modelset  $M'$ , which is made up of  $m$  and the models  $m''$  where  $Val(p) = Val(q) = 0$  and  $m'''$ , where  $Val(p) = 0, Val(q) = 1$ . Their join (the FJDS downdate) is the supermodelset  $m \perp q \models \{m, m', m'', m'''\} = Mod(\mathcal{L})$ .

Thus, an FJDS downdate here results in the empty information state.

### 3.8.3 Partial Join Downdate Semantics (PJDS)

Combining the two approaches above results in partial join downdates: the union of a selected set of appropriate minimal supermodel sets. This avoids the problems with

downdates being either too large or too small.

**Definition 22**  $M_T \parallel p \models \bigcup \beta(M_{T \perp p})$

Where  $\beta$  is a selection function choosing some non-empty **subset** of  $M_{T \perp p}$ .

The result is a compromise between the small information states left after contraction by MDS and the large information states produced by Full Join downdates.

**Claim:** satisfies the postulates and desiderata. Proof: by extension of the MDS case.

As above, there is the expected connection between PJDS and partial meet theory contraction (PMTC), where

$T \dot{-} p = \bigcap \alpha(T \perp p)$ <sup>23</sup> and  $\alpha$  selects a set of maximal elements of  $T \perp p$ , as long as we connect the  $\alpha$  and  $\beta$  selection functions<sup>24</sup> in the correct way.

While  $\beta$  selects a subset from a set of sets of models,  $\alpha$  selects a subset of set of theories. (By the earlier definition 5 and subsequent theorem, such functions are perfectly well defined and behaved.)

To elaborate,  $\beta(M_{T \perp p}) = N \subseteq M_{T \perp p}$

and,  $\alpha(T \perp p) = \mathcal{T} \subseteq T \perp p$

Then connect the selection function in the following way:  $Th(\beta(M_{T \perp p})) = \alpha(T \perp p)$

From the results already established it is easily seen that the same connection holds between PJDS and PMTC, thus;

**Theorem 7** (Completeness of PMTC with respect to FJDS)

$Th(M) \dot{-} p = Th(M \parallel p) \parallel$

$T \dot{-} p = T'$  iff  $M_T \parallel p \models M_{T'}$

Proof: trivial since partial join downdates are merely the functional combination of minichoice and full join downdates.

<sup>23</sup>This function is computationally very costly.

<sup>24</sup>Extended versions of the  $\gamma$  and  $\delta$  functions.

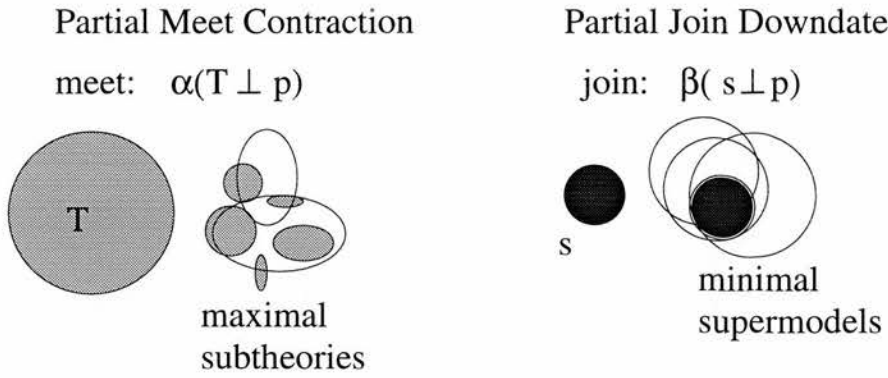


Figure 3.3: Contraction and Downdate

### 3.8.4 Defining Revision

As in the syntax, the following identity can be employed to reduce the problem of finding a suitable semantical revision function to that of finding a downdate function:

where  $\llbracket \cdot \rrbracket$  is a function satisfying DS1-8 and we define  $\llbracket \cdot \rrbracket$  via the identity:

$$s \llbracket p \rrbracket = s \neg p \llbracket \llbracket p \rrbracket \rrbracket$$

then  $\llbracket \cdot \rrbracket$  so defined satisfies RS1-8.

**Proof:** by analogue with the AGM case (the Levi identity).

### 3.8.5 Reprise and critique

So far so good. Given a few simple definitions, in the propositional case, systems of downdate semantics have been found which mirror those syntactic contraction functions explored in the AGM literature. It might be interesting to find a propositional downdate semantics connected to theory contractions constructed by *Epistemic Entrenchment* relations, and various other algorithms (but that is beyond the scope of this work). As far as expressive power goes, the insights gained from the propositional downdate semantics given above are to be extended to first-order dynamic systems.

However, where the modelling of theory revision in communication is at stake we

must recognise that the particular systems presented in this chapter at once fail to confront an important issue in theory revision, and do not go far enough in expressive power. The real issue in theory revision is *obscured* by the use of selection functions and hierarchical theories (as in Epistemic Entrenchment systems), for it is precisely the selection of an appropriate theory (or determination of the hierarchical position of a sentence in a theory) that is at issue in theory revision. Selection functions simply do too much work. Information structures ought to carry with them the properties which allow their contractions and revisions to be determined, rather than have them imposed externally and arbitrarily<sup>25</sup>.

### 3.8.6 A “merely” formal semantics?

A formal semantics ought not to be just another technical system for the classification of strings of symbols. Methodologically speaking, part of the point of giving a semantics is to clarify intuitions underlying symbol processing. If a formal semantics is uninterpretable in any coherent and intuitive way, then it faces the danger of being “merely” formal<sup>26</sup>. So, is “supermodel semantics” *merely* formal? Of course, I shall try to show that the supermodel semantics of theory change is intuitive and coherent.

Think of a set of possibilities (or information state), as an agent’s “commitment slate” — the “sphere” of possibilities which they implicitly claim to inhabit if they espouse a certain theory. Note that there is no necessity to think of these possibilities as possible *worlds* (although they are normally construed in this way); they could just as easily be sets of possible individuals, for example (see the discussion of chapter 6). An agent who is willing to assert a certain theory commits itself to inhabiting a possibility somewhere within the range of possibilities which are not excluded by that theory. Thus, when I claim that “Masja and Tanya are in the kitchen”, I commit myself to inhabiting some set of possibilities, all of which have in common that Masja and Tanya are in the kitchen. An agent which espouses no theory has *S*, the whole universe of possibilities, at its disposal. An agent who espouses a complete theory claims only to inhabit one possibility,

<sup>25</sup>cf. Ryan [88], the iteration problem

<sup>26</sup>The semantics of relevant logics have often had to face this criticism, for example.

which is their version of actuality. The more usual case concerns partial theories and agents who claim to inhabit one of some set of possibilities (they know not which). A supermodel semantics for contraction and revision just describes how this set of possibilities expands and changes under theory contraction and revision. Once I am told that “Masja is not in the kitchen” I change my commitments, to some set of possibilities which only have in common that Tanya is in the kitchen and Masja is not. If an eliminative semantics for theory extension (eg: Update Semantics) is merely formal, then so is revision semantics. If a semantics in terms of “possibilities” is merely formal, then so is revision semantics. However, a general defence of “possible world” semantics (and one would not want to defend *all* such approaches) is somewhat beyond the scope of this chapter!

### 3.9 Application: Revisable Dynamic semantics

Despite their drawbacks, the semantical systems presented above can be used as the core of dynamic semantical systems for retraction and revision of information in dialogue. The bearing of this fruit is the subject of the following chapter.

To anticipate a little, each dialogue contribution is taken to change the individual information states of the agents which hear it. In the following chapters, I concentrate on the minimal case of a two-agent dialogue, although the approach is easily generalizable to cover multi-agent dialogues. If an assertive contribution does not conflict with its hearer’s current information state, then that state is simply updated, via model-elimination, using AGMUS<sup>27</sup>. The interesting work involves cases where either (1) the contribution is not an assertion, but a retraction (triggered by a constituent negation clause, see chapter 7), or (2) where the contribution is an assertion which conflicts with the hearer’s information state.

Case (1) involves downdating, and case (2) involves revision of information states. It will be shown that Downdate semantics (DS) can be added to any existing Update system as a modular addition, preserving all the results of the chosen system. In general,

---

<sup>27</sup>These operations are defined in proper detail at the close of chapter 7.

revisions will be seen as downdate by the negation of the offending sentence followed by update with the new information. This is much in the spirit of the Levi identity. The resulting Revision Semantics has a weak “retraction” negation and provides some new insights into the workings of epistemic modals.

### **3.10 Conclusion**

The main contributions of this chapter have been:

- A review of related research.
- The basic link between theory change and dynamic semantics is forged, explored, and clarified.
- “Downdate Semantics” is defined in general, and desiderata on downdate systems are developed.
- Some motivation is provided for the later application of “Downdate Semantics” in a semantics of dialogue.
- A semantics for theory contraction and revision algorithms is produced.

The completeness results provided here, although simple, give an account of the transformations over information structures underlying AGM theory change. This semantical perspective provides insights for the development of theory change systems, and shows that there are clear semantical ideas underlying ‘belief revision’ systems. Conversely, the theory change perspective dramatically widens the explanatory scope of dynamic systems, and motivates the further exploration of a number of new semantical systems.

Finally I draw a conclusion of a methodological nature on the nature of information states in communication. A central problem is that information states (if they are adequately to meet their explanatory goals) must have an internal structure which allows their (rational) contractions and revisions to be determined (“from the inside”). They must carry with them (rather than have imposed upon them) information which allows

rational information change to be computed. The provision of such structures is beyond the scope of this work, where I shall be content (although not completely happy) to use selection functions. A semantical version of Ryan's "Ordered Theory Presentations" [88], where theories carry enough structure to avoid the use of selection functions, might be a good place to start.



## Chapter 4

# Revision Semantics

AGM - inspired “downdates” open up new possibilities in dynamic semantics.

This chapter explores these new possibilities. It falls into two main sections. In the first part the central aim is to develop, compare, and contrast two approaches in modelling dynamic information structures; “downdating” and paraconsistency, which deal with two problems in standard systems of Dynamic Semantics. This investigation leads to the consideration of growing sets of models (the “supermodels” of the preceding chapter) in a “Revision Semantics”. The focus of the final part of the chapter concerns notions of constructivity relevant to the semantics of communication.

The chapter covers the development of a number of systems of dynamic semantics, and insights into the workings of new epistemic modals. Two problems are raised with regard to the information processing accounts of standard systems of Dynamic Semantics (DS); the “implosion” and “chivalry” problems. In response, two classes of functions over information states are investigated; the *chivalrous* and *robust* functions. The investigation leads to two novel approaches in dynamic systems. The first system (Revision Semantics or ‘RS’), based on the ‘Downdate Semantics’ of the preceding chapter offers revisable information states and a weak ‘retraction’ negation. The other approach (leading to the *Paraconsistent Update Semantics* or ‘PUS’ of chapter 8) is robust under inconsistent input and exhibits a strong negation. These approaches allow distinctions between assertion, denial, and weak denial (retraction) in DS. A consistency preserv-

ing update operation (revision) is also made available in RS, in order to accommodate changing information, from other agents and a changing world. Formal properties of RS are explored in comparison with existing systems of Dynamic Semantics. In chapter 8, which concludes the thesis, the approaches are combined so that both problems find a solution in one system (Paraconsistent *Revision Semantics*.)

In consequence, new insights into the nature of DS information states are offered. In particular a novel treatment of (constituent) negation and epistemic modality is offered. The system also generates a potential logic for (counterfactual) conditionals, and a way of treating “counterfactual” modals.

Consideration of the dynamic account of epistemic modality leads to the final part of the chapter, where various ideas behind a constructive account of epistemic modality are presented.

## 4.1 Two dynamic problems: implosion and chivalry

There are two problems with the account of the dynamics of information states provided by Dynamic Semantics (DS), both of which can be traced back to the dynamic treatments of negation and entailment. These problems need to be tackled if a dynamic semantics is to adequately account for the dynamics of information states under inconsistent input and the retrieval of information from states which contain absurdities. DS needs to be able to process inconsistent input, and more generally, give an account of the dynamics of state *revision*.

Firstly, inconsistencies irretrievably destroy an information state - even when that state contains much information that has no inferential links to sentences responsible for a contradiction. Thus in standard systems of dynamic semantics (US, DPL, EDPL, DMPL etc.), for example, the following dialogue contribution leads to the absurd state.

**Example 11** *“The suspect is tall and male. Sorry, he’s not tall, I meant small.”*

*In US this becomes (something along the lines of)  $s \llbracket \phi \wedge \psi \rrbracket \llbracket \neg\phi \wedge \theta \rrbracket = \emptyset$*

Now notice that as far as the information in this discourse is concerned, obtaining the absurd state, denoted  $\emptyset$ , (due to the state subtraction definition of negation in DS) poses at least two problems. Firstly, continuing on from the above example with the question (interpreted as a test) “Is the suspect blond?”, it so happens that, according to the absurd state, he is. (Since  $\forall \alpha, \emptyset \models \alpha$ .) In other words, the (mis)information “the suspect is blond” is supported by the absurd state, for the absurd state (or empty set of models) supports all propositions<sup>1</sup>. Thus all the previous information (the suspect is male) is swamped by entailments caused by the absurdity. Call this the “Implosion problem” as it is caused by the implosion of information states to  $\emptyset$  under inconsistent input. Implosion is a result of eliminativity (see the preceding chapter). The dynamic account of entailment is responsible for the fact that states support all propositions after implosion. Secondly, the absurd state can not be escaped, or updated with any more information in order to regain consistency. Once the absurd state is reached, no further dialogue contributions can be analysed. In order to solve (or rather bypass) this problem a *robust* way of updating dynamic information states is required. Dynamic Semantics, if it is to model dialogues, needs a robust account of changing information states under inconsistent input.

**Definition 23** (*Robustness*)

Call a function  $f$  on a set of information states  $S$  **implosive** iff.

$$\forall s \in S, f(s) = \emptyset$$

A function is **robust** iff. it is non-implosive ( $f(s) \neq \emptyset$ ).

Dynamic interpretation of all utterances in dialogue (even contradictions) ought to be robust.

The second problem to be addressed in the framework provided by DS concerns a dynamic account of the recovery of a dialogue from contradiction. In eliminative DS, once the absurd state is obtained it cannot be escaped; once a contradiction arises in a dialogue it degenerates irretrievably to absurdity. Thus some kind of dynamic process is required to (non-trivially) rescue a dialogue from a state which contains inconsistencies (call this the “Chivalry problem”). “Chivalrous” functions rescue “states in distress”:

<sup>1</sup> $\emptyset$  supports any formula  $\alpha$ . Due to the definition of entailment in Dynamic Semantics; since  $\emptyset \subseteq \emptyset [\alpha]$ .

**Definition 24** (*Chivalry*)

Call a function  $f$  on a set of information states  $S$  **chivalrous** iff.

$$\exists s \in S \text{ such that } f(\emptyset) = s$$

A chivalrous function is one which retrieves information from states which have suffered implosion.

Obviously, eliminative functions are not chivalrous, and chivalrous functions are non-eliminative. See figure 4.1 for a summary of the general framework here, where the downward arrows indicate the direction of information loss, and the upward arrows show information gain.

That “chivalry is dead in dynamic semantics” is the following fact (of US and EDPL; due to eliminativity of update.)

**Fact 1** (*Chivalry is dead in standard DS*)

Where  $[\cdot]$  is the eliminative update function on information states (of either US or EDPL) and  $p$  is any proposition.

$$\forall p \in \mathcal{L}, \emptyset [p] = \emptyset$$

So there are two problems to be treated if DS is to adequately account for the processing of inconsistent input.

- (1) information states should be *robust* under contradiction, and
- (2) absurdities (once reached) should be escapable (i.e.; some chivalrous functions are required).

It might be thought that a solution to the implosion problem bypasses the issue of chivalry altogether (for then contradictions do not lead to  $\emptyset$ ), but there are dialogue contributions other than straight contradictions which produce the absurd state in DS. These are the following modal tests on states, which result in absurdity (these test are expressible in US and its offspring.)

$$s [\neg p] [\diamond p] = \emptyset \text{ and,}$$

$$\text{when } s \not\models p, s [\square p] = \emptyset$$

So even if a robust handling of contradictions is achieved, chivalry will still be an issue

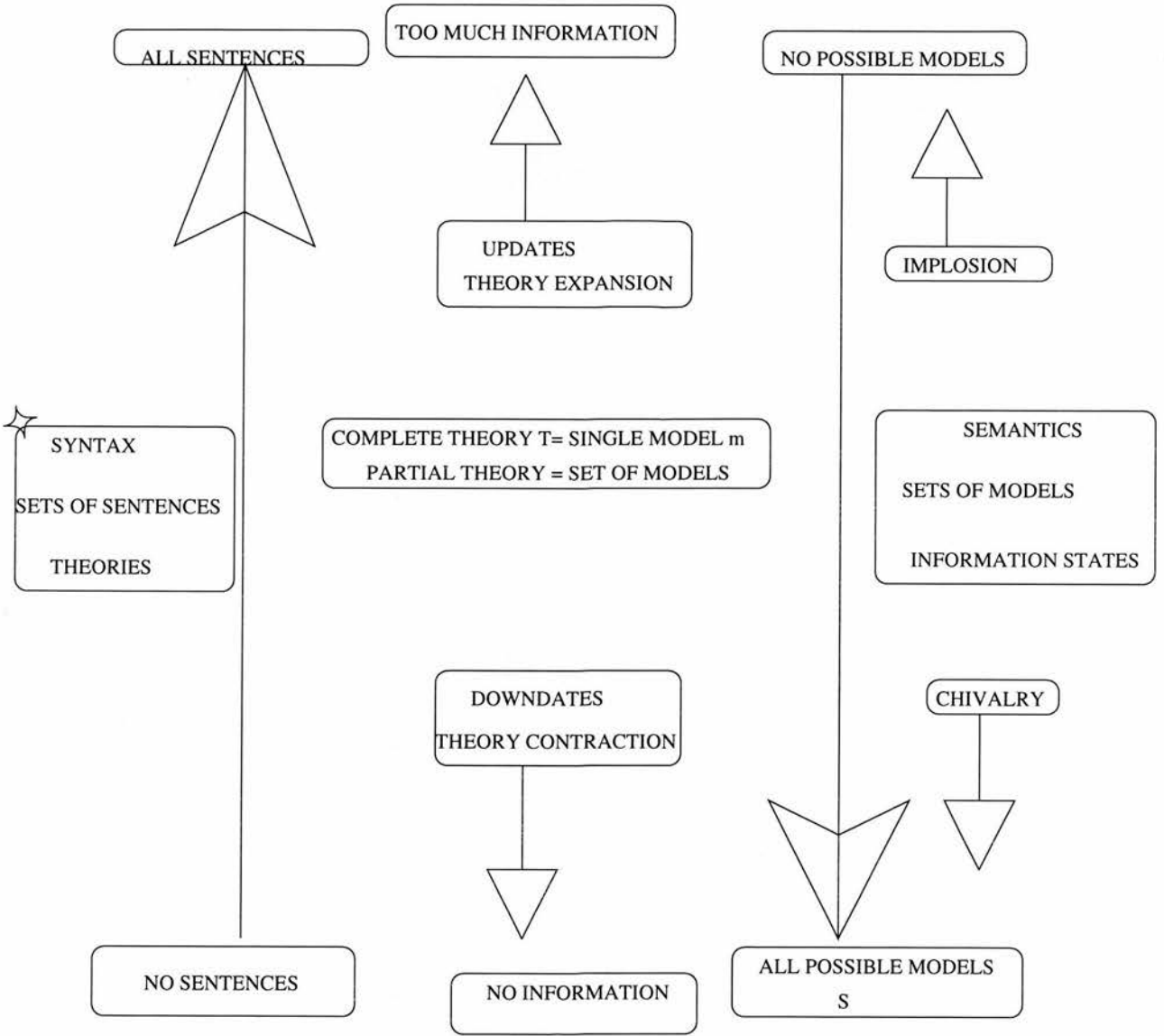


Figure 4.1: Chivalry and Implosion

as regards modal statements. As well as this, and even more importantly perhaps, it seems that an adequate model of dialogue will actually need to accommodate inconsistencies to some degree, rather than totally forbidding them. Conflict and contradiction play a crucial and constructive role in communication (see [38], for example). Information change in dialogue, rather than ignoring inconsistencies, often revolves around contradictions and their resolution. A good model of dialogue ought to be able to detect and resolve inconsistencies, rather than avoiding their occurrence. Some interesting dialogues focus on resolution of inconsistencies. Unless contradictions can be accommodated and then gradually resolved<sup>2</sup>, some dialogues, like the following example, cannot be adequately modelled.

**Example 12** A: "Jack's arriving tonight."

B: "No he's not arriving tonight."

A: "But he promised."

B: "Sorry. He's arriving tomorrow — he just telephoned."

Both the chivalry and implosion problems have a promising potential solution. These solutions make use of ideas from philosophical logic, where they are the research programmes known as "paraconsistent logic" and "belief revision," or "theory change." However, these ideas have not yet been explored in the context of Dynamic Semantics. The first of these approaches advocates avoiding the absurd state (under inconsistent input) by adopting an underlying logic which robustly handles contradictions. This is the solution offered by *paraconsistent logics*, which are investigated in chapter 8. Inconsistencies are *localized* in paraconsistent logics, so as not to infect the state as a whole. However, I postpone the formal treatment of paraconsistent dynamic systems until chapter 8.

The second avenue, that of a dynamic semantics of theory change, leads to a *downdate* operator for dynamic semantics which can, amongst its other virtues, retrieve information states from absurdity.

After a formal treatment of this approach, the extent to which it can solve the above

---

<sup>2</sup>Gradual revisions of inconsistency are treated properly in chapter 8. Of course, agents can also "agree to disagree" and allow a paraconsistent information structure to persist.

problems is ascertained. In addition I determine the degree to which revisable and paraconsistent dynamic systems compete, and finally, how they might be seen to complement each other.

## 4.2 The “Downdating” Approach

What would be the point of constructing a semantics where information can be removed from a state? One specific answer is that such a semantics is required in order to address the chivalry problem. More generally, the answer is that agents often retract (rather than deny) information in conversation.

**Example 13** *“Did I say Kerry was here? No. I actually meant to say that Kelly’s here.”*

*(Retract  $p$ , update with  $q$ .)*

Note that the speaker here is not actually denying that Kerry is present (Kerry might still be there, so the constituent negation is not actually a denial), but merely modifying (correcting) information previously conveyed. The speaker could coherently continue with “Kerry might be here”.

Communication between dialogue partners can be seen as their attempt to converge upon an agreed theory about the world. This convergence requires updating and *revision* of agent-relative information states, the semantical correlates of theories. Thus, if dynamic semantics is to live up to its empirical aims, some notion of information down-date and revision is called for (this has been recognized by [21], and [6].) As Asher mentions, in the context of DRT,

“... the acceptance of the result of interpreting a verbal message may lead to a delineated DRS  $K$  that is inconsistent [ with the hearer’s current information ] ... A further stage of belief integration would consist in a process of “revision” or “contraction” to yield a new, consistent delineated DRS  $K'$  ”

( [6] in [49] p.44)

Perhaps the most obvious case for “downdating” is the chivalry problem; once an absurdity is reached the dialogue must be able to be “rescued” by the retraction of certain statements<sup>3</sup>. In general, however, information is not removed only from the absurd state, but it is removed from arbitrary information states, and arbitrary information states are revised. AGM research suggests criteria that such operations should meet in a syntactic system. The preceding chapter developed semantical correlates of the AGM postulates, as well as systems of downdate semantics meeting them. Such systems can be employed in a semantics which enables an account of a *weak negation*. The notion might also be termed “negation as retraction”, and can be formalized in an Update Semantics with downdates. The resulting system, dubbed “Revision Semantics”, allows DS to account for the kind of dynamics of information states illustrated in the above examples. Consider, also, the following conversation:

**Example 14** *In the course of a long argument John has claimed both that he does, and does not, want to go to a party. The dialogue has reached absurdity.*

*Jackie “Look, do you want to go to this party or not?”*

*John “OK. Forget what I said before. I’d love to go.”*

That standard dynamic accounts cannot handle the above example is a feature of their eliminative account of negation. Negation as state subtraction is too strong for the purposes of information *loss*. To retract commitment to a proposition  $p$  is not the same as strongly denying it. So strong state-subtraction negation in DS needs to be supplemented with a weak contraction-negation, or downdate, which can (amongst its other virtues) rescue dialogues from absurdity. Retraction of information has been studied syntactically (in the propositional case only) in the AGM theory change literature. This is the obvious place to start building a downdate system for DS. It is fairly easy to construct a range of downdate functions ] . [ on information states which are the semantical correlates of the AGM theory contraction systems (this was the burden of the preceding chapter).

---

<sup>3</sup>Often the sentences directly responsible for the contradiction. For example in teaching contexts, where a student is made to realize that she has contradicted herself, and then repairs her model of the problem domain. See eg: the work of [9] on an implementation of Theory Change which teaches (repairs misconceptions in the formulation of) algebraic word problems.



### Chivalry regained

Notice that the initial definition of chivalry is not quite good enough. It is simple to define a chivalrous “reset” function “ $\uparrow$ ” such that  $s \uparrow = S$ , which only downdates a state  $s$  at the expense of trivializing it. However, such an operator would not bring “true chivalry” to DS.

#### Definition 25 (True Chivalry)

Call a function  $f$  on a set of information states  $S$  **truly chivalrous** iff.  $f$  is chivalrous and  $f(\emptyset) = s \neq S$

#### Definition 26 (Downdate)

Information state  $s'$  is a downdate of state  $s$  iff  $s \subset s'$ .

#### Definition 27 (Partial Join Downdate)

Information state  $s'$  is a partial join downdate of state  $s$  iff  $s' = \bigcup \beta(s \perp p)$  for some proposition  $p$ .

#### Fact 2 (Partial Join Downdates are Chivalrous)

$\llbracket \cdot \rrbracket$  is chivalrous.

( $\llbracket \cdot \rrbracket$  is not chivalrous)

This is a trivial consequence of the definition of downdates. DOWNDATING the absurd state by proposition  $p$  results in some state  $s$  which fails to support  $p$ . Thus  $\emptyset \llbracket p \rrbracket = s$  such that  $s \not\models p \neq \emptyset$ .

But what is to stop  $\emptyset \llbracket p \rrbracket = S$ ? The minimality condition on supermodels ensures that (partial join) downdates are in fact *truly* chivalrous.

#### Example 15 Take $\mathcal{L} = \{p, q\}$

Then  $\emptyset \llbracket p \rrbracket = \bigcup \beta(\emptyset \perp p) = p : 0, q : 1$  or  $p : 0, q : 0$  depending on  $\beta^4$ . Whatever, the result is not  $S$ , the empty state.

---

<sup>4</sup>Note that the union of these models would not be a minimal supermodels set of, since it fails to support either  $q$  or its negation.

## Syntax of Revision Semantics

Add a weak negation “ $\neg p$ ” and a revision operator “ $\uparrow p$ ” to update semantics (US). Add both symbols to the syntax of US as meta-logical constants which are banned (via syntactic restrictions) from appearing inside well-formed formulae<sup>5</sup>.

Add the downdate and revision functions  $\Downarrow : S \times FORMS \rightarrow S$  and  $\Uparrow : S \times FORMS \rightarrow S$  respectively, as defined below in the semantics.

Note that the definition of downdate in the semantics developed below relies (in the base case) on systems of *Downdate Semantics* developed in the preceding chapter<sup>6</sup>.

### 4.2.1 Revision Semantics

A family of systems of *Revision Semantics* results from the combination of Update Semantics [101] with “downdates”. Depending on the choice of downdate function, different systems of RS result. The relative merits of different downdate systems were explored in the preceding chapter. The system of RS presented below employs partial join downdates. If the usual update definition of interpretation in DS is replaced by revision as the function which interprets assertions, then contradictions are excluded from information states, and the implosion problem never arises.

#### Definition 28 REVISION SEMANTICS

Where  $s - s' = \{i \in s \mid i \notin s'\}$  (State subtraction)<sup>7</sup>

Models for Revision Semantics are of the form  $M = \langle S, V \rangle$  where  $V(p) \subseteq S$  for each proposition  $p$ .

For  $s \subseteq S$ :

---

<sup>5</sup>This frees the system from unnecessary complications over interpretation of  $\neg \neg p$  and  $\diamond \uparrow p$  for example. A treatment of such expressions can be given relatively easily, but it obscures the main point of the system.

<sup>6</sup>These systems are complete with respect to various AGM theory contraction algorithms

<sup>7</sup>This state subtraction definition leads to an implosive negation in US.

Update clauses	Downdate Clauses
$s \llbracket p \rrbracket = \{i \in s \mid i \in V(p)\}$	$s \llbracket p \rrbracket = \bigcup \beta(s \perp p)$
$s \llbracket \neg p \rrbracket = s - s \llbracket p \rrbracket$	$s \llbracket \neg p \rrbracket = \bigcup \beta(s \perp \neg p)$
$s \llbracket -p \rrbracket = s \llbracket p \rrbracket$	$s \llbracket -p \rrbracket = s \llbracket p \rrbracket$
$s \llbracket \uparrow p \rrbracket = s \llbracket p \rrbracket$	
$s \llbracket p \wedge q \rrbracket = s \llbracket p \rrbracket \llbracket q \rrbracket$	$s \llbracket p \wedge q \rrbracket = s \llbracket p \rrbracket \cup s \llbracket q \rrbracket$
$s \llbracket \diamond p \rrbracket = \{i \in s \mid s \llbracket p \rrbracket \neq \emptyset\}$	$s \llbracket \diamond p \rrbracket = \{i \in s \mid s \llbracket p \rrbracket \neq S\}$
$s \llbracket \Box p \rrbracket = \{i \in s \mid s \llbracket p \rrbracket = S\}$	$s \llbracket \Box p \rrbracket = \{i \in s \mid s \llbracket p \rrbracket = \emptyset\}$
<b>Revision clause</b>	
$s \llbracket p \rrbracket = s \llbracket \neg p \rrbracket \llbracket p \rrbracket$	

Note the two base clauses in the recursion for downdates (the reason for this is that there is no simple way to define  $s \llbracket \neg p \rrbracket$  in terms of an operation on  $p$ <sup>8</sup>) This presents little problem as long as the base logic allows negations to filter through formulae to the right. (The usual de Morgan Laws and modal interdefinability results ensure this for classical systems.)

A possible downdate clause for “-p” might state that to retract the retraction of p is the same as asserting p. That is,  $s \llbracket - - p \rrbracket = s \llbracket p \rrbracket$ . I consider the clauses for the modals, in some detail, shortly.

Figure 4.2 shows how a revision operates. A state which supports *not - p* expands, under retraction of *not - p*, to a new state which includes some “worlds” where *p* holds. Under update, the revised state becomes just those “worlds” in  $s \llbracket p \rrbracket$  which support *p*. Recall the following:

**Definition 29 (US Entailment)**

$$s \models_{US} p \text{ iff } s \subseteq s \llbracket p \rrbracket$$

This is simple; a state already supports a proposition if adding that information to the state does not increase the information in the state (by making it “smaller”).

<sup>8</sup>Although one could try  $s \llbracket \neg p \rrbracket = \downarrow \llbracket p \rrbracket \cup s$ , which is the union of  $s$  with all the points in  $S$  which are non-absurd under update by  $p$ .

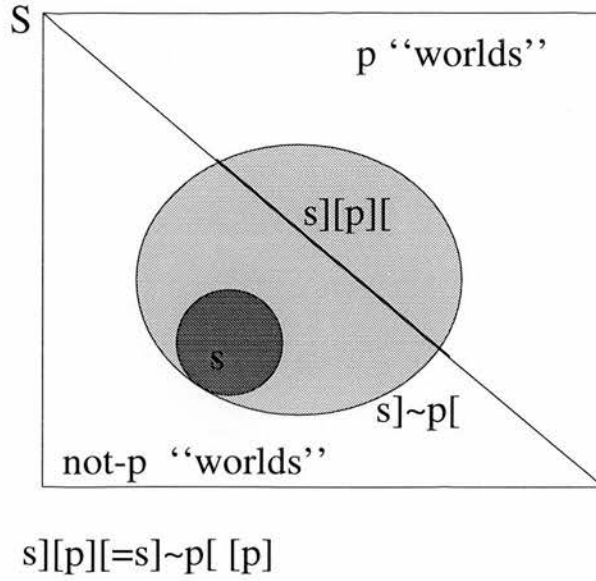


Figure 4.2: Revising  $s \models \neg p$  by  $p$

Revision Semantics (RS) allows a consistency preserving interpretation function for assertions,  $\llbracket \uparrow p \rrbracket$ . This virtue might lead one to consider a “robust” definition of entailment for RS.

