Handbook of Philosophical Logic 16

Dov M. Gabbay
Franz Guenthner *Editors*

Handbook of Philosophical Logic

Second Edition

Springer

HANDBOOK OF PHILOSOPHICAL LOGIC
2ND EDITION

VOLUME 16

HANDBOOK OF PHILOSOPHICAL LOGIC

2nd Edition

Volume 16

edited by D.M. Gabbay and F. Guenthner

Volume 1 – ISBN 0-7923-7018-X
Volume 2 – ISBN 0-7923-7126-7
Volume 3 – ISBN 0-7923-7160-7
Volume 4 – ISBN 1-4020-0139-8
Volume 5 – ISBN 1-4020-0235-1
Volume 6 – ISBN 1-4020-0583-0
Volume 7 – ISBN 1-4020-0599-7
Volume 8 – ISBN 1-4020-0665-9
Volume 9 – ISBN 1-4020-0699-3
Volume 10 – ISBN 1-4020-1644-1
Volume 11 – ISBN 1-4020-1966-1
Volume 12 – ISBN 1-4020-3091-6
Volume 13 – ISBN 978-1-4020-3520-3
Volume 14 – ISBN 978-1-4020-6323-7
Volume 15 – ISBN 978-94-007-0484-8

HANDBOOK OF PHILOSOPHICAL LOGIC

2nd EDITION

VOLUME 16

Edited by

D.M. GABBAY
King's College, London, U.K.

and

F. GUENTHNER
Centrum für Informations- und Sprachverarbeitung,
Ludwig-Maximilians-Universität München, Germany

Springer

Editors
Prof. Dr. Dov M. Gabbay
King's College London
Dept. Computer Science
London
United Kingdom
dov.gabbay@kcl.ac.uk

Franz Guenthner
University of Munich
Centrüm für Informations- und Sprachverarbeitung
Oettingenstr. 67
80538 Munich
Germany
gue@cis.uni-muenchen.de

ISBN 978-94-007-0478-7 e-ISBN 978-94-007-0479-4
DOI 10.1007/978-94-007-0479-4
Springer Dordrecht Heidelberg London New York

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

CONTENTS

PREFACE TO THE SECOND EDITION

It is with great pleasure that we are presenting to the community the second edition of this extraordinary handbook. It has been over 15 years since the publication of the first edition and there have been great changes in the landscape of philosophical logic since then.

The first edition has proved invaluable to generations of students and researchers in formal philosophy and language, as well as to consumers of logic in many applied areas. The main logic article in the Encyclopaedia Britannica 1999 has described the first edition as 'the best starting point for exploring any of the topics in logic'. We are confident that the second edition will prove to be just as good!

The first edition was the second handbook published for the logic community. It followed the North Holland one volume *Handbook of Mathematical Logic*, published in 1977, edited by the late Jon Barwise. The four volume *Handbook of Philosophical Logic*, published 1983–1989 came at a fortunate temporal junction at the evolution of logic. This was the time when logic was gaining ground in computer science and artificial intelligence circles.

These areas were under increasing commercial pressure to provide devices which help and/or replace the human in his daily activity. This pressure required the use of logic in the modelling of human activity and organisation on the one hand and to provide the theoretical basis for the computer program constructs on the other. The result was that the *Handbook of Philosophical Logic*, which covered most of the areas needed from logic for these active communities, became their bible.

The increased demand for philosophical logic from computer science and artificial intelligence and computational linguistics accelerated the development of the subject directly and indirectly. It directly pushed research forward, stimulated by the needs of applications. New logic areas became established and old areas were enriched and expanded. At the same time, it socially provided employment for generations of logicians residing in computer science, linguistics and electrical engineering departments which of course helped keep the logic community thriving. In addition to that, it so happens (perhaps not by accident) that many of the Handbook contributors became active in these application areas and took their place as time passed on, among the most famous leading figures of applied philosophical logic of our times. Today we have a handbook with a most extraordinary collection of famous people as authors!

The table below will give our readers an idea of the landscape of logic and its relation to computer science and formal language and artificial intelligence. It shows that the first edition is very close to the mark of what was needed. Two topics were not included in the first edition, even though

D. Gabbay and F. Guenthner (eds.),
Handbook of Philosophical Logic, Volume 16, vii–ix.
DOI 10.1007/978-94-007-0479-4,

they were extensively discussed by all authors in a 3-day Handbook meeting. These are:

- a chapter on non-monotonic logic
- a chapter on combinatory logic and λ-calculus

We felt at the time (1979) that non-monotonic logic was not ready for a chapter yet and that combinatory logic and λ-calculus was too far removed.[1] Non-monotonic logic is now a very major area of philosophical logic, alongside default logics, labelled deductive systems, fibring logics, multi-dimensional, multimodal and substructural logics. Intensive re-examinations of fragments of classical logic have produced fresh insights, including at time decision procedures and equivalence with non-classical systems.

Perhaps the most impressive achievement of philosophical logic as arising in the past decade has been the effective negotiation of research partnerships with fallacy theory, informal logic and argumentation theory, attested to by the Amsterdam Conference in Logic and Argumentation in 1995, and the two Bonn Conferences in Practical Reasoning in 1996 and 1997.

These subjects are becoming more and more useful in agent theory and intelligent and reactive databases.

Finally, fifteen years after the start of the Handbook project, I would like to take this opportunity to put forward my current views about logic in computer science, computational linguistics and artificial intelligence. In the early 1980s the perception of the role of logic in computer science was that of a specification and reasoning tool and that of a basis for possibly neat computer languages. The computer scientist was manipulating data structures and the use of logic was one of his options.

My own view at the time was that there was an opportunity for logic to play a key role in computer science and to exchange benefits with this rich and important application area and thus enhance its own evolution. The relationship between logic and computer science was perceived as very much like the relationship of applied mathematics to physics and engineering. Applied mathematics evolves through its use as an essential tool, and so we hoped for logic. Today my view has changed. As computer science and artificial intelligence deal more and more with distributed and interactive systems, processes, concurrency, agents, causes, transitions, communication and control (to name a few), the researcher in this area is having more and more in common with the traditional philosopher who has been analysing

[1] I am really sorry, in hindsight, about the omission of the non-monotonic logic chapter. I wonder how the subject would have developed, if the AI research community had had a theoretical model, in the form of a chapter, to look at. Perhaps the area would have developed in a more streamlined way!

such questions for centuries (unrestricted by the capabilities of any hardware).

The principles governing the interaction of several processes, for example, are abstract an similar to principles governing the cooperation of two large organisation. A detailed rule based effective but rigid bureaucracy is very much similar to a complex computer program handling and manipulating data. My guess is that the principles underlying one are very much the same as those underlying the other.

I believe the day is not far away in the future when the computer scientist will wake up one morning with the realisation that he is actually a kind of formal philosopher!

The projected number of volumes for this Handbook is about 18. The subject has evolved and its areas have become interrelated to such an extent that it no longer makes sense to dedicate volumes to topics. However, the volumes do follow some natural groupings of chapters.

I would like to thank our authors are readers for their contributions and their commitment in making this Handbook a success. Thanks also to our publication administrator Mrs J. Spurr for her usual dedication and excellence and to Kluwer Academic Publishers for their continuing support for the Handbook.

Dov M. Gabbay
King's College London

Logic	**IT**			
	Natural language processing	**Program control specification, verification, concurrency**	**Artificial intelligence**	**Logic programming**
Temporal logic	Expressive power of tense operators. Temporal indices. Separation of past from future	Expressive power for recurrent events. Specification of temporal control. Decision problems. Model checking.	Planning. Time dependent data. Event calculus. Persistence through time—the Frame Problem. Temporal query language. temporal transactions.	Extension of Horn clause with time capability. Event calculus. Temporal logic programming.
Modal logic. Multi-modal logics	generalised quantifiers	Action logic	Belief revision. Inferential databases	Negation by failure and modality
Algorithmic proof	Discourse representation. Direct computation on linguistic input	New logics. Generic theorem provers	General theory of reasoning. Non-monotonic systems	Procedural approach to logic
Non-monotonic reasoning	Resolving ambiguities. Machine translation. Document classification. Relevance theory	Loop checking. Non-monotonic decisions about loops. Faults in systems.	Intrinsic logical discipline for AI. Evolving and communicating databases	Negation by failure. Deductive databases
Probabilistic and fuzzy logic	logical analysis of language	Real time systems	Expert systems. Machine learning	Semantics for logic programs
Intuitionistic logic	Quantifiers in logic	Constructive reasoning and proof theory about specification design	Intuitionistic logic is a better logical basis than classical logic	Horn clause logic is really intuitionistic. Extension of logic programming languages
Set theory, higher-order logic, λ-calculus, types	Montague semantics. Situation semantics	Non-well-founded sets	Hereditary finite predicates	λ-calculus extension to logic programs

Imperative vs. declarative languages	**Database theory**	**Complexity theory**	**Agent theory**	**Special comments: A look to the future**
Temporal logic as a declarative programming language. The changing past in databases. The imperative future	Temporal databases and temporal transactions	Complexity questions of decision procedures of the logics involved	An essential component	Temporal systems are becoming more and more sophisticated and extensively applied
Dynamic logic	Database updates and action logic	Ditto	Possible actions	Multimodal logics are on the rise. Quantification and context becoming very active
Types. Term rewrite systems. Abstract interpretation	Abduction, relevance	Ditto	Agent's implementation rely on proof theory.	
	Inferential databases. Non-monotonic coding of databases	Ditto	Agent's reasoning is non-monotonic	A major area now. Important for formalising practical reasoning
	Fuzzy and probabilistic data	Ditto	Connection with decision theory	Major area now
Semantics for programming languages. Martin-Löf theories	Database transactions. Inductive learning	Ditto	Agents constructive reasoning	Still a major central alternative to classical logic
Semantics for programming languages. Abstract interpretation. Domain recursion theory.		Ditto		More central than ever!

Classical logic. Classical fragments	Basic background language	Program synthesis	A basic tool	
Labelled deductive systems	Extremely useful in modelling		A unifying framework. Context theory.	Annotated logic programs
Resource and substructural logics	Lambek calculus		Truth maintenance systems	
Fibring and combining logics	Dynamic syntax	Modules. Combining languages	Logics of space and time	Combining features
Fallacy theory				
Logical Dynamics	Widely applied here			
Argumentation theory games		Game semantics gaining ground		
Object level/ metalevel			Extensively used in AI	
Mechanisms: Abduction, default relevance			ditto	
Connection with neural nets				
Time-action-revision models			ditto	

	Relational databases	Logical complexity classes	The workhorse of logic	The study of fragments is very active and promising.
	Labelling allows for context and control.		Essential tool.	The new unifying framework for logics
Linear logic			Agents have limited resources	
	Linked databases. Reactive databases		Agents are built up of various fibred mechanisms	The notion of self-fibring allows for self-reference
				Fallacies are really valid modes of reasoning in the right context.
			Potentially applicable	A dynamic view of logic
				On the rise in all areas of applied logic. Promises a great future
			Important feature of agents	Always central in all areas
			Very important for agents	Becoming part of the notion of a logic
				Of great importance to the future. Just starting
			A new theory of logical agent	A new kind of model

ODINALDO RODRIGUES, DOV GABBAY AND ALESSANDRA RUSSO

BELIEF REVISION

1 INTRODUCTION AND HISTORICAL PERSPECTIVE

The investigation of how humans reach conclusions from given premises has long been the subject of intense research in the literature. It was the basis of the development of classical logic, for instance. The investigation of how humans change their minds in the face of new contradictory information is however somewhat more recent. Early accounts include the work of Ramsey [Ramsey, 1931; Ramsey, 1990] in his insights into conditional statements, for instance, and subsequently the work on conditionals by Stalnaker [Stalnaker, 1968b] and by Lewis [Lewis, 1973], among others. More recent work on the formalisation of *common-sense reasoning*, sometimes also called *non-monotonic reasoning*, include [McCarthy, 1958; Brewka, 1990; Łukaszewicz, 1990; Reiter, 1980].

The now trademark term AGM is an acronym formed with the initials of the main proposers of this theory of belief change, namely, Carlos Alchourrón, Peter Gärdenfors and David Makinson. Alchourrón and Makinson had worked jointly on *theory change* in the early 80s [Alchourrón and Makinson, 1982; Alchourrón and Makinson, 1985], and, independently, Gärdenfors had been working on *belief change* in the late 70's and early 80s [Gärdenfors, 1978; Gärdenfors, 1982]. After collaborations in various combinations of the three authors, they published together the paper "On the Logic of Theory Change" [Alchourrón *et al.*, 1985], which provided the basis for what is now known as the AGM theory of belief revision.[1]

The main object of study of the theory of belief revision is the dynamics of the process of belief change: when an agent is faced with new information which contradicts her current beliefs, she will have to retract some of the old beliefs in order to accommodate the new belief consistently. In general, this can be done in a number of ways. Belief revision is concerned with how to make the process *rational*. The AGM theory stipulates some rationality principles to be observed — the so-called *AGM postulates for belief change*. These are discussed in more detail in Section 2.