**Definition 30 (RS Entailment)**

$$s \models_{RS} p \text{ iff } s \subseteq s \llbracket p \rrbracket$$

The idea being that a state supports a proposition if and only if *changing* the state so as to accept that proposition doesn’t actually add information to the state. Now notice that this definition has quite similar consequences to that of US. Except that  $s$  revised to include  $p$  is non-absurd (unless  $p$  itself is contradictory), whereas  $s$  updated by  $p$  can result in the absurd state. Therefore, even if  $p$  is inconsistent with  $s$ ,  $s \llbracket p \rrbracket$  exists.

**Theorem 8 ( $US \subseteq RS$ )**

$$s \models_{RS} p \text{ if } s \models_{US} p$$

Proof:  $s \models_{US} p \Rightarrow s \subseteq s \llbracket p \rrbracket \Rightarrow s \subseteq s \llbracket \neg p \rrbracket \llbracket p \rrbracket \Rightarrow s \models_{RS} p$

Note also that  $s \not\models_{US} p \Rightarrow s \not\models_{RS} p$

For if, also,  $s \not\models_{US} \neg p$ , then  $s \llbracket p \rrbracket = s \llbracket p \rrbracket$ . In the case that  $s \models_{US} \neg p$  then  $s \llbracket p \rrbracket = s'$  such that  $s' \cap s = \emptyset$ . Since  $s \not\subseteq s'$ , if  $s \models_{US} \neg p$ , then  $s \not\models_{RS} p$

This shows that using the consistency-preserving update of RS to define entailment (rather than the implosive update of US) does not alter US entailment properties, which is just as it should be.

### 4.2.2 Propositional Theory Change

It follows from the results of the preceding chapter that the system presented above (RS with partial meet contraction) is complete with respect to an AGM (propositional) theory change system (with partial meet contraction.)

**Definition 31** (*Propositional Theory Change*)

*Expansion clauses:*

$$T + p = Cl(T \cup \{p\})$$

$$T + \neg p = Cl(T \cup \{\neg p\})$$

$$T + (p \wedge q) = (T + p) + q$$

$$T + (p \vee q) = T + p \cap T + q$$

$$T + \diamond p = T \text{ iff } T + p \neq \mathcal{L}$$

$$T + \Box p = T \text{ iff } T \dot{-} p = \emptyset$$

$$T + (-p) = T \dot{-} p$$

$$T + (\uparrow p) = T \dot{+} p$$

*Contraction clauses:*

$$T \dot{-} p = \bigcup \alpha(T \perp p)$$

$$T \dot{-} \neg p = \bigcup \alpha(T \perp \neg p)$$

$$T \dot{-} (p \wedge q) = T \dot{-} p \cup T \dot{-} q$$

$$T \dot{-} (p \vee q) = (T \dot{-} p) \dot{-} q$$

$$T \dot{-} \diamond p = T \text{ iff } T \dot{-} p \neq \emptyset$$

$$T \dot{-} \Box p = T \text{ iff } T + p = \mathcal{L}$$

*Revision Clause:*

$$T \dot{+} p = (T \dot{-} \neg p) + p$$

**Theorem 9 (Completeness of RS with respect to Propositional Theory Change)**

$$Th(s_T \llbracket p \rrbracket) = Th(s) \dot{+} p$$

$$T \dot{+} p = T' \text{ iff } Mod(T) \llbracket p \rrbracket = Mod(T')$$

Proof: by composition of the results of the preceding chapter (US clauses are complete with respect to theory extension, and “downdates” model the theory contraction clauses. Then compose contraction/downdate and extension/update to get the result for revision).

### 4.2.3 Epistemic Modals in Revision Semantics

Downdates are interesting in their own right, not just as a treatment of the implosion and chivalry problems, but as the principal component of a fully-fledged RS, which offers a novel treatment of the epistemic modals as well as a weak denial version of negation and revisable information states. Now that there is a non-eliminative downdate function on information states, it can be used to define tests against the *trivialization* of an information state after downdating, just as the eliminative update offered by US allows a test against *absurdity* of a state after update. In this way, the addition of the downdate component to a dynamic semantics allows the expression of a whole new set of constraints on information states in the semantics; constraints invoking *trivialization* (rather than absurdity) of information as undesirable.

The idea that the trivial (or uninformative) information state is just as undesirable as an absurd one could not previously be expressed in dynamic semantics. Moreover, the following ideas can now formally be expressed, in RS:

- $p$  is **indispensible** in  $s$  iff  $s \llbracket \Box p \rrbracket = s$  ( $= \emptyset$  otherwise)

Contracting  $s$  by  $p$  would produce the trivial state.

- $p$  is **disposable** in  $s$  iff  $s \diamond p \llbracket = s (= \emptyset \text{ otherwise})$   
Contracting  $s$  by  $p$  would not lead to triviality.
- $p$  is **prohibited** in  $s$  iff  $s \square p \llbracket = s (= \emptyset \text{ otherwise})$   
Updating  $s$  by  $p$  would lead to absurdity.
- $p$  is **possible** in  $s$  iff  $s \llbracket \diamond p \rrbracket = s (= \emptyset \text{ otherwise})$   
Updating  $s$  by  $p$  does not lead to absurdity.

All of the above operations are *tests* on information states. They do not change the state itself (unless the test fails, in which case the state is reduced to  $\emptyset$ , the absurd state), but are used in checking meta-level properties (consistency and “informativeness”) of a state with respect to addition and retraction of propositions.

The formal definitions of these new epistemic modals are reiterated below.

### Definition 32 (RS epistemic modals)

1.  $s \llbracket \diamond p \rrbracket = \{i \in s \mid s \llbracket p \rrbracket \neq \emptyset\}$
2.  $s \llbracket \square p \rrbracket = \{i \in s \mid s \llbracket p \rrbracket = S\}$
3.  $s \diamond p \llbracket = \{i \in s \mid s \llbracket p \rrbracket \neq S\}$
4.  $s \square p \llbracket = \{i \in s \mid s \llbracket p \rrbracket = \emptyset\}$

Syntactic analogues of these operations are easily expressible in theory change systems.

### Comparison with the standard US modals

A potential point of conflict between RS and US is their respective accounts of “epistemic necessity”. The original idea, in “Data Semantics” (see [100]), was that “must  $p$ ” expresses that there is no extension of the (speaker’s) current theory in which “ $p$ ” turns out to be false. Formally:  $s \llbracket \square p \rrbracket = s$  iff  $\neg \exists s'$  such that  $s' \subseteq s$  and  $s' \models \neg p$ . This condition is met if and only if  $s \models p$ .

In other words, according to US, the statement “must  $p$ ” acts as a test on the current state, to the effect that the state supports proposition  $p$ . Such a construal is supposed to

make “must  $p$ ” a weak epistemic claim; agents can utter “must  $p$ ” when  $\neg p$  is excluded from all possible updates<sup>9</sup> of their current information state. Thus, the account goes, agents can utter “must  $p$ ” when they do not actually know (yet) that  $p$ . However, this weak epistemic claim is lost in US and EDPL for example<sup>10</sup>, where the standard clause is,

$$s \llbracket \Box p \rrbracket = \{i \in s \mid s \models p\}$$

Such modals are better understood in communicative contexts<sup>11</sup>(see chapter 6), as Dekker argues in [21], where “must  $p$ ” is tantamount to a question; an attempt to establish agreement between dialogue participants. All a dialogue partner can do when faced with “must  $p$ ” is agree or disagree. However, the US clause has the effect of classifying contingent statements as “epistemically necessary” if it so happens that they are supported by an agent’s current information. In US an agent who believes that “Masja is in the kitchen” can consistently utter “Masja *must* be in the kitchen”, but this does not satisfy my intuitions. Agents sometimes utter “Must  $p$ ” in contexts where removal of  $p$  would reduce them to an empty state, not just where update by “not –  $p$ ” would lead into absurdity. The RS clause  $s \llbracket \Box p \rrbracket = \{i \in s \mid s \llbracket p \rrbracket = S\}$  covers this context. RS provides a stronger criterion for classifying a proposition as epistemically necessary than does US. Not only must the proposition in question be supported by the current state ( in RS:  $s \not\models p \Rightarrow s \llbracket \Box p \rrbracket = \emptyset$ ), but it must also be indispensable to that state, in the sense that its removal would leave the agent “high and dry” as regards information. RS thus classifies a subset of US’s “epistemically necessary” propositions as “epistemically indispensable”.

## Examples

As examples of each of the above kinds of propositions consider the philosophical system espoused by Descartes in his “Meditations” [25].

Descartes’ “cogito ergo sum” is an example of an *indispensible* proposition, relative

<sup>9</sup>But obviously not revisions.

<sup>10</sup>See eg: [21] p. 202

<sup>11</sup>Indeed, this is already implicit in Data Semantics.



to the philosophical system of the “Meditations”, since all other propositions follow from it (or presuppose it in some sense). The class of indispensable propositions is usually very small (or is even empty) for the majority of information states<sup>12</sup>. Another example, for some thinkers, might be the existence of God.

In contrast, very many propositions seem to be (epistemically) disposable. For example “Mind-body interaction occurs in the pineal gland” is a claim that can be dispensed with in Descartes’ system without trivializing it. Disposable propositions are commonly (but not exclusively) to be found at the periphery of a theory<sup>13</sup>.

Of course there are also plenty of prohibited sentences in a given information state. Any sentence in the complement of the theory will be prohibited. Again for Descartes, for example, take “I do not exist.”

The “possible” or “permissible” sentences form a more interesting class. All the sentences which are not forbidden by a state are possible relative to that state. In complex theories, it is not often easy to decide whether or not a sentence is forbidden or permitted. For example, could “The mind is the brain” be added to the Cartesian Meditations without contradiction? This example requires a first-order treatment, with identity.

Note that the usual interdefinability results are still intuitively correct on this new construal of the modals. For example, consider  $[\neg\Diamond p] \iff [\Box\neg p]$ , which says that not- $p$  is indispensable iff  $p$  is not possible. Also consider, for example,  $[\neg\Diamond p] \iff [\Box\neg p]$ , which says that not- $p$  is prohibited iff  $p$  is not disposable.

The revision operator can also be used to express some quite sophisticated compound tests on information states. For example,  $s \uparrow [\Diamond p] = s \uparrow [\neg\Diamond p] \iff [\Box p] = s$  iff not- $p$  is prohibited and  $p$  is possible.

Various other combinations of the modals shall be explored later in this chapter. Note also that

<sup>12</sup>Disregarding the obvious case of the conjunction of all propositions supported by a state, which, trivially, is always epistemically indispensable.

<sup>13</sup>One could grade propositions in respect of the relative damage that their removal does to a theory.

$$\Box\Box p = \emptyset, \Box\Diamond p = \emptyset, \Diamond\Box p \Rightarrow \Box p, \Diamond\Diamond p \Rightarrow \Diamond p, \Box p \Rightarrow p \Rightarrow \Diamond p, p \not\Rightarrow \Box p$$

#### 4.2.4 Properties of Revision Semantics

Many of the results below hold in general (that is, for systems other than the specific RS given above) as long as a truly chivalrous downgrade function is available. The non-eliminativity (chivalry) of downdates ensures the desired;

**Fact 3** (*Genuine downgrade property*)

$$s \Downarrow p \not\subseteq s \Downarrow \neg p$$

This was one of the most basic desiderata on any useful notion of downdating (see the preceding chapter). That it holds in RS is trivial:  $s \Downarrow p$  is a superset of  $s$ , and  $s \Downarrow \neg p$  is a subset of  $s$ .

Note also, that trivially,  $\forall p, S \Downarrow p \Downarrow S$

So losing information when there isn't any around anyway changes nothing. This is just as it should be.

As far as update properties are concerned, all the desirable features of US (as presented in chapter 2) are also properties of RS since:

**Fact 4** *Revision Semantics is an extension of Update Semantics.*

(This fact follows just from the fact that the update clauses of RS are just those of US.) The next fact shows that downdating only removes the information that it is supposed to.

**Fact 5** *Conservation of Information:  $\forall q$  such that  $s \models q$  and  $p \not\models q$ ,  $s \Downarrow p \models q$*

So downdating with respect to  $p$  only removes  $p$  and its consequences from the state. Interestingly, downdates allow the removal of contradictions from the absurd state.

**Fact 6** *Non-implosive downdates:  $s \Downarrow p \wedge \neg p \Downarrow s \neq \emptyset$*

So removing a contradiction is non-absurd. However, removing a logical law from a state is absurd.

**Fact 7** *Absurd / Implosive downdates: if  $\vdash p$ , then  $s \Downarrow p \Downarrow = \emptyset$*

This shows that, even though downdating is a way of escaping the absurd state, certain downdates (eg: attempting to remove a theorem) do lead to absurdity. Finally, as proven above,

**Fact 8** *RS entailment incorporates US entailment.*

#### 4.2.5 Ups and Downs in Discourse and Dialogue

The following results explore the interaction of up-and-downdates, and explain the un-informativeness, information loss, and absurdity of certain epistemic discourses (absurd or implosive discourses are marked by \*). It is perfectly *consistent* for agents to retract propositions<sup>14</sup> which they asserted earlier, although it is not particularly *coherent* for an agent to assert  $p$  and then retract it (in discourse for example). The operations provided by RS are, of course, best understood as part of a full system of *information exchange*, where an agent uttering a retraction does so in order to disabuse a dialogue partner of perceived mis-information (see chapter 6).

Call a function  $f$  on a state  $s$  *absurd* iff  $f(s) = \emptyset$ . In general, updates potentially lead to absurdity.

Call a dialogue contribution  $f$  on a state  $s$  *incoherent* iff  $f(s) = S$ . In general, only downdates may lead to incoherence.

(Different contributions will be absurd, or incoherent, for different agents<sup>15</sup>.) In a successful dialogue, agents need to steer a course between absurdity and incoherence.

The following results indicate that the epistemic modals interact in the ways one would expect. It is as well to keep in mind that even though epistemic modal formulae appear between assertion and retraction brackets, they are not treated as genuine updates, but are *tests* on the current state. In this sense, the notation is somewhat misleading. Intuition dictates that assertion and retraction of epistemic modals ought to

<sup>14</sup>As long as they are not logical truths.

<sup>15</sup>Although no contribution could be absurd for one agent whilst (at the same time) being incoherent for another.

change the state in some circumstances. This intuition is investigated further shortly. Another point to note in the following examples is that, where one test follows another, we assume that the first test succeeds before evaluating the second one.

$$1. s [\diamond p] ] p [\supseteq \text{ sif } \exists i \in s, i \models p$$

This discourse has the potential to induce information loss.

Discourse: "Maybe p. No, I don't think that p."

Contrast with "Maybe p. In fact, p." — which can lead to information growth.

$$2. s ] p [ [\diamond p] \supseteq \text{ sif } \exists i \in s, i \models p, \exists j \in s, i \models \neg p.$$

Again, possible information loss (due to the initial downgrade).

Discourse: "I don't think that p. Maybe p"

$$3. s [\diamond p] ] \diamond p [= s \text{ as long as } p \text{ is consistent with } s \text{ and } p \text{ is not indispensable in } s.$$

Discourse: "Maybe p. I can remove p."

$$4. s ] p [ [\square p] = \emptyset$$

Discourse: "\*I can remove p. It must be that p."

$$5. s [\neg p] [\square p] = \emptyset$$

Discourse: "\*Not p. Must be that p."

$$6. s [\square p] ] \diamond p [= \emptyset$$

Discourse: "\*It must be that p. I can forget p."

$$7. s ] \diamond p [ [\square p] = \emptyset$$

Discourse: "\*I can forget p. It must be that p."

$$8. s [\square p] ] \square p [= \emptyset$$

Discourse: "\*Must be p. Cannot be p."

$$9. s [\diamond p] ] \square p [= \emptyset$$

Discourse: "\*Maybe p. Cannot be p."

$$10. s [\square p] [\diamond p] = s \text{ if } p \text{ is supported by } s:$$

Discourse: " Must be p. Maybe p."

11.  $s \Vdash \Diamond p \Vdash \Box p \Vdash s$  if  $p$  prohibited in  $s$ .  
Discourse: "Can forget  $p$ . Must not be that  $p$ ."
12.  $s \Vdash \Diamond p \Vdash [\Diamond p] = s$  if  $p$  dispensable in  $s$  and consistent with  $s$ .  
Discourse: "Can forget  $p$ . Maybe  $p$ ."
13.  $s \Vdash [\Diamond p] \Vdash \Diamond p \Vdash s$  if  $p$  consistent with  $s$ , but not indispensable to  $s$ .  
Discourse: "Maybe  $p$ . Can forget  $p$ ."
14.  $s \Vdash [\Diamond p] \Vdash [\Box p] = s$  if  $p$  supported by  $s$ .  
Discourse: "Maybe  $p$ . Must be  $p$ ."

The examples given above have been *discourses* (ie: uttered by one agent), but the epistemic modals come into their own during communication. If the above examples were *dialogues*, then rather than rendering absurdity, they would establish the extent to which the participants *agree* on certain information. Modal expressions used in dialogue have the effect of testing the hearer's state for certain properties. Thus (as sketched in Dekker [21]), use of modals allows a speaker to become informed about the current information state of the hearer.

#### 4.2.6 A Logic for Conditionals

The idea of revision can be used to generate a logic of counterfactual conditional statements. The basic idea is that "if  $p$  then  $q$ " can be construed as a test to the effect that " $q$  is a consequence of the theory after revision by  $p$ ." If  $\neg p$  is not supported by the state, then this conditional reduces to that of US, where  $s \Vdash [p \rightarrow q] = s$  iff  $s \Vdash [p] \models q$ . However, the interesting case arises when the state itself holds that  $\neg p$ , for then US maintains that  $p \rightarrow q$  for all  $q$ <sup>16</sup> Thus, where the antecedent of the US conditional is inconsistent with the current state, US classifies the conditional as "true". However, this prescription does not meet with common intuition about the interpretation of conditionals in natural language. Agents can evaluate conditionals whose antecedents are contrary to their current information. The revision clause of RS allows a simple treatment of both normal US conditionals, and counterfactuals too (in the case that the antecedent is not

<sup>16</sup>Since then  $s \Vdash [p] = \emptyset$  and  $\forall q, \emptyset \models q$ .

contrary to the current state, the revision function reduces to update, so the conditional is evaluated as a normal US conditional).

Consider adding to RS the following clause for a counterfactual conditional:

$$s \llbracket p \mapsto q \rrbracket = s \text{ iff } s \llbracket p \rrbracket \models q$$

Syntactically, in the theory extension system, the clause looks like this;

$$T \vdash p \mapsto q \text{ iff } q \in T \dot{+} p$$

Comparison with other approaches to counterfactuals (eg: [74], [100]) might prove to be interesting, but it is unfortunately beyond the scope of the current work.

### 4.3 Paraconsistency versus Chivalry

Adoption of a paraconsistent base logic (or “internal logic”) for theories increases the complexity of dynamic information states (by invoking positive and negative classifications of formulae). However, a paraconsistent semantics does avoid the implosion problem altogether (by localizing inconsistencies). While a paraconsistent approach has the virtue of avoiding implosion, and thus the chivalry problem, virtue never tested is no virtue at all; a paraconsistent semantics simply allows contradictions to persist. Now this feature of paraconsistent systems is not really adequate for a semantics of dialogue as intersubjective theory manipulation. A paraconsistent semantics still requires down-dates in order for modal inconsistencies (failed tests) to be resolvable, and the general point about information retraction still holds; an empirically adequate DS ought to allow for repair, and retraction of prior input. Presumably some way of reducing the number of contradictions in a (paraconsistent) information state is required, rather than just letting them accumulate. As well as being robust, a paraconsistent semantics needs to at least try to maximize consistency of its information states; and for this it also requires chivalry (again, see chapter 8).

On the other hand, RS is (truly) chivalrous and preserves (and even extends) the US treatment of the epistemic modals. Rather than requiring a change in the whole under-

lying logic of a DS, the downdating approach is a module which can be “plugged into” any existing system of DS (the clause for  $\llbracket -p \rrbracket$  is the “socket”). This means that,

**Fact 9** *Revision Semantics is an extension of Update Semantics.*

In the sense that all results that hold for US also hold in RS (since the US account of  $\llbracket \diamond p \rrbracket$  is preserved in RS).

Unfortunately, RS is still implosive under contradictions; licensing any inference at all after an absurdity has been reached (it’s just that, by using revision rather than update, the absurd state is a lot harder to come by). Thus, just as a paraconsistent dynamic semantics requires chivalry, so a full Revision Semantics for dialogues ought to localize inconsistencies.

#### 4.4 Constructivity in Dynamic Semantics

There is a potential confusion to be avoided, between the concept of constructivity as employed in Intuitionistic and Nelson logics, and the notion intended here. “Constructivity” in the sense of downdates and revisions of information states, refers to the construction of models (or possibilities) in a semantical system. This feature ought not to be confused with the debate over a constructive negation, which is a feature of the paraconsistent systems developed later (in chapter 8).

A constructive (or “possibility-creating”) account of conditionals and epistemic modality does justice to the intuitions and philosophical motivations behind Dynamic Semantics. Such an account both complements and supersedes the eliminative approach of recent tradition.

Meaning, for DS, is identified with the information *change* potential of a sentence. Thus DS currently prescribes only utterances which serve to eliminate epistemic possibilities as informative or, indeed, meaningful. But this is obviously inadequate, for language is more often than not used to *construct*, or bring to light, new possibilities for the consideration of other agents. We could distinguish (along with Kripke) between the *epistemic modals* (those which inform about the beliefs of other agents) and modals which inform about how the world might be (or have been) different; the “*Metaphysical*

*Modals.*" Earlier, I argued against importing this distinction into formal semantics. Systems of dynamic semantics which rely on model elimination as their sole method of information growth, or utterance interpretation, actually rule out such utterances as "Maybe it's Jackie" as being properly meaningful (in the sense of being "informative"). Such statements are interpreted as *tests* on information states, which either succeed (with no effect on the current information state), or fail and reduce the state to absurdity. Thus, "Maybe it's Jackie" means, according to standard DS, either nothing at all or results in absurdity. But this prescription is plainly wrong. Why say, "Maybe I'll see you at the party" if the utterance is going to be either empty or incoherent? The treatment of modals offered by DS is clearly at odds with semantical and linguistic intuition; there are "*metaphysical*" modal statements<sup>17</sup> whose content relates to how the world might be, as well as the standard epistemic modals treated by DS. That purely eliminative accounts have sufficed until now has perhaps been due to the absence of any real alternative to the eliminative perspective.

The problems here have much to do with the focus on discourse rather than dialogue. In discourses (where information accumulates), the eliminative treatment of modals seems fairly well motivated, but once DS is extended to model dialogue (where information changes and is exchanged), a purely eliminative interpretation of modals becomes untenable. Recall, however, that the Kripkean distinction between epistemic and metaphysical modality was brought into question (in chapter 1). As far as a semantics of dialogue is concerned, it is important to realize that all the modals are epistemic, in an important sense. For even those supposedly "*metaphysical*" modals which express how the world might have been, only do so with respect to an agent's information state. When an agent states "He might have survived" the intention is that, as far as the agent knows or believes, things may have turned out differently. Of course, the big difference between this and the standard case is that the agent knows that "not p" is the case, while stating "might have been p." The case seems very similar to the counterfactual conditionals mentioned earlier. One might think of some occurrences of  $\diamond$  as a "counterfactual modal". Treatment of such statements does not require an extra set of "*metaphysically possible worlds*", but an approach to some modals (eg: those which appear

---

<sup>17</sup>[47] recognizes this.



in constituent negation clauses or apply to the past tense) as conjectures about consequences of theory change, or state revision.

This shortcoming of standard DS reveals a deeper problem with the current dynamic perspective. The semantics may be dynamic, but its subject matter is a static world. DS assumes that once an update is accepted in a state there is no question of removing its effect or updating with contrary information. However, a changing world throws out changing information. An adequate DS needs to be able to deal with this fact, and the “metaphysical” modals play an important part here for, rather than conveying information about an agent’s information state (the job of the epistemic modals), they convey info about how the world might be or might change, relative to an agent’s information state.

The claim here is not that there is no role for eliminativity to play in formal semantics, but that there is constructive work to be done at the same time.

#### 4.4.1 A constructive account of epistemic modality

Update Semantics handles growing information with eliminative functions, but chivalrous functions in Revision Semantics can cope with *changing* information<sup>18</sup>. Work can now be done against the following (current) situation in dynamic semantics:

“our approach to information change is eliminative rather than constructive. And we don’t know yet how to do better than that.”  
(Groenendijk, Stokhof, and Veltman, [47] p.2)

The “chivalry” of downdates allows a more radical treatment of information change and modality in RS. Rather than preserving the *test* account of epistemic modals from eliminative DS one can begin to give the kind of *constructive* account of information change alluded to in the above quotation. For example, rather than seeing “might p” as an uninformative test on a current state, it can be construed as an action, a genuine assertion to the effect that p is possible, by using downdates.

---

<sup>18</sup>Not just temporally dependent changes in the environment, but changing information supplied by fallible communicants

Let CRS (Constructive Revision Semantics) be RS supplemented with the following (constructive) account of epistemic possibility.  $s \llbracket \diamond p \rrbracket = s \llbracket \neg p \rrbracket \parallel p \llbracket$

Some assertions of “maybe p” in s may now be viewed as constructing an *extension* of s in which neither “p” nor “not p” are supported by the state. In a sense, this construal of “maybe p” performs the function of “opening” the state for p. Thus a subsequent update by either p or  $\neg p$  will be permissible. In US however, if p is supported by s, then the discourse “Not p. Maybe p.” is absurd. CRS, on the other hand, construes “Maybe p” as *resetting* the information in the state with regard to p, so that even with prior knowledge that not-p “Maybe p.” is non-absurd.

Such a step is not meant to supplant the treatment given in US, but to complement it. It seems that there is a variety of ways in which “might”, “maybe”, and “must” are used in dialogues. There ought to be a corresponding range of formal treatments of these epistemic modals: some as tests, and some as actions. Use of the (“dynamic”) CRS modal, for example, might be restricted to cases where “maybe p” appears in the scope of constituent negation or in a “counterfactual” context.

**Example 16** “John’s not here. No, actually, he might be here.”

Whilst US classifies this example as absurd, CRS allows it to be informative. Thus, CRS allows some epistemic modals to reflect *temporal change*.

This approach constitutes a new slant on “informative” statements, which previously were taken to be simply the eliminative ones. Downdates allow chivalrous or constructive statements to be treated as “informative” in the sense that processing them *changes* the current information state. This procedure is, after all, closer to the original philosophical slogan of DS of meaning as *change* in information state.

## 4.5 Conclusion

This chapter makes the following contributions to the thesis.

1. Two problems for DS are raised.

2. Chivalrous and Robust functions are defined.
3. Revision Semantics is presented and explored in some detail.
4. Novel ideas about negation and epistemic modalities are presented and explored.
5. The paraconsistent approach is motivated, and its relationship to systems of theory change and dynamic semantics established.
6. A notion of constructive DS is discussed.

The construction of RS provides new (semantical) insights into the nature of the information states required for an empirically adequate dynamic semantics. New properties of information states (and functions on them) have become formally expressible through the employment of the supermodel semantics of the preceding chapter. In particular the treatment of epistemic modals in RS employs the new dynamic dimension which downdates supply .

Downdates and paraconsistent systems have been seen to complement each other. The approaches can be combined to produce "Paraconsistent Revision Semantics"; a robust DS which accommodates gradual repair of localized absurdities (the main subject of chapter 8).

In general terms, the chapter demonstrates that an empirically adequate dynamic semantics must attend to the fact that information structures conveyed in communicative utterances are *paraconsistent* and obey instructions regarding the *downdate and revision* of information structures as well as their update.

The problems raised here, and the systems investigated, highlight serious issues for the structure and nature of information states in dynamic semantics. The approach commends further systems for future study.

The issue of paraconsistency shall become the focus of chapter 8, where paraconsistency is considered in tandem with the rational *autonomy* of agents who can be sceptical about accepting new information.

## Chapter 5

# First-Order Theory Change Systems with Semantics

Cognitively relevant information processing is often argued to have at least the complexity of first-order logic (for example see Kamp [62], Asher [49]). Accordingly, predicates and variables are distinguished in the processing of information in natural language interpretation. In order to meet the demands of such applications systems of theory change and dynamic semantics must operate over structures expressed in first-order logic.

As Peter Gärdenfors mentioned in discussion [43]<sup>1</sup>,

*“So far, I have only been working with propositional languages. However, the epistemic semantics could be extended to quantificational languages as well. The main problem would be to handle revisions of the domain of individuals.”*

Of course, for the purposes of dialogue modelling, it is essential that theory change systems are extended to the first-order case.

While the work in earlier sections has been confined to the dynamics of propositional systems, the formalism now increases in expressive power, to incorporate first-order dynamics.

The chapter is organized as follows. Firstly it is demonstrated that that Gärdenfors’ completeness result (in [42], reviewed in the second chapter) can be extended to the first-

---

<sup>1</sup>With Frank Veltman, Jim Delgrande, Michael Clarke, and Bob Moore. Reported in [43], page 237.

order case. The result is that AGM-style first-order theory extension (FOTE) is complete in relation to the semantics of the Intuitionistic Predicate Calculus (IPC). This fact leads to a consideration of how downdates might be defined in an intuitionistic semantics. Secondly, intuitions are given about a set of first-order rationality postulates for contraction, covering quantifiers, names, variables, and identity. Some of the intuitions are formalized. Following this, a variety of systems for first-order theory change (FOTC) are given (some employing Fuhrmann's *Package Contractions* from [35], introduced in chapter 2), along with their semantics. The supermodel semantics of chapter 3 is extended to the first-order case.

In general, the class of problems involved in a first-order revision semantics is found to be similar to that encountered in the construction of a Quantified Modal Logic<sup>2</sup>, and a semantics for counterfactuals<sup>3</sup>.

A hierarchy of systems is then explored, using variations on the theme of changing dialogue-internal and external information (anaphoric and truth-conditional information), and changing information about identity statements. For example, a revisable DPL is shown to be connected with a particular FOTC system. Employing the same methodology, after examining a variety of systems, a revisable FUL (van Eijck's First-Order Update Logic [96]) is connected with a different system of (First-Order) Theory Change.

## 5.1 First-Order Extension and Intuitionistic Semantics

Firstly it is demonstrated that, before one becomes entangled with the dynamics of anaphora, a simple first-order theory *extension* (FOTE) system is complete with respect to the intuitionistic predicate calculus (IPC). This should come as no great surprise when one recalls Gärdenfors' proof<sup>4</sup> that Propositional Theory Extension is complete with respect

---

<sup>2</sup>Otherwise known as Modal Predicate Logics. See James Garson's "Quantification in Modal Logic", in [37], for a useful review of these issues.

<sup>3</sup>Recall that Grove's sphere semantics [48] for revision is similar to Lewis's semantics for counterfactuals [74].

<sup>4</sup>Pages 138 - 142 of [42].

to IL. (This result was reviewed in the second chapter.) An explicitly verificationist semantics is adopted. To call a proposition ‘true’ is simply to have observed it to be so, or to have been told that it is so (agents are gullible in the sense that they uncritically adopt new information and continue to believe it come what may). Therefore, resulting in *monotonicity*, it is assumed that verifications *persist*; that received truths remain true. Of course, as regards theory *change*, non-persistent systems (where agents are gullible in the sense that they uncritically accept new information at the expense of older information<sup>5</sup>) are later investigated.

Syntactically speaking, a *theory*  $T$  held by an agent is simply a deductively closed set of first-order sentences which have been adopted by an agent. For the moment, assume that theories simply make reference to names (of course, theories in dialogue use variables too, but they are more the subject of the dynamic systems which I investigate in the body of the chapter). Such a theory grows by addition of sentences and closure under logical consequence; as usual:

**Definition 33 AGM First-Order Theory Extension (AGMFOTE)**

$$T + \phi = Cl(T \cup \{\phi\})$$

Semantically, the epistemic states of an agent are modelled by way of a *partially ordered* set  $\mathcal{T}$ , elements of which correspond to possible knowledge situations of an agent. Lower points in the order encode less knowledge. Thus the elements  $t \in \mathcal{T}$  serve to index theories in the order of their growth. With each point  $t \in \mathcal{T}$  we also associate a set  $Cons_t$  of names of individuals referred to by state  $t$  (sometimes I refer to this set of names as the *domain of the theory*  $T_t$ ), and a *total* valuation function  $V_t$ . Thus  $t = \langle Cons_t, V_t \rangle$ <sup>6</sup>. A theory  $T_t$  is also associated with each point in  $\mathcal{T}$ ; the set of sentences verified at  $t$ . Thus it makes little difference whether one talks of the state corresponding to  $T_t$  or the state at  $t$ . Both are defined, for either  $T_t$  or  $t$ , as the set of points  $t'$  such that  $t \leq t'$ . Write the state at  $t$  as  $s_t$ . The state at  $t$  is not just the characteristic function of membership of  $T_t$ , but includes all the possible extensions of  $t$ .

<sup>5</sup>A characterization of *scepticism* is given in chapter 7, where agents choose whether or not to accept new information.

<sup>6</sup>This structure is much like a DRS; it has a domain of referring expressions and conditions on them.

**Definition 34** (*Atomic Valuations*)

$V_t : ATOM_{\mathcal{L}} \Rightarrow \mathbf{2}$  assigns truth values to atomic propositions, relative to each possible epistemic state  $t$ .  $V_t(\phi) = 1$  when  $\phi$  has been accepted by the agent whose state it is. Thus,  $V_t$  is the characteristic function of  $T_t$ .

Consider the following functions:

**Definition 35**  $Th(s_t) = \{\alpha \mid \forall i \in s_t, V_i(\alpha) = 1\}$

$s_t = Mod(T_t) = \{t \in \mathcal{T} \mid \forall \alpha \in T_t, V_t(\alpha) = 1\}$

This means that  $s_t \subseteq \mathcal{T}$ . Now it is simply the case that  $Th(s_t) = T_t$ . As far as the *growth* of such knowledge states is concerned I assume that agents are able to verify<sup>7</sup> more and more sentences involving more and more objects. Once a sentence is verified it remains so (*Truth Persistence*), and the set of names grows with theory extension (agents increase the set of referring expressions). Formally then, impose the following conditions;

(1) for all atomic sentences  $\phi$ ,  $\forall t' \geq t, V_{t'}(\phi) \geq V_t(\phi)$  (*Truth Persistence*)

(2)  $\forall t \forall t'$ , if  $t \leq t'$  then  $Cons_t \subseteq Cons_{t'}$  (*Vocabulary Growth*)

To sum up, a sequence of growing theories (sets of sentences),  $T_1 \subseteq T_2 \subseteq \dots$ , is modelled by a corresponding sequence of shrinking states (sets of models),  $s_1 \supseteq s_2 \supseteq \dots$

**Evaluation clauses**

Recall that, in Intuitionistic Logic,  $\neg Pa$  is defined as  $Pa \rightarrow \perp$ . Intuitionistic Predicate logic has the following valuation clauses for complex formulae. Define  $Val_t : FORMS_{\mathcal{L}} \Rightarrow \mathbf{2}$  so that;

<sup>7</sup>For alternatives here, see Milne 1991, [75]

**Definition 36 Semantics of Intuitionistic Predicate Calculus<sup>8</sup>**

1.  $\forall t, V_t(\perp) = 0$
2. for atomic  $\phi$ ,  $Val_t(\phi) = V_t(\phi)$
3.  $Val_t(\phi \vee \psi) = \max\{Val_t(\phi), Val_t(\psi)\}$
4.  $Val_t(\phi \wedge \psi) = \min\{Val_t(\phi), Val_t(\psi)\}$
5.  $Val_t(\phi \rightarrow \psi) = 1$  iff  $\forall t' \geq t, Val_{t'}(\phi) \leq Val_{t'}(\psi)$
6.  $Val_t(\forall x Px) = 1$  iff  $\forall t' \geq t, \forall c \in Cons_{t'}, Val_{t'}(Pc) = 1$
7.  $Val_t(c = d) = 1$  iff  $\forall t' \geq t, \forall P \in PRED(\mathcal{L}), V_{t'}(Pc) = V_{t'}(Pd)$

The last three definitions invoke meta-conditions on the structure, which impose constraints on possible extensions of the current state  $t$ .

Alternatively, write the system in terms of the sentences supported by states. Write

$t \models \phi$  when state  $t$  supports sentence  $\phi$ .

$t \models \phi$  iff  $Cons(\phi) \subseteq Cons_t$  and  $Val_t(\phi) = 1$

**Definition 37** Where  $Cons(\phi)$  is the set of names used in formula  $\phi$ .

1.  $t \models \phi \vee \psi$  iff  $t \models \phi$  or  $t \models \psi$
2.  $t \models \phi \wedge \psi$  iff  $t \models \phi$  and  $t \models \psi$
3.  $t \models \phi \rightarrow \psi$  iff  $\forall t' \geq t, t' \models \neg\phi$  or  $t' \models \psi$
4.  $t \models \forall x Px$  iff  $\forall t' \geq t, \forall c \in Cons_{t'}, t' \models Pc$
5.  $t \models (c = d)$  iff  $\forall t' \geq t, \forall P \in PRED(\mathcal{L}), t' \models Pc$  iff  $t' \models Pd$

This way of putting things enables the definition of an update function on states:

**Definition 38** (*Update*)

$$s_t \llbracket \phi \rrbracket = \{t' \in s_t \mid t \leq t' \text{ and } t' \models \phi\}$$

<sup>8</sup>As presented in [75]



Thus  $T_t + \alpha = T_{t'}$  such that  $Val_{t'}(\alpha) = 1$  and  $t \leq t'$

Intuitively, addition of expression  $\alpha$  to a theory has the effect, in the semantics, of moving up to the next node which supports  $\alpha$  in the intuitionistic information growth tree.

Define *survival in structures*,  $S_{\mathcal{T}}$ , for IPC in the following way;

**Definition 39** (*Structures and Survival*)

$S_{\mathcal{T}}$  is a structure iff.  $S_{\mathcal{T}} = \langle \mathcal{T}, \leq, \{V_t : t \in \mathcal{T}\} \rangle$

Following Milne [75], say that  $\phi$  survives in  $S_{\mathcal{T}}$  when  $\forall t \in \mathcal{T}$ , if  $Val_t(\phi) = 1$  when all names in  $\phi$  are in  $Cons_t$ <sup>9</sup>.

$\phi$  \*-survives if  $\forall S_{\mathcal{T}}, \phi$  survives.

Let  $\Gamma$  be a finite set of sentences. Then  $\Gamma \models \phi$  survives in  $S_{\mathcal{T}}$  when  $\forall t \in \mathcal{T}$ , if  $\forall \psi \in \Gamma, Val_t(\psi) = 1$  then  $Val_t(\phi) = 1$  (again, when all names in  $\Gamma$  and  $\phi$  are in  $Cons_t$ .)

$\Gamma \models \phi$  \*-survives if  $\forall S_{\mathcal{T}}, \Gamma \models \phi$  survives.

The following result is based on Milne [75].

**Fact 10**  $\Gamma \models \phi$  \*-survives if and only if  $\Gamma \vdash_{IPC} \phi$  where  $\vdash_{IPC}$  is the deducibility relation of Intuitionistic Predicate Calculus.

**Theorem 10** (*Completeness of First-Order Extension wrt Intuitionistic Information Growth*)

Where  $+$  is the AGM theory extension operation for first-order theories and  $[[\cdot]]$  is information growth (update) as defined above for the semantics for IPC:

$$(1) Th(s_t [[\phi]]) = T_t + \phi$$

$$(2) T' = T + \phi \text{ iff } s_{T'} = s_T [[\phi]]$$

(Proof)

First part: by definition,  $Th(s_t [[\phi]]) = Th\{t' \in s_t \mid t \leq t' \text{ and } t' \models \phi\}$   
 $= Th\{t' \mid t' \models \phi \text{ and } t' \models \alpha, \forall \alpha \in T_t\}$

---

<sup>9</sup>Dropping this condition results in a free logic.

$$= Cl(T_t \cup \{\phi\}) = T_t + \phi$$

Second part:

$$T' = T + \phi \text{ iff } Th(s_T) + \phi = Th(s_{T'}) \text{ iff (by part 1) } Th(s_T \llbracket \phi \rrbracket) = Th(s_{T'}) \text{ iff } s_{T'} = s_T \llbracket \phi \rrbracket$$

This completes the basic case. The other clauses ought not to present further problems.

Then for theory expansion, as an easy consequence of the above:

**Theorem 11**  $T + \phi \vdash_{IPC} \alpha$  iff  $Th(s_t \llbracket \phi \rrbracket) \models \alpha$  \*-survives.

Proof: assume that  $Th(s_t \llbracket \phi \rrbracket) \models \alpha$  \*-survives. Then by the above fact  $Th(s_t \llbracket \phi \rrbracket) \vdash_{IPC} \alpha$  and so by the preceding theorem  $T_t + \phi \vdash_{IPC} \alpha$

For the converse, simply reverse the argument.

### 5.1.1 Intuitionistic Downdates

The above system can easily be used to define an intuitionistic notion of downdate. When instructed to remove  $\phi$  one could “look back” down  $\mathcal{T}$  to the first point  $t$  where  $Val_t(\phi) = 0$ . Revision could then be described as “jumping” from branch to branch of an intuitionistic tree. For example, one could find a “lower” node where the contracted information does not hold, and then ascend some branch leading off that node to a point where as much as possible of the information prior to downdate is regained (a rather arduous process). Thinking of each branch of  $\mathcal{T}$  as a “stack” of information states, there is an analogy with the work of researchers such as Zeevat [111], Vermeulen [104], and Jaspars [60], where similar structures are employed. These approaches advocate storing an information growth history with which agents can retrace their steps. However, such an approach would restrict agents to considering only the theories held prior to a contraction, thus losing all intermediate information. For an extreme example, if  $\alpha$  is the first

sentence in my theory<sup>10</sup> (at point  $t_1$ ) and by  $t_n$  I have also learnt  $\phi_1 \dots \phi_n$ , and I then have to contract by  $\alpha$ , the naive approach of simply retracing my steps will put me in  $t_0$ , where I know nothing at all. To repeat part of the argument of chapter 3, this is simply not rational (failing the maxim of conservation of information). Agents should be able to conserve (most of)  $\phi_1 \dots \phi_n$ . (In fact they should be able to conserve all the  $\phi_i \not\vdash \alpha$ ). The problem here has to do with the basic monotonicity of IL, and the fact that the *order* of information growth is encoded in an Intuitionistic semantics. However, in AGM research, the order in which expressions are added to a theory is irrelevant, save that a current revision instruction overrides all preceding information. In order to escape these limitations, AGM-style contractions are now employed.

## 5.2 Levels, Systems, and Rationality Postulates

Both the syntactical work on theory change for theories defined over first-order languages, and systems for modelling such change, are taken care of a little later. The general approach might be thought of as constructing an “experientable logic” based on investigation of a changing world. That is a logic where statements can be verified, falsified, or undecided, and where such classifications are defeasible, in the sense that verifications and falsifications are not (necessarily) persistent (ie: there is non-monotonicity of valuations). One might also think of such a logic as describing, not a changing world, but an agent’s changing information about the world, perhaps as conveyed by other, differently informed, agents.

In the propositional case, of course, AGM theory revision systems meet many of the requirements on such a logic. However, thinking about predicate logical theories also forces the development of new rationality postulates for names, the quantifiers, and identity statements. One trivial way of producing a revision system over first-order theories is to restrict contractions and revisions to atomic propositions, and simply continue to use propositional AGM methods. Indeed, this is the method used in [9], where theory change is implemented as an educational tool. However, such an approach merely sidesteps some real issues in first-order theory change. The system which I present em-

---

<sup>10</sup>A favourite example, due to Peter Milne, is the case of forgetting one’s own name.

bodies a similar approach. Theories are taken to be defined over a fixed domain of constants, or names. Thus, universal quantification reduces to conjunction of propositions, and (non-anaphoric) existential quantification is construed as disjunction. Rather than computing such contractions and revisions in a piecemeal fashion (eg: for universals, as a series of individual contractions), the theory of *general contractions* is used to calculate them all at once (eg: as contraction of a *set* of sentences).

### Levels

First though, it pays to be explicit about the different levels at which the various systems operate, and the connections between those levels. In general, three stages of formalism are employed in theory change and revision semantics, and with two particular operations<sup>11</sup>. The three levels are,

- (1) the *observation or input language*, in which the theories are expressed,
- (2) the *change language*, in which the operations of *expansion* and *contraction* are invoked over sets of input language sentences, and
- (3) the *semantics*, which describes the effect that instructions in the change language have on the contents of theories.

The object language connectives  $\wedge$  and  $-$  (dynamic conjunction and retraction) correspond to the change language operations  $+$  and  $\dot{-}$  (expansion and contraction) respectively, and the semantical functions ‘up’ and ‘down’  $\llbracket \cdot \rrbracket$  and  $\lceil \cdot \rceil$  on information states, respectively.

### Systems

FOTC systems allow the formal expression of many different intuitions about change. Indeed the range of possible systems is a little bewildering, given the number of pa-

<sup>11</sup>Revision can be treated as a derived operation, which could be seen as the “meta-language” analogue of a consistency preserving conjunction. Such a conjunction (call it  $\hat{\wedge}$ ) could be defined  $A \hat{\wedge} B = (A - (\neg B)) \wedge B$ .

<i>Extension</i>	<i>Semantics</i>	restrictions	Information Growth
FOTE	DPL	single model	anaphoric information
FOTE <sub>III</sub>	DMPL	exhaustive and full over Cons	anaphoric & external information
FOTE <sub>III</sub>	FUL	Cons exhaustive, canonical models	anaphoric & external & identity

Table 5.1: Dynamic semantics vs Theory Extension

rameters available for change. This chapter navigates a particular course through the landscape of FOTC systems, using van Eijck’s pioneering work [96] on first-order theory extension systems as a guide. In order to chart this course some of van Eijck’s basic assumptions are adopted. The prospects for different FOTC systems and their semantics are explored at the close of the chapter.

To begin with, consider the propositional TC and RS systems of the preceding chapters. There, changing partial theories were modelled by changing sets of models, or worlds. The semantics is rather more involved in the first-order case, due to the complication of variable assignments. In DPL, for example, information increase is modelled by way of eliminating possible variable assignments (rather than possible worlds) relative to a *single* model. If attention is restricted to one model in this way, information change relative to the model can be described by way of growing and shrinking sets of variable assignment functions. Where the model is fixed, there is a connection (due to van Eijck) between DPL and an FOTE system (see table 5.1), which I extend to a connection between Revisable DPL (RDPL) and a FOTC system (see table 5.2). However, restriction to one model does not allow evaluation of modal expressions, or update by “discourse external” information. For that to be possible, information structures must be made up of both sets of assignments (DPL states) and of “worlds” (US states)(see eg: Dekker [21]). In this hybrid case, different FOTE systems are connected with DMPL and FUL (see van Eijck [96]), and a FOTC system with a revisable DMPL and a “first-order revision logic” (FRL). Again, see table 5.2.

There seems to be a “missing” system in the above progression(table 5.1): a first-order update semantics (FUS), dealing solely with growing *discourse-external* information in the semantics of (modal) predicate logic. Such a system would define updates

<i>Theory Change</i>	<i>Semantics</i>	Information Change
FOTC	RMPL	external information
FOTC <sub>I</sub>	RDPL	anaphoric information
FOTC <sub>II</sub>	RDMPL	anaphoric & external information, "static" identity
FOTC <sub>II</sub>	FRL	anaphoric & external information, "dynamic" identity

Table 5.2: Dynamic semantics vs Theory Change: information processing

<i>Theory Change</i>	<i>Semantics</i>	restrictions
FOTC	RMPL	single assignment functions
FOTC <sub>I</sub>	RDPL	single model
FOTC <sub>II</sub>	RDMPL	exhaustive and full over Cons, Canonical models
FOTC <sub>II</sub>	FRL	exhaustive over Cons, Canonical Models

Table 5.3: Dynamic semantics vs Theory Change: formal restrictions

over sets of first-order models, with a shared domain, and single input and output assignment functions. FUS would be complete with respect to an FOTE system which ignores the dynamics of anaphora. In fact, FUS and its corresponding theory extension system are the "information growth" part of the systems RMPL and FOTC given below.

Concerns over anaphora, although important in dynamic semantics, are not central to a (first-order) logic of theory change. Thus, the first system developed disregards "dialogue internal" information, instead concentrating on processing changing information about the world.

### 5.2.1 Intuitions behind the Rationality Postulates

The basic motivations behind the development of first-order contraction postulates are discussed below.

- **(Names)**

The set of naming expressions of a theory ( $Cons(T)$ ) grows under utterance interpretation. Thus, the set of names in a theory grows and persists, although it can be the case that some names are found to refer to the same object, or even to be non-referring. Even in the case of non-existence claims, the naming expression involved is not lost. All in all then, contractions and revisions can only *add* to the set of names of a theory. Names can become redundant (or “fall into disuse”) through acceptance of identity statements, but they are never lost. So,  $Cons(T)$  can only grow, even during theory contraction.

Names are assumed to designate rigidly (they are assigned to the same individuals across models) in any information state, considered statically. However, rigid designation also holds for names under theory change; they always denote the same individual after change as they did in the preceding theory. In other words, one can revise a theory like so: “John is not in Greece, he’s in Canada”, which changes the properties associated with John, but not the actual individual referred to. This models changing knowledge of the reference of theoretical terms (more shall be said about such cases in chapter 6).

- **(Variables and Anaphora)**

Variables, in dynamic semantics, are associated with natural language’s pronouns. In theories, each variable is associated with a name from the set of names of the theory. Under change, a variable can become associated with a different name, thus modelling change in anaphoric information (the information about which name or individual a pronoun stands for.) The denotation of variables is thus, as expected, non-rigid under change.

Anaphoric change is effected either by simple sentential revision (in a revisable Dynamic Predicate Logic, for example), or by revision of identity statements. As an example of the latter, consider, “One of them was ill. It was John. No, it was Jackie.”

- **(Predicates)**

Theories could be said to have growing sets of predicates (descriptive terms, perceptual capacities, experimental techniques), as well as growing sets of names. The same kinds of consideration apply to  $Pred(T)$ , the set of predicates of a theory  $T$ , as apply to  $Cons(T)$  (apart from rigid designation). Thus, even under contraction, the set of predicates of a theory may grow.

- **(Existential Quantifiers)**

Think of the contraction  $T - \exists xPx$  as removal of the disjunction  $Pa \vee Pb \vee \dots \vee Pn$  where  $Cons(T) = \{a, b, \dots, n\}$ , so that theory  $T$  no longer implies  $\exists xPx$ . This amounts to removal of *all* the atoms of the disjunct, and so is clearly analogous to Fuhrmann's notion of *Package contraction*<sup>12</sup>. Thus, take *retraction* of existential statements to have a finite domain of application; the constants currently employed in the theory.

However, in dynamic semantics, existential quantification is taken to introduce a fresh variable as the referent of subsequent pronouns. In this case, contraction of an existential statement is taken to remove the variable as a candidate for anaphoric reference. Thus there are two strategies to choose from. Where dialogue-external information is at stake, contraction of existential claims is interpreted as Package Contraction. Where anaphoric (or dialogue internal) information is of interest, retraction of an existential effectively "kills" its variable. Where both types of information are of concern, Package Contraction can be used to interpret statements like "No, none of them were ill," while the anaphoric "death" of a variable can be used to interpret utterances such as "No, there is no thief." The first of these utterances could continue, "They were all hallucinating" (because the individuals referred to by "they" are still available for reference), while the second could not be added to by "He is innocent."<sup>13</sup>

- **(Universal Quantifiers)**

Similarly, contraction of  $\forall xPx$  is removal of the conjunction  $Pa \& Pb \& \dots \& Pn$ , so that  $T$  no longer licences the inference of  $\forall xPx$ . This requires the removal of at least one of

---

<sup>12</sup>See [35] chapter 3

<sup>13</sup>Unless "he" is interpreted as referring to some other individual.



the atoms of the conjunct<sup>14</sup>, and is clearly analogous to Fuhrmann's notion of *Choice Contraction*.

- **(Identity)**

Rather than statements concerning absolute or metaphysical identity, *identity on the basis of available information* is of central interest in the following systems. That is, identity statements are *relative* to the descriptive resources (predicates) of the current theory. Recall that the focus is on theories as epistemic approximations for an agent, so that rather than having any direct metaphysical import, theoretical terms refer, in the first place, to "objects" in epistemic states<sup>15</sup>. Identity claims between names are thought of as licencing the inference that the named objects share properties. Thus identity is relative to a set of predicates, modelling a theory's descriptive capabilities (the more predicates at our disposal, the more fine-grained our distinctions can be.) Thus retraction of an identity statement requires that the terms involved no longer fall under all the same predicates. However retraction of an identity statement is not to be equivalent to the assertion of non-identity, so it is simply required that one or other of the terms involved becomes underspecified for some property or properties<sup>16</sup>, in comparison with the other term<sup>17</sup>. This all means that the 'objects' that are of interest here ("dialogue referents") are *partial* and may change their identity relations to other individuals. These "objects" are more properly investigated towards the close of the thesis.

### Name and Predicate Postulates

- N1  $Cons(T \dot{-} \phi) \supseteq Cons(T)$  (**Inclusion**)
- N2 if  $\phi \in T$  then  $Cons(T \dot{-} \phi) = Cons(T)$  (**Vacuity**)

---

<sup>14</sup>For an example along the lines of Quine-Duhem theory revision, take the case where a new observation contradicts a universal claim.

<sup>15</sup>This fits in with the more philosophical considerations adduced in chapter 1, and shall be elaborated upon later.

<sup>16</sup>Removal of a bare identity statement obviously leaves an agent in quite an epistemic predicament. They have to make some (arbitrary) choice of contraction of a predication for one of the terms involved.

<sup>17</sup>This can be achieved by "contracting" one term in respect of a certain property.

N1 says that the set of names does not shrink under theory contraction. This becomes plausible when one considers the following example;

Where  $T = \{\forall xPx\}$ , with the contraction  $T \dot{-} Pc$

Then  $c \in Cons(T \dot{-} Pc)$

However, removal of existential claims might seem to warrant the removal of a name or referent. This possibility is considered later in the chapter.

N2 says that names are invariant under contraction by sentences which are already in  $T$ .

For predicates, and for very similar reasons;

- P1  $Pred(T \dot{-} \phi) \supseteq Pred(T)$  (**Inclusion**)
- P2 if  $\phi \in T$  then  $Pred(T \dot{-} \phi) = Pred(T)$  (**Vacuity**)

### Quantifier Postulates

Inclusions ( $T \dot{-} \forall xPx \subseteq T, T \dot{-} \exists xPx \subseteq T$ ), Vacuity, and Success are all covered by the general case (ie: the AGM postulates). However, I make the treatment explicit by noting the following Success requirements.

- Q1  $\forall c \in Cons(T), Pc \notin T \dot{-} \exists xPx$  (**Success**)
- Q2  $\exists c \in Cons(T), Pc \notin T \dot{-} \forall xPx$  (**Success**)

Q1 says that contraction by an existential quantifier amounts to removal of all names from the denotation of the relevant predicate. (Package Contraction)

Q2 says the converse for universal quantification (Choice Contraction).

### Identity Postulates

Again, Inclusion, Vacuity, Success ( $c = d \notin T \dot{-} (c = d)$ ), and Recovery are each covered by the general (propositional) case.

- I1  $T \dot{-} (c_1 = c_2) \not\vdash \neg(c_1 = c_2)$ , unless  $\neg(c_1 = c_2) \in T$  (**Partial Identity**)

- I2  $\exists P$  such that  $Pc_1 \in T^{\dot{-}}(c_1 = c_2)$  and  $Pc_2 \notin T^{\dot{-}}(c_1 = c_2)$ , or  $Pc_2 \in T^{\dot{-}}(c_1 = c_2)$  and  $Pc_1 \notin T^{\dot{-}}(c_1 = c_2)$  (**Distinguishability**)

I1 states that contraction of an identity statement doesn't lead to the stronger claim of non-identity (unless that claim was already part of the theory.) For example;

**Example 17** "I used to think that Batman was Clark Kent, but now I'm not sure."

This statement does not commit the speaker to the claim that Batman is *not* Clark Kent.

I2 (Distinguishability) says that retraction of identity requires that there is an available predicate  $P \in Pred(T)$  which distinguishes the constants (one of them has property  $P$ , but the other doesn't).

### 5.3 First-Order Theory Change

I demonstrated earlier that first order theory *extension*, disregarding anaphora, brings about Intuitionistic Predicate Calculus. Unfortunately, the intuitionistic framework was found to be unsuitable for modelling rational theory change, due to its basic monotonicity. Again, before any entanglement with anaphora, a semantics for first-order theory change systems can be developed by adding to (modal) predicate-logical semantics so that verifications and falsifications are now *defeasible* (or non-persistent).

The language under consideration at first is first-order logic without identity, but with strong and weak (contraction) negations, denoted " $\neg$ " and " $\dot{-}$ " respectively (the former is in the "theory-internal" logic, the latter is an operator in the "external logic" of change). Recall that theories are *partial*; it makes little sense to talk of contractions (or expansions) in complete theories, for then  $T^{\dot{-}}p = T + \neg p$ . If, in the semantics, attention is restricted to assignments in a single model, then the only information that can be gained is about variable assignments relative to that model (eg: in DPL). In order to avoid this limitation for a basic FOTC system, I shall consider sets of models, defined over a fixed shared domain  $D$ .

First-order theories will be given in terms of **configurations**. A configuration is a pair consisting of  $T$ , a deductively closed set of first-order sentences, and a set  $n$  of *naming* functions  $j \in n, j : Vars \rightarrow Cons(T)$ , called a *naming*.

If  $c \in \text{Cons}(\mathcal{L})$ ,  $j(c) = c$ .

Write  $S$  for the set of all assignments. Write  $N$  for the set of all namings. Following van Eijck, make the following crucial assumption, which brings about a correspondence between a naming  $n$  and sets of assignment functions  $i \in s$  (sets of which make up DPL information states):

**(NA) Naming Assumption**

$\forall d \in D(t), \exists c \in \text{Cons}(\mathcal{L})$  such that  $s(c) = d$ , where  $s$  is a state (set of assignment functions.)

In other words, every object in the domain of the model has a name. This makes a naming (a set of naming functions) the syntactic counterpart of a DPL-state or set  $s$  of variable assignments  $i \in s$ , where  $i : \text{Vars} \rightarrow D$ . Thus, every variable is associated with a name, and every name with an individual in the domain of the model. This means that namings and states are interchangeable; mapping a variable to a constant is tantamount to mapping it to an individual.

The naming assumption is reflected in the further requirement of *constant exhaustiveness*:

Closed Theory  $T$  is **constant exhaustive** (or “exhaustive over  $\text{Cons}$ ” if  $\forall c \in \text{Cons}, \phi(c) \in T \Leftrightarrow \forall x \phi \in T$ )

This first system (i.e. FOTC) handles predicate logical theory change with regard to discourse-external information only. FOTC defines a function from an input configuration (theory and singleton naming), to an output configuration, without change in anaphoric information, with the following assumptions (from van Eijck [96]).

**(NC) Names Constant** The set of names  $\text{Cons}(T)$  remains the same under extension, contraction, and revision.

**(SCD) Shared Constant Domains:** domains of models are fixed under extension,

contraction, and revision. Where more than one model is under consideration, all models share the same domain  $D$  of individuals<sup>18</sup>.

Now, simply extend the AGM partial-meet contraction function to the first order case. This involves ensuring that  $T, n$  contracted by  $\phi$  fails to support  $\phi$ , under the association  $n$  of variables with names.

**Definition 40 First-Order (partial meet) contraction**

For  $\phi$  a first-order formula, where  $T \perp_n \phi$  is the set of all maximal subsets of  $T$  which fail to imply  $\phi$  under naming  $n$ , and  $\alpha$  is a selection function on that set, delivering a subset of  $T \perp_n \phi$ .

$$T, n \dot{-} \phi = \bigcap \alpha(T \perp_n \phi), n$$

**Fact 11** *First -order partial meet contractions obey the first-order rationality postulates.*

**Proof:**

Firstly, partial-meet contractions already obey all the propositional-level postulates (the AGM postulates), so that the quantifier postulates are met.

Then, N1 and N2 are trivial due to the assumption NC.

P1 and P2 can be treated similarly.

$$I1: T \dot{-} (c_1 = c_2) = \bigcap \alpha(T \perp (c_1 = c_2)) = T' \text{ such that } T' \not\vdash (c_1 = c_2) \not\equiv T' \vdash \neg(c_1 = c_2)$$

$$I2 \text{ is trivial so long as } \forall P \in Pred(T), (a = b) \in T \Rightarrow Pa \in T \Leftrightarrow Pb \in T$$

**5.3.1 FOTC; syntax**

This first system (FOTC) does not deal with the dynamics of anaphora, but with the dynamics of dialogue-external information.

Where  $n(x | c)$  is the set of naming functions just like  $n$  but that  $x$  might get mapped to  $c$ .

---

<sup>18</sup>As Veltman et al. [47] recognise, this assumption is problematic, since it implies that agents have prior knowledge of the domain of discussion.

Assume that there are syntactic restrictions in force (as in the propositional case) which ensure that  $-$  and  $\uparrow$  only ever appear on the far left of well-formed formulae (this rules out the intricacies of " $\neg - p$ " and so on). Unlike the propositional case, interpret extension by negative information as exclusion of positive information (this move both makes the recursion neater and interfaces better with van Eijck's systems.) Unfortunately, contraction of negative information cannot be so neatly dealt with; as in the propositional case, contracted negative information engages in the same partial meet algorithm as positive information.

**Extension:**

$$T, n + Pt = Cl(T \cup \{Pnt\}), n$$

$$T, n + \neg\phi = \bigcap \{T' \mid T' \supseteq T \text{ and } \exists w \in N : T', n + \phi = \mathcal{L}, w\}, n$$

$$T, n + \exists xPx = T + Pc, n(x \mid c), \exists c \in Cons(T)$$

$$T, n + \forall xPx = (\dots (T, n + Pc_1) + \dots + Pc_m), \forall c_1 \dots c_m \in Cons(T)$$

$$T, n + (\phi \wedge \psi) = (T, n + \phi) \cup (T, n + \psi)$$

$$T, n + -\phi = T, n \dot{-} \phi$$

$$T, n + \uparrow \phi = T, n \dot{+} \phi$$

**Contraction:**

$$T, n \dot{-} \phi = \bigcap \alpha(T \perp_n \phi), n$$

$$T, n \dot{-} \neg\phi = \bigcap \alpha(T \perp_n \neg\phi), n$$

$$T, n \dot{-} \exists xPx = T \underline{\forall} \{Pa \dots Pm\}, n \text{ for } a \dots m \in Cons(T)$$

$$T, n \dot{-} \forall xPx = T, n \dot{-} Pc, \exists c \in Cons(T)$$

$$T, n \dot{-} (\phi \wedge \psi) = (T, n \dot{-} \phi) \cup (T, n \dot{-} \psi)$$

**Revision:**

$$T, n \dot{+} \phi = (T, n \dot{-} \neg\phi) + \phi$$

### 5.3.2 A Semantics of FOTC

The basic semantical framework for the family of systems is an extension of the model-theoretic approach of preceding chapters to (sets of) predicate logical models with (sets of) assignment functions.