[1] A more comprehensive recapitulation of the early history of AGM done by Makinson himself can be found in [Makinson, 2003]

D. Gabbay and F. Guenthner (eds.),
Handbook of Philosophical Logic, Volume 16, 1–114.
DOI 10.1007/978-94-007-0479-4_1,

Many other articles followed the initial proposal of the AGM theory analysing properties of belief change operations and how they relate to each other. Subsequently, Gärdenfors published a book entitled "Knowledge in Flux" [Gärdenfors, 1988], which is an excellent reference to the early work in the area. Since then, the work on belief revision has flourished and diversified into several different subareas.

One of the first specialisations was related to the status given to certain beliefs held by an agent. Following the usual terminology, we will call the collection of beliefs of an agent at a given moment in time her *belief set*. According to the *coherentist view*, an agent has no mechanism other than logical coherence for keeping track of the 'reasons' why a given belief is supported. Consequently, belief change operations need only to describe the relationship between belief sets at two adjacent moments in time. In the *foundationalist view*, however, beliefs can be held by an agent only if they have a proper *justification* — if a justification becomes untenable and is retracted, then all beliefs that rely on it must also be retracted. Therefore, belief change operations need to specify a mechanism for maintaining the justifications for the beliefs. In the simplest form of foundationalism, some beliefs are regarded as requiring no justification and called *basic* (sometimes also called *foundational* [Harman, 1986, page 31]). A variation of this approach with special interest to computer science makes a simple distinction between the set of beliefs supported by an agent (her *belief set*) and the set of beliefs from which these are derived (her *belief base*). Changes are made to the belief base, which, in general, is a finite set. The reader is referred to [Gärdenfors, 1990; Doyle, 1992] for a more comprehensive discussion of the differences between the two paradigms. Obviously, the problem of belief revision can be approached from different perspectives as well. We will present some of these throughout this chapter.

A related area of investigation on theory change is concerned with the formalisation of the effects of the execution of actions in the real world. Operations of this types are usually called *updates* and this problem is related with other areas of artificial intelligence, including planning, logical databases and robotics. In a very influential article [Fagin *et al.*, 1983], Fagin et al. investigated how to update logical databases presenting many of the principles of theory revision and update now widely accepted. One of these is the idea of preservation of old information through the definition of some minimality criteria. They also realised early on the importance of considering the *logical consequences* associated with a database.

Some other early work in that area is also worth mentioning. For instance, Ginsberg and Smith's articles on the formalisation of the reasoning about actions [Ginsberg and Smith, 1988a; Ginsberg and Smith, 1988b] and the well known article written by Winslett "Reasoning about action using a possible models approach" [Winslett, 1988a]. The latter highlighted the importance of semantical considerations in the achievement of rational changes of information caused by the execution of actions.

Even though belief revision and updates are clearly distinct, they have similarities between them. In particular, the so-called *principle of informational economy.* After all, it does not seem rational for an agent to discard all of the knowledge accumulated about the world in the face of new contradictory information. The similarities between updates and belief revision (as well as those between other forms of non-monotonic reasoning) have been extensively investigated (see, for instance [Makinson, 1993; Katsuno and Satoh, 1991; Makinson and Gärdenfors, 1989]). Analogously, the differences between the two were emphasised by Katsuno and Mendelzon [Katsuno and Mendelzon, 1991a; Katsuno and Mendelzon, 1992]. In an analogy to the AGM trio, they proposed some postulates for update operations. Further investigation on specialised types of update operations appeared in [Brewka and Hertzberg, 1993].

We note that several formalisms deal independently with belief revision [Dalal, 1988a; Gärdenfors, 1988; Alchourrón and Makinson, 1985], updates in databases [Brewka and Hertzberg, 1993; Winslett, 1988a; Ginsberg and Smith, 1988a; Ginsberg and Smith, 1988b], default reasoning [Reiter, 1980; Poole, 1988a; Brewka, 1989b; Brewka, 1991], conditional reasoning [Lewis, 1973; Nute, 1984; Stalnaker, 1968b; Grahne, 1991a; Grahne, 1991b], argumentation [Bench-Capon and Dunne, 2007; Besnard and Hunter, 2008; Modgil, 2009], etc. However, there is very little work on the combination of these. Such integration is arguably essential to the modelling of a rational agent that needs to deal with multiple forms of common-sense reasoning.

One issue that often arises in the problem of belief revision is the choice of what beliefs to give up during a revision operation. One approach relies on the representation of how strongly an agent feels about her beliefs so that when a choice needs to be made, those beliefs on which she less strongly believes will go first. In order to support this approach, some of the work on belief revision includes the investigation of the representation of priorities associated with the beliefs in the belief set (or base). The usual mechanism is a preference relation associated with the beliefs. The reasoning about the preferences can itself be quite complex, because in the general (and interesting) case, an agent has only partial information about these preferences.

Another issue is the investigation of the relation between belief sets obtained after successive belief change operations. The original AGM theory had little to say about the iteration of the process of belief change. The study of the properties of *iterated revision* started in the mid 90's and is, of course, of great importance to both computer science and philosophy in general. More profound considerations started with [Darwiche and Pearl, 1994; Freund and Lehmann, 1994; Lehmann, 1995; Friedman and Halpern, 1996; Darwiche and Pearl, 1997; Rodrigues, 1998; Friedman and Halpern, 1999] and more recent work includes [Konieczny and Pérez, 2000; Herzig *et al.*, 2003; Rodrigues, 2005]. The problems of iteration will be discussed in Section 4.

Finally, some effort has also been directed towards the study of the complexity involved in the implementation of belief revision systems. Some results can be found in [Gärdenfors and Rott, 1995, page 98] and in [Eiter and Gottlob, 1992a; Eiter and Gottlob, 1993; Nebel, 1991a; Nebel, 1992; Nebel, 1998; Gärdenfors and Rott, 1995; Eiter and Gottlob, 1992b; Nebel, 1992; Nebel, 1998]. As we shall see in Section 2, one of the postulates for belief revision operations stipulates that the resulting belief set is consistent provided that the revising information is not itself contradictory. Thus, the problem of belief revision is at least as hard as the problem of deciding the satisfiability of a set of formulae. Reasoning about preferences can also add to the complexity of the problem and, as a consequence, many belief revision formalisms constrain themselves to a fragment of first-order logic or, in most cases, to propositional logic [Dalal, 1988a; Dargam, 1996; Katsuno and Mendelzon, 1991b; Gabbay and Rodrigues, 1997; Rodrigues, 2003]. Some complexity results will be briefly presented in Section 6.

In the sections that follow we provide a review of the AGM framework, discuss some philosophical problems of the process of belief change not addressed by the theory, and present some representative work proposed to tackle the issues.

This chapter is structured as follows: alternative formalisations of the problem of belief revision are presented and discussed in Section 2. Some well known belief revision operators are presented in Section 3. In Section 4, we discuss the problems of iterated revision and alternative ways of dealing with it. Section 5 presents some special types of revision formalisms. This is followed by a survey of complexity issues associated with the revision operation in Section 6; some applications of belief revision in Section 7 and a discussion of challenges and open issues in Section 8.

2 FORMALISATION OF THE PROBLEM OF BELIEF REVISION

In order to discuss belief revision in more detail, it will be useful to introduce some terminology first. Let K be a set of formulae representing the beliefs of an agent in the language of some logic L, with consequence relation Cn.[2] K is called a *belief set* when it is closed under Cn, i.e., $K = \mathrm{Cn}(K)$. Given a belief set K and a belief φ, we say that φ is *accepted* in K when $\varphi \in K$.

As mentioned previously, the framework of belief revision was developed around some desiderata of the operations on belief sets, called the *AGM postulates for belief change*, whose main purpose is to model *rational changes of belief*. The AGM theory defines three main types of belief change:

- *Expansion*: the incorporation of a new belief φ into a belief set K. The new belief set is represented by $K{+}\varphi$ and defined simply as $\mathrm{Cn}(K \cup \{\varphi\})$. Notice that $K{+}\varphi$ will be inconsistent if K is inconsistent; or if φ is contradictory; or if they are both independently satisfiable although $K \cup \{\varphi\}$ is not jointly satisfiable. Since belief sets are closed under the consequence relation, the inconsistent belief set is unique and equivalent to the set of all formulae in the language. The inconsistent belief set will be denoted by $K_\perp$.

- *Contraction*: the retraction of a belief from a belief set. Since belief sets are closed under the consequence relation, in order to retract a belief φ from K, it is also necessary to remove other beliefs in K that imply φ. A contraction of K by φ is represented by $K{-}\varphi$.

- *Revision*: the incorporation of a belief φ into a belief set K in such a way that the resulting belief set is consistent unless φ is itself contradictory. The interesting case is when φ is not contradictory but inconsistent with K. The main issue in this case is to determine which of the beliefs in K to retract in order to consistently accept φ. The revision of a belief set K by a belief φ is represented by $K{\circ}\varphi$.

As can be seen, the interesting belief change operations are contractions and revisions. In fact, it is possible to define one of the operations in terms of the other. The *Levi identity* defines revisions in terms of contractions and the *Harper identity* defines contractions in terms of revisions.[3]

Levi identity: $K{\circ}\varphi{=}(K{-}\neg\varphi){+}\varphi$

[2] We will use the terms "belief" and "formula" interchangeably.

[3] Note that these operations are all defined at the metalevel. For an investigation on how to bring contraction to the object level see [Gabbay *et al.*, 2002; Gabbay *et al.*, 2004].

Harper identity: $K - \varphi = K \cap (K \circ \neg\varphi)$

In this chapter, we will concentrate only on the revision process although we state the following important result [Gärdenfors, 1988] that will prove useful in the next sections:

THEOREM 1. *If a contraction function verifies the AGM postulates for contraction and an expansion function verifies the AGM postulates for expansion and a revision function $\circ$ is defined in terms of both according to the Levi identity above, then $\circ$ verifies the AGM postulates for revision (presented in the next section).*

The intuition behind revisions defined via the Levi identity is that one should first give up all beliefs that are in conflict with the new information before adding it (if consistency is to be maintained). Naturally, when forced to give up some of the beliefs, one should try and minimise the loss of information involved in the process. This requirement is commonly referred to as the *principle of minimal change* or the *principle of informational economy* [Gärdenfors, 1988, page 49]:

> "... when we change our beliefs, we want to retain as much as possible of our old beliefs — information is not in general gratuitous, and unnecessary losses of information are therefore to be avoided."

Without considering preferences between beliefs, in general there will be several possibilities of contracting a belief set K in such a way that it can consistently accept a new belief φ. In order to comply with the principle of minimal change, we will be interested only in those contractions that minimise the loss of beliefs. This can be formalised in the following way:

DEFINITION 2 (Maximal subsets that fail to imply a sentence). Let K be a belief set and $\neg\varphi$ a belief. A set K' is a maximal subset of K that fails to imply $\neg\varphi$ iff the following conditions are met:

- $K' \subseteq K$
- $\neg\varphi \notin \mathrm{Cn}(K')$
- $\forall K''$, $K' \subset K'' \subseteq K$ implies $\neg\varphi \in \mathrm{Cn}(K'')$

In other words, any subset of K larger than K' (and containing it) would result in the derivation of $\neg\varphi$. It should always be possible to find such subsets unless $\neg\varphi$ is a tautology. Following the usual convention found in the literature, the set of all subsets of K that do not imply $\neg\varphi$ will be denoted $K_\perp\neg\varphi$. A *maxichoice contraction* of K by $\neg\varphi$ is an operation that returns one of the elements of $K_\perp\neg\varphi$ when there is at least one or K itself if $\neg\varphi$ is a tautology. A *full meet contraction* of K by $\neg\varphi$ is an operation that returns the intersection of all elements of $K_\perp\neg\varphi$ or K itself if $K_\perp\neg\varphi$ is empty. Finally, a *partial meet contraction* of K by $\neg\varphi$ is an operation that returns the intersection of *some* appropriately selected elements of $K_\perp\neg\varphi$ if it is non-empty[4] or as before K itself, otherwise. Based on these contraction functions, a number of revision functions can be defined via the Levi identity. However, Alchourrón and Makinson showed that maxichoice contraction functions produce belief sets that are too large [Alchourrón and Makinson, 1982] and, as a result, revision operations defined in terms of these contractions will be *maximal*:[5]

THEOREM 3. *If a revision function* $\circ$ *is defined from a maxichoice contraction function* $-$ *via the Levi identity, then for any belief* φ *such that* $\neg\varphi \in K$, $K \circ \varphi$ *will be maximal.*

On the other hand, full meet contractions are too restrictive, and analogously Alchourrón and Makinson showed that revisions defined in terms of this kind of contractions will in general produce belief sets that are too "small":

THEOREM 4. *If a revision function* $\circ$ *is defined from a full meet contraction function* $-$ *via the Levi identity, then for any belief* φ *such that* $\neg\varphi \in K$, $K \circ \varphi = \mathrm{Cn}(\{\varphi\})$.

Thus, on the one hand, if maxichoice revisions are used, the arrival of some conflicting information makes an agent omniscient. On the other hand, if full meet revisions are used, the arrival of new conflicting information causes the agent to lose all previous information. So it seems that the only realistic revisions that can be defined in terms of contractions and the Levi identity are the ones that use partial meet contractions. The difficulty with this type of revisions is that they rely on a selection mechanism that is *external* to the agent's own representation of beliefs. Although arguably not as elegant from the philosophical point of view, the need for some extra information supporting the beliefs of an agent will become evident in the sections that follow. In particular, a discussion to motivate the employment of an agent's *epistemic state* (as opposed to her *belief* state) is presented in Section 4.

[4] The selection function may take other criteria such as epistemic entrenchment (see Section 2.5) into consideration during the process.

[5] A belief set K is maximal if for any belief φ, either $\varphi \in K$ or $\neg\varphi \in K$.

The postulates for the revision operation as given in [Gärdenfors, 1988, pp. 54–56] are now presented. In the following presentation φ and ψ denote beliefs and the symbol $K_\perp$ denotes the inconsistent belief set as before. From now on, $\circ$ will be subscripted to denote a specific belief revision operation. In particular, $\circ_a$ will be used to denote a revision operation complying with the AGM postulates.

2.1 AGM postulates for belief revision

(K°1) $K \circ_a \varphi$ is a belief set
(K°2) $\varphi \in K \circ_a \varphi$
(K°3) $K \circ_a \varphi \subseteq \mathrm{Cn}(K \cup \{\varphi\})$
(K°4) If $\neg\varphi \notin K$, then $\mathrm{Cn}(K \cup \{\varphi\}) \subseteq K \circ_a \varphi$
(K°5) $K \circ_a \varphi = K_\perp$ only if φ is contradictory
(K°6) If $\varphi \equiv \psi$, then $K \circ_a \varphi \equiv K \circ_a \psi$
(K°7) $K \circ_a (\varphi \wedge \psi) \subseteq \mathrm{Cn}((K \circ_a \varphi) \cup \{\psi\})$
(K°8) If $\neg\psi \notin K \circ_a \varphi$, then $\mathrm{Cn}(K \circ_a \varphi \cup \{\psi\}) \subseteq K \circ_a (\varphi \wedge \psi)$

Postulate (K°1) requires the result of the revision operation to be a belief set, i.e., that the revised set be closed under the consequence relation Cn. In more general terms, (K°1) requires that the operation preserves the defining properties of the original belief set.

Postulate (K°2) is known as the *success postulate*, but sometimes also referred to as the *principle of the primacy of the update* [Dalal, 1988b]. It basically requires the revision process to be successful in the sense that the new belief is effectively accepted after the revision operation is applied. The controversy is that the new belief may be itself contradictory, in which case (K°2) requires the new belief set to be inconsistent. Since the logic used to model the belief sets is classical and AGM adopts the coherentist view, all beliefs become accepted after such a revision is performed. The reliance of AGM framework on the consistency notion is discussed in more detail in [Gabbay *et al.*, 2000; Rodrigues *et al.*, 2008].

(K°3) says sets an expansion as the upper bound of a revision operation. (K°4) on the other hand, specifies that provided that the new belief is not inconsistent with the current belief set, the revision operation will include all of the consequences of the old belief set together with the new belief. Thus, it sets a lower bound for the operation in the case when the new belief is consistent with the current belief set.

(K°5) is sometimes referred to as the *recovery postulate*. It guarantees that the result of a revision is consistent provided that the revising sentence itself is non-contradictory. To understand what (K°3)–(K°5) say, two cases need to be considered:

Case 1: $K \cup \{\varphi\}$ is consistent.

In this case, (K°3) and (K°4) require that $K \circ_a \varphi = \mathrm{Cn}(K \cup \{\varphi\})$, since by (K°3), $K \circ_a \varphi \subseteq \mathrm{Cn}(K \cup \{\varphi\})$ and by (K°4), $\mathrm{Cn}(K \cup \{\varphi\}) \subseteq K \circ_a \varphi$. (K°5) is vacuously true.

Case 2: $K \cup \{\varphi\}$ is inconsistent.

In this case, (K°3) does not say much about $K \circ_a \varphi$. If $K \cup \{\varphi\}$ is classically inconsistent, then any theory whatsoever is included in $\mathrm{Cn}(K \cup \{\varphi\})$, because this theory is simply $K_\perp$. Similarly, (K°4) says little about $K \circ_a \varphi$. Since $K \cup \{\varphi\}$ is inconsistent, it follows that $\neg\varphi \in K$ (since K is a closed theory), and hence (K°4) is satisfied vacuously. As for (K°5), two subcases can be considered:

1. φ is non-contradictory. In this case, (K°5) requires $K \circ_a \varphi$ to be consistent, but gives us no clue as to what $K \circ_a \varphi$ should look like — minimal requirements are given by (K°1) and (K°2).

2. φ is contradictory. In this case, (K°5) says nothing about $K \circ_a \varphi$. However, (K°1) and (K°2) jointly force $K \circ_a \varphi = K_\perp$.

The above case analysis shows that the AGM postulates (K°3)–(K°5) have something to say only when $K \cup \{\varphi\}$ is consistent, or when it is inconsistent even though φ is non-contradictory. The particular way of writing the postulates given above makes use of technical properties of classical logic (the way inconsistent theories prove everything). Also notice that classical inconsistency is the (only) trigger of the revision process — if the new belief and the current belief set are jointly consistent, the revision simply amounts to an expansion.

When considering the AGM postulates for logics other than classical logic, the notion of *acceptability* needs to be employed instead of consistency whenever the latter is missing. In that case, one needs to decide when a revision is required according to what is reasonable in the non-classical logic. In classical logic, the postulates do not give any clue beyond (K°5) as to what to require when $K \cup \{\varphi\}$ is inconsistent. These issues have been analysed in detail in [Gabbay *et al.*, 2000; Rodrigues *et al.*, 2008; Gabbay *et al.*, 2010].

To summarise, postulates (K°3)–(K°4) effectively mean the following:

($\mathrm{K}^\circ_{3,4}$) If φ is consistent with K, then $K \circ_a \varphi = \mathrm{Cn}(K \cup \{\varphi\})$.

If K is finitely representable, it can be taken as a formula and the postulate above corresponds to postulate (R2) in Katsuno and Mendelzon's rephrasing of the AGM postulates for belief sets represented by finite bases [Katsuno and Mendelzon, 1992, p. 187] (see also Section 2.4 below).

(K°6) specifies that the revision process should be independent of the syntactic form of the sentences involved. It is called the *principle of irrelevance of syntax* by many authors, including [Dalal, 1988b].

(K°7) and (K°8) are the most interesting and controversial postulates. They try to capture the informational economy principle outlined before. In order to understand these postulates, consider the following semantical interpretation and assume one has some mechanism to evaluate similarity between models (i.e., valuations of the logic L). In order to keep as much as possible of the *informational content* of a belief set K, we need to look at the valuations that most resemble the models of K itself (in symbols, $\mathrm{mod}(K)$). If a new belief φ is also to be accepted, we will then be looking at the models of φ that most resemble *some* model of K. (K°7) says that if a model I of φ is among the valuations that are most similar to models of K and it happens that I is also a model of a belief ψ, then I should also be among the models of $\varphi \wedge \psi$ which are most similar to models of K.

Similarly, to understand the intuitive meaning of (K°8) consider the following situation: suppose that $(K \circ_a \varphi) \wedge \psi$ is satisfiable. It follows that some models of φ which are most similar to models of K are also models of ψ. These models are obviously in $\mathrm{mod}(\varphi \wedge \psi)$, since by (K°1), $\mathrm{mod}(K \circ_a \varphi) \subseteq \mathrm{mod}(\varphi)$. Now, every model in $\mathrm{mod}(\varphi \wedge \psi)$ which is most similar to a model of K must also be a model of $(K \circ_a \varphi) \wedge \psi$. This situation is depicted in Figure 1, where valuations are represented around K according to their degree of similarity. The closer it is to $\mathrm{mod}(K)$, the more similar to K is a valuation (the exact nature of the similarity notion is irrelevant to the understanding of the postulate now). The figure provides an illustration of (K°8) using Grove's modelling of theory change [Grove, 1988] presented in Section 2.3.

Another way of seeing (K°7) and (K°8) is by considering the restrictions they impose on the acceptance of beliefs φ and ψ as a sequence (revising by φ, then expanding by ψ), as compared to revising by φ and ψ at the same time (i.e., revising by $\varphi \wedge \psi$). One of the main criticisms to the AGM framework is the fact that they do not constrain enough properties of sequences of revisions. (K°7) and (K°8) impose the bare minimum restrictions (see Section 4 below). We distinguish the following three cases:

Case 1: φ is consistent with K.

In this case, $K \circ_a \varphi = \mathrm{Cn}(K \cup \{\varphi\})$ (by previous postulates). Three possible subcases with respect to the sentence ψ are considered.

1. ψ is consistent with $K \circ_a \varphi$. In this case, the antecedent of (K°8) holds and (K°7) and (K°8) together effectively say that $\mathrm{Cn}((K \circ_a \varphi) \cup \{\psi\}) = K \circ_a (\varphi \wedge \psi)$. A more thorough analysis reveals more about AGM in this case, namely, that $(K \circ_a \varphi) \circ_a \psi = \mathrm{Cn}(K \circ_a \varphi \cup \{\psi\})$.

2. ψ is inconsistent with $K \circ_a \varphi$, but ψ itself is non-contradictory. In this case, $\mathrm{Cn}((K \circ_a \varphi) \cup \{\psi\})$ is $K_\perp$. (K°7) holds because the right hand side of the inclusion is the set of all well-formed formulae and any set of formulae is included in this set. (K°8) holds vacuously, since the antecedent of the implication is false.

3. ψ is itself contradictory. The postulates effectively say nothing new in this case, since $K \circ_a (\varphi \wedge \psi) = \mathrm{Cn}((K \circ_a \varphi) \cup \{\psi\}) = K_\perp$. (K°7) holds trivially and (K°8) holds vacuously.

Case 2: φ is not consistent with K, but φ is itself non-contradictory.

In this case, $K \circ_a \varphi$ can be any consistent theory (by previous postulates), such that $\varphi \in K \circ_a \varphi$. As before, there are three possibilities:

1. ψ is consistent with $K \circ_a \varphi$.

2. ψ is inconsistent with $K \circ_a \varphi$, but ψ itself is non-contradictory.

3. ψ is itself contradictory.

These three cases follow, respectively, the same reasoning of cases (1.1), (1.2) and (1.3) above.

Case 3: φ is itself contradictory.

In this case, $K \circ_a \varphi = K_\perp$. Whether or not ψ is contradictory is irrelevant in the postulates in this case. $\mathrm{Cn}(K \circ_a \varphi \cup \{\psi\}) = K_\perp$ and as for case (1.2) above (K°7) holds because any set of formulae is included in $K_\perp$. (K°8) holds vacuously, since the antecedent of the implication is false.

Summary of (K°7)–(K°8)

Postulates (K°7)–(K°8) do not tell us anything new (beyond what we can deduce from earlier postulates), except in the case where ψ is consistent with $K \circ_a \varphi$ (case 1.1), when (K°7) and (K°8) together are equivalent to the postulate below:

($K^\circ_{7,8}$) If ψ is consistent with $K \circ_a \varphi$, then $\mathrm{Cn}((K \circ_a \varphi) \cup \{\psi\}) = K \circ_a (\varphi \wedge \psi)$

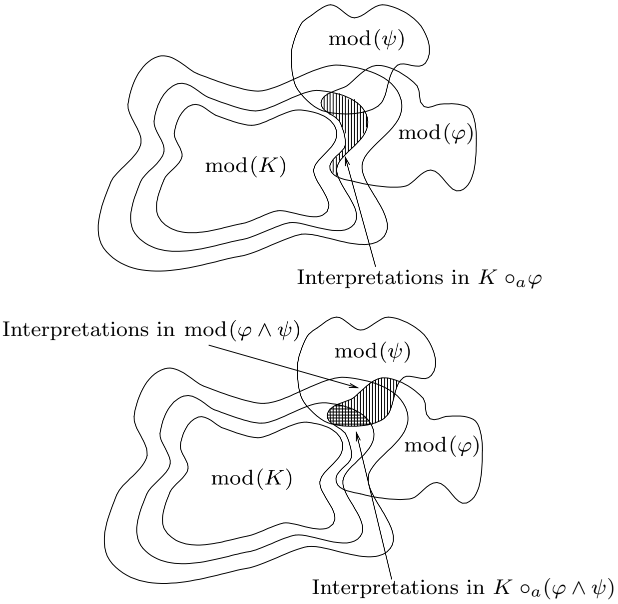


Figure 1. Illustrating postulate (K°8).