$s_T$  now denotes a set of predicate logical models which supports the theory  $T$ .

Entanglement with assignment functions also complicates the interpretation functions  $\llbracket \cdot, \cdot \rrbracket$ ,  $\llbracket \cdot, \cdot \rrbracket$ , and  $\llbracket \cdot, \cdot \rrbracket$ . They now carry with them (sets of) input and output assignment functions to the state, denoted by superscript and subscript respectively, for example  $s \llbracket \phi \rrbracket_u^n$  is a downdate function on state  $s$  taking input state  $n$  (a set of assignment functions  $i : Vars \cup Cons \rightarrow D$ ) and returning output state  $u$ . The general framework here is that of Quantified Modal Logic, or Modal Predicate Logic. Updates by existential quantifiers effect a random assignment to their variables, and in (later) dynamic systems, those assignments are passed over sentential boundaries (interpreted as sequential conjunctions).

#### Supermodel Semantics for General Contraction

The Choice and Package contractions of Fuhrmann ([35], presented in chapter 2) are to be put to good use here (where contraction of existentially quantified statements is not interpreted as “killing” variables), but as yet they have not been given any semantical analysis. It turns out that the “ignorant supermodel” semantics for contraction (from chapter 3) can be used to model general contractions, but now the supermodels have to be modified so that they fail to support *sets* of sentences.

The case is quite simple for package contraction:

In the syntax (recall chapter 2),  $T \forall B = \bigcap \alpha(T \perp B) = \bigcap \alpha(T \perp \{\alpha_1, \dots, \alpha_m\})$

So that semantically, one can define,  $s_T \llbracket B \rrbracket_{\forall} = \bigcup \beta(s \perp B)$

where the concept of supermodel sets is generalized (as follows) so as to fail to support  $B$ , a set of sentences. It is possible to model Fuhrmann's "General Contractions" by way of "General Ignorant Supermodels"; supermodels which fail to support whole sets of sentences.

### General Ignorant Supermodels

Generalise the definition given in chapter 3, to cover downgrade by sets of sentences over sets of models with variable assignments:

**Definition 41 General Ignorant Supermodels:**  $M'_T \in M_T \perp_n B$

For  $B$  a set of formulae of  $\mathcal{L}$ ,  $M'_T$  is a minimal supermodel set of  $M_T$  which fails to support  $B$  under assignment  $n$  (or simply "supermodel set of  $T$  ignoring  $B$  under  $n$ ") iff.

1.  $M_T \subseteq M'_T$  (supermodels)
2.  $M'_T \not\models_n p, \forall p \in B$  (general ignorance)
3. for any  $M_{T''}$  such that  $M_T \subseteq M_{T''} \subset M'_T, M_{T''} \models_n p, \exists p \in B$  (minimality)

where  $\models_n$  is the standard relation of satisfaction under assignment  $n$ .

The AGM case occurs where the set  $B$  is a singleton. Thus, the set of minimal supermodel sets of  $T$  ignoring  $B$  (under assignment  $n$ ), written  $M_{T \perp_n p}$ , is the set of (minimally larger than  $M_T$ ) sets of models which (under  $n$ ) support as much of  $T$  as they can without also supporting all the propositions in  $B$ .

### Semantical Package Contraction

As anticipated, the above definition allows a semantical version of Fuhrmann's *Package Contraction* [35]:

**Definition 42**  $s_T \llbracket B \rrbracket_{\forall} = \bigcup \beta(s_T \perp B)$



The definition can be employed in the downdate semantics for existential claims:

$$s_T \models \exists x Px \iff s_T \models Pa, \dots, Pm \text{ for all } a, \dots, m \in \text{Cons}(T).$$

A semantical version of *Choice Contraction* could similarly be constructed in order to deal directly with the contraction of universal claims.

$$T, n \dashv \forall x Px = T \exists \{Pa \dots Pm\}, n \text{ for } a \dots n \in \text{Cons}(T)$$

and then simply,  $s_T \models \forall x Px \iff s_T \models Pa \text{ for } \exists a \in \text{Cons}(T)$ .

### 5.3.3 Revisable Modal Predicate Logic

The semantics presented below is a version of Modal Predicate Logic (presented as an update semantics), augmented with a new operator  $\perp$  which expands the set of possible models supporting a partial (first-order) theory. Dub this new system RMPL (for “Revisable Modal Predicate Logic.”) The system does not deal with changing anaphoric relations, but with changing information states (dynamics of anaphora are tackled later). The existential quantifier still makes use of random assignment, but the assignments are not passed on as inputs to further interpretations. The semantics of RMPL defines a function from sets  $s$  of first-order models over a single domain  $D$  and an input assignment function  $n$  to sets of first-order models over  $D$  and an output assignment function  $u$ .

Where  $s \perp_n \phi$  is the set of minimal supermodels of  $s$  which fail to support  $\phi$  under assignment  $n$ , and  $\beta$ , as in the propositional case, is a selection function on  $s \perp_n \phi$ , choosing just the set of sets of models supporting the theories chosen by  $\alpha$  from  $T \perp_n \phi$  in the syntax.

Models  $i \in s$  consist of a fixed domain  $D$  of individuals and an interpretation function  $F$  assigning sets of n-tuples of elements of  $D$  as the denotations of n-place predicates. Thus  $i = \langle D, F \rangle$

**Definition 43** (RMPL Semantics)

$u = n(x \mid d)$  is the assignment function just like  $n$  but for the possible mapping of  $x$  to some  $d \in D$

$n = u$  in all cases except where existential quantification is involved.

**Updates**

$$\begin{aligned} s \llbracket P(t_1 \dots t_m) \rrbracket_u^n &= \{i \in s \mid i \models_n P(t_1 \dots t_m)\} \\ s \llbracket \neg\phi \rrbracket_u^n &= \{i \in s \mid n = u \text{ and there is no } r \text{ such that } i \in s \llbracket \phi \rrbracket_r^n\} \\ s \llbracket \exists x Px \rrbracket_u^n &= \{i \in s \mid i \models_u Pc, \exists c \in \text{Cons}(T), u = n(x \mid d)\} \\ s \llbracket \forall x Px \rrbracket_u^n &= s \llbracket Pc_1 \rrbracket_u^n \dots \llbracket Pc_m \rrbracket_u^n, \forall c_1 \dots c_m \in \text{Cons}(T) \\ s \llbracket \phi \wedge \psi \rrbracket_u^n &= s \llbracket \phi \rrbracket_u^n \cap s \llbracket \psi \rrbracket_u^n \\ s \llbracket \neg\phi \rrbracket_u^n &= s \llbracket \phi \rrbracket_u^n \\ s \llbracket \uparrow \phi \rrbracket_u^n &= s \llbracket \phi \rrbracket_u^n \\ s \llbracket \neg\phi \rrbracket_u^n &= s \llbracket \phi \rrbracket_u^n \end{aligned}$$

**Downdates**

$$\begin{aligned} s \llbracket P(t_1 \dots t_m) \rrbracket_u^n &= \bigcup \beta(s \perp_n P(t_1 \dots t_m)) \\ s \llbracket \neg P(t_1 \dots t_m) \rrbracket_u^n &= \bigcup \beta(s \perp_n \neg P(t_1 \dots t_m)) \\ s \llbracket \phi \wedge \psi \rrbracket_u^n &= s \llbracket \phi \rrbracket_u^n \cup s \llbracket \psi \rrbracket_u^n \\ s \llbracket \exists x Px \rrbracket_u^n &= s \llbracket Pa, \dots, Pm \rrbracket_u^n \llbracket \forall = s \rrbracket Pa \llbracket u, \dots \rrbracket Pm \llbracket u \rrbracket_u^n \\ s \llbracket \forall x Px \rrbracket_u^n &= s \llbracket Pc, \llbracket u \exists c \in \text{Cons}(T) \rrbracket \end{aligned}$$

**Revision**

$$s \llbracket \phi \rrbracket_u^n = s \llbracket \neg\phi \rrbracket_u^n \llbracket \phi \rrbracket_u^n$$

**An example theory contraction/state downdate**

Take the following theory growth:

$$T = \{Pa, Pb, Qa\}$$

$$T + Qb = T'$$

$$T' = \{Pa, Pb, Qa, Qb\}$$

Semantically things looks like this:  $s_T = \{i \in S \mid i \models Pa \& Pb \& Qa\} =$   
 $s_T \llbracket Qb \rrbracket = \{i \in s_T \mid i \models Qb\} = s_{T'}$

—the set of models becomes smaller so as to support  $Qb$ .

Now try the contraction:  $T'' = T' \dot{-} Pa = \bigcap \alpha(T' \perp Pa) = \{Pb, Qa, Qb\}$

Semantically:

$s_{T''} = s_{T'} \llbracket Pa \rrbracket = \bigcup \beta(s_{T'} \perp Pa) = s_K \supseteq s_{T'}, \exists K$  such that  $s_K \not\models Pa$  and  $s_K \models$   
 $\Delta, \forall \Delta \subseteq T''$  such that  $\Delta \not\models Pa$

—the state becomes larger again, expanding to incorporate some models which fail to support  $Pa$ .

### 5.3.4 Completeness of FOTC wrt RMPL

**Theorem 12** (1)  $Th(s), n + \phi = Th(s \llbracket \phi \rrbracket_u^n), u$

(2)  $T, n + \phi = T', u$  iff  $Mod(T') = Mod(T) \llbracket \phi \rrbracket_u^n$

(3)  $Th(s), n \dot{-} \phi = Th(s \llbracket \phi \rrbracket_u^n), u$

(4)  $T, n \dot{-} \phi = T', u$  iff  $Mod(T') = Mod(T) \llbracket \phi \rrbracket_u^n$

By composition, this leads to the result that

(5)  $Th(s), n \dot{+} \phi = Th(s \llbracket \phi \rrbracket_u^n), u$  and

(6)  $T, n \dot{+} \phi = T', u$  iff  $Mod(T') = Mod(T) \llbracket \phi \rrbracket_u^n$

**Proof:**<sup>19</sup>

The update/expansion clauses first (items 1 and 2):

<sup>19</sup>By induction on the structure of  $\phi$ , similarly to the propositional case. Not all of the clauses are given here.

$$\begin{aligned} & \text{For } n = u, Th(s), n + P(t_1 \dots t_m) = \{\psi \mid Th(s) \cup P(nt_1 \dots nt_m) \vdash \psi\}, n \\ & = Th(\{i \in s \mid i \models_n P(t_1 \dots t_m)\}), n = Th(s \llbracket P(t_1 \dots t_m) \rrbracket_u^n), u \end{aligned}$$

$$\begin{aligned} & \text{For } n=u, Th(s), n + \neg\phi = \bigcap \{T' \mid T' \supset Th(s) \text{ and } \exists w : T', n + \phi = \mathcal{L}, w\}, n \\ & = Th(\{i \in s \mid i \llbracket \phi \rrbracket_u^n = \emptyset\}) = Th(s \llbracket \neg\phi \rrbracket_u^n), u \end{aligned}$$

$$\begin{aligned} & \text{For } u = n(v \mid d) \text{ for some } d \in D, \\ & Th(s), n + \exists x Px = \{\psi \mid Th(s) \cup P(c) \vdash \psi, \exists c \in Cons(T), u = n(x \mid c)\}, u \\ & = Th(\{i \in s \mid i \models_u Pc\}), u = Th(s \llbracket \exists x Px \rrbracket_u^n), u \end{aligned}$$

$$\begin{aligned} & Th(s), n + \phi \wedge \psi = Th(s), n + \phi \cup Th(s), n + \psi \\ & = (\text{Ind hyp}) Th(s \llbracket \phi \rrbracket_w^n), w \cup Th(s \llbracket \psi \rrbracket_y^n), y \\ & = (\text{Ind hyp}) = Th(\{i \in s \mid i \models_w \phi\}), w \cup Th(\{i \in s \mid i \models_y \psi\}), y \\ & = Th(\{i \in s \mid i \models_w \phi\} \cap \{i \in s \mid i \models_y \psi\}), u \\ & = Th(s \llbracket \phi \rrbracket_w^n \cap s \llbracket \psi \rrbracket_y^n), u \\ & = Th(s \llbracket \phi \wedge \psi \rrbracket_u^n), u \end{aligned}$$

Part(2),

$$\begin{aligned} & T, n + \phi = T', u \text{ iff } Th(Mod(T)), n + \phi = Th(Mod(T')), u \\ & \text{iff by part (1)} Th(Mod(T)), n + \phi = Th(Mod(T) \llbracket \phi \rrbracket_u^n), u \\ & \text{iff } Mod(T') = Mod(T) \llbracket \phi \rrbracket_u^n \end{aligned}$$

**Downdates, (items 3 and 4)**

The following subproofs make use of the basic fact that, by definition,

$$\text{Lemma: } \bigcap \alpha(T \perp \phi) = Th(\bigcup \beta(s_T \perp \phi))$$

$$\begin{aligned} & \text{For } n = u, \\ & Th(s), n \dot{-} P(t_1 \dots t_m) = \bigcap \alpha(Th(s) \perp_n P(t_1 \dots t_m)) \\ & = \bigcap \alpha(\{T' \mid T' \subseteq Th(s) \text{ and } T' \not\vdash P(t_1 \dots t_m)\}) \end{aligned}$$

$$\begin{aligned}
&= \bigcap \alpha(\{T' \mid T' = Th\{N \mid N \supseteq s \text{ and } N \not\models_n P(t_1 \dots t_m)\}\}) \\
&= Th(\bigcup \beta(s \perp_n P(t_1 \dots t_m))), u \\
&= Th(s) \llbracket P(t_1 \dots t_m) \rrbracket_u^n, u
\end{aligned}$$

For  $n = u$ ,

$$\begin{aligned}
Th(s), n \dot{-} \neg P(t_1 \dots t_m) &= \bigcap \alpha(Th(s) \perp \neg P(nt_1 \dots nt_m)) \\
&= \bigcap \alpha(\{T' \mid T' = Th\{N \supseteq s \mid N \not\models_n \neg P(t_1 \dots t_m)\}\}) \\
&= Th(\bigcup \beta(s \perp_n \neg P(t_1 \dots t_m))), u \\
&= Th(s) \llbracket \neg P(t_1 \dots t_m) \rrbracket_u^n, u
\end{aligned}$$

For  $n = u$ , and  $B = \{Pa, \dots, Pk\}$  for  $a, \dots, k \in Cons(T)$ ,

$$\begin{aligned}
Th(s), n \dot{-} \exists x Px &= Th(s), n \dot{-} \{Pa, \dots, Pk\} \\
&= \bigcap \alpha(Th(s) \perp B), u \\
&= Th(\bigcup \beta(s \perp B)), u \\
&= Th(s) \llbracket B \rrbracket_u^n, u \\
&= Th(s) \llbracket \exists x Px \rrbracket_u^n, u
\end{aligned}$$

$$\begin{aligned}
Th(s), n \dot{-} (\phi \wedge \psi) &= Th(s), n \dot{-} \phi \cap Th(s), n \dot{-} \psi \\
&= (\text{Ind hyp}) Th(s) \llbracket \phi \rrbracket_w^n, w \cap Th(s) \llbracket \psi \rrbracket_y^n, y \\
&= (\text{Ind hyp}) = Th\{i \in s \mid i \not\models_w \phi\}, w \cap Th\{i \in s \mid i \not\models_y \psi\}, y \\
&= Th(\{i \in s \mid i \not\models_w \phi\} \cup \{i \in s \mid i \not\models_y \psi\}), u \\
&= Th(s) \llbracket \phi \rrbracket_w^n \cup \llbracket \psi \rrbracket_y^n, u \\
&= Th(s) \llbracket \phi \wedge \psi \rrbracket_u^n, u
\end{aligned}$$

Part(4),

$$\begin{aligned}
T, n \dot{-} \phi = T', u &\text{ iff } Th(Mod(T)), n \dot{-} \phi = Th(Mod(T')), u \\
&\text{ iff by part (3) } Th(Mod(T)), n \dot{-} \phi = Th(Mod(T)) \llbracket \phi \rrbracket_u^n, u \\
&\text{ iff } Mod(T') = Mod(T) \llbracket \phi \rrbracket_u^n
\end{aligned}$$

Revision: by composition of the preceding results.

### 5.3.5 Comparisons

As discussed above, in relation to Intuitionistic semantics, the semantics presented above differs from the spirit of Jaspars [60] presentation of up-and-downdates in Kripke structures. There Jaspars uses a “look back” definition of downdate in multi-modal logics, where agents retrace their information growth history in order to effect a downdate (he also only considers propositional systems). Note that the semantics that I present employs a different conception of rational downdating; one that does not just look back to previous information states, but instead has the capacity to bring about entirely novel states in which the retracted information is absent. In other words, the Multiple Extensions Problem makes a semantical point too; there are a number of information states which fail to support a given formula, and one needs (in the most rational way) to specify which is the most appropriate downdate. Whilst Jaspars’ “look-back” approach embodies the conservative maxim “only consider states that you were in before” (similarly to the intuitionistic case), genuine AGM-inspired downdating allows for consideration of previously unencountered states.

Using the “downdate” function that I have defined, stacks or sequences of information states (growth histories) are not necessary in order to perform revisions; they can be computed on individual states. This has the consequence that downdating can produce new states rather than simply re-instating old ones.

The avoidance of stacks of states is also, ontologically speaking, more conservative.

## 5.4 Adding Dynamics of Anaphora

Information conveyed in linguistic communication has two important facets, as recent developments in dialogue and discourse analysis illustrate. There is information about the world or how it might be (external, truth-conditional information), and information about the possible values of variables, or anaphor bindings. The latter is “discourse internal” information, or a “mere” book-keeping device. The former has been the province of US and epistemic logics in general, which are superseded by RS when it is taken into account that both the world itself, and our beliefs about it, are dynamic. Information about variable bindings has been the subject matter of DRT and DPL, and brings into

semantics the notion of information about partial and abstract objects, discourse referents, pegs, indefinite objects, (eg: Karttunen [63], Kamp [61], Heim [53], Landman [69]). Both of these conceptions of information (the internal and the external) are required in a semantics adequate for the purposes of a theory of communication. What's more, for a semantics of dialogue, the information states which they produce need to be revisable. That revision of variable bindings actually does go on in dialogue will be shown by a few examples (see chapter 6). An interesting problem is that of how to model arbitrary *change* of information about the values of variables, so that sentence interpretation involves genuine information change. In this sense a Revision Semantics reformulation of DPL is required, or an adaptation of the underlying logic of DPL to that of RS, in order to accommodate repair and change of discourse internal information. The combination will allow the modelling of information about "partial objects" which can be changed (and later exchanged) via sentence utterance and interpretation (see chapter 6).

For DPL, information growth (or theory extension) in the context of a fixed model, has to do with changing  $s$ , the set of possible assignments to variables. Interpretations of atomic formulae are simply tests to the effect that the current state (set of assignments) supports the formula in question. Since the model under consideration in DPL is static, all that changes is the association of variables with individuals in the domain of the model. For example;  $\exists x Px \wedge Qx$  is true with respect to fixed model  $m$  and sets of assignments  $s$  iff there is an assignment  $i \in s$  which maps  $x$  to some  $d \in D$  such that  $d \in F(P)$  and  $d \in F(Q)$ .

A state of total information arises when only one possible assignment is left; each variable is associated with just one name, and thus (via the naming assumption) with one individual each in the domain. Syntactically speaking, in a total theory this is the same as finding out which names the variables are associated with; settling on just one naming function.

**Example 18**  $Cons(T) = \{a, b\}$

$T, n = \{Pa, \neg Pb\}, \{i \in N \mid i(a) = a, i(b) = b\}$

$T, n + \exists xPx = \{Pa, \neg Pb, \}, \{j \in n \mid j(x) = a\}$

or,  $T, n + \exists xPx = \{\mathcal{L}\}, \{j \in n \mid j(x) = b\}$

### Anaphoric relations under downdate

A characteristic of DPL is that sentential close is interpreted as a dynamic conjunction, which passes variable assignments from the first to the second conjunct (ie: across sentential boundaries.) Note that there is an interesting parallel case in retraction, or downdate. The following example illustrates that anaphoric relations survive revision (and therefore downdate).

**Example 19** *“He’s old and happy. Well, he’s not happy. He’s my friend.”*

The pronoun *he* is still available for reference after the revision in the second sentence.

The next system explored is a variant of DPL where states (sets of variable assignments) are revisable. I later outline similar systems which model genuine theory change, and provide an interesting treatment of changing knowledge about identities. Agents do not often ‘dump’ subjects in dialogue; they can repair and recover wayward anaphoric relations. For example:

**Example 20** *“A man and a boy were running along the beach. He started to laugh. Not the man, the boy.”*

The above example cannot be dealt with using standard negation, for then the story degenerates into absurdity. Rather one assignment is to be retracted, and replaced by another. This kind of consideration motivates the provision of revisable DPL states.

#### 5.4.1 First-Order theory extension with Anaphora

Van Eijck [96] presents two different first-order theory **extension** systems, the first of which is taken as the point of departure for a revisable DPL. I choose to rewrite van Eijck’s system in more familiar notation:



**Definition 44 (FOTE, first system)**

Here I present a system of theory extension for theories consisting of sets of sentences expressed in first-order logic with identity.

- $T, n + Rt_1 \dots t_m = T \cup \{Rnt_1 \dots nt_m\}, n$
- $T, n + \neg\phi = \bigcap \{T' \mid T' \supseteq T, \exists w; T', n + \phi = \mathcal{L}, w\}$
- $T, n + (t_1 = t_2) = T \cup \{nt_1 = nt_2\}, n$
- $T, n + \exists x = T, n(x \mid c), c \in \text{Cons}(T)$
- $T, n + \phi \wedge \psi = (T, n + \phi) + \psi$ <sup>20</sup>

Note that conjunctions passes namings from the first to the second conjunct, and that the only clause which changes the naming part of a configuration concerns the existential quantifier; assigning some constant from the theory to the relevant variable. Note also that the negation clause of van Eijck's system is intuitionistic (it acts so as to exclude positive information), which marries well with the state-subtraction negation of DPL.

**5.4.2 Semantics of DPL**

The semantics for DPL (here formulated as an update semantics, as in Dekker [21], page 150) defines a function from an input state  $s$  (a set of variable assignments) to an output state. The states are sets of functions from the set of variable  $Vars$  to the domain  $D$  of the (single) model,  $m = \langle D, F \rangle$

Where  $s[x] = \{i[x/d] \mid i \in s \text{ and } d \in D\}$  (the state just like  $s$  except that it can assign  $d$  to  $x$ .)

---

<sup>20</sup> $B, u$  such that  $T, n + \phi = C, w$  and  $C, w + \psi = B, u$

and,

$\downarrow \llbracket \phi \rrbracket = \{i \mid \{i\} \llbracket \phi \rrbracket \neq \emptyset\}$  (all the assignments which are non-absurd under update by  $\phi$ )

**Definition 45** (DPL Semantics<sup>21</sup>)

$$s \llbracket Rx_1 \dots x_n \rrbracket = \{i \in s \mid \langle i(x_1) \dots i(x_n) \rangle \in F(R)\}$$

$$s \llbracket x = y \rrbracket = \{i \in s \mid i(x) = i(y)\}$$

$$s \llbracket \neg \phi \rrbracket = s - \downarrow \llbracket \phi \rrbracket$$

$$s \llbracket \phi \wedge \psi \rrbracket = s \llbracket \phi \rrbracket \llbracket \psi \rrbracket$$

$$s \llbracket \exists x \phi \rrbracket = s[x] \llbracket \phi \rrbracket$$

Note that DPL has total information states<sup>22</sup>; that is,

**Fact 12** for all DPL formulas  $\phi$ ,

$$\vdash_{dpl} \phi \vee \neg \phi$$

FOTE is provably complete with respect to the semantics of DPL under the following assumptions;

1. theories are **complete**,
2. **the Naming Assumption**; every individual is named (is in the set  $Cons$ ),
3. and theories are **Constant Exhaustive**

A closed theory  $T$  is *exhaustive* over  $Cons$  if  $\phi(c) \forall c \in Cons$  implies that  $\forall x \phi \in T$ .

<sup>21</sup>The “-” here is state subtraction; all the assignments in  $s$  but not in  $\downarrow \llbracket \phi \rrbracket$

<sup>22</sup>see Emiel Kraemer [64]

### Completeness

I state van Eijck's result below (translated into the notation used here) . Where  $m$  is a predicate-logical model,  $s$  the input state,  $u$  the output state.

**Theorem 13** (*Completeness of FOTE wrt DPL (van Eijck 1993)*)

$m, s, u \models_{dpl} \phi$  iff  $Th(m), s + \phi = Th(m), u$

and

$T, s + \phi = T, u$  iff  $Mod(T), s, u \models_{dpl} \phi$

Note that, since attention is restricted to one model, the theory part of the configuration is static in the above results, and it is only the naming, or state which changes (in fact namings and states can be used interchangeably because of the naming assumption). But these results aren't really about *theory* extension then, for theory extension involves a change in *sets* of models - adding external, truth conditional information about the world. The restriction above to *complete* theories (ie: where  $T = Th(i)$  and  $Mod(T) = i$ ) actually means that the result is about associating variables with names and names with individuals, rather than adding sentences to a theory. DPL tracks change in discourse-internal information. Strengthening these results to incorporate genuine theory change (discourse-external information) involves moving away from DPL, towards "modal" DPL, an issue which I postpone for the moment.

#### 5.4.3 Revisable DPL

Discourse-internal information is open to revision. Consider the following (see also the example of chapter 6):

**Example 21** "A man and a boy came in. He was tall.

No, not the boy . I meant the man."

$s[x][y] \llbracket Tx \rrbracket \llbracket \neg Tx \rrbracket \llbracket Ty \rrbracket$

Here discourse internal information is being repaired (or corrected, or revised). It seems, in general, that most constituent negations are revision instructions (see chapter

7), rather than downdates. However, note that in the above example, *the boy* is not excluded from being tall. Rather, a previously supplied constraint on the set of variable assignments (of “the boy” to tall individuals) is being retracted. In order to deal with this type of repair, a Revisable DPL is required.

In DPL, agents can only learn about the values of variables. Genuine update of discourse external information is to be handled by diminishing sets of models. However, loss of information in RDPL can be modelled by growing sets of assignment functions. For example, to retract  $Px$  from a state  $s$  in RDPL, is to move to a superstate  $s'$  of  $s$  in which variable  $x$  is not assigned to  $d \in F(P)$  for all the  $i \in s'$ . Thus,  $s \perp Px$  is the set of minimal superstates of  $s$  which do not (under the new set of assignments) support  $Px$ . This requires a slight change in the supermodel semantics for contraction; where information is encoded as sets of variable assignments, a *superstate semantics* is required.

The basic idea is that  $s \perp \phi$  is the set of all (minimal) superstates of  $s$  which fail to support  $\phi$ . That is, some assignments in  $s \perp \phi$  map variables in  $\phi$  to  $d \in D$  which do not respect their predications in  $\phi$ . As in the first-order AGM case, select some of the minimal superstates, and join, to form a downdate.

Since DPL can only gain information about indefinites, RDPL can only lose and revise information about indefinites. There are no clauses here for retraction of formulae involving names (eg: “John isn’t ill.”), since DPL interprets assertions of such formulae as *tests* on the current state. It would be a simple matter to interpret retraction of definite information in RDPL as a test that the current state does not support the retracted information. In order to deal with retraction of existential claims, some ideas from *free logic* can be employed. Distinguished elements  $\emptyset \in Cons$  and  $* \in D$  stand for the “null” individual (the individual with no properties) and its name. Thus,  $\forall n \in N, n(\emptyset) = *$ , and also  $\forall P \in Pred, * \notin F(P)$ . Using a “null” name and individual for the interpretation of non-existence claims is (at least) better than van Eijck’s *FOTE<sub>III</sub>*, where interpretation of the denial of an existential leads to absurdity. Here, such utterances do not trigger implosion of the state; they are processible.

Thus, modelling repair of anaphoric information in RDPL, add the following clauses to DPL.

**Definition 46** (RDPL clauses)

$$s \llbracket \neg Px \rrbracket = s \llbracket Px \rrbracket$$

$$s \llbracket Px \rrbracket = \bigcup \beta(s \perp Px)$$

$$s \llbracket (x = y) \rrbracket = \bigcup \beta(s \perp (i(x) = i(y)))$$

$$s \llbracket \neg \phi \rrbracket = s \cup \downarrow \llbracket \phi \rrbracket$$

$$s \llbracket \phi \wedge \psi \rrbracket = s \llbracket \phi \rrbracket \cup s \llbracket \psi \rrbracket$$

$$s \llbracket \exists x \rrbracket = s \llbracket x \rrbracket$$

$$s \llbracket x \rrbracket = \{n = i(x \mid *) \mid i \in s\}$$

(the state just like  $s$  except that it assigns  $x$  to  $*$ .)

$$s \llbracket \uparrow Px \rrbracket = s \llbracket Px \rrbracket$$

$$s \llbracket Px \rrbracket = s \llbracket \neg Px \rrbracket \llbracket \llbracket Px \rrbracket \rrbracket$$

(One could also give a clause for disjunction,  $s \llbracket \phi \vee \psi \rrbracket = s \llbracket \phi \rrbracket \cup s \llbracket \psi \rrbracket$ , but since DPL lacks one, so does RDPL.)

The basic downdate clause has the effect of expanding the set of assignments to one which does not support  $Px$ . Conversely to the DPL update clause for negations, down-dating by, or “forgetting”,  $\neg\phi$  just adds to  $s$  all those assignments which are non-absurd under update by  $\phi$ .

The example given above can now be dealt with as a case of revision.

“A man and a boy came in. He was tall. No, not the man, the boy.”

$s[x][y] \llbracket Tx \rrbracket \llbracket \neg Tx \rrbracket \llbracket Ty \rrbracket = \{i \in s \mid i(x) \in F(T)\} \llbracket Tx \rrbracket \llbracket Ty \rrbracket = s' \llbracket Tx \rrbracket$  for some  $s' \supset s$  such that  $s' \not\models Tx = s''$  such that  $\forall i \in s'', i \models Ty$

The downdate clause for the existential needs some elaboration. It is intended to interpret utterances in dialogue such as,

**Example 22** *“Someone at home will feed the cat.”*

*“No, there isn’t anyone at home.”*

$\llbracket \exists x Cx \rrbracket \rrbracket \exists x \llbracket$

Such a statement is interpreted as retracting a prior claim in the dialogue that some  $x$  exists, not just that it has a particular property. Importantly, the second sentence in the above example cannot be interpreted as “ $\neg Cx$ ” (retraction of information about an indefinite). Rather, it seems to be a retraction of the indefinite expression from the current dialogue. It’s not that there is someone who has the property of not being at home, but that the previously introduced indefinite cannot be employed any further. This has the effect of “killing” the variable  $x$ . For example, one cannot continue the above dialogue with “They can cook the dinner.” The indefinite is no longer available for anaphoric reference after being the subject of the retraction of an existential claim. However, such a statement does not reduce the dialogue to absurdity (as prescribed in  $FOTE_{II}$ ). Formally, this is achieved by mapping  $x$  to a “null” element  $* \in D$ , whose name is  $\emptyset$ . Once a variable is associated with this name, it is no longer available for further use. I shall return to these issues at the close of the chapter.

Retraction of identity information moves to a superstate of  $s$  in which the identity statement is not supported - that is, where one of the variables in question is possibly assigned to some new individual or individuals, and thus found to lack some property which the other retains. The clause is meant to account for the following type of example (where “the boy” and “the suspect” are both indefinites).

**Example 23** *“The boy had dark hair, the suspect had dark hair”*

*“No, the suspect is not the boy.”*

$s \rrbracket x = y \llbracket$

Due to distinguishability (I2) , it is not the case that both the boy and the suspect retain their properties, but one is forced to lose some property which the other retains. Retracted identity statements can be re-imposed if it is learnt (via update) that the “lesser” of the two individuals has the additional property after all, or if the other individual is discovered (via downdate or revision) not to have the additional property after all. These issues are more properly the subject of chapter 6.