Several representation theorems exist for the AGM postulates, for more details the reader is referred to [Grove, 1988; Katsuno and Mendelzon, 1992; Boutilier, 1994].

The realisation of a belief set as an infinite set of formulae poses some problems for computer science applications. In order to overcome this, many authors concentrate instead on a set of basic beliefs from which the belief set is derived. In this case, the basic set of beliefs is called the *belief base* and the belief change process called *base revision* instead. A reformulation of the AGM postulates for finite belief bases will be discussed in Section 2.4.

2.2 Counterfactual statements and the Ramsey Test

There are references to the work on information change since the early 30's [Ramsey, 1931] as well as in subsequent decades [Chisholm, 1946; Stalnaker, 1968a; Stalnaker and Thomason, 1970; Lewis, 1973]. In particular, there is work on the so-called *counterfactual statements*. The best way to introduce the intuition behind counterfactual statements is by presenting an example borrowed from Lewis' book on the subject [Lewis, 1973]:

"If kangaroos had no tails, they would topple over."

Since the antecedent of the sentence is false, its evaluation as an implication in classical logic is trivially *true*. However the *intended* meaning of such a sentence is, as described by Lewis, something like "in any possible state of affairs in which kangaroos have no tails, and which resembles our actual state of affairs *as much as kangaroos having no tails permits it to*, kangaroos topple over" [Lewis, 1973].

The evaluation of such statements has been the object of investigation by many authors [Stalnaker, 1968a; Adams, 1975; Pollock, 1981; Nute, 1984]. Speaking generically, a counterfactual is a sentence of the form

(CNT) "If it were the case that φ, then it would also be the case that ψ."

Following Lewis, we will represent (CNT) by the expression $\varphi \Box\!\!\rightarrow \psi$. It is natural to ask how one should evaluate the truth-values of such sentences. The intended meaning described above suggests that one should accept the belief in φ, changing as little as possible one's current state of beliefs in order to maintain consistency, and then check whether ψ follows from the resulting belief set. This corresponds to the well known *Ramsey Test*, inspired by one of Ramsey's philosophical papers [Ramsey, 1990; Ramsey, 1931], and generalised to its present form by Stalnaker [Stalnaker, 1968b]. One could be easily mislead to think that belief revision could be the operation employed in the Ramsey Test, by taking $\varphi \Box\!\!\rightarrow \psi$ as accepted in a belief set K whenever ψ is accepted in $K \circ \varphi$. In symbols,

(RT) $K \vdash \varphi \Box\!\!\rightarrow \psi$ iff $K \circ \varphi \vdash \psi$

However, it is well known that belief revision cannot be used to evaluate counterfactual statements [Gärdenfors, 1986] [Gärdenfors, 1988, Section 7.4]. *Gärdenfors's impossibility theorem* showed us that whereas (RT) forces the belief change operation to be monotonic, belief revision is intrinsically non-monotonic. To see the first, assume (RT) is accepted, that $\circ$ is the belief change operation used to evaluate counterfactual statements and suppose that for belief sets K_1 and K_2, it is the case that $K_1 \subseteq K_2$. We show that $K_1 \circ \varphi \subseteq K_2 \circ \varphi$. Take any ψ such that $\psi \in K_1 \circ \varphi$. By (RT), $\varphi \Box\!\!\rightarrow \psi \in K_1 \therefore \varphi \Box\!\!\rightarrow \psi \in K_2$, and by (RT) again $\psi \in K_2 \circ \varphi$. To see that belief revision is incompatible with monotonicity, recall the following postulates and consider Example 5.

($K^\circ 2$) $\varphi \in K \circ_a \varphi$
($K^\circ_{3,4}$) If φ is consistent with K, then $K \circ_a \varphi = \mathrm{Cn}(K \cup \{\varphi\})$
($K^\circ 5$) $K \circ_a \varphi = K_\perp$ only if φ is contradictory

EXAMPLE 5. Consider three formulae φ, ψ and (non-contradictory) $\neg\varphi \vee \neg\psi$ and three belief sets K_1, K_2 and K_3, such that $K_1 = \mathrm{Cn}(\{\varphi\})$, $K_2 = \mathrm{Cn}(\{\psi\})$, and $K_3 = \mathrm{Cn}(\{\varphi, \psi\})$. It can be easily seen that $K_1, K_2 \subseteq K_3$. By $(\mathrm{K}^{\circ}_{3,4})$, $K_1 \circ_a (\neg\varphi \vee \neg\psi) = \mathrm{Cn}(\{\varphi, \neg\varphi \vee \neg\psi\}) = \mathrm{Cn}(\{\varphi, \neg\psi\})$; $K_2 \circ_a (\neg\varphi \vee \neg\psi) = \mathrm{Cn}(\{\psi, \neg\varphi\})$; and since $\neg\varphi \vee \neg\psi$ is non-contradictory, $K_3 \circ_a (\neg\varphi \vee \neg\psi)$ is satisfiable. However,

i) $\varphi \in K_1 \circ_a (\neg\varphi \vee \neg\psi)$

ii) $\neg\varphi \in K_2 \circ_a (\neg\varphi \vee \neg\psi)$

and hence, either $K_1 \circ_a (\neg\varphi \vee \neg\psi) \not\subseteq K_3 \circ_a (\neg\varphi \vee \neg\psi)$ or $K_2 \circ_a (\neg\varphi \vee \neg\psi) \not\subseteq K_3 \circ_a (\neg\varphi \vee \neg\psi)$, since $\{\varphi, \neg\varphi\} \not\subseteq K_3 \circ_a (\neg\varphi \vee \neg\psi)$.

In semantical terms, the reason can be understood by recalling Lewis's formulation of satisfiability of counterfactuals via systems of spheres. Let us first introduce the notion of a *centred system of spheres* [Lewis, 1973]:

DEFINITION 6 (Centred system of spheres). Let $\mathcal{I}$ be a set of worlds. A centred system of spheres \$ is an assignment from $\mathcal{I}$ to a set of subsets of $\mathcal{I}$, $\$_I$, where for each $I \in \mathcal{I}$:

① $\{I\} \in \$_I$. (*centring*)

② For all $S, T \in \$_I$, either $S \subseteq T$ or $T \subseteq S$. (*nesting*)

③ $\$_I$ is closed under unions.

④ $\$_I$ is closed under nonempty intersections.

Systems of spheres are used to represent the degree of similarity between worlds. The smaller a sphere containing a world J in $\$_I$ is, the closer to world I world J is. The centring condition ① can be interpreted as "there is no world more similar to world I than I itself". ① can be replaced by

①′ For all $S \in \$_I$, $I \in S$.

This condition is often called *weak centring*.[6] If, in addition to conditions ①–④ above, we also have that for all I, $\bigcup \$_I = \mathcal{I}$, then we say that \$ is *universal*.

In terms of a system of spheres \$, a world I satisfies $\varphi \Box\!\!\rightarrow \psi$, according to the following rules:

[6] Update operations as semantically characterised by Katsuno and Mendelzon in [Katsuno and Mendelzon, 1991b], require strong centring (i.e., the innermost sphere in $\$_I$ contains just I itself).

DEFINITION 7 (Satisfiability of counterfactuals via systems of spheres). Let $\mathcal{I}$ be a set of worlds, $I \in \mathcal{I}$ and \$ a centred system of spheres for $\mathcal{I}$:

$I \Vdash_{\$} \varphi \Box\!\!\rightarrow \psi$ iff

1. either $\forall S \in \$_I$ $\mathrm{mod}(\varphi) \cap S = \emptyset$;
2. or $\exists S \in \$_I$ such that $\mathrm{mod}\,(\varphi) \cap S \neq \emptyset$, and $\forall I \in S,\ I \Vdash \varphi \rightarrow \psi$.

In case (1) above, we say that φ is not *entertainable* at I. That is, there is no sphere around I which intersects any worlds where φ is true. If \$ is universal, this happens only if $\mathrm{mod}(\varphi) = \emptyset$. The set of models of a counterfactual $\varphi \Box\!\!\rightarrow \psi$ can be defined as $\mathrm{mod}(\varphi \Box\!\!\rightarrow \psi) = \{I \in \mathcal{I} \,|\, \forall S \in \$_I\ (\mathrm{mod}(\varphi) \cap S \neq \emptyset$ implies $\forall J \in S\ J \Vdash \varphi \rightarrow \psi)\}$. As for case (2), since $\$_I$ is nested, it is sufficient to check whether $\varphi \rightarrow \psi$ is satisfied by every world in the innermost sphere S for which $S \cap \mathrm{mod}(\varphi)$ is non-empty. Intuitively, this intersection corresponds to the models of φ which are more similar (or closer) to I. Now, if we want to evaluate whether a counterfactual $\varphi \Box\!\!\rightarrow \psi$ is entailed by a belief set K, we have to check whether for each $I \in \mathrm{mod}(K)$, (2) holds, that is, whether the models of φ that are more similar to *each* of the models of K are also models of ψ.

It is not surprising that belief revision cannot be used to evaluate counterfactuals, since it fails to consider each model of K individually — which the operation of *update* does. Indeed, the relationship between counterfactual statements and updates has been pointed out many times [Grahne, 1991b; Rodrigues *et al.*, 1996; Ryan and Schobbens, 1996]. Updates can be used to evaluate conditional statements, and the properties of the resulting conditional logic will depend on the properties of the specific update operation considered.

An alternative way of evaluating counterfactuals is by employing an ordering relation on the set of worlds that determines similarity with respect to a given world I. This is done via the definition of a *comparative similarity system* [Lewis, 1973]:

DEFINITION 8 (Comparative similarity system). A *comparative similarity system* is a function that assigns to each world I a tuple $\langle \leq_I, S_I \rangle$, where S_I is a set of worlds, representing the worlds that are accessible from I; and $\leq_I$ is a binary relation on worlds, representing the comparative similarity of worlds with respect to I, such that

① $\leq_I$ is transitive

② $\leq_I$ is strongly connected

③ $I \in S_I$

④ For any world J, $J \neq I$ implies $I <_I J$

⑤ $K \notin S_I$ implies K is $\leq_I$-maximal

⑥ For any J, K, $J \in S_I$ and $K \notin S_I$ implies $J <_I K$.

The intended meaning for $\leq_I$ is the following: if $J \leq_I K$, then world J is at least as similar to world I as world K is.

Lewis proved that there is a correspondence between the satisfiability of counterfactuals via a system of spheres and their satisfiability via a comparative similarity system. If we consider only *universal comparative similarity systems* (i.e., $S_I = \mathcal{I}$), the truth conditions for counterfactuals can be simplified as follows:

DEFINITION 9 (Satisfiability of counterfactuals via comparative similarity systems). Let $\mathcal{I}$ be a set of worlds and take $I \in \mathcal{I}$. Given a comparative similarity system as in Definition 8:

$I \Vdash \varphi \Box\!\!\rightarrow \psi$ iff

1. either $\mathrm{mod}(\varphi) = \emptyset$;
2. or $\exists M \in \mathrm{mod}(\varphi)$ such that for any $N \in \mathcal{I}$, $N \leq_I M$ implies $N \Vdash \varphi \rightarrow \psi$

These early results were very influential on the work of theory change carried out in the 80's and beyond.

We now turn to another very important semantical characterisation of belief change operations.

2.3 Grove's systems of spheres

In a very influential paper, Grove proposed a semantical characterisation of the AGM theory based on the so-called *systems of spheres* [Grove, 1988]. The idea is similar to that of Lewis' own systems of spheres presented in Section 2.2 above, except for a few modifications. Firstly, the spheres in Lewis' systems contain *worlds*, whereas in Grove's formulation they contain *theories*. In addition, Grove's systems of spheres can contain a *collection* of theories in their centre (a form of weak centring).

Interestingly enough, Grove was one of the first to notice that Lewis' formulation was incompatible with belief revision [Grove, 1988]. The relationship between the types of system of spheres proposed by Lewis and

Grove on the one hand and formalisms for theory change on the other was explored in more detail in [Katsuno and Mendelzon, 1991a; Katsuno and Mendelzon, 1992; Rodrigues *et al.*, 1996] and as it turns out only strongly centred systems of spheres can be used to model *updates* of a knowledge base in the reasoning about the effects of actions [Winslett, 1988b; Katsuno and Mendelzon, 1992].