It turns out that RDPL is complete with respect to FOTE supplemented with the following theory contraction clauses (dub this system  $FOTC_I$ ). The result deals with theory change through growing and shrinking sets of naming functions  $n : Vars \rightarrow Cons$ , whose semantical counterparts are DPL states.

**Definition 47** ( $FOTC_I$  contraction clauses)

$$T, n \dot{-} Pt = \bigcap \alpha(T \perp Pnt), n$$

$$T, n \dot{-} \neg \phi = \bigcap \{T' \mid T' \subseteq T \text{ and } \exists w \in N : T', n + \phi \neq \mathcal{L}, w\}$$

$$T, n \dot{-} \exists x = T, u \text{ such that } u = n(x \mid \emptyset)$$

$$T, n \dot{-} (\phi \wedge \psi) = (T, n \dot{-} \phi) \cup (T, n \dot{-} \psi)$$

$$T, n \dot{-} (t = t') = \bigcap \alpha(T \perp nt = nt'), n$$

#### 5.4.4 Completeness of $FOTC_I$ wrt RDPL

**Theorem 14** *The update and extension clauses are the same as the preceding theorem (for DPL and FOTE). Here  $m$  denotes the single static model, and  $I$  and  $\dot{-}$  give the result in terms of updates and downdates on sets of namings/ variable assignments  $s$  and  $u$  in that theory/ model.*

$$(1) Th(m), s + \phi = Th(m), s \llbracket \phi \rrbracket$$

$$(2) T, s + \phi = T, u \text{ iff } u = s \llbracket \phi \rrbracket$$

$$(3) Th(m), s \dot{-} \phi = Th(m), s \llbracket \phi \llbracket$$

$$(4) T, s \dot{-} \phi = T, u \text{ iff } u = s \llbracket \phi \llbracket$$

**Proof:**

Update/extension clauses (1 and 2) follow from van Eijck's proof connecting DPL and FOTE (see [96] page 10, in [23] page 258). The other parts are very similar, but that they connect contraction and downdate clauses.

For example;

$$Th(m), s \dot{-} \exists x = Th(m), s(x \mid \emptyset) = Th(m), s \llbracket x [= Th(m), s \llbracket \exists x \llbracket$$

The other clauses are similar, if less interesting.

## Part (4)

$$T, s \dot{-} \phi = T, u \text{ iff } Th(Mod(T)), s \dot{-} \phi = Th(Mod(T)), u$$

$$\text{iff by part (3)} Th(Mod(T)), s \dot{-} \phi = Th(Mod(T)), s \llbracket \phi \llbracket$$

$$\text{iff } u = s \llbracket \phi \llbracket$$

## 5.5 Dynamic Modal PL, FUL, and $FOTE_{III}$ ,

Overall I attempt to give a reformulation of Modal DPL adapted to the logic of Revision Semantics. The project has much in common with EDPL (Dekker [21]) and DMPL [97], where DPL is given an underlying Update Logic. In preparation for the full change systems, systems of extension and update are investigated first.

In  $FOTE_{III}$  (van Eijck's second system) theories can be extended with negative information (denoted  $T - \phi$ ) as well as positively. In general, this amounts to extending by the negation of the sentence in question, so that;  $T, n - \phi = T, n + \neg\phi$

**Definition 48** ( $FOTE_{III}$ )

$$T, n + Rt_1 \dots t_m = Cl(T \cup \{Rnt_1 \dots nt_m\}), n$$



$$T, n - Rt_1 \dots t_m = Cl(T \cup \{\neg Rnt_1 \dots nt_m\}), n$$

$$T, n + \neg\phi = \bigcap \{T' \mid T' \supseteq T, \exists w \in N; T', n + \phi = \mathcal{L}, w\}$$

$$T, n - \neg\phi = T', n \text{ iff } \exists w : T, n + \phi = T', w$$

$$T, n + (t_1 = t_2) = Cl(T \cup \{nt_1 = nt_2\}), n$$

$$T, n - (t_1 = t_2) = Cl(T \cup \{\neg(nt_1 = nt_2)\}), n$$

$$T, n + \exists x = T, n(x \mid c), c \in Cons$$

$$T, n - \exists x = \mathcal{L}, u$$

$$T, n + \phi \wedge \psi = B, u \text{ such that } T, n + \phi = C, w \text{ and } C, w + \psi = B, u \\ = (T, n + \phi) + \psi$$

$$T, n - \phi \wedge \psi = (T - \phi) \cap (\bigcup \{T' \mid T', w = (T, n + \phi) - \psi\}), n$$

$$T, n + \diamond\phi = T, n \text{ iff } T, n + \phi \neq \mathcal{L} \\ = \mathcal{L}, n \text{ iff } \forall w, T, n + \phi = \mathcal{L}, w$$

$$T, n - \diamond\phi = T, n \text{ iff } T, n + \phi \neq \mathcal{L} \\ = \mathcal{L}, n \text{ iff } \forall w, T, n + \phi = \mathcal{L}, w$$

Van Eijck proves that if the shared domain of the models under consideration is *fixed*,  $FOTE_{III}$  is complete with respect to DMPL (Dynamic Modal Predicate Logic). When

the domain is canonical (ie: changes with respect to identity statements), then the system is linked to First-Order Update Logic. Note some peculiarities of the above (as translated from the original): clauses for positive and negative extension by modals are identical<sup>23</sup>, and denial of the existential leads to absurdity<sup>24</sup>.

The semantics of DMPL (see van Eijk and Cepparello [97]) defines an interpretation function over sets of models sharing a fixed domain  $D$ , an input set of assignments  $n$ , and an output set of assignments  $u$ , to sets of first-order models over  $D$ . The relation  $\models_n$  is the standard "satisfaction in state  $n$ " relation for FOL. Identity is a test on the current state, negation is state subtraction, conjunction passes along variable assignments, and existential quantification randomly assigns individuals to variables. Updates actually eliminate possible models which do not support the new information.

**Definition 49** (*Semantics of DMPL*)

$$s \llbracket Rt_1 \dots t_m \rrbracket_u^n = \{i \in s \mid n = u \text{ and } i \models_n Rt_1 \dots t_m\}$$

$$s \llbracket t_1 = t_2 \rrbracket_u^n = \{i \in s \mid n = u \text{ and } i \models_n (t_1 = t_2)\}$$

$$s \llbracket \neg\phi \rrbracket_u^n = \{i \in s \mid n = u \text{ and there is no } r \text{ such that } i \in s \llbracket \phi \rrbracket_r^n\}$$

$$s \llbracket \phi \wedge \psi \rrbracket_u^n = \{i \in s \mid \exists w \text{ such that } i \in s \llbracket \phi \rrbracket_w^n \llbracket \psi \rrbracket_u^w\}$$

$$s \llbracket \exists x \rrbracket_u^n = s \text{ iff } u = n(x \mid d), \exists d \in D$$

$$s \llbracket \diamond\phi \rrbracket_u^n = s \text{ iff } n = u \text{ and } \exists r : s \llbracket \phi \rrbracket_r^n \neq \emptyset$$

---

<sup>23</sup>But surely, denial of a possibility is not equivalent to its assertion. Rather, the clause should be:  $T, n - \diamond\phi = T, n$  iff  $T, n + \phi = \mathcal{L}$   
 $= \mathcal{L}, n$  iff  $\forall t', wT, n + \phi \neq \mathcal{L}, w$

<sup>24</sup>As noted earlier, this is surely too drastic a measure.

**Completeness of DMPL with respect to  $FOTE_{III}$** **Theorem 15** (van Eijck 1993)

(1)  $Th(s), n + \phi = Th(s \llbracket \phi \rrbracket_u^n), u$

(2)  $Th(s), n - \phi = Th(s \llbracket \neg\phi \rrbracket_u^n), u$

and if  $T$  is closed, and full and exhaustive over  $Cons$ , then also:

(3)  $T, n + \phi = T', u$  iff  $Cmod(T) \llbracket \phi \rrbracket_u^n = Cmod(T')$

Where  $Cmod(T)$  is the set of all models for  $T$  over the canonical domain  $Cons_{\equiv}$ , constructed with respect to the identity statements holding in  $T$  (construct the domain by setting  $a \equiv b$  iff  $(a = b) \in T$ ). Since  $T$  is full over  $Cons$ , this operation defines an equivalence relation on  $Cons$ . As van Eijck shows, states can be defined for  $Cons_{\equiv}$ , but there is no need to distinguish them from states for  $Cons$ . (The main point here is that  $T = Th(Cmod(T))$  for closed  $T$  which is exhaustive and full over  $Cons$ , so that (3) follows.)

If the restriction to full theories, or a single domain for the models is dropped, the system  $FOTE_{III}$  is then complete with respect to a system of **First-Order Update Logic** or **FUL**.

**5.5.1 First-order Update Logic (FUL)**

The system FUL (see van Eijck's [96]) allows for growth of knowledge about identities<sup>25</sup>. Its semantics is given in terms of sets of *canonical models* (as described above), an input state, and an output state. Identity statements provide constraints on the equivalence components of the canonical models.

The system allows for an analysis of growth of knowledge about identity relations. The set of equivalence classes on  $Cons$  (the domain of "proto-individuals") is updated through interpretation of identity statements. If  $\theta$  is an equivalence class on  $Cons$ , then  $Cons_{\theta}$  is a domain of individuals. Now restrict attention to canonical models, with domains  $Cons_{\theta}$ .

<sup>25</sup>DMPL does not since there true identities are "epistemically necessary";  $s \llbracket x = y \rrbracket_n^n = \{i \in s \mid nx = ny\} = s$  or  $\emptyset$ . Thus, as van Eijck mentions, in DMPL,  $s \llbracket x = y \rrbracket_n^n = s \llbracket \Diamond(x = y) \rrbracket_n^n = s \llbracket \Box(x = y) \rrbracket_n^n$

Canonical models<sup>26</sup> are ordered pairs  $i = \langle i_m, i_e \rangle$ . Where  $i_m$  is a “proto-model” over  $Cons$ , called the *proto-model component* of  $i$ , and  $i_e$  is the *equivalence component* of  $i$ . The members of each  $i_e$  class have the same properties (they respect the signature of the model).

**Definition 50 (FUL)**

- $s \llbracket P(t_1 \dots t_n) \rrbracket_u^n = \{i \in s \mid n = u \text{ and } i_m \models_n P(t_1 \dots t_n)\}$
- $s \llbracket t_1 = t_2 \rrbracket_u^n = \{i \in s \mid n = u \text{ and } (nt_1 = nt_2) \in i_e\}$
- $s \llbracket \neg\phi \rrbracket_u^n = \{i \in s \mid n = u \text{ and there is no } w \text{ such that } i \in s \llbracket \phi \rrbracket_w^n\}$
- $s \llbracket \phi \wedge \psi \rrbracket_u^n = \{i \in s \mid \exists w \text{ such that } i \in s \llbracket \phi \rrbracket_w^n \llbracket \psi \rrbracket_u^w\}$
- $s \llbracket \exists x \rrbracket_u^n = s$  iff  $\exists c \in Cons$  such that  $u = n(x \mid c)$
- $s \llbracket \diamond\phi \rrbracket_u^n = s$  iff  $n = u$  and  $\exists w$  such that  $s \llbracket \phi \rrbracket_w^n \neq \emptyset$

Here, update by identity information is no longer a test on the current state, but imposes constraints on the equivalence components of the canonical models. Note also that the clause for the existential no longer deals with individuals in the domain of the models, but in names. This reflects the move away from reference to “real individuals”, to reference to “proto-individuals” instead.

FUL and  $FOTE_{III}$  are connected in the following way, where  $s$  is a set of *canonical models*;

**Theorem 16 (van Eijck 1993)**

$$Th(s), n + \phi = Th(s \llbracket \phi \rrbracket_u^n), u$$

---

<sup>26</sup>See [96] for full details.

$$Th(s), n - \phi = Th(s \llbracket \neg\phi \rrbracket_u^n), u$$

and if  $T$  is closed, and exhaustive over  $Cons$ , then also (as for DMPL):

$$T, n + \phi = T', u \text{ iff } Cmod(T) \llbracket \phi \rrbracket_u^n = Cmod(T')$$

### 5.5.2 Adding Downdates and Revision

Here, I indicate how van Eijck's  $FOTE_{III}$  can be made into a genuine FOTC system, and might be given a semantics along the lines of a Revisable DMPL. The result would be a semantics of theory change which handles revision of anaphoric information as well as dialogue-external (truth-conditional) information.

Later, I show how FUL can be given the same treatment.

### 5.5.3 Revisable DMPL

Revisable DMPL is simply DMPL with a mixture of downdate clauses from RDPL and RMPL, and a revision clause. Unsurprisingly, it can be connected to a theory change system based on  $FOTE_{III}$ . Updates and downdates by atomic formulae (reflecting dialogue-external inputs) now no longer (as in DPL and RDPL) just test the current state, but actually change it.

#### Definition 51 (RDMPL)

Add the following clauses to DMPL:

$$s \llbracket Pt_1 \dots t_m \rrbracket_u^n = \bigcup \beta(s \perp_n Pt_1 \dots t_m), n = u$$

$$s \llbracket \neg\phi \rrbracket_u^n = s \cup \downarrow \llbracket \phi \rrbracket_u^n$$

$$s \llbracket x = y \rrbracket_u^n = \{i \in s \mid n = u \text{ and } s \not\vdash_n x = y\}$$

$$s \llbracket \phi \wedge \psi \rrbracket_u^n = s \llbracket \phi \rrbracket_u^n \cup s \llbracket \psi \rrbracket_u^n$$

$$s \llbracket \exists x \rrbracket_u^n = s \text{ iff } u = n(x \mid *), * \in D$$

$$s \llbracket \diamond\phi \rrbracket_u^n = s \text{ iff } s \llbracket \phi \rrbracket_u^n \neq S$$

#### Revisions:

$$s \llbracket \phi \rrbracket_u^n = s \llbracket \neg\phi \rrbracket_w^n \llbracket \phi \rrbracket_u^w$$

Note the clause for downdates of identity statements in the above, defined as a *test* that the current state does not support the identity information. Analogously with interpretation of “discourse-external” information in DPL and RDPL, update and downdate by identity information in DMPL and RDMPL is simply a *test* on the current state. That is, in (R)DMPL, either  $s \llbracket x = y \rrbracket_u^n = s$  or  $\emptyset$  and in RDMPL  $s \llbracket x = y \rrbracket_u^n = s$  or  $\emptyset$ . FRL, on the other hand, incorporates genuine update and downdate of identity information. This dynamic treatment of identity information can be applied to the semantics of various dialogues.

RDMPL is complete with respect to the following theory change system.

**Definition 52** (*FOTC<sub>II</sub>*)

Add the following clauses to *FOTE<sub>III</sub>*

$$T, n \dot{-} Pt_1 \dots t_m = \bigcap \alpha(T \perp Pt_1 \dots t_m), n$$

$$T, n \dot{-} \neg \phi = \bigcap \{T' \mid T' \subseteq T \text{ and } \exists w \in N, T', n + \phi \neq \mathcal{L}, w\}$$

$$T, n \dot{-} \exists x = T, u = n(x \mid \emptyset)$$

$$T, n \dot{-} (\phi \wedge \psi) = (T, n \dot{-} \phi) \cup (T, n \dot{-} \psi)$$

$$T, n \dot{-} (x = y) = T, n \text{ iff } (nx = ny) \notin T$$

$$T, n \dot{-} \diamond \phi = T, n \text{ iff } T, n \dot{-} \phi \neq \emptyset$$

**Revision:**

$$T, n \dot{+} \phi = (T, n \dot{-} \neg \phi) + \phi$$

**Theorem 17** (*RDMPL and FOTC<sub>II</sub>*)

In addition to the usual connection between state update and theory expansion (due to van Eijck, presented above),

$$(1) Th(s), n \dot{-} \phi = Th(s \llbracket \phi \rrbracket_u^n), u$$

$$(2) Th(s), n \dot{+} \phi = Th(s \llbracket \phi \rrbracket_u^n), u$$

and if  $T$  is closed, and full and exhaustive over *Cons*, then also:

$$(3) T, n \dot{-} \phi = T', u \text{ iff } Cmod(T) \llbracket \phi \rrbracket_u^n = Cmod(T')$$

$$(4) T, n \dot{+} \phi = T', u \text{ iff } Cmod(T) \llbracket \phi \rrbracket_u^n = Cmod(T')$$

Proof:

items 2 and 4 are trivial consequences of items 1 and 3, together with van Eijck's results. The other proofs are available by combination of the results for RMPL and RDPL, for example:

Item(1):

as for RMPL and FOTC, but for the following clauses (involving negation and anaphoric information):

$$\begin{aligned} Th(s), n \dot{-} \neg Px &= \bigcap \{T' \mid T' \subseteq Th(s) \text{ and } \exists w \in N, T', n + \phi \neq \mathcal{L}, w\} = Th(s \cup \downarrow \\ & \llbracket Px \rrbracket), n \\ &= Th(s) \llbracket \neg Px \rrbracket, n \end{aligned}$$

$$\begin{aligned} Th(s), n \dot{-} \exists x &= Th(s), u = n(x \mid \emptyset) \\ Th(\{i \in s \mid i(x) = *\}) &= Th(s)x[] = Th(s) \exists Px [], u \end{aligned}$$

Item(3) as for DMPL, but with contraction and downdate:

if  $T$  is closed, full and exhaustive over  $Cons$ , then

$$T = Th(Cmod(T)), \text{ and}$$

$$T, n \dot{-} \phi = T', u \text{ iff } Th(Cmod(T)), n \dot{-} \phi = Th(Cmod(T')), n \text{ iff by part (1)}$$

$$Th(Cmod(T)), s \dot{-} \phi = Th(Cmod(T)) \phi \llbracket \frac{n}{u} \rrbracket, u \text{ iff}$$

$$Cmod(T) \llbracket \phi \llbracket \frac{n}{u} \rrbracket = Cmod(T')$$

#### 5.5.4 Adding some modals

It is a simple enough matter to extend the account of the epistemic modals ("possible, indispensable, disposable, prohibited") offered in chapter 4 to the first-order case. Again, the notation here is somewhat misleading, because appearance of a formula between "dynamic" interpretation brackets usually signals a change in information state. However, as usual, these modals are static tests which only change the state if they fail.

$$1. s \llbracket \Diamond p \rrbracket \frac{n}{u} = \{i \in s \mid n = u \text{ and } \exists r, s \llbracket p \rrbracket \frac{n}{r} \neq \emptyset\}$$

$$2. s \llbracket \Box p \rrbracket_u^n = \{i \in s \mid n = u \text{ and } \exists r, s \llbracket p \rrbracket_r^n = S\}$$

$$3. s \llbracket \Diamond p \rrbracket_u^n = \{i \in s \mid n = u \text{ and } \exists r, s \llbracket p \rrbracket_r^n \neq S\}$$

$$4. s \llbracket \Box p \rrbracket_u^n = \{i \in s \mid n = u \text{ and } \exists r, s \llbracket p \rrbracket_r^n = \emptyset\}$$

Of course, one could also give a first-order constructive “dynamic modal” along the lines of CRS (from chapter 4).

## Results

In RDMPL, consider the “theory” that “Ann, Bill and Cathy were on holiday.” Then  $Cons = \{a, b, c\}$ ,  $s = 1 \in s \mid i \models Ha \wedge Hb \wedge Hc$ , and  $\forall m \in Cons, n(m) = m$ . When only names are involved here, suppress the assignments/ namings.

“Maybe someone was ill” is consistent, so long as the variable  $x$  for ‘someone’ can consistently be associated with the name of a (possibly) ill person.

$$s \llbracket \Diamond Px \rrbracket_u^n = s \text{ iff } \exists m \in Cons, u = n(x \mid m), \text{ and } s \llbracket Pm \rrbracket_u^n \neq \emptyset$$

“Everyone might have been ill. Ann wasn’t ill” is consistent:

$$s \llbracket \forall x \Diamond Ax \rrbracket \llbracket \neg Aa \rrbracket = s \llbracket \Diamond Aa \wedge \Diamond Ab \wedge \Diamond Ac \rrbracket \llbracket \neg Aa \rrbracket = s \llbracket \neg Aa \rrbracket \neq \emptyset$$

Quantifying in: “Maybe someone was ill. Ann wasn’t ill” is consistent:

$$s \llbracket \Diamond \exists x Ax \rrbracket \llbracket \neg Aa \rrbracket = s \llbracket \neg Aa \rrbracket \neq \emptyset$$

“Ann was ill. There must have been someone who was ill” is consistent:

$$s \llbracket Aa \rrbracket \llbracket \exists x \Box Ax \rrbracket = s \llbracket Aa \rrbracket \llbracket \exists x Ax \rrbracket = s \llbracket Aa \rrbracket \llbracket Aa, Ab, Ac \rrbracket \llbracket \forall = s \llbracket Aa \rrbracket \neq \emptyset$$

“Ann was ill. Someone must have been ill” is consistent:

$$s \llbracket Aa \rrbracket \llbracket \exists x \Box Ax \rrbracket = s \llbracket Aa \rrbracket \llbracket \Box Aa \vee \Box Ab \vee \Box Ac \rrbracket = s \llbracket Aa \rrbracket \neq \emptyset$$

“Somebody wasn’t ill. Maybe everybody was ill” is inconsistent:

$$s \llbracket \exists x \neg Ax \rrbracket \llbracket \Diamond \forall x Ax \rrbracket = S \llbracket \neg Aa \vee \neg Ab \vee \neg Ac \rrbracket \llbracket \forall x Ax \rrbracket = \emptyset$$



Now try a revision:

“Everyone might have been ill. Ann wasn’t ill. No, Ann was ill.” is consistent:

$$s \llbracket \forall x \diamond Ax \rrbracket \llbracket \neg Aa \rrbracket \llbracket Aa \rrbracket = s \llbracket \diamond Aa \wedge \diamond Ab \wedge \diamond Ac \rrbracket \llbracket \neg Aa \rrbracket \llbracket Aa \rrbracket = s \llbracket \neg Aa \rrbracket \llbracket Aa \rrbracket = s \llbracket \neg Aa \rrbracket \neg Aa \llbracket Aa \rrbracket = s \llbracket Aa \rrbracket \neq \emptyset$$

### 5.5.5 First-Order Revision Logic

“First-order Revision Logic” (FRL) is an extension of van Eijck’s FUL with first-order downdates and revisions. FRL allows for genuine growth, loss, and revision of knowledge about identities (as well as the usual dialogue-internal and external information change).

The basic idea is that contractions of identity statements downdate the equivalence classes  $i_e$  on  $Cons$ . DOWDATING by identity statements relaxes conditions on the equivalence components of the canonical models. All other downdates and revisions work in the same way as in RDMPL.

**Definition 53** (FRL clauses)

Add the RDMPL downdate and revision clauses to FUL, but for the following exceptions:

$$s \llbracket x \doteq y \rrbracket_u^n = \bigcup \beta(s \perp_n (x \doteq y)), n = u$$

$$s \llbracket \exists x \rrbracket_u^n = s \text{ iff } u = n(x \mid \emptyset), \emptyset \in Cons$$

So, as in FUL, existential quantification is about “proto-individuals” or names. Downdate by an identity statement  $x = y$  now increases the set of canonical models under consideration, to include some models whose equivalence components do not support the identity statement  $nx \doteq ny$ .

**Theorem 18** (FRL and FOTC<sub>II</sub>)

Where  $s$  is a set of canonical models for  $Cons$ , and  $\dot{-}$  and  $\dot{+}$  are the theory contraction and revision operations of FOTC<sub>II</sub> respectively,

$$(1) Th(s), n \dot{-} \phi = Th(s) \phi \llbracket_u^n \rrbracket, u$$

$$(2) Th(s), n \dot{+} \phi = Th(s \parallel \phi \parallel_u^n), u$$

and if  $T$  is closed, and exhaustive over  $Cons$ , then also:

$$(3) T, n \dot{-} \phi = T', u \text{ iff } Cmod(T) \parallel \phi \parallel_u^n = Cmod(T')$$

$$(4) T, n \dot{+} \phi = T', u \text{ iff } Cmod(T) \parallel \phi \parallel_u^n = Cmod(T')$$

Proof: by composition of the RDMPL and FUL results, but for the above two clauses. The reasoning is analogous to the preceding results.

This concludes the exposition of the hierarchy of revisable systems of dynamic semantics based on van Eijck [96]. The result is a variety of systems for the processing of different types of information change. FRL, the last of the systems, can process changing information about the world, about (provisional) identity statements, and about anaphoric relations. However, there is a price to be paid for the “dynamic” identity statements of FUL and FRL; models no longer contain individuals, but some less “ontologically pure” objects instead. Chapter 6 investigates a variety of possible approaches to such “dialogue referents”.

## 5.6 Prospects and Conclusions

To conclude the chapter, I sketch some possible further developments, which depart from some of the assumptions of the preceding systems.

### 5.6.1 FOTC with growing domains

Agents commonly introduce new names and variables in dialogue, in the hope (or knowledge) that they refer.

Formally, modelling this is fairly trivial. Simply drop the assumption that the set of names ( $Cons$ ) is constant, and assume instead that it grows only under theory expansion, or state update, in the following way:

$$(D+) Cons(T, n + \phi) = Cons(T) \cup Cons(\phi)$$

where  $Cons(\phi)$  is just the set of constants appearing in  $\phi$ . The resulting Theory Change system would be complete with respect to RMPL where individuals in the domain are employed as denotations of newly introduced names.

### 5.6.2 Free Logic for Existential contractions

More interestingly, consider the issue of *losing* names and variables from a theory (as in the downdate clause for the existential in RDPL). To ask for an example of a name that is now no longer in use generates a paradox. Nevertheless, there are names that are no longer in common usage, for example, "braddle", "bubblecar", or "chivalry". More dramatically, there are names which people use in full knowledge that they fail to refer ("Batman"), and even names which are non-fictional and non-referring ("phlogiston"). On yet other names the jury is still out (consider, "superstrings," or "the antichrist.") Formally, the question becomes: Under what conditions does the domain of a theory shrink? The classic answer is that terms which are found to be non-denoting should be removed from a theory (eg: phlogiston, ether, witches). Whilst in the case of scientific theories such cases are quite rare, in the case of dialogue-based theories such terms can be quite common, for example "the person who stole my hat" ceases to be an operative discourse referent whenever I find that I've just misplaced it, and this constitutes genuine growth of knowledge. In the systems above, a "null" individual with no properties was used to accommodate such cases.

There are various candidates for conditions under which domains ought (rationally) to be contracted. Non self-identical objects are presumably to be excluded from theories, for example.

The case of objects with contradictory properties is less clear. Certain cases in the history of science might be mentioned here, but a focus on dialogue brings up similar cases; agents do not stop referring to Mary when they are told that she both is and is not coming to their party, they want to know which it is.

Agents can discover that a name is non-referring, or at least somehow dubious in some way, when it accumulates contradictory properties. In such a situation there are at least three options. The first is to just remove the offending term from the domain of discourse (but this is too brutal); another is to adopt an *outer domain semantics* in a free-

logic, to which the offending terms are exiled (a Meinongian approach.) I consider the latter option anon. Finally paraconsistent predicate logics may be explored, the intuition being that inconsistent objects are exiled to an outer domain of “paraconsistent” or impossible objects. For the moment, though, let us stay classical.

Consider a new rationality postulate: it says that under removal of a name from a theory,  $T \dot{-} \exists c$ , all equivalent constants should also be removed.

$$c_1 = c_2 \in T \Rightarrow c_2 \notin T \dot{-} \exists c_1$$

But this is far too strong for our purposes, for example: “Smith is the murderer”  
 “But there is no murderer - Jackie is alive and well!”

$$S \llbracket m = s \rrbracket \rrbracket \exists m \llbracket \llbracket A_j \wedge W_j \rrbracket \rrbracket$$

One ought not to conclude from this that Smith no longer exists, unless, perhaps, the only way in which Smith were involved in the theory is by virtue of the “murderer property”. Rather, it seems, removal of a name should result in the removal of all statements involving that name (including identities of course) from the theory.

So now take the new postulate to be:

$$(N3) T \dot{-} \exists c \not\vdash \alpha, \forall \alpha \in T \text{ such that } c \in Cons(\alpha)$$

Formally then, let existential claims be part of a theory, so that intuitively, if  $\exists c \in T$  then  $T$  contains the claim that the object named by constant  $c$  exists. Modify the syntax and semantics accordingly. Any formula in which  $c$  appears implicitly contains the claim that  $d$ , where  $n(c) = d$ , exists. There are now expressions such as  $\exists c \vee \neg \exists c_1$  to deal with.

Thus,  $T \vdash \exists c \Leftrightarrow c \in Cons(T)$  and semantically,  
 $s \models \exists c \Leftrightarrow \forall i \in s, i(c) \in D$

This has the following ramifications in a theory revision system;

$$T, n + \exists c = Cl(T \cup \{\exists c\}), n$$

$$(D-) T, n \dot{-} \exists c = \bigcap \alpha(T \perp \exists c), n(c | c') \exists c' \neq c \in Cons(T)^{27}$$

$$(D+) Cons(T + \exists c) = Cons(T) \cup \{c\}$$

Note that when a name is lost it can no longer be used in the naming part of configurations. In the semantics then, simply;

$$s \llbracket \exists c \rrbracket = \{i \in s \mid i(c) \in Dom(s)\}$$

(where  $Dom(s)$  denotes the shared domain of the set of models  $s$ ), and,

$$(D-) s \llbracket \exists c \rrbracket = \bigcup \beta(s \perp \exists c)$$

Conjecture: (Completeness of  $FOTR_{dom-}$  wrt  $RMPL_{dom-}$ )

$$Th(s \llbracket \exists c \rrbracket) = T + \exists c \text{ and } Th(s \llbracket \exists c \rrbracket) = T \dot{-} \exists c$$

$$Mod(T + \exists c) = s \llbracket \exists c \rrbracket \text{ and } Mod(T \dot{-} \exists c) = s \llbracket \exists c \rrbracket$$

It seems to be the case (see chapter 6) that people do not cease to be able to use discourse referents which they nevertheless believe are non-referring. To this end, one could partition the domain of discourse referents as in a free logic;  $Dom(i)$  is the domain of “operational” referents, while  $D'(i)$  is the domain of “dubious” or generally sketchy, referents. Thus  $RMPL_{free}$  models are of the form

$$i = \langle Dom(i), D'(i), F \rangle$$

The intuition here is that certain names are found to denote in  $D'(i)$ , and that this is genuine growth of knowledge which nevertheless does not result in the eradication of the referring expression (from the theory) or the referent (from the model). In fact, for communication purposes, it seems vital that the discourse “referent” is still operative, albeit non-referring, so that such information can be shared with other agents.

“No, there is no murderer. Jack isn’t dead after all!”

Thus, retraction of existential claims now no longer involves a diminishing domain but instead the assignment of objects in  $D'(i)$  to variables and names.

First ensure that:

$$D'(i) \cap Dom(i) = \emptyset \text{ (Disjoint sets)}$$

Then replace (D-) with ( $D_{free}$ ):

---

<sup>27</sup> $n(c | c')$  is the naming just like  $n$  but that  $c$  is replaced by  $c'$

$\exists i \in s_T$  such that  $i(c) \in D' \Rightarrow \exists c \notin T$

Of course, quantification now ranges only over  $Dom(i)$ .

### 5.6.3 Summary and Conclusions

I hope to have made the following contributions in this chapter.

1. The exploration of intuitionistic approaches to modelling first-order theory growth and change.
2. The development of several first-order systems of theory change.
3. The demonstration that FOTC systems are associated with interesting systems of dynamic semantics.

After investigating the possibility of a non-truth-persistent Quantified Intuitionistic Logic, and analysing of the capacity of Intuitionistic Predicate Logic to accommodate cases where information not only grows, but changes (is defeasible), I argued that the best way to accommodate information loss in semantics is through genuine AGM-style rational contraction.

Some basic rationality constraints on first-order theory change systems were then developed.

Next, after reviewing connections between dynamic semantics and systems of first-order theory extension, a variety of systems of first-order theory change were connected to revisable systems of dynamic semantics. Different strategies were called for in order to model changing information about the world, about anaphoric relationships, and about identities.

The approach taken in this chapter has the following virtues:

- The systems avoid stacks of information states when computing information revision.
- The proposals are technically, as well as philosophically, interesting. Extending AGM theory change and Revisable Dynamic Semantics to the predicate logical case raises

important issues to do with the (semantical and syntactical) behaviour of quantifiers, names, identities, and anaphoric relations under change.

- The approach commends a family of new systems for further study.

This concludes the purely formal development of the thesis. The remainder of the work consists of applying the systems developed herein in an analysis of corrections and a formal model of communication (chapter 7), and a discussion of possible variants of those systems (paraconsistency and autonomy, in chapter 8). The idea of changing information about names, variables, and “proto-individuals” emerging from the systems of this chapter, motivates the discussion of the next.

**Part III**  
**Applications, Prospects, and**  
**Discussion**



## Chapter 6

# Linguistic Data: analysing theory changes in communication

“A prerequisite to a theory of the way agents understand speech acts is a theory of how their beliefs and intentions are revised as a consequence of events.”  
(Appelt and Konolige, [4], p. 1)

### 6.1 Overview

The basic picture from preceding chapters should by now be evident. Dialogues are taken to modify the epistemic states of agents; the contents of their theories. Communication is to be described as theory change, or iterated belief revision, such that agents converge upon a mutually agreeable theory about the world; how it is, was, may be, or may become.

This chapter is not intended to provide a full formal account of constituent negation, but to show how the systems of previous chapters may be applied in such an analysis. A full account is beyond the scope of this work, but I claim that it would have to employ a framework much like the one offered in this thesis – where information states and interpretations are revisable. Here I provide distinctions and classifications in an analysis of correction data, and point out the route that a more detailed account would take. The axioms for speech acts at the close of the chapter indicate quite formally how the systems may be applied in a theory of communication.

In this chapter, an analysis of various linguistic phenomena (corrections, repairs, negations, editing terms, intonation) is provided, which is further evidence for the thesis that theory change is an integral part of communicative behaviour. The examples given also serve as data that a dynamic semantics of communication ought to be able to cover. Many of the claims made, although explicitly about dialogues, also hold for a full semantics of discourse – due to phenomena of *self-repair*.

The linguistic evidence to be presented in this chapter illustrates the necessity (if the goal of empirical adequacy is to be met) of the *first-order* theory change systems and Revision Semantics developed in the preceding chapters. Thus the transition of the preceding chapters, from propositional to first-order theory change, is seen to be required by the linguistic evidence.

This chapter also raises some more pragmatic issues concerning rational agenthood in general, and speech acts by which theory-revision can be supported. In general, I explore the structure of information states manipulated in dialogue by rational agents. This is an empirical enterprise to the extent that we have intuitions as to the likely (and warranted) theory-transitions of an agent within a particular dialogue context. Indeed, that we have such intuitions is the very fact that makes strategic dialogue, and communication, possible at all. An important source of empirical support will be mini-dialogues involving *corrections* in both discourse and dialogue, employing *constituent negation clauses* in particular. These examples result in an analysis of (some instances of) negation (and words like ‘but’ and ‘except’) as triggers, cues, or instructions for theory contraction and revision. Intonation is also examined as an important cue as to the correction intended by a speaker.

In addition, some more general dialogue strategies and contributions are analysed as contractions or revisions of information states:

1. Suspension of Disbelief - contraction ‘for the sake of argument’
2. Negative Definitions of concepts.  
For example; “Red ants are exactly the same as black ants - except that they’re red.”
3. Qualified Universals - defaults.

For example; “Everybody likes the Beatles - except people who have never heard them.”

#### 4. Counterfactual Conditionals

#### 5. Future Conditionals

## 6.2 Theory manipulation in dialogue

I have argued for a view of communication as construction and revision of agent-relative theories, or their contents: epistemic states. A next step is to explore the internal structure of these information states as exhibited by their manipulation in dialogue. Suffice it to say that the burden of preceding chapters has been to provide algorithms which, given utterances analysed into logical form, construct and revise theories and information states. I assume that such theories and states are partial, revisable, and may be paraconsistent on occasion. How might the systems developed in the preceding chapters be employed in a model of communication?

### 6.2.1 General features of the model

Each agent is assumed to employ two interdependent information ‘spaces’. The most important one for the analysis presented here is their current interpretation of the dialogue as it stands; this has been the province of formal semantics to date. The other vital domain of interpretation enters the picture when one begins to consider theory change processes instigated by *belief integration*. Communication has to take into account the fact that interpretations of utterances may not “fit” informationally with the hearer’s own epistemic state; the actualities or possibilities towards which the hearer has epistemic attitudes. The fact that some kind of integration (unification or revision) happens between information constructed from discourse and dialogue interpretation and an agent’s own information (memory, knowledge, beliefs) leads to two conclusions. The first being that the information structures derived from utterances and those characterising “beliefs” (or background information) must be (at some level) comparable<sup>1</sup>. Also see Kamp’s (1985) Unity of Thought and Information [62] and compatible. The

---

<sup>1</sup>see Asher in [49]

second conclusion is that an adequate semantics of communication must take account of background information and revisions relative to it, because they have a direct effect on dialogue interpretation, communicative strategy, and linguistic behaviour in general (see the arguments in chapter 1).

It was noted that the two “information spaces” are interdependent; in that inferences available in the “background” space influence dialogue interpretation, and that revisions instigated by dialogue interpretation may change the content of the background space. As an idealization, agents are taken to (attempt to) incorporate dialogue information into the background space at the close of each sentence. This process of “suspension of disbelief” whilst processing each sentence can, in reality, obviously go on much longer (for example, while considering an argument presented by a speaker). The issue of the incrementality of revisions and interpretations is an interesting one, but it is assumed in this simple model that a sentence close triggers the revision process. A different model might take each conversational turn as the revision trigger, or a finer-grained incremental approach might try to revise after every sentential sub-clause.

Two important processes now clearly present themselves. The construction (but not the revision) of dialogue interpretations in the “dialogue space” has, to a large extent, been dealt with by dynamic semantics (though theory change will have important lessons for the information structures developed therein.)

An important point to note is that each dialogue space is an *agent-relative* interpretation of the utterance information. Each agent has their own stock of “dialogue referents” and inferences which enable them to construct an interpretation. Thus there is no objective agent external “dialogue content space” into which information is placed in order to be “grasped”. This allows agents to talk at cross purposes, mix up referents, and make other common mistakes, as well as repair them. Indeed, part of the force of chapter 1 was the argument that *agent-relative theories* allow, and even necessitate, an account of communication.

### 6.2.2 Game-theoretical Dialogue Semantics

Dialogue has been analysed via the tools of game-theory (eg: Carlson [16], Houghton and Isard [57]). In a game-theoretical analysis communicative interactions are classified in terms of a limited repertoire of conversational moves open to each conversant. Each utterance or “move” made in a dialogue “game” commits the participant making it to a number of options or obligations; the *public content* of the dialogue. Thus, assertions are taken to commit conversants to certain claims about the world. The underlying idea is something like that employed in eliminative dynamic semantics; each utterance commits the speaker to “inhabit” or endorse a certain set of possibilities.

The treatment of dialogue offered by game-theoretical analysis is not, then, fundamentally at odds with the project of this thesis. The research here modifies and extends the approach offered by game-theory, rather than challenging it directly. In addition to an analysis of assertions, a Dynamic Semantics of dialogues employing information state revisions offers an analysis of retraction and correction in dialogue.

In particular, however, the notion of *public content* of a dialogue in a game-theoretical analysis is troublesome. That such a common-ground or shared information state is not actually directly available to participants (ie: does not actually exist for them) was the upshot of the deliberation over the “problem of communication” in chapter 1. Instead, I employ each participant’s version of the dialogue-theory (their minimally derivable theories produced by the dialogue), to do something like the same work. This was part of the “solution” to the problem. In practice I also assume that each participant can actually remember all the utterances made in the dialogue (the idealization is that nobody mis-hears, mis-remembers, or fails to infer.) To complicate matters a little, agents also have “dormant” private doxastic states, which constitute their background information about the world. When there is a mismatch between the background state and the state derived from the dialogue theory, repair, revision, and querying occurs.

## 6.3 Repairs and Corrections in Dialogue

Carletta, Caley, and Isard [15] make the following claim;

“Repairs can be used to investigate the cognitive processes underlying language production because they are a visible trace of where those processes break down, showing us the limits of those processes and something of how they interact with each other.”

Furthermore, as van Leusen recognises in her analysis [99] of corrections,

“an implementation that rises above the level of a toy example will put heavy requirements on a formal theory of discourse interpretation . . . since corrections induce a revision of the contextual information, it should be possible to represent a change of commitment of a participant towards a piece of contextual information.”

This is, of course, precisely the formal job that the various systems of Revision Semantics developed earlier are supposed to do. I now turn to an analysis of corrections (as well as other communicative phenomena) in terms of revision semantics.

Here I present and analyse some important data; corrections or repairs in discourse and dialogue. A correction occurs when an utterance (the “correction” or “repair”) is contrastive with a previous statement (the “correctum”<sup>2</sup>), such that the new information supersedes or overrides information given in the previous statement. Prior utterances set up the context in which the correction makes sense; a correction cannot just be made *a propos* of nothing. There is also often an important *structural parallelism* between the correction and the sentence which it repairs.

Four kinds of correction data will be distinguished; self-correction, hearer-correction, content-correction, and non-content correction. I am most interested in *content corrections* of the hearer (where the information conveyed by a preceding speaker’s statement is under dispute). For example (throughout I use upper case letters to denote stress, and letters A,B . . . to denote distinct agents):

**Example 24** (*van Leusen 1994*)

A: *They gave Tim a blue kite.*

B: *No, they gave Tim a YELLOW kite.*

Here the correction rejects certain information in the correctum and replaces it. The predicate in focus (“yellow”) is a contrastive parallel element to the predicate “blue”, so that the kite is now understood to be yellow, rather than blue *and* yellow. Note that

<sup>2</sup>In [15] this is termed the “OU”, for “original utterance.”

the correction relies implicitly on shared background (“conceptual”) beliefs about kites and colours; the assumption that agents have background ‘theories’ about aspects of the world which they may access freely during dialogue.

There is a delicate issue in deciding whether it is the predicate “yellow”, or the identity of the object (the kite), which is the focus of the correction here. Van Leusen assumes that both the predicate and the referent are corrected (ie: that a different object of a different colour is being referred to in the correction). However, it seems clear that there is a genuine ambiguity here, which ought to be treated either one way or the other. My intuitions are that where the referent is an indefinite object it is the predicate that is corrected, and where the referent is definite it is the identity of the object itself which is at stake (again see chapter 8 for more discussion of these issues).

Again, the above example also indicates that one must distinguish between dialogue referents and what is predicated of them.

Of related interest are “non-content corrections”, which repair syntactic form, stylistic features, and general lexical confusion, but do not affect the content of the corrected utterance.

**Example 25** (*van Leusen 1994*)

A: Alan trapped the mongeese.

B: He trapped the monGOOSES, you mean.

There is no disagreement about the content of the correctum here.

Both content and non-content corrections obviously have a crucial part to play in any theory of *language acquisition*. I shall not pursue such a theory here, but note that it might employ techniques much like those made available by a Revision Semantics.

There are two kinds of content-correction data of major interest; *self-correction* or *self-repair* (in discourse) and *hearer-correction* (in dialogue). Self-correction occurs when an agent wishes to repair one of its own (previous) utterances, or even part of a sentence which is currently being produced. *Hearer-correction* is the repair of another speaker’s

preceding utterance by the current speaker.

### 6.3.1 Self-correction

Levelt [73], Blackmer and Mitton [14], and Hieke [56], have collected self-repair data in Dutch, English, and German. Carletta Caley and Isard [15] discuss the following example of Levelt's;

**Example 26** *"Go from left again to /uh .../ from pink again to blue"*

In the above example of self-repair the speaker realises that the word "left" is incorrect and backtracks to make a correction.

Temporary memory failure can require a speaker to repair one of his or her own utterances;

**Example 27** *"I've read a few pages of your book. Ah, no, actually I read the whole first chapter late last night."*

Whatever the psychological processes (memory failure, confusion) giving rise to such episodes, one thing is clear; speakers are at liberty to retract and revise previously conveyed information. Such speakers do not mean actually to contradict themselves at all (self-repair does not invoke self-contradiction). Rather, we interpret them as *correcting* themselves, retracting a previously made commitment, and replacing it with a new one. The mechanisms of self-correction appear to be exactly the same as for multi-agent (speaker) corrections; primarily constituent negation clauses. Thus the speaker appears to treat his or her own prior utterances in the same way as utterances made by another dialogue participant. Once made, all utterances are treated as "public property", as commonly available (shared) information<sup>3</sup>.

<sup>3</sup>Speakers assume that participants understand our utterances unless it is made obvious that they don't.



### 6.3.2 Hearer Correction

Hearer corrections are perhaps slightly less common than self-repairs. They arise from the wish of a dialogue participant to revise information presented by another speaker. Presumably a collection of hearer-corrections could also be culled from the Map Task corpus. Noor van Leusen ([99], pp. 523-532) discusses the following dialogue correction;

**Example 28** *A - The journalists are interviewing Arafat.*

*B - No, they're interviewing RABIN.*

Obviously B's utterance is not to be interpreted as saying that the interviewers are *also* interviewing Rabin. Nor is it the case that the dialogue degenerates into absurdity ( $p \wedge \neg p$ ; the interviewers are and are not interviewing Rabin.) Rather, the point is that information in B's utterance is supposed to *replace* that provided by A. Note also that the repair here is specifically directed to one part of the predicate-argument structure of A's utterance, so that the correction makes use of the form of the previous sentence. Thus a *structural parallelism* exists between the correction and the correctum. However, as van Leusen shows, the parallelism is not between the surface syntactic forms of the correctum and correction:

**Example 29** (*van Leusen 1994*)

*A: Peter hit Michael.*

*B: No, HENRY was hit by him.*

Here the corrected element (Michael) is direct object, while its correction is given as subject (Henry). Thus, it seems that it is the *thematic structure*, rather than the syntactic structure, that is shared by correction and correctum; 'Henry' and 'Michael' both take the 'theme' role in the sentences in which they occur.

This analysis leads to the following conjectures (see van Leusen [99]):

(1) Corrections utilize the thematic structure of the correctum.

(2) Where correction and correctum share the same predicates, we compare and contrast the arguments of those predicates which share thematic roles.

(3) Where correction and correctum share the same “objects” or predicate arguments, we compare and contrast those predicates which share objects.

Note that in the Blue/Yellow Kite example it is the predicate itself which is the contrastive parallel element since the predicate is in focus, rather than the object. This example thus falls under case (3) above. Van Leusen loosely terms this a “conceptual parallelism”, presumably because the correction relates to the properties of some ‘conceptual object’. A first-order revision semantics treats this as a case of change in the properties of a “theoretical entity.” Again, this was the subject of chapter 6.

Another kind of parallelism is illustrated by the following infelicitous example;

**Example 30** (*van Leusen 1994*)

*A: Mary taught Peter Latin during last year’s holiday.*

*# B: No, she taught him Latin at the beach.*

The correctum specifies a period in time, while the attempted correction specifies a location. This example shows that the required parallelism is not structural, but semantical.

## 6.4 Analysis

Self-repairs and Hearer-repairs are to be analysed as information state revisions, in the style of the systems of Revision Semantics given in preceding chapters.

In contrast to an approach in terms of “sequence semantics” [104], which assumes that agents keep track of “stacks” of *all* their previous information states, an account in terms of revision semantics does not require agents to manipulate anything other than their current information states. The constituent negation clauses analysed here only make use of information conveyed in utterances preceding the repair.

In terms of theory change, first-order logic is required to describe most of the above examples. Consider the Arafat/Rabin example given above.

**Example 31** Let  $Cons = \{a, r, l, m, n\}$ , the names for Arafat, Rabin, and 3 journalists. Assume also that background knowledge dictates that Arafat is not Rabin, and that no-one can interview themselves.

Analyse A's assertion as an update ("The journalists are interviewing Arafat");

$$T = \{ \forall x \neg Ixx, \neg(a = r), \neg Ja, \neg Jr, Jl, Jm, Jn \} + \forall x (Jx \rightarrow Ixa) = \{ \forall x \neg Ixx, \neg(a = r), \neg Ja, \neg Jr, Jl, Jm, Jn, Ila, Ima, Ina \}$$

B's instruction is to revise  $T$  in the following way: contract the information that any journalist is interviewing Arafat, and add the statement that they are all interviewing Rabin.

$$T \dot{-} \exists x (Jx \wedge Ixa) + \forall x Jx \rightarrow Ixr = T \forall \{ Ila, Ima, Ina \} + \forall x Jx \rightarrow Ixr = \{ \forall x \neg Ixx, \neg(a = r), \neg Ja, \neg Jr, Jl, Jm, Jn, Ilr, Imr, Inr \}$$

An analysis of this dialogue in dynamic semantics would make use of the system RDMPL (tracking revisable dialogue-external information, without dynamic identity.) In terms of theory change, the first system of FOTC could be used.

### Change of Identity under correction

In some cases new information is accommodated by an existing dialogue referent; this certainly is the case for simple updates, where new information is given as a predication of the existing dialogue referent. During some repairs the referent survives correction even though its properties have changed, and it inherits modifiers and arguments from the correctum. However, as illustrated by the blue/yellow kite example, where predicates are the focus of correction, the issue can arise as to whether or not a new dialogue referent comes into play. Sometimes it is more appropriate to assume that a new object has been introduced to carry the correction information.<sup>4</sup> In these cases, an indefinite introduced in a prior utterance is either found to change its properties (eg: from blue to

<sup>4</sup>Considering the analogy with scientific theories as long multi-agent dialogues the same issue crops up; was Fresnel actually referring to the electromagnetic field in his theory of optics based on the "elastic ether"? In other words, once Fresnel's theory underwent repair (via Maxwell's correction of his mistakes about the nature of the ether), do we (liberally) accept that Fresnel was talking about the electromagnetic field all along (even though he was wrong about its properties), or do we insist on a replacement of the theoretical entity?

yellow:  $Bx \dot{+} Yx$  ), or to refer to a different individual.

These questions are dealt with by way of rationality constraints on objects under change. Again, these issues are more the subject of chapter 6.

## 6.5 Mechanisms of Correction: Linguistic Cues for Theory Revision

One major task of this chapter is to reveal the kinds of theory revision strategies that linguistic agents actually use, by dint of some linguistic analysis. This analysis will support the argument for the necessity of interpreting certain NL expressions as instructions (or requests) to perform particular types of theory revision. These revision operations will be found to be considerably more fine-grained than the account offered by AGM theory<sup>5</sup>.

I investigate the following contraction and revision triggers:

- Constituent Negation
- Suspension of Disbelief - contraction 'for the sake of argument'
- Negative Definitions
- Qualified Universals - defaults
- Counterfactual Conditionals
- Future Conditionals
- Contractions of Sub-theories

### 6.5.1 Constituent Negation as Revision Cue

The capacity to interpret and employ constituent negation clauses for repairing linguistic information, in either discourse or dialogue, is not a particularly sophisticated one.

---

<sup>5</sup>This ought not to be a surprise. AGM theory change was never intended to deal with the phenomena considered here.

In fact we acquire these skills relatively early on in our linguistic development, if indeed we acquire them at all. It may even be a prerequisite for language acquisition that corrections are interpreted properly by an agent. (It is certainly a prerequisite for language teaching that corrections be produced properly by the teacher<sup>6</sup>.) For example, the following behaviour, an example of constituent negation for self-correction, was exhibited by a child of 46 months (pointing at a coloured block);

**Example 32** *“That one is yellow. No, that one is orange.”*

Nick Asher [5] mentions the use of constituent negation following an assertion by a dialogue partner<sup>7</sup>. For example,

**Example 33** *A: “Zebedee and Ermintrude are going to visit Dougal.”*

*B: “No, it’s not Ermintrude who’s going with Zebedee, but Dylan.”*

Notice that a truth-conditional view of the meaning of B’s utterance (that Dylan and Zebedee are going to see Dougal) does not capture its real character; that it is an attempt to correct A’s assertion. Current dynamic accounts do not fare much better however (in fact many fare much worse, since they classify B’s utterance as rendering absurdity). A dynamic account of linguistic interpretation can only account for the above dialogue if it offers an account of revision.

The above example allows more specific inferences to be drawn. That the negation in B’s utterance only has scope over the proper name “Ermintrude” tells us something about the information structure that is being manipulated here. B’s instruction to A is to replace just one part of his previous assertion, indeed to change one of the arguments of one predicate. This analysis supports the claim that, in order to describe B’s utterance adequately, we must take the information manipulation involved to exploit at least predicate-argument structure. In addition, this information content must be linked to the surface syntactic form of B’s utterance, from which underlying process of theory

---

<sup>6</sup>See [9]

<sup>7</sup>Apparently Hans Kamp brought this phenomenon to Asher’s attention.

revision is derived. Revision by constituent negation then exploits the grain of theory structure underlying NL information states.

Constituent negations can be interpreted in two ways (depending on their force) – as constituent *retractions* of information (via “downdate”) or as revision instructions.

For example, in “No, it’s not Ermintrude who’s going with Zebedee, but Dylan”, the negation clause is a constituent retraction instruction, and “Dylan” fills the resulting ‘gap’. However, in “No, DYLAN went with Zebedee” the force of the constituent negation is that of a revision instruction. Notice that B’s revision instruction need not be in response to an explicit assertion from A, for agent B may have inferred that A has some information that needs revision. In addition, of course, A’s *autonomy* means that he is not forced to accept B’s instruction. A understands what B intends as an appropriate revision, and then judges whether or not to carry out his instruction. Autonomy implies that acceptance of the revision instruction is not automatic. This is a complexity that is ignored in most work on revision (AGM revision just gives priority to the most recent information), but I shall have more to say on the matter in the next chapter<sup>8</sup>.

The investigation proceeds by finding expressions that can occur in the scope of constituent negation. I claim that each of these expression types can be interpreted as the argument of a theory change or state revision function.

As expected, propositions can appear in the scope of constituent negation,

**Example 34** *“No, arithmetic is incomplete.”*

This example licenses the AGM unit of revision and contraction (the proposition) as applicable to NL revision. However, more fine-grained units of revision are common.

Predicates (adjectival expressions) can also appear in the scope of constituent negation,

**Example 35** *A: “Whoever smashed that window was an idiot”*

*B: “No, he wasn’t an idiot, he was just unlucky.”*

---

<sup>8</sup>see Chapter 8: Paraconsistency, Autonomy, and Change

Determiners (and quantifiers) can also be the arguments of a revision function,

**Example 36** A: *"Everyone hates that idiot who smashed the window"*

B: *"No, not everyone, just a few people."*

**Example 37** A: *"Only a few people here eat meat."*

B: *"No, actually everyone here eats it."*

Modifiers appear in corrections too,

**Example 38** (*van Leusen 1994*)

A: *"Mary taught Peter Latin during last year's holiday."*

B: *"No, she taught him Latin 3 YEARS AGO."*

The following example also illustrates inheritance of the modifier from the correctum:-

**Example 39** (*van Leusen 1994*)

A: *"Soon after twelve o'clock, everybody left."*

B: *"Well, MOST people left." (soon after twelve o'clock.)*

Agents can dispute universals and conditionals, which have a more radical effect on the content of a dialogue,

**Example 40** A: *"If John's got grandchildren, then his kids must be adults."*

B: *"No, his kids might be dead."*

**Example 41** A: *"Every animal has a heart."*

B: *"No, not every animal. Slugs don't."*

To summarize, then, I have observed that corrections using constituent negation have scope over the following grammatical units:

1. propositions
2. determiners

3. quantifiers
4. modifiers
5. conditionals

Finally, the analysis presented above begs the question of limits on the types of expression that can participate in revisions. In other words, which elements *cannot* fall inside the scope of constituent negation?

Asher<sup>9</sup> claims that grammar rules out sentences and ordinary conjunctions from the scope of constituent negation. However, some of the examples given above use propositions in the scope of constituent negation, and are perfectly well-formed.

All in all, the above examples suggest an analysis in terms of contraction and revision of partial theories, and their underlying information structures. The process relies on contraction of a certain expression from the hearer's current theory, and its replacement by a different expression of the same semantical type. Constituent negation is analysed as theory contraction specified by the focus of the correction, ascertained by way of structural parallelism with the correctum, leaving a "gap" which may or may not require filling. The combination of contraction and subsequent update of that expression type results in revision or correction

Note that constituent negation is really a *request or instruction* to contract, and that it need not be obeyed;

**Example 42** A: "Jackie went to the cinema with John last night."

B: "No, he didn't go with Jackie."

If A accepts B's instruction, or carries out his request to contract, the result will be a "gap" in A's information. Agent A will be more likely to carry out B's request if this gap can be consistently filled with information offered by B; agents do not like to lose information if they can avoid it (by conservation of information). Thus, there seems to be room for a process of "suspension of disbelief" by the hearer, before he or she decides either to accept or reject the requested contraction or correction.

---

<sup>9</sup>See [7] page 312.



A's autonomy means that he might need to be convinced that the retraction that B wants him to accept is coherent, or consistent with his current theory. Suppose B detects A's hesitancy in accepting the retraction; he might add,

B: "He went went with Kate."

So long as this correction is consistent with A's current theory, A (rationally) ought to accept B's repair. Thus we can make the conjecture that an agent accepts a contraction instruction on the proviso that the resulting revision is consistent with other information.

Another point to note is that revision instructions can be shared in a multi-agent context:

**Example 43** A: "*Everyone hates that idiot who smashed the window*"

B: "*No, not everyone does. Just a few people.*"

C: "*And he's not an idiot, just unfortunate.*"

Finally, the recent phenomenon of *post-sentential negation*, often employed for comic effect, might be analysed as a (playful) revision instruction.

**Example 44** "*You are so intelligent. . . . not!*"

This is another example of the proposition as a unit of revision. A sentence is asserted, followed swiftly by its negation. So, corrections can also be used to comic effect.

### **6.5.2 Intonation and "Editing Terms" as revision cues**

Intonation and stress patterns are often signals about what sort of revision the speaker considers appropriate, while so called "editing terms" (such as "uh", "sorry", "erm", "no", "but", "I mean", "you mean") can be used both by speakers in self-correction or by hearers who wish to interrupt and correct the current speaker.

**Example 45** *“HE didn’t write “Hamlet” (it was Marlowe.)”*

The revision here has the effect of focussing on the subject as the candidate for repair (rather than the name of the play itself.) People also use intonation to focus on an object as repair candidate:

**Example 46** *“Jane didn’t kiss SAM, it was Jack.”*

The next example shows how stress can be used to focus on predicates ;

**Example 47** *“You may not believe it, but I didn’t KISS her.”*

- this kind of denial disputes not the individuals involved, but the predication used to relate them.

## **6.6 Other Theory Contractions**

Andre Fuhrmann mentions some interesting examples of theory contractions dictated by certain kinds of natural language constructions. Here I elaborate upon Fuhrmann’s considerations (in [35] p. 71 - 73), and discuss how they might be accommodated in a fully fledged first-order Revision Semantics. I also add some classes of examples of my own.

### **6.6.1 Contraction “for the sake of argument”**

A strong motivation for considering theory change is the analysis of suspension of disbelief, or contraction ‘for the sake of argument.’

**Example 48** *“Just suppose for the moment that Sartre is wrong about the contingency of existence. Then what follows?”*

Here, the hearer is instructed to imagine the result of contraction from a theory, even though that contraction might not ultimately be warranted.

A full analysis of such constructions will have much in common with that offered for counterfactual conditionals in Revision Semantics (chapter 4).

### 6.6.2 Negative Definitions of Concepts, and definition by revision

Broadly speaking, assume a ‘concept’ of something to be similar to a theory about its properties and relations to other theoretical entities. People commonly use figures of speech in order to define new ‘concepts’ in terms of others;

**Example 49** *“Modern classical music is just ambient music, but without all the electronic instruments.”*

The new concept (or theory) to be defined is given by contracting a property or set of properties from another ‘concept’. Concepts can also be defined by way of revision (again, such examples could be handled by a system like RDMPL):

**Example 50** *“Red ants are black ants, except they’re red.”*

Such definitions are used when the speaker believes that an audience understands one particular concept, some of the properties of which can be ‘removed’ in order to define a more general concept, or revised in order to specify a slightly different concept. One might imagine that such a definition can be captured by way of conjunction, but as Fuhrmann points out (in [35]), such a construal would lead to contradictory definiens for all negative definitions  $\alpha = \beta \wedge \gamma$ , since  $\gamma$  is intended to deny part of whatever  $\beta$  affirms. Thus I intend such definitions to be captured by way of theory contraction and revision.

### 6.6.3 Universal Claims with Exceptions

As in the literature on default reasoning, there are the following kinds of everyday examples of qualified universal claims,

**Example 51** *All that food is Naoko’s, except maybe that melon.*

In RMPL, this sentence could be analysed as:

$[[\forall xNx]] Nm [$

**Example 52** *No mammals lay eggs, except the platypus.*

**Example 53** *Everyone is going to vote, but Sam might not.*

Again, these claims cannot properly be captured conjunctively, for contradictions immediately ensue. Rather, I take qualifications of universal claims to be interpreted as (first-order) theory contractions and revisions.

#### 6.6.4 Counterfactual and Future Conditionals

Recall the treatment of counterfactual conditionals sketched in Revision Semantics (chapter 4). In addition, it seems that some future conditional statements are tests about theory revision.

**Example 54** *“If Major had lost his seat, Tony Blair would be prime minister now.” (Counterfactual Conditional)*

**Example 55** *“If you tell her the truth about last night, she will leave you.” (Future Conditional)*

In the first case the consequent of the conditional needs to be contained in the hearer’s theory (about politics) revised by the proposition in the antecedent.

Counterfactual Conditionals:  $\phi \mapsto \psi$  iff  $T \dot{+} \phi \models \psi$

Future conditionals can, it seems, be treated in the same way, but (strictly speaking) using *updates* rather than revision, because the antecedent adds to, rather than contradicts, the current “state of affairs”.

#### 6.6.5 Removal of sub-theories

Sometimes, in particularly drastic situations, an agent is instructed to remove whole sub-theories from a larger theory.

**Example 56** *Everything you believe about last night is wrong.*

**Example 57** *Everything I told you about quantum mechanics is a lie.*

These cases require the theory of general contractions developed by Fuhrmann; in particular, the above examples can be analysed as instances of package contraction (see chapter 5).

This completes the investigation of the scope of repair and contraction expressions in dialogues. I claim that the semantics of such expressions can be formally analysed using First-Order systems of Revision Semantics.

## 6.7 Revision Semantics in a formal theory of communication

Towards the close of [21], Paul Dekker provides the beginnings of a theory of information exchange in dynamic semantics. He argues that the standard epistemic modals are better understood in a system of information exchange, where they serve to establish agreement between agents. However, Dekker's system only accounts for information growth between communicants (he is quite explicit about this limitation.) Recall, the problem (from chapter 1),

"If the speaker attempts to exchange the information that  $\phi$ , and the hearer has information to the contrary, then the exchange is simply taken to come to a halt. For the exchange to proceed in such a situation, a higher order discussion may be required ... as well as some method of belief revision. Since ... belief revision fall[s] beyond the scope of the present undertaking, we just have to settle for expelling the occurrence of inconsistency of information."

(Paul Dekker, [21], pages 211-2.)

The problem is that, using EDPL, agents can only communicate about information that they agree upon, or can consistently add to their information states. The prize offered by the RDMPL account of information states is an account of communication where agents can disagree and resolve conflicts.

Let  $s_a$  and  $s_b$  be the information states of two agents  $a$  and  $b$ . Following Dekker, pages 209-218 of [21], take the ordered pair  $\langle s_a, s_b \rangle$  as the information state of a two-agent system<sup>10</sup>. Assertion of  $\phi$  by either agent, results in the new state  $\langle s_a, s_b \rangle [\phi] = \langle s_a [\phi], s_b [\phi] \rangle$ . By the Gricean maxim [45] of sincerity, whoever made that assertion already has that information, so either  $s_a [\phi] = s_a$  or  $s_b [\phi] = s_b$  (or both if both agents already believe  $\phi$ ). In Dekker's system one also has to make the consistency assumption that, if agent  $b$  is the hearer,  $s_b \not\models \neg\phi$ , otherwise  $s_b$  implodes and the conversation is over.