The starting point in Grove's formulation is the set $M_L^\top$ of all maximal consistent sets of L. These in fact correspond to all (consistent) complete theories of L. Amongst these, some are of particular interest for a given (not necessarily complete) belief set K — the ones that *extend* it. The set of all such extensions is denoted by $|K|$ and formally defined as $\{m \in M_L^\top \mid K \subseteq m\}$. Notice that if K is $K_\perp$, then $|K|$ is simply $\emptyset$. Analogously, given a set S of maximal consistent sets of L, the set $t(S)$ is defined as $\bigcap\{S_i \in S\}$ or $K_\perp$ if $S = \emptyset$. It follows that $t(S)$ is also closed under logical consequence.[7]

In semantical terms, one can think of the set $\mathcal{I}$ of all valuations instead of $M_L^\top$. Analogously, $|K|$ would correspond to $\mathrm{mod}(K)$ and for a given set of valuations $S \subseteq \mathcal{I}$, $t(S) = \{\varphi \mid I \vDash \varphi \text{ for all } I \in S\}$. According to this view of the formulation, if K is $K_\perp$, then $|K| = \mathrm{mod}(K) = \emptyset$ and if $K = \mathrm{Cn}(\emptyset)$, then $|K| = \mathcal{I}$, as expected. However, viewing the revision process in terms of sets of formulae as done by Grove makes the relationship with the AGM postulates immediate, whereas viewing it semantically, i.e., in terms of valuations, gives us an interesting insight into the process.[8]

DEFINITION 10 (Grove's systems of spheres). Let $\mathcal{S}$ be a collection of subsets of $M_L^\top$ and take $S \subseteq M_L^\top$. $\mathcal{S}$ is a called a *system of spheres centred on* S if it satisfies the following conditions:

(S1) $\mathcal{S}$ is totally ordered by $\subseteq$

(S2) S is the $\subseteq$-minimum of $\mathcal{S}$

(S3) $M_L^\top$ is the $\subseteq$-maximum of $\mathcal{S}$

(S4) For any wff φ, if $M_L^\top \cap |\varphi| \neq \emptyset$, then there is a smallest sphere in $\mathcal{S}$ intersecting $|\varphi|$.

[7] It is possible to construct a lattice by ordering theories of a logic L under set inclusion. In this case, $\mathrm{Cn}(\emptyset)$ will be the minimum and $K_\perp$ will be the maximum. The only inconsistent theory in the lattice is $K_\perp$ itself with the elements of $M_L^\top$ sitting immediately below it. The only way to extend an element of $M_L^\top$ and retain closure under logical consequence is to jump to $K_\perp$.

[8] The semantical view is essentially the basis for the work by Katsuno and Mendelzon seen in Section 2.4 [Katsuno and Mendelzon, 1991a; Katsuno and Mendelzon, 1992].

Provided a formula φ is non-contradictory, there is always a maximal consistent extension of $\mathrm{Cn}(\{\varphi\})$. Since by (S3) the outermost layer in every system of spheres is $M_L^\top$ itself (i.e., *all* maximal theories of L), some maximal consistent extension of $\mathrm{Cn}(\{\varphi\})$ will intersect $\mathcal{S}$ at some sphere. Condition (S4) only ensures that there are not infinitely many smaller spheres in $\mathcal{S}$ whose intersection with $|\varphi|$ is non-empty. We use S_φ to denote the *smallest* sphere around S in $\mathcal{S}$ whose intersection with $|\varphi|$ is non-empty if it exists and if it does not, we define $S_\varphi = M_L^\top$. $f_S(\varphi)$ is used to denote the intersection itself, i.e., $|\varphi| \cap S_\varphi$. Notice that if φ is contradictory $f_S(\varphi) = \emptyset$. Intuitively, the function f_S selects the φ-worlds which are closest to S. This can be used to define a revision operation as follows [Grove, 1988]:

DEFINITION 11 (Revision in terms of systems of spheres). Let $\mathcal{S}$ be a system of spheres centred on $|K|$.

$$K \circ_{\mathcal{S}} \varphi = t(f_K(\varphi))$$

PROPOSITION 12. *If K and φ are consistent with each other, then $K \circ_{\mathcal{S}} \varphi = \mathrm{Cn}(K \cup \{\varphi\})$.*

Proof. Notice that if K and φ are consistent with each other, then $|K| \cap |\varphi|$ is non-empty, and hence $K_\varphi = |K|$. Therefore, $f_K(\varphi) = \{m \in M_L^\top \mid K \subseteq m \text{ and } \mathrm{Cn}(\varphi) \subseteq m\}$. It follows that $t(f_K(\varphi)) = \mathrm{Cn}(K \cup \{\varphi\})$. The fact that $\mathrm{Cn}(K \cup \{\varphi\}) \subseteq t(f_K(\varphi))$ comes directly from the definition of $f_K(\varphi)$. To see that $t(f_K(\varphi)) \subseteq \mathrm{Cn}(K \cup \{\varphi\})$, suppose $\gamma \in t(f_K(\varphi))$, but $\gamma \notin \mathrm{Cn}(K \cup \{\varphi\})$. γ is neither contradictory itself (for it belongs to $t(f_K(\varphi))$, by assumption); nor is it a tautology (for it does not belong to $\mathrm{Cn}(K \cup \{\varphi\})$, ditto), therefore it is possible to have two extensions m_1 and m_2 of $\mathrm{Cn}(K \cup \{\varphi\})$ in $f_K(\varphi)$ such that $\gamma \in m_1$ and $\neg\gamma \in m_2$. Since m_1 and m_2 are both consistent, $\gamma \notin m_2$, but this is a contradiction since $\gamma \in t(f_K(\varphi))$. ∎

The proof above establishes that $\circ_{\mathcal{S}}$ verifies $(\mathrm{K}^\circ_{3,4})$. In fact, $\circ_{\mathcal{S}}$ verifies all of the AGM postulates for revision. Grove has proved the following important results:

THEOREM 13. *For any system of spheres $\mathcal{S}$ with centre on $|K|$, if $K \circ_{\mathcal{S}} \varphi$ is given as in Definition 11, then $\circ_{\mathcal{S}}$ verifies the AGM postulates for belief revision.*

THEOREM 14. *If $\circ$ verifies the AGM postulates for belief revision, then for any belief set K, there is a system of spheres $\mathcal{S}$, with centre on $|K|$ such that $K \circ \varphi = t(f_s(\varphi))$.*

A system of spheres around $|K|$ is depicted in Figure 2.

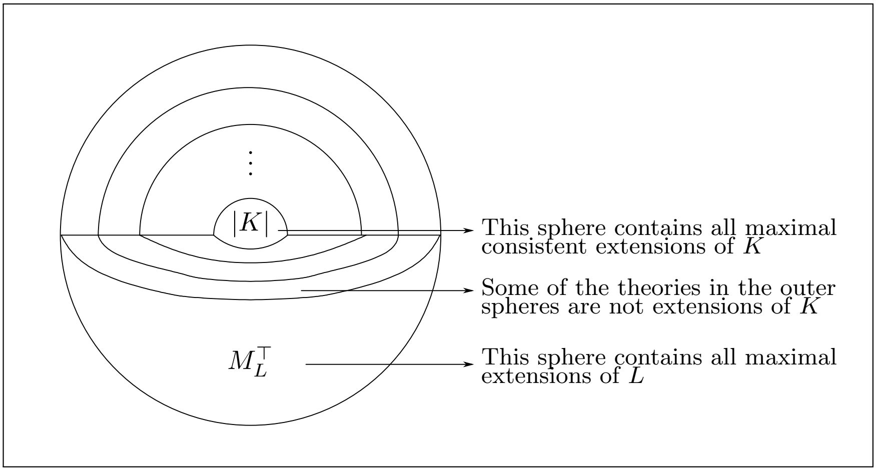


Figure 2. A system of spheres around $|K|$

An alternative representation theorem is given in terms of total pre-orders $\leq_G$ on $\mathsf{wff}(\mathcal{L})$ satisfying the following conditions (we call such pre-orders *Grove relations*):

($\leq_G$ 1) $\leq_G$ is total
($\leq_G$ 2) $\leq_G$ is transitive
($\leq_G$ 3) If $\varphi \rightarrow (\psi \vee \gamma)$, then either $\psi \leq_G \varphi$ or $\gamma \leq_G \varphi$.
($\leq_G$ 4) φ is $\leq_G$ minimal if and only if $\neg\varphi \notin T$.
($\leq_G$ 5) φ is $\leq_G$ maximal if and only if $\neg\varphi$ is a tautology.

It is also possible to define a revision operation in terms of a relation $\leq_G$ as follows ($<_G$ is the strict counterpart of $\leq_G$):

DEFINITION 15 (Revision obtained from $\leq_G$).

$$T \circ_G \varphi = \{\psi \mid (\varphi \wedge \psi) <_G (\varphi \wedge \neg\psi)\}.$$

Grove has shown the following correspondences:

THEOREM 16. *Let $\leq_G$ be any pre-order satisfying ($\leq_G$ 1)–($\leq_G$ 5). If a revision operation $\circ_G$ is defined according to Definition 15, then $\circ_G$ also satisfies* (K°1)–(K°8)*.*

THEOREM 17. *Any revision operator $\circ$ satisfying the AGM postulates can be defined in terms of some relation $\leq_G$ according to Definition 15.*

It is not surprising that $\leq_G$ can also be defined in terms of a system of spheres $\mathcal{S}$:

DEFINITION 18 (Relation between Grove relations and systems of spheres). Let $\mathcal{S}$ be a system of spheres with centre in $|S|$. $\varphi \leq_G \psi$ if and only if $S_\varphi \subseteq S_\psi$.

PROPOSITION 19. *If $\leq_G$ is given according to Definition 18, then it will also verify ($\leq_G 1$)–($\leq_G 5$).*

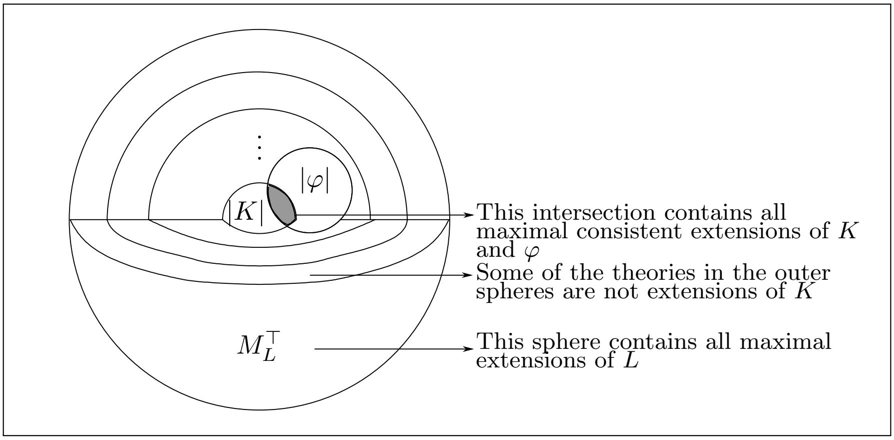


Figure 3. A system of spheres when K and φ are consistent

There is a relationship between Grove relations and the notion of *epistemic entrenchment*. This will be discussed in Section 2.5.

We now turn to a characterisation of the AGM postulates for finite belief bases.

2.4 AGM revision for finite belief bases

In [Katsuno and Mendelzon, 1992], Katsuno and Mendelzon provided a reformulation of the AGM postulates for belief sets represented by finite sets of propositional formulae. We shall refer to these here as *belief bases*. Finite belief bases can be associated with the conjunction of their formulae. In the presentation below, the language $\mathcal{L}$ is assumed to be constructed from a finite set of propositional variables $\mathcal{P}$. K is the *formula* representing the current belief base and φ and ψ are formulae representing new beliefs to be accepted; valuations are constructed by assigning $\{\top, \bot\}$ to each element of $\mathcal{P}$ and extended to complex formulae in the usual way. We use capital letters from the middle of the alphabet (sometimes with a prime symbol), e.g., I (I'), to denote valuations and the symbol $\mathcal{I}$ to denote the set of all valuations. The set mod (K) denotes the set of models of K as before (more formally, $\mathrm{mod}(K) = \{I \in \mathcal{I} \mid I \Vdash K\}$).