The virtue of systems of Revision Semantics is that no such consistency assumption

<sup>10</sup>This can easily be extended to a multi-agent model.

need be made. If  $a$  makes an utterance that conflicts with  $b$ 's information,  $b$  can use a constituent negation, or other dwndate trigger, to remove that information from  $a$ 's state. Of course, this means that whoever is speaking has the power to change the beliefs of a hearer almost at will. Such a system models a gullible hearer and a charismatic or authoritative speaker. Not too much is wrong with this as long as the speaker is well-informed and sincere. However, ultimately, agents ought to be autonomous in their response to utterances.

Consider the following account, where  $s_a$  is a sincere speaker:

$$\begin{aligned} \langle s_a, s_b \rangle \llbracket \phi \rrbracket &= \langle s_a \llbracket \phi \rrbracket, s_b \llbracket \phi \rrbracket \rangle = \langle s_a, s_b \llbracket \phi \rrbracket \rangle \\ \langle s_a, s_b \rangle \llbracket \phi \rrbracket &= \langle s_a, s_b \llbracket \phi \rrbracket \rangle \\ \langle s_a, s_b \rangle \llbracket \phi \rrbracket &= \langle s_a, s_b \llbracket \phi \rrbracket \rangle \\ \langle s_a, s_b \rangle \llbracket \diamond \phi \rrbracket &= \langle s_a, s_b \rangle \text{ iff } s_a \llbracket \phi \rrbracket \neq \emptyset \text{ and } s_b \llbracket \phi \rrbracket \neq \emptyset \\ \langle s_a, s_b \rangle \llbracket \square \phi \rrbracket &= \langle s_a, s_b \rangle \text{ iff } s_a \llbracket \phi \rrbracket \neq S \text{ and } s_b \llbracket \phi \rrbracket \neq S \end{aligned}$$

The above clauses establish the basics of a system of information exchange (one could also add a variety of modals and conditionals, from both RS and US). The following axioms provide a more detailed treatment.

$U_a^+(\phi)$  and  $U_b^+(\phi)$  stand for assertions of formula  $\phi$  by the agents. Similarly,  $U_a^-(\phi)$  formalises *retraction* of  $\phi$  by agent  $a$  ("No, I didn't say that X"). The sentence  $B_a(\phi)$  stands for  $a$ 's belief that  $\phi$ .  $s'_b$  is  $b$ 's information state immediately after change from  $s_b$ .

Agent  $a$  "believes that"  $\phi$  if and only if their current information state supports  $\phi$ .

In other words

$$B_a(\phi) \Leftrightarrow s_a \models \phi$$

### 6.7.1 Axioms for assertion (update)

The minimal system models sincere and gullible/credulous communicants, who cannot contradict each other without collapse of the dialogue.

- U1  $U_a^+(\phi) \Rightarrow s_a \models \phi$  (Assertion theory of belief/ Sincerity)
- U2  $U_a^+(\phi) \Rightarrow s'_b = s_b \llbracket \phi \wedge B_a(\phi) \rrbracket$  (Integrity Assumption) (Gullibility)

- U3  $U_a^+(\phi)$  and  $s_b \models \phi \Rightarrow s'_b = s_b$  (Vacuity)

U1 states that speakers only assert what they believe to be the case.

U2 states that the hearer models the speaker's beliefs/information, and that the hearer unquestioningly accepts new information at the expense of old.

U3 ensures that telling hearer something he already believes has no effect (a simple consequence of the definition of update).

### 6.7.2 Axioms for retraction (downdate)

Agents can also remove a hearer's commitment to information.

- D1  $U_a^-(\phi) \Rightarrow s_a \not\models \phi$  (Sincerity)
- D2  $U_a^-(\phi) \Rightarrow s'_b = s_b \parallel \phi \vee B_a(\phi)$  [(Integrity Assumption) (Gullibility)]
- D3  $U_a^-(\phi)$  and  $s_b \not\models \phi \Rightarrow s'_b = s_b$  (Vacuity)

### 6.7.3 Axioms for modals

Agents can use epistemic modals to test each other's states, and establish shared information.

- M1  $U_a^+(\diamond\phi) \Rightarrow s_a \not\models \phi$  (Sincerity)
- M2  $U_a^+(\diamond\phi)$  and  $s_b \models \phi \Rightarrow U_b^+(\phi)$  (Co-operation/Sharing)
- M3  $U_a^+(\diamond\phi)$  and  $s_b \models \neg\phi \Rightarrow U_b^+(\neg\phi)$  (Co-operation/Sharing)
- M4  $U_a^+(\diamond\phi)$  and  $s_b \not\models \phi$  and  $s_b \not\models \neg\phi \Rightarrow U_b^+(\diamond\phi)$  (Vacuity)

M2 and M3 have the effect of ensuring that dialogue participants respond helpfully to modal utterances. The modals are almost like questions here; they operate so as to elicit any relevant information the hearer has.

M4 simply says that a hearer echoes the modal expression if they are unable to respond more helpfully.

#### 6.7.4 Axioms for (teacher/pupil) repair

Dropping the gullibility assumption, and installing some level of autonomy, agents can disagree and correct each other. Here, update is taken to revision in all cases for the pupil  $a$ , who can learn from the teacher  $b$ , who does not believe anything that the pupil states. In other words,  $a$  is gullible, while  $b$  is sceptical and authoritative.

- R1  $U_a^+(\phi) \Rightarrow s_a \models \phi$  (Sincerity)
- R2  $U_a^+(\phi) \Rightarrow s'_b = s_b \parallel B_a(\phi) \parallel$  (Integrity Assumption)
- R3  $U_b^+(\phi) \Rightarrow s'_a = s_a \parallel \phi \wedge B_b(\phi) \parallel$  (Integrity Assumption)
- R4  $U_a^+(\phi)$  and  $s_b \models \neg\phi \Rightarrow U_b^+(\uparrow \neg\phi)$  (Correction/Repair instruction)
- R5  $U_b^+(\uparrow \phi) \Rightarrow s'_a = s_a \parallel \phi \wedge B_b(\phi) \parallel$  (Correction/Repair acceptance)
- R6  $U_a^+(\phi)$  and  $s_b \not\models \phi$  and  $s_b \not\models \neg\phi \Rightarrow U_b^-(\phi)$  (Downdate instruction)
- R7  $U_b^-(\phi) \Rightarrow s'_a = s_a \parallel \phi \parallel$  (Downdate acceptance)

R1 states that the pupil never lies. R2 states that the teacher believes that the pupil believes whatever he utters. R3 states that the pupil believes whatever the teacher utters.

R4 states that the teacher corrects utterances made by the pupil which she considers to be incorrect.

R6 states that the teacher admits (and encourages!) ignorance, and R7 states that the pupil becomes ignorant of whatever the teacher is too (perhaps this models a teacher who disabuses the pupil of their, presumably unwarranted, beliefs).

#### Installing genuine autonomy

The above pupil/teacher system highlights a limitation of the theory change approach. New information is always believed, according to the AGM tradition. Applying AGM techniques results in a theory of communication in which whoever is speaking can brainwash their unfortunate audience. Genuine autonomy would result in a system where



rational agents judge whether or not to take new information on board. This issue is the subject of the next chapter.

## **6.8 Conclusion**

In this chapter I hope to have given some empirical motivation for the formal systems of preceding chapters, brought to light some interesting linguistic phenomena, and advanced the beginnings of a formal theory of communication. The theory allows agents to interpret sentences which conflict with their current information (both about the world and about the dialogue), and to dispute and repair information provided by other agents in a dialogue. The following chapter explores how considerations about autonomous communicating agents and paraconsistent dynamic systems might interact.

## Chapter 7

# Paraconsistency, Autonomy, and Change

This chapter is intended to further projects initiated in chapters 4 and 6; the investigation of the potential for a fruitful combination of paraconsistent logics with systems of theory change and of dynamic semantics. The marriage is explored with respect to the application of the resulting systems in modelling the changing information states of *autonomous* agents in dialogue (see eg: [38]). As defined in the first chapter, an agent is *autonomous* if it does not automatically accept new information, but *judges* whether or not to do so on the basis of information available to it. AGM theory assumes that new information automatically overrides current information, partly because its concern is with changing *knowledge* rather than *belief*. In some contexts this assumption makes sense (in scientific experimentation perhaps), but it is not applicable in a semantics of dialogue (as argued in the preceding chapters).

The basic idea here is to modify the base logic (the “internal logic”) of theories and information states, so that inconsistencies can be localized. The theory change component then acts so as to minimize the number of inconsistencies in a theory or information state. Agents are taken to accept revision information in dialogue if the proposed information state is “better” than their current state. This begs the question of how one is to judge between competing theories, a (very tough!) general question which is beyond the scope of this work. I offer a modest way of grading competing information states in

terms of their “information quality.” A measure of the “Information Quality” (IQ) of a state is developed and employed in a system of autonomous first-order theory change. The result is also applied in an embryonic formal system of communication between autonomous rational agents, in the style of the preceding chapter.

## 7.1 Motivation: the case for equal rights

Combinations of *partial* logics (those where the law of excluded middle is dropped) and DPL have been investigated by Kraemer[64]<sup>1</sup>. Such a union is motivated by consideration of semantic presupposition (Strawson [93]). The possible combination of partial logics with systems of dynamic semantics was discussed in chapters 3 and 5.

Related to this issue, it ought to be investigated how *paraconsistent* logics (those non-trivial logics which lack the law of non-contradiction and are non-explosive under addition of information) might be combined with Dynamic Semantics (DS). It was argued (in chapter 4) that DS might have much to gain from such a marriage. A Paraconsistent DS is (unsurprisingly enough) able to deal with phenomena which require a paraconsistent semantics. The phenomena in question are *self-contradictions* in discourse and resolution of *disagreements* in dialogue. In general, natural language interpretation is taken to build up a partial (and possibly paraconsistent) theory about the world. The argument for a paraconsistent basis for semantics turns upon particular notions of acceptance, rejection, assertion, denial, and the interpretation of negation. Taking the epistemic, or informational, perspective on semantics (which was argued for in chapter 1) it seems quite natural to accord the denial of a proposition the same primitive status as an assertion (see eg: Pearce and Rautenberg [78], Wansing [106]). Rather than interpreting negation classically, or intuitionistically in terms of assertion, implication, and falsum ( $\neg p = p \rightarrow \perp$ ), it is more natural, in the epistemic setting, to view  $\neg p$  as the *rejection* of  $p$ . Epistemic logic has concentrated on positive information to the exclusion of negative information, but rejection ought not to have such a secondary status. There is more to negative information than that which is excluded by the positive informa-

---

<sup>1</sup>where the possible combination of paraconsistent and dynamic logics is acknowledged, but otherwise ignored.

tion. The classical approach suffices in certain contexts (where information is consistent and exhaustive), but in the context of a dynamic semantics (dynamic in the “discourse-external” sense) of dialogue, where there are very often conflicting statements under discussion, a robust explicit account of negative information processing is called for.

Standardly, as discussed throughout the thesis, systems of dynamic semantics take inconsistencies to render information states absurd (the “implosion problem”). However, as far as dialogue modelling is concerned, there is an important distinction to be made between the *consistency* of a dialogue and its *coherence*. In a paraconsistent semantics, an inconsistent dialogue can still be coherent, or interpretable. If contradictions do not lead to state implosion, then inconsistencies are interpretable and repairable, and an inconsistent dialogue remains coherent. Recall that, as Rescher and Brandom [83]<sup>2</sup> argue,

“We must envisage the realm of the coherently discussable as broader than that of the logically self-consistent.”  
[83], page 33.

As has been demonstrated in preceding chapters, conflict, disagreement, contradiction and inconsistency all play a positive and pivotal role in evolving dialogues.

If any example of the importance of paraconsistency in non-trivial theories is required, recall that even in the history of mathematics and logic (paragons of consistency) there have been non-trivial inconsistent theories. Frege’s second order logic was found to be inconsistent, as were set theory (Russell’s paradox) and Quine’s version of it. Of course, these same examples point out that consistency is still of primary value in determining a “good” theory. However, localized inconsistencies offer the promise of intelligible and useful inconsistent theories, to which theory change systems can make improvements.

---

<sup>2</sup>There are, of course, logical systems which fulfill Wittgenstein’s prophecy, “Indeed, even, at this stage, I predict a time when there will be mathematical investigations of calculi containing contradictions, and people will actually be proud of having emancipated themselves from consistency.” [110] p. 332.

## 7.2 Rationality and Paraconsistency

The concern of this thesis is not to describe what agents actually do in their revision procedures (even though such data certainly supports the argument for paraconsistency), but to explore what an ideal rational agent could and should (logically and rationally) do during theory change, particularly change triggered by dialogue interpretation. Paraconsistency is rational, because it takes conservation of information seriously. However, there is a sense in which paraconsistent systems take conservation of information a little *too* seriously.

As a methodological point, stipulations of what are to count as rational revisions should be kept to a minimum, so as to avoid unnecessary prescriptiveness. Consistency preservation under change has always been just such a prescription. This constraint, embodied in the AGM literature, can be relaxed. Not only is it possible to escape the constraints of consistency, but it is productive to do so.

Rationality postulates play the role of the regulating body of information management, and crucially, it is logic that provides the notion of *consistency* as a minimal measure of the quality of information structures. How, then, could paraconsistency possibly qualify as a feature of rationally managed information states? It might seem at first that paraconsistency and rationality are fundamentally at odds. This conclusion is, indeed, inescapable if the rational options are identified with logical options (but I argued against such an identification in the first chapter). However, once rationality is tied to good information management, the argument for paraconsistency as a useful strategy becomes clear. Paraconsistent states and theories are *robust* under inconsistent input, or inconsistencies derived from input. Thus, in any environment where input is not of guaranteed consistency and where conservation of information is a primary consideration, paraconsistency is a decidedly rational virtue of a system. Communicative contexts, of course, constitute just such an environment.

### 7.2.1 Paraconsistency and Dynamic Semantics

Paraconsistent logics have been discussed in the field of philosophical logic for a number of years (see eg; Priest [78].) Those systems sought to tackle the classical “paradox”

of implication  $\forall q, p \wedge \neg p \models q$ ; the problem being that classical theories “explode” to absurdity under contradiction. That paraconsistency has not yet found its way into dynamic semantics can perhaps be explained by the general avoidance of the issues of correction, repair, and recovery from absurdity in a formal semantics of dialogue. An empirically adequate formal semantics must meet these challenges. While dynamic semantics based on partial logics (close cousins of paraconsistent systems<sup>3</sup>) have been explored in relation to presupposition (Krahmer [64], Muskens [76]), a fusion of paraconsistency and dynamics is unexplored territory, perhaps because the problem raised here with regard to the *implosive* nature of DS has not before been focussed upon, and partly because DS has not attempted to treat the repair and correction data which is analysed in chapter 6.

To “go paraconsistent” in dynamic semantics is fairly simple. US, DPL, and their extensions, can be based on a partial logic which allows the denotations of formulae to be overdefined (simultaneously true and false) in the semantics. The resulting systems each have a “double-barrelled” semantics (separate clauses for positive and negative evaluation of each formula). Recent work in the logic of theory change (Pearce and Rautenberg [78]) suggests that the notions of assertion and rejection should be taken as independent epistemic primitives. An extension of this stance into dynamic semantics proves to be productive.

It is a relatively simple matter to construct a Paraconsistent Update Semantics (PUS) with the following properties:

**Robustness under contradiction in PUS**

$s \llbracket \phi \wedge \neg \phi \rrbracket \neq \emptyset$ , and

for  $\phi \neq \psi$ , and  $s \not\models_{PUS} \psi$ ,  $s \llbracket \phi \wedge \neg \phi \rrbracket \not\models_{PUS} \psi$

I develop this system shortly.

### 7.2.2 Paraconsistency and Theory Change

At first blush it might seem that the motivations behind paraconsistent logics and theory change systems are severely at odds. For the whole point of theory revision is to main-

---

<sup>3</sup>Paraconsistent logics are partial logics without a constraint ruling out overdefinedness on propositional interpretations.

tain consistency under inputs which contradict the initial theory (or which lead to contradictions in the theory via inference). In contrast, paraconsistent logics simply allow any contradictions to persist, without trivializing the theory. However, an attempt to combine the two perspectives is not destined to failure, but at once points out problems and limitations in standard approaches to paraconsistency and theory change, and suggests a new approach to the modelling of interpretation in dialogue in terms of paraconsistent and partial information states, where agents can repair each other's utterances and robustly "agree to disagree".

The classical logician's disquiet over paraconsistency might be traced to the following point. Even though paraconsistent logics have succeeded in the robust accommodation of inconsistencies, inconsistencies in a theory remain undesirable. There is very little point in allowing contradictions into theories if it is then the case that "anything goes" in terms of theory change. If paraconsistent theories can be non-trivially updated with any proposition (even those which lead to inconsistency) then, as far as modelling theories goes, this is in itself a kind of triviality. While theories ought to be non-explosive and contradictions processible, constraints of some kind are required to ensure that their number is kept to a minimum. These constraints are to be provided by a suitable (paraconsistent) theory change system, which seeks to minimize the number of contradictions in a paraconsistent theory.

In general, any system which is to deal with phenomena of change has lessons to learn from paraconsistent logic, and conversely, any paraconsistent logic ought to be equipped with a change component. However, since the whole motivation for AGM theory change systems was to preserve consistency under change at all costs, it might still seem that the motivations behind paraconsistent logic and theory change systems are fundamentally at odds.

One of the lessons of chapter 4 was that if a dynamic semantics is to incorporate the insights of theory change systems then paraconsistency is a reasonable interface requirement on their combination. There are (at least) two possible arguments for this conclusion.

(1) Contradictions must be recognised as such (without inducing triviality) before they can enter a revision process. This recognition process would require that, at some point,

contradictions are represented in the logical formalism before they are removed. This much is true; if the system is to reason about inconsistencies, then they must be represented. However, if the system is to reason under inconsistent input, the requirement need not be met. One can easily imagine an algorithm which checks for the consistency of new information relative to a current theory without making the requirement that any contradictions are explicitly represented in the theory. For example, the US test  $s \llbracket \diamond p \rrbracket$  could be carried out prior to updating by  $p$ , the update occurring conditional on success of the test.

Nevertheless, a second line of argument is more compelling in the context of dialogue modelling.

(2) Contradictions must be allowed to persist, and must be processed (in the case of theories built up during NL interpretation) without “infecting” unrelated propositions. It is clear that dialogues continue non-trivially even when there have been mutually inconsistent utterances. Part of the point of dialogue is to resolve such inconsistencies, or disagreements between participants.

To suppose that there is simply some “paraconsistent antechamber” where interpretations arrive to be integrated by the consistency preserving revision procedure (the assumption behind (1)), is unwieldy and unnecessary. Rather, the revision procedure itself should incorporate paraconsistency (option (2)), minimizing the number of inconsistencies whilst preserving information under contradiction.

Thus there is a trade-off between consistency of a theory and the amount of information that it contains. The following definition states that rational agents seek to gain information which maximizes the consistency and informativeness of their current theory.

**Definition 54** (*Rational Revision with local inconsistencies*)

*Rational revision of a theory  $T$  by sentence  $p$  is modification so as to minimize inconsistencies in  $T \dot{+} p$  while maximizing the information in  $T \dot{+} p$ .*

This rather sketchy definition will be given some content by the formal work in the remainder of the chapter. The difficult point comes, of course, in deciding whether a



less consistent, more informative theory is better than a trivial but consistent one.

### 7.2.3 Autonomy and Paraconsistency

Paraconsistency is also desirable for autonomous theory change systems (see the preceding chapter), where agents judge the quality of a new theory before accepting or rejecting the revision instruction which would lead to it. Paraconsistency is used in order to make a comparison between competing theories, without just ruling out inconsistent ones. If theories do not contain inconsistencies, then any revised theory is as “good” as the theory preceding it (some even contain less information).

$T = \{a, b, c\}$  contains just as much information as  $T \dot{+} \neg a = \{\neg a, b, c\}$  (Of course, extended or consistently updated theories are always more informative than their predecessors.) An autonomous agent ought not, by conservation of information, to accept a pure downgrade, for without further evidence (ie: in the form of a revision) the agent is simply being instructed to discard previously believed information. For sceptical agents, this is not a rational option.

## 7.3 TC and DS with inconsistencies

A deductively closed (or just “closed”) theory  $T$  is a set of sentences of some language  $\mathcal{L}$  closed under the consequence relation of that language.  $T$  is *paraconsistent* if  $\exists \phi \in \mathcal{L}$  such that  $\phi \wedge \neg \phi \in T$  and  $T \neq \mathcal{L}$ .

Alternatively,  $T$  is paraconsistent when if  $T \vdash p$  then  $T + \neg p \neq \mathcal{L}$

Note that a paraconsistent closed theory (where both  $p$  and  $\neg p$  are in the theory) can still be complete.

### Extension

For paraconsistent and partial theories, the basic propositional extension clause remains the same, but closure under logical consequence does not lead to the absurd state in the presence of contradictions. Thus, paraconsistent theory extension meets the following conditions,

$T + (p \wedge \neg p) \neq \mathcal{L}$  (Non-explosive)

which in terms of the dynamic semantics is;

$s \llbracket p \wedge \neg p \rrbracket \neq \emptyset$  (Non-implosive)

The semantical condition is met by way of defining a new paraconsistent consequence relation, in the semantics developed below. Syntactically the condition is met by using a paraconsistent deducibility relation<sup>4</sup> in the base logic for theories. Crucially, this change can be made without loss of standard AGM results (see Restall and Slaney [84] for full details).

### Contraction and Revision

In a classical non-partial and consistent logic the contraction of a theory with respect to a proposition  $p$  is tantamount to expansion with  $\neg p$ . (Denial that  $p$  is assertion that not  $p$ .) Van Eijck's second system for the extension of propositional theories (chapter 5's "*FOTE<sub>III</sub>*", from [96]) works in just this way; he is able to provide a translation of that system into Veltman's Update Logic.

However, these are obviously not constraints that ought to be imposed upon theories manipulated during natural dialogue. Since paraconsistent theories simply allow inconsistencies to accumulate, they benefit from contraction methods which allow contradictions to be resolved. Removal of a sentence from a paraconsistent theory is achieved in the same way as for standard, consistent, theories. As demonstrated in [84], paraconsistent theories have the same well-defined notion of "maximal subtheories failing to support  $p$ " as classical theories do. Localized inconsistencies can be removed through theory contraction, by the removal of one of the contradictory conjuncts.

The idea of revision as a consistency preserving update need not be discarded either. Inconsistent *updates* are permitted, but revision is (still) consistency preserving. Assertions which contradict current information (unbeknownst to the speaker, presumably<sup>5</sup>) can be interpreted as (paraconsistent) updates, while constituent negation clauses are interpreted as revision instructions from a speaker who is aware of the hearer's "mis-

---

<sup>4</sup>see Priest[78]

<sup>5</sup>Otherwise they ought to utter a revision instruction, such as a constituent negation clause.

information"<sup>6</sup>.

### 7.3.1 Paraconsistent Update Semantics (PUS)

$[[\cdot]]^+$  is the assertion function, for updating a state with positive information.  $[[\cdot]]^-$  is the denial function, which updates a state with negative information. The notation  $[[\cdot]]$  is used where the polarity of update is irrelevant.

$M = \langle S, V^+, V^- \rangle$  where  $V^+(p), V^-(p)$  are (not necessarily disjoint) subsets of  $S$ .

Assertion	Denial
$s [[p]]^+ = \{i \in s \mid i \in V^+(p)\}$	$s [[p]]^- = \{i \in s \mid i \in V^-(p)\}$
$s [[\neg p]]^+ = s [[p]]^-$	$s [[\neg p]]^- = s [[p]]^+$
$s [[p \wedge q]]^+ = s [[p]]^+ \cap s [[q]]^+$	$s [[p \wedge q]]^- = s [[p]]^- \cup s [[q]]^-$
$s [[p \vee q]]^+ = s [[p]]^+ \cup s [[q]]^+$	$s [[p \vee q]]^- = s [[p]]^- \cap s [[q]]^-$
$s [[\Diamond p]]^+ = \{i \in s \mid \forall \phi, s [[p]]^+ \not\subseteq \phi \wedge \neg \phi\}$	$s [[\Diamond p]]^- = \{i \in s \mid \exists \phi \text{ such that } s [[p]]^- \not\subseteq \phi \wedge \neg \phi\}$

The system construes negation as denial, where denial has an independent status in the semantics. Moreover, non-modal paraconsistent systems never reach the absurd state. While this result may be a pleasing one in terms of philosophical logic, it produces obvious problems for the standard definitions of the epistemic modals in US and its extensions. For in dynamic semantics an update by  $\Diamond p$  is standardly a test against absurdity under update of  $p$ . There can be no such test without an accessible absurd state. The only change necessary, though, is to refine the usual US definition of  $[[\Diamond p]]$  to take account of the new localized inconsistencies which paraconsistent logic provides.

$s [[\Diamond p]]^+ = s$  iff  $\forall \phi, s [[p]]^+ \not\subseteq \phi \wedge \neg \phi$  ( $= \emptyset$  otherwise.)

Alternatively (as above):

$s [[\Diamond p]]^+ = \{i \in s \mid \forall \phi, s [[p]]^+ \not\subseteq \phi \wedge \neg \phi\}$

So, contradictions are still undesirable, but they can be accommodated.

Denial of possibility is treated similarly:

<sup>6</sup>Of course, quite whose information is the mis-information depends on your point of view.

$$s [\Diamond p]^- = \{i \in s \mid \forall \phi, s [p]^- \not\models \phi \wedge \neg \phi\}$$

The upshot is that non-modal contradictions no longer implode the information state; only modal statements (as tests) can do that. For example,

$$\text{in PUS; } s [\neg p]^+ [\Diamond p]^+ = \emptyset,$$

$$\text{which is another way of saying, } s [p]^- [\Diamond p]^+ = \emptyset$$

### 7.3.2 Paraconsistent Revision Semantics (PRS)

Paraconsistency and revisability (as explored in chapter 4) complement each other; they are solutions to the same general kind of problem. Here the two approaches are combined in a natural way.

There are 3 basic operations available in PRS (1)  $[p]^+$ , (assert p), (2)  $[p]^-$ , (deny p), and (3)  $p \llbracket$ , (retract p). Revision by p is again defined as retraction of  $\neg p$  followed by assertion of p.

Assertion clauses	Denial Clauses
$s [p]^+ = \{i \in s \mid i \in V^+(p)\}$	$s [p]^- = \{i \in s \mid i \in V^-(p)\}$
$s [\neg p]^+ = s [p]^-$	$s [\neg p]^- = s [p]^+$
$s [-p]^+ = s ] p \llbracket$	$s [-p]^- = s ] \neg p \llbracket$
$s [p \wedge q]^+ = s [p]^+ [q]^+$	$s [p \wedge q]^- = s [p]^- \cup s [q]^-$
$s [\Diamond p]^+ = \{i \in s \mid \forall \phi, s [p]^+ \not\models \phi \wedge \neg \phi\}$	$s [\Diamond p]^- = \{i \in s \mid \exists \phi \text{ such that } s [p]^- \not\models \phi \wedge \neg \phi\}$
$s [\Box p]^+ = \{i \in s \mid s ] p \llbracket = S\}$	$s [\Box p]^- = \{i \in s \mid s ] \neg p \llbracket \neq S\}$

#### Downdate clauses

$$s ] p \llbracket = \cup \beta(s \perp p)$$

$$s ] \neg p \llbracket = \cup \beta(s \perp \neg p)$$

$$s ] - p \llbracket = s [p]^+$$

$$s ] p \wedge q \llbracket = s ] p \llbracket \cup s ] q \llbracket$$

$$s ] \Diamond p \llbracket = \{i \in s \mid s ] p \llbracket \neq S\}$$

$$s ] \Box p \llbracket = \{i \in s \mid s [p]^+ \models q \wedge \neg q, \exists q\}$$

#### Revision clause

$$s \Vdash p \Vdash s \neg p \Vdash [p]^+$$

**Conjecture:** PRS is complete with respect to a Theory Change system based upon Restall and Slaney's [84]<sup>7</sup>.

### Repairing local inconsistencies

Revision Semantics allows repair of the absurd state (of course, in paraconsistent semantics there are many stable inconsistent states). This is because, in RS,  $\emptyset \Vdash p \Vdash s$  such that  $s \models p$ . DOWDATING the absurd state removes all contradictions from that state, since their retention would support  $p$  (it being the case that even one contradiction in an RS state entails all propositions). However, simply using RS to effect such a repair does not do justice to the idea of conservation of information. Information ought to be conserved even in the face of absurdity, so that if a proposition  $q$  is supported by a state which later becomes absurd via addition of some proposition  $p$ , then  $q$  ought to be supported by the state after repair. Put formally, this results in the following desideratum:

$$\emptyset \Vdash p \Vdash q, \forall q \text{ such that } p \not\models q, \text{ and } s \models q, \text{ and } s \Vdash p \Vdash \emptyset$$

RS fails to meet this requirement since if  $s \Vdash p \Vdash \emptyset$  then  $\emptyset \Vdash p \Vdash s'$  and it is not excluded that  $s \subseteq s'$ . (ie: it may be the case that  $s' \models p$ )

PRS, with its localized inconsistencies, can repair absurdities in a way which preserves information present prior to the absurdity.

For example, in PRS take  $s \Vdash p \wedge q \wedge \neg p \Vdash s' \neq \emptyset$ . Then  $s' \Vdash p \Vdash \neg p \wedge q$

All in all then, in order to be able to rationally recover dialogues from absurdity, while preserving prior information, PRS is required.

One could produce a version of RS that carries a consistency check with  $\Vdash \diamond p \Vdash$  before

<sup>7</sup>And see also Pearce and Rautenberg [78].

allowing  $p$  to be asserted. Alternatively all updates could be treated as revisions, so that consistency is always preserved. However, as illustrated in chapter 6, agents interpreting natural language do not always, or even often, seem to do this.

### Paraconsistent Information States

A semantics for a paraconsistent logic must at least allow that some propositions can be considered as both true and false simultaneously. Semantically (first-order) paraconsistent models  $M = \langle D, d^+, d^- \rangle$  are classical models furnished with a denotation and anti-denotation for each  $n$  place predicate  $P$ ,  $d^+P$  and  $d^-P \in D^n$ . Denotation and anti-denotation may overlap, producing overdefinedness for some individuals. In order to avoid partiality, impose the following constraint on the models:

**(No underdefinedness)** for all  $P^n : d^+P \cup d^-P = D^n$ .

Partial theories are modelled by sets of total models considered disjunctively. As discussed in chapter 3, further complicating the semantics by way of partial models is largely superfluous. One could give a paraconsistent semantics by way of considering sets of total models conjunctively (ie: where one model makes  $p$  true and the other false then  $p$  is both true and false). However, partiality would then have to be modelled in some other way.

### 7.3.3 Paraconsistent Predicate Logic (PPL)

The semantics of PPL is defined in terms of assignments in a model.  $S$  is a set of total assignments such that  $i : VAR \rightarrow D$  for every  $i \in S$ .  $S$  is the set of all assignments.

Let  $[[\phi]]^+ = \{i \in S \mid i \models \phi\}$  and

$[[\phi]]^- = \{i \in S \mid i \not\models \phi\}$

so that  $[[\phi]]^+$  and  $[[\phi]]^-$  are the sets of assignments which respectively support and reject  $\phi$  (with respect to a suppressed model  $M$ .)