AGM postulates for belief revision rewritten for finite belief bases

(R1) $K \circ_a \varphi$ implies φ
(R2) If $K \wedge \varphi$ is satisfiable, then $K \circ_a \varphi \equiv K \wedge \varphi$
(R3) If φ is satisfiable, then $K \circ_a \varphi$ is also satisfiable
(R4) If $K_1 \equiv K_2$ and $\varphi_1 \equiv \varphi_2$, then $K_1 \circ_a \varphi_1 \equiv K_2 \circ_a \varphi_2$
(R5) $(K \circ_a \varphi) \wedge \psi$ implies $K \circ_a (\varphi \wedge \psi)$
(R6) If $(K \circ_a \varphi) \wedge \psi$ is satisfiable, then $K \circ_a (\varphi \wedge \psi)$ implies $(K \circ_a \varphi) \wedge \psi$

The reader will notice that the above reformulation contains two fewer postulates than AGM's original formulation. (K°1) does not make sense for belief bases and (K°3) and (K°4) were combined into (R2). This is no surprise, as we saw in the previous section that in the case that $K \wedge \varphi$ is satisfiable, (K°3) and (K°4) amount to ($\mathrm{K}^{\circ}_{3,4}$) — which in the finite case can be rewritten as (R2); and in the case that $K \wedge \varphi$ is not satisfiable, neither (K°3) nor (K°4) provide any useful information.

In the same paper [Katsuno and Mendelzon, 1992], Katsuno and Mendelzon provided a semantical characterisation in terms of pre-orders of all revision operators satisfying the AGM postulates. This characterisation is useful as a tool to abstract from the postulates and instead concentrate on the semantical properties of the operations. The ideas used are reminiscent of Grove's [Grove, 1988] (see Section 2.3).

The first step in the characterisation is to define minimum requirements the pre-orders have to satisfy. The pre-orders are used to compare the relative "similarity" of valuations with respect to a given belief base. A pre-order satisfying the requirements is said to be *faithful*. In general, for valuations I and I' and a belief base K, $I \leq_K I'$ denotes the fact that valuation I is at least as similar to K as I' is with respect to $\leq$. As usual, $I <_K I'$ denotes $I \leq_K I'$ and $I' \not\leq_K I$.

DEFINITION 20 (Faithful assignment for belief revision). A *faithful assignment for belief revision* is a function mapping each propositional formula K to a *pre-order* $\leq_K$ on $\mathcal{I}$, such that

(F1) If $I, I' \in \mathrm{mod}(K)$, then $I <_K I'$ does not hold.

(F2) If $I \in \mathrm{mod}(K)$ and $I' \notin \mathrm{mod}(K)$, then $I <_K I'$ holds.

(F3) If $K \leftrightarrow \varphi$, then $\leq_K = \leq_\varphi$.

A faithful assignment is *total* if its associated pre-orders are total. What Definition 20 effectively says is that provided K is satisfiable *i)* any two distinct models of K are either incomparably or equivalently similar to each other with respect to $\leq_K$; *ii)* models of K are strictly more similar to K than any non-model of K and *iii)* pre-orders assigned to logically equivalent bases are equivalent. Note that no constraints on the granularity of $\leq_K$ for non-models of K are imposed even though a restriction is imposed on *not* preferring some models of K over other models. The strongest requirement is the strict preference of models of K over valuations that do not satisfy it.

Notation 1. Let $\leq$ be a pre-order on a set S, and $M \subseteq S$. The expression $\min_{\leq}(M)$ will denote the set $\{m \in M \mid \neg\exists m' \in M \text{ such that } m' < m\}$.

The following theorem given in [Katsuno and Mendelzon, 1992] establishes the correspondence between faithful assignments and revision operators satisfying the AGM postulates:

THEOREM 21. *A revision operator* $\circ$ *satisfies postulates* (R1)–(R6) *if and only if there exists a* total *faithful assignment* $\leq_K$ *for each formula* K, *such that* $\text{mod}(K \circ \varphi) = \min_{\leq_K}(\text{mod}(\varphi))$.

Since the assignments associated with operators verifying the AGM postulates are in fact total, (F1) requires equivalence between any two models of K. Because of (F2), total faithful assignments require that $\min_{\leq_K}(\mathcal{I})$ form an equivalence class consisting of all the valuations in $\text{mod}(K)$.

Katsuno and Mendelzon's characterisation is perhaps the most intuitive way to understand the revision process. In order to verify (R1), the models of $K \circ \varphi$ must be included in the models of φ (if any). Intuitively, one expects these to be those which preserve as much as possible of the informational content of K. The measurement of similarity to K is given by the ordering $\leq_K$. The minimum requirements $\leq_K$ must fulfil in order for the associated revision operator to verify the information preservation requirements given by (R1)–(R6) are specified in Definition 20 and hence the minimal elements in mod (φ) are exactly the models of φ which *best* preserve the informational content of K (with respect to $\leq_K$).

Notice that if K is consistent with φ, it is easy to see by Definition 20 that $\min_{\leq_K}(\text{mod}(\varphi)) = \text{mod}(K) \cap \text{mod}(\varphi)$. According to Theorem 21, these are the models of $K \circ \varphi$ — meeting exactly the requirements imposed by (R2). Notice also that K is considered as a whole in the similarity measurement $\leq_K$. This reflects well the coherentist view adopted by AGM.

As seen in this section, a key concept in the process of belief revision is the measurement of information change, that is, how much of the old information is lost during a revision operation. One way of evaluating the change is by comparing the relative informative value of the beliefs held by an agent. This will be discussed in the next sections.

2.5 Epistemic entrenchment

One of the main criticisms against the AGM postulates is that although they define general properties of rational changes of belief, they do not actually provide an explicit construction of the belief change operations themselves. The notion of *epistemic entrenchment* helps to address this issue. Epistemic entrenchment can be used to guide the contraction operation. Since expansions can be trivially constructed, and revisions can be defined from contractions and expansions via the Levi identity (see page 5), the definition of a contraction operation is sufficient to determine the three basic types of belief change.

The idea behind epistemic entrenchment is to retain the more *informative* propositions during a contraction operation. This involves a comparison of the relative strengths of propositions and is modelled by a so-called *epistemic entrenchment relation* $\leq_K$ for a given belief set K. Desirable properties of such relations are given in the form of the postulates presented below. In what follows, φ, ψ and γ are formulae of $\mathcal{L}$ and $\varphi \leq_K \psi$ denotes the fact that ψ is at least as (epistemologically) entrenched as φ as far as K is concerned.

(EE1) $\varphi \leq_K \psi$ and $\psi \leq_K \gamma$ imply $\varphi \leq_K \gamma$
(EE2) If $\varphi \vdash \psi$, then $\varphi \leq_K \psi$
(EE3) Either $\varphi \leq_K \varphi \wedge \psi$ or $\psi \leq_K \varphi \wedge \psi$
(EE4) If $K \neq K_\perp$, then $\varphi \notin K$ iff $\varphi \leq_K \psi$ for all ψ
(EE5) If for all ψ, $\psi \leq_K \varphi$, then $\vdash \varphi$.

(EE1) simply stipulates that epistemic entrenchment relations should be transitive. (EE2) represents the principle of minimal change. Since belief sets are closed under the consequence relation $\vdash$, in the choice between giving up φ or ψ (given that $\varphi \vdash \psi$), it makes more sense to give up φ first, since giving up ψ would ultimately require φ to be given up as well. Note that (EE2) implies reflexivity of $\leq_K$. (EE3) postulates that the loss of information incurred in giving up $\varphi \wedge \psi$ is equivalent to that of giving up either φ or ψ. This follows from the fact that (EE2) already constrains $\leq_K$ so that $\varphi \wedge \psi \leq_K \varphi$ and $\varphi \wedge \psi \leq_K \psi$. (EE1)–(EE3) jointly imply that $\leq_K$ is total. (EE4) expresses a minimality condition for $\leq_K$. Since $\leq_K$ is total, (EE4) stipulates that all beliefs that are not in K are equivalent (modulo

$\leq_K$). Also notice that the proviso is necessary, because if $K = K_\perp$, then $\varphi \in K$ for all φ, in which case there would be an infinite descending chain of beliefs $\varphi >_K \psi_1 >_K \psi_2 > \ldots$ in K. (EE5) says that only the tautologies are maximal in $\leq_K$. The converse follows trivially from (EE2) and hence all tautologies are equivalent modulo $\leq_K$.

What is needed now is a way to define a revision function in terms of an epistemic entrenchment relation. This is done indirectly via a contraction function and the Levi identity (see page 5).

DEFINITION 22 (Contractions and epistemic entrenchment).

$$\psi \leq_K \varphi \text{ iff } \psi \notin K-(\varphi \wedge \psi)$$

In particular,

THEOREM 23 (Properties of epistemically entrenched contraction). *Let $\leq_K$ and $-$ be defined according to Definition 22. It follows that $\leq_K$ satisfies (EE1)–(EE5) iff $-$ satisfies the AGM postulates for contraction.*

These results are similar to the correspondences arising from revision operations defined in terms of Grove relations. This is not surprising. In fact, it is possible to define epistemic entrenchment relations and Grove relations in terms of each other.

DEFINITION 24. For all formulae φ and ψ in $\mathcal{L}$, $\varphi \leq_K \psi$ iff $\neg\varphi \leq_G \neg\psi$.

The properties of $\leq_G$ and $\leq_K$ are related as given in the following theorem [Gärdenfors, 1988, page 96].[9]

THEOREM 25. *A total pre-order $\leq_G$ satisfies ($\leq_G$ 1)–($\leq_G$ 5) iff $\leq_K$, as given in Definition 24, satisfies (EE1)–(EE5).*

A number of revision formalisms using epistemic entrenchment were proposed. The reader is referred to [Rott, 1992; Wobcke, 1992; Nayak, 1994a] for more details.

2.6 Discussion

The relationship between the several representations seen in this section is depicted in Figure 4. Roughly speaking, an arrow $\longrightarrow\!\!\!\!\triangleright$ indicates that the object at its source has the properties stated by the object at its target; an arrow $\dashrightarrow$ indicates that there exists an object of the type of its target for each object of the type of its source; and finally, an arrow $\longrightarrow$ indicates that

[9]This result appears as Lemma 4.26 and Lemma 4.27 in that reference.

the object at its target can be defined in terms of the object at its source. We discuss some of these relationships next.

A relation verifying properties ($\leq_G$ 1)–($\leq_G$ 5) can be defined from a system of spheres $\mathcal{S}$ and vice-versa via Definition 18. Theorem 25 establishes the correspondence between Grove relations and epistemic entrenchment relations defined according to Definition 24. An epistemic entrenchment relation $\leq_E E$ can be used to define a contraction function $-$. According to Theorem 23, $-$ will satisfy the AGM postulates for contraction. Via the Levi identity, it is possible to define a revision function $\circ_a$, which according to Theorem 1 will satisfy the AGM postulates for revision. Theorem 21 establishes the correspondence between faithful assignments $\leq_{KM}$ (for a belief base KM), revision operators $\circ_{KM}$ defined in terms of these and the AGM postulates. The dashed arrow is to be interpreted as "given a revision operator $\circ_{KM}$ satisfying the AGM postulates for revision, there exists an associated faithful assignment $\leq_{KM}$." The rest of the diagram can be read in an analogous form.

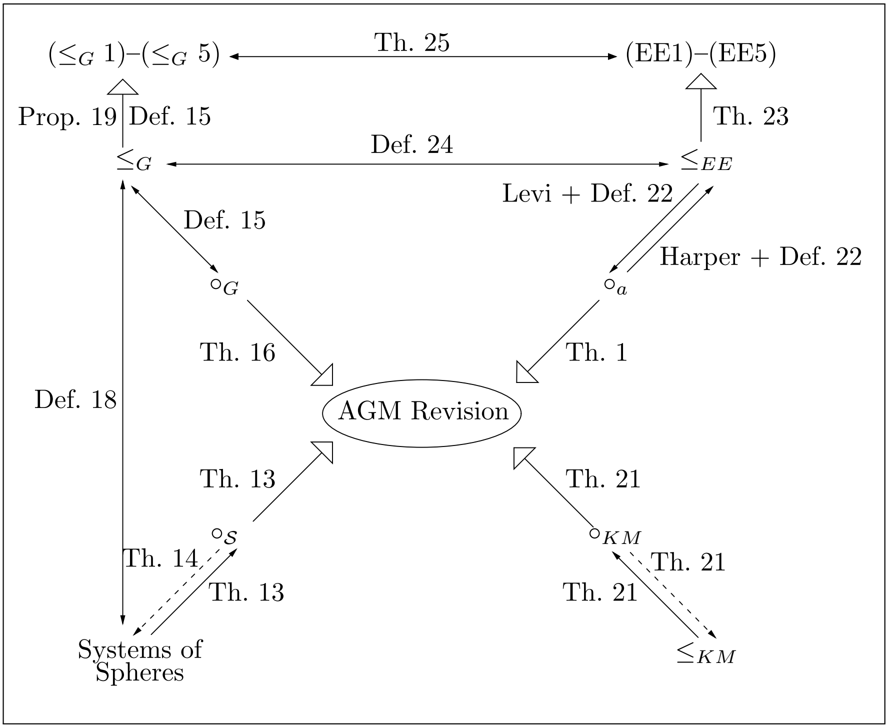


Figure 4. Relationships between different revision formalisms.

3 BELIEF REVISION OPERATORS

Up to now, the ideas presented for the evaluation of change in belief sets were based on general principles a belief change operation should verify or on requirements of an underlying preference relation. However, they did not form the basis for an algorithmic mechanism for computing and/or evaluating change between belief sets. In this section, we present some quantitative and qualitative mechanisms for change measurement. In general, quantitative measurements of change result in total similarity orderings, whereas qualitative ones yield partial orderings. As we have seen in Section 2.4, AGM operations are associated with total similarity measurements and hence quantitative procedures have been used in a number of belief change operations [Dalal, 1988a; Winslett, 1988a; Gabbay and Rodrigues, 1996a].

3.1 Measuring information change

In the case of propositional logic, one possibility of evaluating the degree of change between belief sets is by quantifying the amount of disagreement between the truth-values of propositions in different valuations. This section will describe how this has been used in belief revision.

Let us call the "distance" between two valuations M and N as the number of propositional variables with different truth-values in M and N. This measurement of change was initially proposed in [Dalal, 1988a]. Following [Katsuno and Mendelzon, 1991a], we will assume the language $\mathcal{L}$ of the logic L to be defined from an arbitrarily large but finite set of propositional variables $\mathcal{P}$. We denote the set of all valuations of L by $\mathcal{I}$.

DEFINITION 26. Let M and N be two elements of $\mathcal{I}$. The *distance d between M and N* is the number of propositional variables p_i, for which $M(p_i) \neq N(p_i)$.

The function d satisfies a number of interesting properties [Rodrigues, 1998]:

PROPOSITION 27.

(M1) $d(M,N) \geq 0$

(M2) $d(M,N) = 0$ *iff* $M = N$

(M3) $d(M,N) = d(N,M)$

(M4) $d(M,O) \leq d(M,N) + d(N,O)$

Proof. Let M, N, O be elements of $\mathcal{I}$. (M1) and (M2): Clearly, $d(M, N)$ cannot be negative. If $d(M, N) = 0$, then $M(p_i) = N(p_i)$ for all $p_i \in \mathcal{P}$, and then $M = N$. If $M = N$, then they obviously agree w.r.t. every propositional variable, and hence $d(M, N) = 0$.

(M3) follows directly from Definition 26.

(M4): Let us identify interpretations M, N, O with the sets M_l, N_l and O_l of (positive or negative) literals that they satisfy, respectively. We then define X to be $M_l - O_l$, that is the set of literals l_i such that $l_i \in M_l$ and $l_i \notin O_l$, Y to be $M_l - N_l$ and Z to be $N_l - O_l$. It is easy to see that, for a given propositional variable p_i and interpretation M, either $p_i \in M_l$ or $\neg p_i \in M_l$ (but not both), and that there is a correspondence between the cardinalities of X, Y and Z and the distances $d(M, O)$, $d(M, N)$ and $d(N, O)$, respectively. We show that $|X| \leq |Y| + |Z|$, by proving that $X \subseteq Y \cup Z$.

Suppose $p_i \in X$, but $p_i \notin Y \cup Z$. If $p_i \in X$, then $p_i \in M_l$ and $\neg p_i \in O_l$. Either $p_i \in N_l$ or $\neg p_i \in N_l$. If $p_i \in N_l$, then $p_i \in Z$. On the other hand, if $\neg p_i \in N_l$, then $p_i \in Y$. In either case, $p_i \in Y \cup Z$, a contradiction. ■

Using the distance d to evaluate the degree of change to a belief set

The interesting case in belief revision is when the new belief contradicts the agent's current belief set and hence consistency can only be maintained by giving up some of her old beliefs. This is where the notion of minimal change comes into play, since generally one wants to minimise the loss of information incurred in the process.

Obviously, what constitutes "change" itself varies according to the type of operation being performed. For the case of belief revision, the change is measured with respect to the belief set as a whole. In semantical terms, this can be seen as the minimal distance associated to any pair of models of the belief set and the new belief. It will prove useful to extend the distance d defined above to classes of valuations, taking into account the special case when either of the two classes is empty (i.e., when either the belief set is inconsistent or the new belief is contradictory).

DEFINITION 28. Let $\mathcal{M}$ and $\mathcal{N}$ be two classes of valuations of $\mathcal{L}$. The *set distance D between $\mathcal{M}$ and $\mathcal{N}$*, $D(\mathcal{M}, \mathcal{N})$, is defined as

$$D(\mathcal{M}, \mathcal{N}) = \begin{cases} \inf\{d(M, N) \mid M \in \mathcal{M} \text{ and } N \in \mathcal{N}\} & \Rightarrow \textit{if } \mathcal{M} \textit{ and } \mathcal{N} \textit{ are both nonempty} \\ \infty & \Rightarrow \textit{otherwise} \end{cases}$$

REMARK 29. If the class $\mathcal{M}$ is composed solely by a valuation I, that is, if

$\mathcal{M} = \{I\}$, then D in fact computes the minimum distance between a class of models and a single valuation, i.e., a *world*. This notion is used in the definition of *update* operations by Katsuno and Mendelzon [Katsuno and Mendelzon, 1992].

PROPOSITION 30. *D is symmetric.*

Proof. This follows directly from Definition 28 and Proposition 27, since by (M3), d is symmetric. ■

Intuitively, the greater the distance between two classes of valuations, the more they disagree with respect to the truth-value of propositions. When combining the information represented by these two classes one will try and minimise the distance between any two pairs of models taken from each class. If the new belief φ is consistent with the belief set K, then $\mathrm{mod}\,(K) \cap \mathrm{mod}(\varphi) \neq \emptyset$. Thus, they will have a model M in common. By (M1), $d(M, M) = 0$ (see Proposition 27) and hence $D(\mathrm{mod}(K), \mathrm{mod}(\varphi)) = 0$. In this case, it is possible to combine K and φ without any loss of information to either. On the other extreme, if either the new belief is contradictory or the current belief set is inconsistent, then one of them (or both) is not really informative and we are free to interpret the loss of information in the combination as we wish. In this case, ∞ seems an appropriate value.

Now, if one wants to choose the models of φ which preserve as much of the informational content of K as possible (according to d), the natural solution is to pick those with a minimum distance to any model of K. With this idea in mind, it is possible to define the following similarity ordering on valuations with respect to a belief set K. The only technicality is to decide how to compare valuations when K is inconsistent (i.e., $\mathrm{mod}\,(K) = \emptyset$). We can arbitrarily consider all valuations equivalent in this case (remember from Definition 20, that no constraints on the granularity of non-models of K are actually imposed).

DEFINITION 31. For any $M, N \in \mathcal{I}$ and each propositional formula K, $M \leq_K N$ iff

① $\mathrm{mod}\,(K) = \emptyset$; or

② $\exists I \in \mathrm{mod}\,(K)$ such that $\forall I' \in \mathrm{mod}\,(K)$ $d(M, I) \leq d(N, I')$

That is, M is at least as good as N at preserving the information in K if there is some model I of K such that $d(M, I)$ is less than or equal to the distance between N and any model of K.

PROPOSITION 32. *For each propositional formula K, $\leq_K$ is a pre-order.*

Proof. Reflexivity follows trivially when $\mathrm{mod}(K) = \emptyset$ and to see that condition ② ensures that $M \leq_K M$ for all M when $\mathrm{mod}(K) \neq \emptyset$, just take $I = M$, then $d(M, I) = d(M, M) = 0$ and by (M1) of Proposition 27, this is the minimum value for d. For transitivity, the interesting case is as follows: suppose $M \leq_K N$, and $N \leq_K O$. If $M \leq_K N$, then $\exists I_1 \in \mathrm{mod}(K)$, such that $\forall I' \in \mathrm{mod}(K)$ $d(M, I_1) \leq d(N, I')$. If $N \leq_K O$, then $\exists I_2 \in \mathrm{mod}(K)$ such that $\forall I'' \in \mathrm{mod}(K)$ $d(N, I_2) \leq d(O, I'')$. $I_2 \in \mathrm{mod}(K)$. Thus, in particular, $d(M, I_1) \leq d(N, I_2)$, and hence $M \leq_K O$. ■

Now consider the following function χ which assigns to every propositional formula K a pre-order $\leq_K$ on $\mathcal{I}$ as defined above.

DEFINITION 33. For each formula K of propositional logic $\chi(K) = \leq_K$.

It should not surprise the reader that χ is in fact a faithful assignment for belief revision in Katsuno and Mendelzon's terms.

PROPOSITION 34. *χ is a faithful assignment for belief revision.*

Proof. By Proposition 32, for each propositional formula K, $\leq_K$ is a pre-order. It can be easily seen that $\leq_K$ is also total. We are left to prove:

1. If $M, N \in \mathrm{mod}(K)$, then $M <_K N$ does not hold.

   If $M \in \mathrm{mod}(K)$, then $M \leq_K N$, because $\forall I' \in \mathrm{mod}(K)$ $d(M, M) \leq d(N, I')$. Also, $N \leq_K M$, for the same reason, since $N \in \mathrm{mod}(K)$, and hence $M \equiv_K N$.

2. If $M \in \mathrm{mod}(K)$ and $N \notin \mathrm{mod}(K)$, then $M <_K N$.

   Clearly, $M \leq_K N$ (see previous item). Since $N \notin \mathrm{mod}(K)$, then every model of K will will have at least one disagreement in a propositional letter with respect to N and hence the distance with be a positive natural number greater than zero. However, $d(M, M) = 0$, therefore $\neg\exists I \in \mathrm{mod}(K)$ such that $d(N, I) \leq d(M, M)$.

3. If $K \leftrightarrow \varphi$, then $\leq_K = \leq_\varphi$.

   This follows directly, since d and $\leq_K$ are defined semantically.

■

In [Katsuno and Mendelzon, 1991a; Katsuno and Mendelzon, 1992], Katsuno and Mendelzon used the same distance function d to define a faithful assignment for belief revision and prove that Dalal's revision operator (see Section 3.2) verifies the AGM postulates. The idea above is similar but corrects a few problems. Specifically, they define a total pre-order $\leq_K$ on $\mathcal{I}$ as

$$I \leq_K J \text{ if and only if } \mathrm{dist}(\mathrm{mod}(K), I) \leq \mathrm{dist}(\mathrm{mod}(K), J)$$

where

$$\mathrm{dist}(\mathrm{mod}(K), I) = \min_{J \in \text{ mod } (K)} d(J, I).$$

This definition does not cover all possibilities, as K may be inconsistent and hence $\mathrm{mod}(K) = \emptyset$. As it turns out, a similar problem occurs with Dalal's revision operator when K is inconsistent [Dalal, 1988a].

Now suppose K is indeed consistent and consider the following collection $\mathcal{S}$ of subsets S_0, S_1, …, $S_{|\mathcal{P}|}$ of $M_L^\top$: $S_i = \{m \in M_L^\top \mid D(\mathrm{mod}(m), \mathrm{mod}\ (K)) \leq i\}$.

PROPOSITION 35. *The collection $\mathcal{S}$ defined above is a system of spheres centred on $|K|$ (in Grove's sense).*

Proof. First, notice that it is easy to see that $\mathcal{S}$ is totally ordered. Similarly, $S_0 = \{m \in M_L^\top \mid D(\mathrm{mod}(m), \mathrm{mod}(K) = 0\}$, and then $K \subseteq m$, for all $m \in S_0$, and hence $S_0 = |K|$. Since D is based on d and the minimum value for d is 0, then $|K|$ is the minimum in $\mathcal{S}$. The maximum number of propositional variables any two valuations in $\mathcal{I}$ can differ is $|\mathcal{P}|$ (since we consider a finite number of propositions), therefore $S_{|P|}$ will contain all maximal consistent extensions of $\mathcal{L}$. Condition (S4) is trivially satisfied since there are only finitely many spheres in $\mathcal{S}$. ■

If we consider valuations in each sphere instead of their corresponding theories, we will have a semantical characterisation as proposed by Katsuno and Mendelzon. The innermost sphere contains all of the models of K, the next sphere extends the former to include all valuations that differ from any model of K by at most the truth-value of one propositional variable, and so forth (see Figure 5). This inclusion of valuations proceeds until *all* valuations are eventually included (in Grove's sense, this corresponds to $M_L^\top$ — the $\subseteq$-maximum in $\mathcal{S}$ — see Definition 10).