#### Definition 55 PPL<sup>8</sup>

---

<sup>8</sup>Based on what is called the “relevantist” approach to paraconsistency in [79]

1.  $\llbracket P(t_1 \dots t_n) \rrbracket^+ = \{i \mid \langle i(t_1) \dots i(t_n) \rangle \in d^+ P\}$   
 $\llbracket P(t_1 \dots t_n) \rrbracket^- = \{i \mid \langle i(t_1) \dots i(t_n) \rangle \in d^- P\}$
2.  $\llbracket \neg \phi \rrbracket^+ = \llbracket \phi \rrbracket^-$   
 $\llbracket \neg \phi \rrbracket^- = \llbracket \phi \rrbracket^+$
3.  $\llbracket \phi \wedge \psi \rrbracket^+ = \{i \mid i \in \llbracket \phi \rrbracket^+ \text{ and } i \in \llbracket \psi \rrbracket^+\}$   
 $\llbracket \phi \wedge \psi \rrbracket^- = \{i \mid i \in \llbracket \phi \rrbracket^- \text{ or } i \in \llbracket \psi \rrbracket^-\}$
4.  $\llbracket \phi \vee \psi \rrbracket^+ = \{i \mid i \in \llbracket \phi \rrbracket^+ \text{ or } i \in \llbracket \psi \rrbracket^+\}$   
 $\llbracket \phi \vee \psi \rrbracket^- = \{i \mid i \in \llbracket \phi \rrbracket^- \text{ and } i \in \llbracket \psi \rrbracket^-\}$
5.  $\llbracket \exists x \phi \rrbracket^+ = \{i \mid \exists j \text{ and } i =_x j \text{ and } j \in \llbracket \phi \rrbracket^+\}$   
 $\llbracket \exists x \phi \rrbracket^- = \{i \mid \forall j, i =_x j \Rightarrow j \in \llbracket \phi \rrbracket^-\}$
6.  $\llbracket \forall x \phi \rrbracket^+ = \{i \mid \forall j, i =_x j \Rightarrow j \in \llbracket \phi \rrbracket^+\}$   
 $\llbracket \forall x \phi \rrbracket^- = \{i \mid \exists j \text{ and } i =_x j \text{ and } j \in \llbracket \phi \rrbracket^-\}$

Here  $i =_x j$  means that assignment  $j$  is exactly like  $i$ , but (possibly) for the value it assigns to  $x$ .

This allows a definition of the **evaluation function**  $Val : FORM \times D^V \rightarrow \pi$ .

### Definition 56 Assertibility, Rejectibility, Overdefinedness

- $\phi$  is **assertible** in  $M$  under assignment  $i$ , or  $Val(\phi, i) = \{1\}$ , iff  $i \in \llbracket \phi \rrbracket^+$
- $\phi$  is **rejectible** in  $M$  under assignment  $i$ , or  $Val(\phi, i) = \{0\}$ , iff  $i \in \llbracket \phi \rrbracket^-$
- $\phi$  is **overdefined** in  $M$  under assignment  $i$ , or  $Val(\phi, i) = \{0, 1\}$ , iff  $i \in \llbracket \phi \rrbracket^-$  and  $i \in \llbracket \phi \rrbracket^+$  alternatively;  $\phi$  is overdefined by  $i$  in  $M$  iff  $\exists i$  such that  $i \in \llbracket \phi \rrbracket^+ \cap \llbracket \phi \rrbracket^-$

Note that for all PPL models which obey the no-underdefinedness constraint  $\llbracket \phi \rrbracket^+ \cup \llbracket \phi \rrbracket^- = S$

**Entailment:**  $\Sigma \models \alpha$  iff it is the case in all models  $M$ , under all assignments  $i$ , that if for all  $\beta \in \Sigma, 1 \in Val(\beta, i)$ , then  $1 \in Val(\alpha, i)$ .

This is equivalent to: if  $\forall \beta \in \Sigma, i \in \llbracket \beta \rrbracket^+$  then  $i \in \llbracket \alpha \rrbracket^+$

**s supports**  $\alpha$  if  $\forall i \in s, i \in \llbracket \alpha \rrbracket^+$

For the purpose of modelling robust information states, note that;

**Fact 13**  $\alpha \wedge \neg\alpha \not\models \beta$

$\alpha, \neg\alpha \vee \beta \not\models \beta$ .

### Paraconsistent DPL, MDPL, RMDPL, and FRL

Given a paraconsistent predicate logic, as shown above, it is a simple matter to base a dynamic predicate logic upon it, as well as the other first-order systems of chapter 5. Indeed, once the base logic is fixed, its paraconsistency percolates up to all the systems which stem from it. A range of systems akin to those of chapter 5, with completeness results, could fairly easily be generated using a base logic which localizes inconsistencies. In the remainder of the chapter I assume such a series of systems to have been developed. These systems would differ from the systems of chapter 5 only in that they have “double-barrelled” semantics for assertion and denial.

## 7.4 Information Quality Measures

Information Quality (IQ) measures are defined so as to provide a way of adjudicating between competing theories. They are to allow agents to exhibit rational autonomy in theory change, or information revision, in the sense that agents choose whether or not to accept new information, on the basis of the “information quality” which its acceptance would result in. AGM research simply models gullible agents; those assuming new information to be true. Here, sceptical or autonomous agents judge the quality of the incoming information; the merits it would bring to their current theory.

The basic idea is that more information is better information, but that inconsistencies are undesirable (although not completely ruled out.) Thus, as long as an update or revision is consistent with a rational agent’s current theory, they will accept it. If a downgrade would remove an inconsistency, but no other information, the rational agent will carry out the downgrade. The interesting cases arise in weighing up whether to accept an informative new statement which induces inconsistency, and whether to accept down-



dates which remove contradictions at the expense of other information.

### Coherence and Informativeness

I now define two simple informational (as opposed to logical) measures, which can be used to guide theory change processes.

#### Definition 57 (PRS coherence)

Where  $\#(T)$  is the number of propositional variables overdefined in  $T$ .

$$\text{coh}(T) = \frac{1}{1 + \#(T)}$$

#### Definition 58 (PRS Informativeness)

$$I(T) = |\{p \mid p \in T \text{ or } \neg p \in T\}|$$

The informativeness of a theory is simply the number of propositional variables which it contains information about.

#### Definition 59 (PRS Information Quality)

$$IQ(T) = n\text{coh}(T) \times mI(T)$$

Where  $n$  and  $m$  are parameters controlling the relative weightings of coherence and informativity. A gullible agent, (perhaps a database for example), will value informativity over coherence, and vice versa for a skeptical agent (say, a pure mathematician)<sup>9</sup>. Then define theory change to be conditional on the IQ measure of the proposed theory.

---

<sup>9</sup>One could easily imagine agents changing these parameters with respect to the status of their dialogue partner.

**Definition 60** (*Autonomous Rational Theory Change*)

$$T + p = Cl(T \cup \{p\}) \text{ iff } IQ(T + p) \geq IQ(T)$$

$$T \dot{-} p = \bigcap \alpha(T \perp p) \text{ iff } IQ(T \dot{-} p) \geq IQ(T)$$

$$T \dot{+} p = (T \dot{-} \neg p) + p \text{ iff } IQ(T \dot{+} p) \geq IQ(T)$$

**Example 58** Take  $T = \{p, q, r\}$

Then should a rational agent accept the assertion  $+(q \rightarrow \neg p \wedge \neg r \wedge t \wedge u)$  ?

$$\text{Put } n = m = 1, \text{ then } coh(T) = 1/1 + 0 = 1$$

$$coh(T') = coh(\{p, \neg p, q, r, \neg r, t, u\}) = 1/(1 + 2) = 1/3$$

$$I(T) = 3, I(T') = 5$$

$$IQ(T) = 3, IQ(T') = 5/3$$

A rational agent, who values coherence and informativity equally ( $n = m = 1$ ), would not accept the proposed assertion, since  $IQ(T') \leq IQ(T)$ .

Neither would they accept “-p” (the instruction to forget p), without any compensating information.

### 7.4.1 Axioms for resolving disagreement

Dropping the gullibility assumption of the preceding chapter, and installing autonomy, agents can disagree and correct each other, depending on how their “parameters” are set for coherence vs. informativity.

- A1  $U_a^+(\phi) \Rightarrow s_a \models \phi$  (Sincerity)
- A2  $U_a^+(\phi) \Rightarrow s'_b = s_b \llbracket B_a(\phi) \rrbracket$  (Integrity Assumption<sup>10</sup>)
- A3  $U_a^+(\phi)$  and  $s_b \models \neg\phi \Rightarrow U_b(\uparrow \neg\phi)$  iff  $IQ(s_b \llbracket \phi \rrbracket) \leq IQ(s_b)$  (Autonomous Correction/Repair)
- A4  $U_a^+(\uparrow \phi)$  and  $IQ(s_b \llbracket \phi \rrbracket) \geq IQ(s_b) \Rightarrow s'_b = s_b \llbracket \phi \rrbracket$  (Rational Autonomous Acceptance)
- A5  $U_a^+(\uparrow \phi)$  and  $IQ(s_b \llbracket \phi \rrbracket) \leq IQ(s_b) \Rightarrow U_b(-\phi)$  (Rational Autonomous Downdate)

<sup>10</sup>Hearer models speaker’s beliefs.

- A6  $U_a^+(-\phi)$  and  $IQ(s_b) \llbracket \phi \rrbracket \geq IQ(s_b) \Rightarrow s'_b = s_b \llbracket \phi \rrbracket$  (Rational Downtdate Acceptance)

Agents disagree only if incorporating the new information leads to a state of less information quality.

One could even imagine a system where A2 is altered so that a hearer can detect “lies”. That is, if A2 were changed to

$$A2' U_a^+(\phi) \Rightarrow s'_b = s_b \llbracket B_a(\phi) \rrbracket \text{ iff } IQ(s'_b) \geq IQ(s_b)$$

then the hearer could even refuse to accept that the speaker believes what they have just uttered.

#### 7.4.2 Information Quality in DS states

The IQ measures presented above work in the syntax of propositional paraconsistent theories. Here I show how the definition can be extended to cover the information states of Dynamic Semantics: first-order IQ measures over sets of models and assignments.

The coherence of a predicate  $P$  can be measured inversely to the number of variables that are assigned to fall in  $d^+(P) \cap d^-(P)$ . This set of the individuals overdefined for  $P$  (the glut of  $P$ ) is written  $\#(P) \subseteq D$ . Then  $\#(P)_i \subseteq V$  is the set of variables and names overdefined for a predicate  $P$  relative to an assignment  $i$  (and a model  $m$ );

**Definition 61** (*Glut of  $P$* )

$$\#(P)_i = \{x \in Vars \cup Cons \mid i(x) \in \#(P)\}$$

**Definition 62** (*Coherence of a Predicate*)

*If information about a predicate  $P$  has no overdefined variables or names, then that predicate is coherent. The more overdefined terms a predicate amasses, the less coherent it becomes.  $1 \leq coh(p) < 0$*

$$coh(P)_i = \frac{1}{1 + |\#(P)_i|}$$

The coherence of a model is just the sum of the coherence measures of the predicates which its interpretation function ranges over, in proportion to the number of those predicates. Coherence is intended to give a measure of the quality of information in a paraconsistent model (classical models always have  $coh(M) = 1$ .)

**Definition 63** (*Coherence of a Model*)

$$coh(M) = \frac{\sum_{P \in PRED} \frac{1}{1 + |\#(P)_i|}}{|PRED|}$$

We say a predicate is **coherent** when  $|(d^+(P) \cap d^-(P))| = 0$ , and then write  $coh(P) = 1$ . A model is coherent ( $coh(M) = 1$ ) iff all the predicates that it models are.

### Informativeness

The informativeness  $I(P)_i$  of a predicate  $P$  (in model  $I$ ) is the number of variables assigned to  $d^+(P) \cup d^-(P)$ , the number of terms that the predicate applies to (either positively or negatively).

**Definition 64** (*Informativeness of a Predicate*)

$$I(P)_i = |x \in V \text{ such that } i(x) \in d^+(P)| + |x \in V \text{ such that } i(x) \in d^-(P)|$$

The informativeness of a model is just the sum of the informativeness of its predicates. Thus the more predicates a model interprets, and the more terms are classified by a predicate, the more informative a model is.

**Definition 65** (*Informativeness of a Model*)

$$I(M) = \sum_{P \in PRED} I(P)_i$$

Note that informativeness can increase while coherence decreases (the case where predicates become increasingly overdefined). We aim to give a combined measure which will deal with the informational considerations in a given belief revision. Such a measure could be thought of as a first-order analogue of the Epistemic Entrenchment (EE) hierarchy for propositional theory change. The approach here has a number of advantages over the Gärdenfors EE ordering. Rather than just being an ad hoc hierarchy of beliefs, this finer grained set of measures provides a more principled and explanatory framework for handling the informational considerations that arise in adjudicating between alternative logically possible revisions<sup>11</sup>. The point is that paraconsistent first-order theories might be fine-grained enough to be able to *generate* a suitable hierarchy of subtheories for contractions, using IQ measures to grade competing subtheories. The information quality of a model  $M$ , written  $IQ(M)$  is simply the weighted product of its coherence and informativity measures. Where  $n$  and  $m$  are parameters reflecting gullibility, scepticism, and (in combination) rigor of the agent.

**Definition 66** (*Information Quality*)

For real numbers  $n$  and  $m$ ;

$$IQ(M) = mcoh(M) \times nI(M)$$

## 7.5 Summary

This chapter has discussed prospects for an integration of paraconsistent semantics with systems of theory change, arguing that the two approaches complement each other, particularly in the application of modelling communication. Perhaps surprisingly, it turns out that neither theory change systems nor systems of dynamic semantics suffer too much under the adoption of a paraconsistent base logic. Dynamic systems gain “double - bared” interpretation functions, and TC systems adopt a notion of “change on the basis of increased Information Quality”. The benefits of these changes are varied.

---

<sup>11</sup>Again, see the multiple extensions problem.

Systems of Paraconsistent Revisable dynamic semantics can model robust dialogue interpretation for (a variety of) rational autonomous agents. This allows an analysis of dialogues in which contradictions may persist, but can also be resolved.

These considerations bring the thesis to a close. Various suggestions for further research have been made in the development of these chapters:

- An investigation of the use of *structured theories* and states for iterated revision.
- Further exploration of a range of epistemic modalities employing model construction.
- A full account of the counterfactual conditional of Revision Semantics.
- A further investigation of the range of revisable dynamic systems suggested by chapter 5, and the properties of their “dialogue referents”.

# Bibliography

- [1] Carlos Alchourron, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial Meet Contraction Functions and their Associated Revision Functions . *Journal of Symbolic Logic*, 50:510–30, 1985.
- [2] A. Anderson, M. Bader, E. Bard, E. Boyle, G. Docherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. Thompson, and R. Weinert. The HCRC Map Task Corpus. *Language and Speech*, 34(4):351–66, 1992.
- [3] Alan R. Anderson and Nuel D. Belnap. *Entailment , the logic of relevance and necessity*, volume 1. Princeton University Press, 1975.
- [4] Doug Appelt and Kurt Konolige. A practical nonmonotonic theory for reasoning about speech acts. Technical report, SRI International, California, 1988.
- [5] Nicholas Asher. Belief in discourse representation theory. *Journal of Philosophical Logic*, 15(2):127 – 189, 1986.
- [6] Nicholas Asher. Verbal Information, Interpretation, & Attitudes . In P. Hanson, editor, *Information, Language, and Cognition*, Vancouver Studies in Cognitive Science, vol 1. University of British Columbia Press, Vancouver, 1990.
- [7] Nicholas Asher. Discourse representation theory and belief dynamics. In *The Logic of Theory Change*, volume 465 of *Lecture Notes in Artificial Intelligence*, pages 282–321. Springer Verlag, Berlin, 1991.
- [8] Nicholas Asher. *Reference to Abstract Objects in Discourse*. Studies in Linguistics and Philosophy. Kluwer Academic Publishers, 1993.
- [9] Normiah Aziz, Helen Pain, and Paul Brna. Modelling and mending student’s misconceptions in translating algebra word problems using a belief revision system in taps. In Greer, editor, *Artificial Intelligence in Education 1995 — Proceedings of AI-ED 95, World Conference on AI in Education*, pages 107 – 114. AACE, Virginia, 1995.
- [10] Jon Barwise. Constraints, channels, and the flow of information. In *Situation Theory and its applications*, volume 3, pages 3 –27. CSLI, Stanford, 1993.
- [11] Jon Barwise and John Etchemendy. Model-Theoretic Semantics. In M. I. Posner, editor, *Foundations of Cognitive Science*, pages 207–43. MIT Press, Cambridge, MA, 1989.

- [12] Jon Barwise and John Etchemendy. Information, Infons, and Inference. In Robin Cooper, Kuniaki Mukai, and John Perry, editors, *Situation Theory and Its Applications, I*, CSLI Lecture Notes Number 22, pages 33–78. CSLI Publications, Stanford, 1990.
- [13] Nuel D. Belnap. A Useful Four-valued Logic . In G. Epstein and J. M. Dunn, editors, *Modern Uses of Many-valued Logic*, pages 8–37. Reidel, 1977.
- [14] E. R. Blackmer and J. L. Mitton. Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, 39:173–194, 1991.
- [15] Jean Carletta, Richard Cayley, and Steven Isard. A collection of self-repairs from the map task corpus. Technical report, Human Communication Research Centre, Edinburgh, 1994.
- [16] Lauri Carlson. *Dialogue Games: An Approach to Discourse Analysis*. Reidel, 1984.
- [17] Brian F. Chellas. *Modal Logic: An Introduction*. Cambridge University Press, Cambridge, 1980.
- [18] Herbert Clark and C. Marshall. Definite reference and mutual knowledge. In *Elements of Discourse Understanding*, pages 10–63. Cambridge University Press, 1981.
- [19] Paul Cohen and Hector Levesque. Rational Interaction as the Basis for Communication. Technical Report CSLI-87-89, CSLI, Stanford, 1987.
- [20] Charles Cross and Richamond Thomason. Conditionals and Knowledge Base Update . In P. Gärdenfors, editor, *Belief Revision*, Cambridge Tracts in Theoretical Computer Science, 29. Cambridge University, Cambridge, 1991.
- [21] Paul Dekker. *Transsentential Meditations*. PhD thesis, ILLC/ Department of Philosophy, Amsterdam, 1993.
- [22] Paul Dekker. Predicate logic with anaphora. In *Integrating Semantic Theories II*, DYANA-2 Deliverable R2.1.B. ILLC/Department of Philosophy, University of Amsterdam, 1994.
- [23] Paul Dekker and Martin Stokhof, editors. *Proceedings of the 9th Amsterdam Colloquium*. ILLC, Amsterdam, 1993.
- [24] Daniel Dennett. *The Intentional Stance*. MIT Press, 1987.
- [25] Rene Descartes. *Meditations on first Philosophy*. Penguin, 1642.
- [26] John Doyle. A truth maintenance system. *Artificial Intelligence*, 12(2):231 – 272, 1979.
- [27] Fred Dretske. *Knowledge and the Flow of Information*. MIT Press, 1981.
- [28] R. Elio and J. Pelletier. The effect of syntactic form on simple belief revisions and updates. In *16th annual conference of the Cognitive Science Society*, pages 260–5, Atlanta, Georgia, 1994.



- [29] Fagin, Ullman, and Vardi. On the semantics of updates in databases. In *Proceedings of Second ACM SIGACT-SIGMOD*, pages 352–365, Atlanta, 1983.
- [30] Hartry Field. Theory change and the indeterminacy of reference. *Journal of Philosophy*, 70, 1973.
- [31] Kit Fine. A defense of arbitrary objects. In Fred Landman and Frank Veltman, editors, *Varieties of Formal Semantics*, pages 123–42. Kluwer, Dordrecht, 1984.
- [32] Kit Fine. *Reasoning with Arbitrary Objects*. Basil Blackwell, London, 1985.
- [33] Gottlob Frege. On sense and reference. In Peter Geach and Max Black, editors, *Translations from the Philosophical Writings of Gottlob Frege, 3rd Edition*. Blackwell, 1892.
- [34] Andre Fuhrmann. Belief revision. Technical report, Saarbrücken Summer School, 1991.
- [35] Andre Fuhrmann. *An Essay on Contraction*. Studies in Logic Language and Information. CSLI, Stanford, 1995. (forthcoming).
- [36] Andre Fuhrmann and Michael Morreau, editors. *The Logic of Theory Change*, volume 465 of *Lecture notes in Artificial Intelligence*. Springer-Verlag, Berlin, London, 1991. (Workshop, Konstanz, FRG, October 13-15, 1989).
- [37] Dov Gabbay, editor. *Handbook of Philosophical Logic volume II: Extensions of Classical Logic*. Reidel, Dordrecht, Netherlands, 1984.
- [38] Julia Rose Galliers. Belief Revision and a theory of communication. Technical Report 193, University of Cambridge Computer Laboratory., Cambridge, 1990.
- [39] Julia Rose Galliers. Autonomous belief revision and communication. In P. Gärdenfors, editor, *Belief Revision*, Cambridge Tracts in Theoretical Computer Science 29. Cambridge University Press, 1992.
- [40] Peter Gärdenfors. Dynamics of belief as a basis for logic. *British Journal of the Philosophy of Science*, 35:1–10, 1984.
- [41] Peter Gärdenfors. Propositional logic based on the dynamics of belief. *Journal of Symbolic Logic*, 50(35):390–394, 1985.
- [42] Peter Gärdenfors. *Knowledge in Flux*. MIT Press, Cambridge MA, 1986.
- [43] Peter Gärdenfors. Conditionals and belief revision. *Journal of Applied Non-Classical Logics*, 1(2):237, 1991.
- [44] Jonathan Ginzburg. Interrogatives: Questions, facts, and dialogue. Technical report, Centre for Cognitive Science, Edinburgh, 1994.
- [45] Paul Grice. Logic and conversation. In Peter Cole and Jerry L. Morgan, editors, *Syntax and Semantics: Speech Acts*. Academic Press, 1975.

- [46] Jeroen Groenendijk and Martin Stokhof. Dynamic Predicate Logic . *Linguistics and Philosophy*, 14:39–100, 1991.
- [47] Jeroen Groenendijk, Martin Stokhof, and Frank Veltman. This Might be it. In Robin Cooper and Jeroen Groenendijk, editors, *Integrating Semantics Theories II* , Dyana -2 deliverable 2.1.B. University of Amsterdam, Amsterdam, 1994.
- [48] Adam Grove. Two Modellings for Theory Change. *Journal of Philosophical Logic*, 17:157–170, 1988.
- [49] Philip P. Hanson, editor. *Information, Language, and Cognition*, volume 1 of *Vancouver Studies in Cognitive Science*. University of British Columbia Press, Vancouver, 1990.
- [50] Sven Ove Hansson and Wlodek Rabinowicz, editors. *Logic for a Change* . Uppsala Prints and Preprints in Philosophy 1995:9. Department of Philosophy, Uppsala, 1995.
- [51] Gilbert Harman. *Change in View - Principles in Reasoning*. Bradford books, MIT Press, Cambridge, Mass, 1986.
- [52] G. W. F. Hegel. *Preface to the Phenomenology of Mind*. Clarendon Press, Oxford, 1807. (trans: A.V. Miller).
- [53] Irene Heim. *The Semantics of Definite and Indefinite Noun Phrases in English*. PhD thesis, University of Massachusetts, Amherst, 1982.
- [54] Carl Hempel. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. Free Press, New York, 1965.
- [55] Heinrich Hertz. *The Principles of Mechanics presented in a new form*. Macmillan, London, 1899. quoted in "Wittgenstein's Vienna" / Allan Janik and Stephen Toulmin London : Weidenfeld and Nicolson, 1973.
- [56] A. E. Hieke. A content-processing view of hesitation phenomena. *Language and Speech*, 24(2):147–160, 1981.
- [57] G. Houghton and S. Isard. Why to speak, what to say and how to say it: Modelling language production in discourse. In P. Morris, editor, *Modelling Cognition*. Wiley United Kingdom, 1987.
- [58] William James. *A Pragmatism — a new name for some old ways of thinking*. Longmans, Green and Company, New York, 1907.
- [59] Allan Janik and Stephen Toulmin. *Wittgenstein's Vienna*. Weidenfeld and Nicolson, London, 1973.
- [60] Jan Jaspars. *Calculi for Constructive Communication*. PhD thesis, ILLC / ITK, Amsterdam, 1994.
- [61] Hans Kamp. A Theory of Truth and Semantic Representation. In Geroen Groenendijk, Theo Janssen, and Martin Stokhof, editors, *Formal Methods in the Study of Language*. Mathematisch Centrum, Amsterdam, 1981.

- [62] Hans Kamp. Context, Thought, and Communication . *Proceedings of the Aristotelian Society*, 85:239–61, 1985.
- [63] Lauri Karttunen. Discourse referents. In J. Mc Cawley, editor, *Syntax and Semantics 7*. Seminar Press, 1976.
- [64] Emiel Krahmer. Partiality and dynamics. In *Proceedings of the 9th Amsterdam Colloquium*, pages 391 – 411, Amsterdam, 1993. ILLC.
- [65] Saul A. Kripke. Semantical Considerations on Modal Logic. *Acta Philosophica Fennica*, 16:83–94, 1963.
- [66] Saul A. Kripke. Semantical analysis of intuitionistic logic i. In *Formal Systems and Recursive Functions*. North Holland, 1965.
- [67] Saul A. Kripke. *Naming and Necessity*. Harvard University Press, Cambridge, MA, 1972.
- [68] Saul A. Kripke. A puzzle about belief. In Avishai Margalit, editor, *Meaning and Use*, number 3 in Synthese Language Library, pages 239–283. D. Reidel, London, 1979.
- [69] Fred Landman. *Towards a Theory of Information: the Status of Partial Objects in Semantics*. GRASS 6. Foris, 1986.
- [70] Michael C. Lemon. *The Discipline of History and the History of Thought*. Routledge, London/New York, 1995.
- [71] Oliver Lemon. First Order Theory Change systems and their Dynamic Semantics. In *Tbilisi Symposium on Language, Logic, and Computation*, CSLI series. Centre for the Study of Language and Information, Stanford, 1995. (forthcoming).
- [72] Oliver Lemon. First-Order Theory Change Systems with Intuitionistic and Dynamic Semantics. In *Logica '95: proceedings of the 9th international symposium*, Prague, 1995. Philosophia Academic Publishing.
- [73] Willem J. M. Levelt. Monitoring and self-repair in speech. *Cognition*, 14:41–104, 1983.
- [74] David Lewis. *Counterfactuals*. Basil Blackwell, Oxford, 1973.
- [75] Peter Milne. Verification, falsification, and the logic of enquiry. *Erkenntnis*, 34:23–54, 1991.
- [76] Reinhard Muskens. *Going partial in Montague Grammar*. PhD thesis, ILLC, Amsterdam, 1988.
- [77] David Nelson. Constructible falsity. *Journal of Symbolic Logic*, 14:16 –26, 1949.
- [78] David Pearce and Wolfgang Rautenberg. Propositional Logic Based on the Dynamics of Disbelief . In *The Logic of Theory Change* , volume 465 of *Lecture Notes in Artificial Intelligence*. Springer-Verlag, Berlin, 1991.

- [79] Graham Priest. *Paraconsistent Logic: essays on the inconsistent*. Philosophia Verlag, Munchen, 1989.
- [80] Willard Quine. *The Ways of Paradox and Other Essays*. Harvard University Press, Cambridge, Mass., 1966.
- [81] Willard Quine. *Quiddities: an intermittently philosophical dictionary*. Penguin, London, 1987.
- [82] Nicholas Rescher. *Hypothetical Reasoning*. North-Holland, Amsterdam, 1964.
- [83] Nicholas Rescher and Robert Brandom. *The Logic of Inconsistency: a study in non-standard possible-world semantics and ontology*. Blackwell, Oxford, 1980.
- [84] Greg Restall and John Slaney. Realistic Belief Revision. Technical Report TR-ARP-2-95, Automated Reasoning Project, Australian National University, Canberra, 1995.
- [85] Elanor Rosch. Family resemblance: studies in the internal structure of categories. *Cognitive Psychology*, 7:573–605, 1975.
- [86] Hans Rott. Preferential belief change using generalized epistemic entrenchment. *Journal of Logic, Language, and Information*, 1(1):45–78, 1992.
- [87] Bertrand Russell. On denoting. *Mind*, 14:479–493, 1905.
- [88] Mark D. Ryan. *Ordered Presentations of Theories*. PhD thesis, Department of Computing, Imperial College, London, 1992.
- [89] Andreas Schöter. *The Computational Application of Bilattice Logic to Natural Reasoning*. PhD thesis, Centre for Cognitive Science, Edinburgh, 1995.
- [90] Krister Segerberg. Some Questions about Hypertheories. In Sven Ove Hansson and Wlodek Rabinowicz, editors, *Logic for a Change*, Uppsala Prints and Preprints in Philosophy 1995:9. Department of Philosophy, Uppsala, 1995.
- [91] Claude Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, 1949.
- [92] Robert Stalnaker. Pragmatics. In *Semantics of Natural Language*, Synthese Library, pages 380–397. Reidel, 1972.
- [93] Peter Strawson. On referring. *Mind*, 59:320–44, 1950.
- [94] Johan van Benthem. Semantic parallels in natural language and computation. In *Logic Colloquium, Granada*. North-Holland, Amsterdam, 1987.
- [95] Jaap van der Does. The dynamics of sophisticated laziness. In Jeroen Groenendijk, editor, *Plurals and Anaphora*, DYANA-2 Deliverable R2.2.A, Part I. ILLC/Department of Philosophy, University of Amsterdam, 1993.

- [96] Jan van Eijck. The Dynamics of Theory Extension. In Paul Dekker and Martin Stokhof, editors, *9th Amsterdam colloquium proceedings*, pages 249–67, Amsterdam, 1993. ILLC.
- [97] Jan van Eijck and Giovanna Cepparello. Dynamic Modal Predicate Logic. In Kanazawa and Pinon, editors, *Dynamics, Polarity and Quantification*, pages 251–276. CSLI, Stanford, 1994.
- [98] Jan van Eijck and Giovanna Cepparello. Dynamic Modal Predicate Logic. Technical report, Research Institute for Language and Speech (OTS), Utrecht University, 1994.
- [99] Noor van Leusen. The interpretation of corrections. Technical report, ILL, Heidelberg, 1994.
- [100] Frank Veltman. *Logics for Conditionals*. PhD thesis, Faculteit der Wiskunde en Natuurwetenschappen, University of Amsterdam, Amsterdam, 1985.
- [101] Frank Veltman. Defaults in Update Semantics . In Hans Kamp, editor, *Conditionals, Defaults and Belief Revision*, DYANA Report R2.5A. Centre for Cognitive Science, Edinburgh, 1990. (To appear in the Journal of Philosophical Logic).
- [102] Frank Veltman. Defaults in Update Semantics. *Journal of Philosophical Logic*, 25(3):221–261, 1996.
- [103] Frank Veltman, Ewan Klein, and Marc Moens. Default reasoning and dynamic interpretation of natural language. In *Esprit 90: Proceedings of the 7th Annual ESPRIT Conference*, pages 52–65, 1990.
- [104] Cees F. M. Vermeulen. Sequence semantics for dynamic predicate logic. Technical report, Department of Philosophy, University of Utrecht, 1990.
- [105] Gerd Wagner. *Vivid Logic* . Lecture Notes in Artificial Intelligence v. 764. Springer-Verlag, Berlin : London, 1994.
- [106] Heinrich Wansing. *The Logic of Information Structures* . Lecture Notes in Artificial Intelligence v. 681. Springer-Verlag, Berlin : London, 1993.
- [107] David Wiggins. *Sameness and Substance*. Blackwell, Oxford, 1980.
- [108] Ludwig Wittgenstein. *Tractatus Logico Philosophicus*. Routledge and Kegan Paul, London, 1921.
- [109] Ludwig Wittgenstein. *Philosophical Investigations*. Blackwell, 1953. (trans G. E. M. Anscombe).
- [110] Ludwig Wittgenstein. *Remarks on the foundations of Mathematics*. Basil Blackwell, Oxford, 1956. ( trans G.E.M. Anscombe).
- [111] Henk Zeevat. A treatment of belief sentences in discourse representation theory. In *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*, Groningen-Amsterdam Studies in Semantics (GRASS), pages 189 – 215. Foris, 1987.