Intuitively, the valuations in the innermost sphere are the ones that most preserve the informational content of K (according to d). In fact, any valuation in the innermost sphere preserves all of K's informational content, since these are K's *models*. The farther away from the centre, the bigger the loss of information, because other valuations which do not satisfy K are included as well. Provided a formula φ is non-contradictory, there will be some non-empty intersection between the models of φ and some sphere

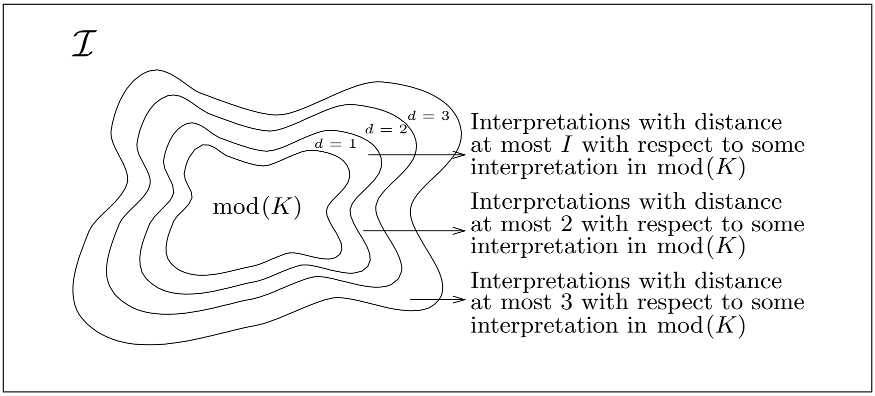


Figure 5. A system of spheres around mod(K).

around $|K|$. A revision of K by φ in semantical terms corresponds to finding exactly those models of φ that are as close to the centre, i.e., $|K|$, as possible.

We now give a concrete example of these spheres for a propositional formula $p \wedge q$ according to the similarity measurement d. The centre of the system of spheres will be constituted by mod$(p \wedge q)$, the next sphere will contain these and all valuations that differ from them with respect to at most the truth-value of either p or q (i.e., the models of $p \vee q$) and the last sphere will contain all valuations of $\mathcal{L}$. This is depicted in Figure 6. The intuitive meaning can be grasped as follows: if an agent believes in $p \wedge q$[10] and she is told that it is not the case that both p and q are true, then the next best thing is to think that at most one of p or q is false, hence $p \vee q$. Subsequently, any further loss of beliefs requires that both p and q be false. These three possibilities include all possible valuations. Of course, the similarity measurement provided by d is one of the simplest possible. It is presented here mainly for historical reasons, but also for the easy visualisation of the revision process it provides.

Qualitative measurements of change

The function d seen before provides a quantitative measure of the degree of similarity between valuations and classes of valuations. One advantage of this is that it leads directly to the definition of a total ordering on $\mathcal{I}$.

Even though the distance d provides an indication of the magnitude of the difference between two valuations, it has little to say about the qualitative

[10]Note that AGM is coherentist and hence believing in $p \wedge q$ and believing in p *and* believing in q (or any variation of beliefs implying $p \wedge q$) amounts to exactly the same.

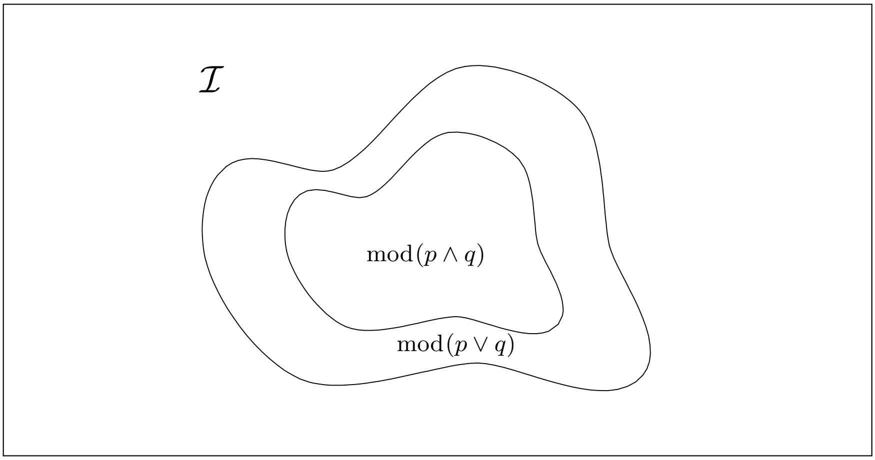


Figure 6. A system of spheres around mod$(p \wedge q)$ based on $\leq_{p \wedge q}$.

value of the change in information itself.

In this section, we consider another way of comparing valuations by analysing the sets of propositional variables which differ with respect to the truth-values assigned by these valuations. The function presented here was initially proposed by Borgida [Borgida, 1985].

DEFINITION 36. Let M and N be two elements of $\mathcal{I}$. The *difference set between M and N*, in symbols, $\mathit{diff}(M, N)$ is the set of propositional variables p_i, for which $M(p_i) \neq N(p_i)$.

It is easy to see that *diff* is a symmetric function. That is, $\mathit{diff}(M, N) = \mathit{diff}(N, M)$. Also notice that for any valuations M and N, $|\mathit{diff}(M, N)| = d(M, N)$.

The function *diff* was used in a number of formalisms both for belief revision and updates, including Borgida's revision operator, extended in [Dalal, 1988b], Winslett's *possible models approach* [Winslett, 1988a], Weber's revision operator [Weber, 1987] and, to some extent, in Satoh's formalism too [Satoh, 1988].

If we want to define closeness with respect to a given valuation I via the function *diff* presented above, in symbols $\sqsubseteq_I$, it can be done in the following way:

DEFINITION 37 (Closeness of valuations using the function *diff*).

$$M \sqsubseteq_I N \text{ iff } \mathit{diff}(M, I) \subseteq \mathit{diff}(N, I)$$

PROPOSITION 38. $\sqsubseteq_I$ *is a pre-order.*

Proof. It is straightforward to see that $\sqsubseteq_I$ is reflexive. For transitivity, suppose that $M \sqsubseteq_I N$. It follows that $\mathit{diff}(M, I) \subseteq \mathit{diff}(N, I)$. Similarly, if $N \sqsubseteq_I O$, then $\mathit{diff}(N, I) \subseteq \mathit{diff}(O, I)$. It follows that $\mathit{diff}(M, I) \subseteq \mathit{diff}(O, I)$, and hence $M \sqsubseteq_I O$. ■

The reader can easily check that $\sqsubseteq_I$ is also antisymmetric. However, unlike $\leq_I$, $\sqsubseteq_I$ is not a total ordering on $\mathcal{I}$, as illustrated below.

EXAMPLE 39. Consider three valuations I, M and N, such that $I(p) = M(p) = \mathit{true}$, $I(q) = N(q) = \mathit{true}$ and $M(q) = N(p) = \mathit{false}$. I, M and N agree in the truth values of all other propositional variables.

It follows that $\mathit{diff}(M, I) = \{q\}$ and $\mathit{diff}(N, I) = \{p\}$. Neither $\mathit{diff}(M, I) \subseteq \mathit{diff}(N, I)$, nor $\mathit{diff}(N, I) \subseteq \mathit{diff}(M, I)$, and therefore M and N are incomparable with respect to $\sqsubseteq_I$.

Using the function *diff* directly to evaluate similarity between valuations according to Definition 37 has some implications for belief revision. The example above demonstrates that $\sqsubseteq_I$ defines a partial order on $\mathcal{I}$. In [Katsuno and Mendelzon, 1991b, Theorem 5.2], Katsuno and Mendelzon showed that partial pre-orders on $\mathcal{I}$ can be associated with revision operators which only verify a weaker version of the AGM postulates. The weakening applies basically to postulate (R6) which is dropped in favour of the rules below:

(R6′) If $K \circ_a \varphi \to \psi$ and $K \circ_a \psi \to \varphi$, then $K \circ_a \varphi \leftrightarrow K \circ_a \psi$
(R6″) $(K \circ_a \varphi) \wedge (K \circ_a \psi) \to K \circ_a (\varphi \vee \psi)$

Note that (R6′) and (R6″) are very similar to the following *update* postulates below, where $K \diamond \varphi$ denotes the update of K by φ.

(U6) If $K \diamond \varphi \to \psi$ and $K \diamond \psi \to \varphi$, then $K \diamond \varphi \leftrightarrow K \diamond \psi$
(U7) If K is complete, then $(K \diamond \varphi) \wedge (K \diamond \psi) \to K \diamond (\varphi \vee \psi)$

(R6′) and (U6) are identical. In (U7), the requirement that K is complete is to make sure that there is just one valuation satisfying it, if any. In fact, the completeness requirement in (U7) is related to the requirement of a language with a finite number of propositional variables in Katsuno and Mendelzon's formulation.

Not surprisingly, all of the early revision operators defined via *diff* [Borgida, 1985; Satoh, 1988; Weber, 1987] fail to verify (R6), even though they are shown to verify weaker versions of it. It must be pointed out that in the early 80s, there was still some confusion about the appropriate roles of the operations of belief revision and updates in common-sense reasoning.

Thus, some of the "revision" postulates cited above in fact could be seen as *update* ones. An important contribution in the clarification of the roles of these two operations came in 1992 when Katsuno and Mendelzon published a follow-up to their 1991 paper [Katsuno and Mendelzon, 1991a; Katsuno and Mendelzon, 1992].

The function *diff* can, however, be used as a measurement of change for update operations, since the similarity orderings required for that kind of operation need not be total. In fact, Rodrigues showed that $\sqsubseteq_I$ can be used in the definition of a so-called *faithful assignment for updates.* The reader is referred to [Rodrigues, 1998] for further details on this.

3.2 Dalal's revision operator

As we discussed previously, some of the early revision operators proposed in the literature turned out to actually be *update* operators. The confusion arose because at the time work in belief revision started to emerge, there was some expectation that work initially done to model updates in databases could fit within the belief revision perspective. Later on, it was made clear that, in spite of many similarities, the two kinds of theory change have some fundamentally distinct characteristics: in a belief revision, an agent sees the world as a *static* entity about which she has incomplete or inaccurate information; whereas in an update the agent wants to bring up to date an internal representation of the world after it changed (e.g., by the execution of an action).

Thus, it is not reasonable to classify Borgida's "revision" operation [Borgida, 1985] as a belief revision operation, since it was first devised to model the handling of exceptions in database systems. The same criticism equally applies to Weber's revision operator [Weber, 1987]. In this section we will present Dalal's revision operator, one of the earliest operators known to verify the AGM postulates for belief revision [Dalal, 1988a]. A full description of Dalal's formalism can be found in [Dalal, 1988b; Dalal, 1988a].

Dalal starts by defining systems of spheres around valuations of the given logic.

DEFINITION 40 ([Dalal, 1988a]). For a given $I \in \mathcal{I}$,

$$g(I) = \{J \in \mathcal{I} \mid d(I, J) \leq 1\}$$

And then the definition is extended to classes of interpretations:

DEFINITION 41 ([Dalal, 1988a]). Let $\mathcal{M}$ be a class of interpretations

$$g(\mathcal{M}) = \bigcup_{I \in \mathcal{M}} g(I)$$

For a formula ψ, $G(\psi)$ is defined in terms of its models as

$$\mathrm{mod}(G(\psi)) = g(\mathrm{mod}(\psi))$$

DEFINITION 42 ([Dalal, 1988b, Definition 4.4]). $G^k(\psi)(k \geq 0)$ is defined recursively as follows:

$$G^0(\psi) = \psi$$
$$G^k(\psi) = G(G^{k-1}(\psi))$$

The revision operator is then defined semantically as

DEFINITION 43 ([Dalal, 1988b, Definition 4.3]). Let ψ and φ be propositional formulae. The revision of ψ by φ, $\psi \circ_d \varphi$, is defined as:

$$\mathrm{mod}(\psi \circ_d \varphi) = \mathrm{mod}(G^k(\psi)) \cap \mathrm{mod}(\varphi)$$

where k is the least value for which $\mathrm{mod}(G^k(\psi)) \cap \mathrm{mod}(\varphi) \neq \emptyset$.

By using the result below given in [Weber, 1987], Dalal provides a way to compute $\psi \circ_d \varphi$ syntactically.

LEMMA 44. *Let ψ be a formula and p a propositional symbol. There exists formulae ψ_p^+ and ψ_p^- such that*

- *ψ_p^+ and ψ_p^- do not contain p; and*
- *$\psi \equiv (p \wedge \psi_p^+) \vee (\neg p \wedge \psi_p^-)$*

According to Weber [Weber, 1987], ψ_p^+ and ψ_p^- can be obtained by replacing the propositional variable p in ψ by $\top$ and $\bot$ respectively. These constants can then be eliminated through standard simplifications of classical logic. What Lemma 44 actually does is to "isolate" the symbol p from the formula ψ. The next two definitions originally appeared in [Weber, 1987].

DEFINITION 45. Let ψ be a formula which may contain the propositional variable p. If ψ_p^+ and ψ_p^- are formulae obtained according to Lemma 44, then

$$res_p(\psi) = \psi_p^+ \vee \psi_p^-$$

is called the *resolvent of ψ with respect to p.*

Intuitively, ψ and $res_p(\psi)$ admit change with respect to the truth-value of p only. That is, the models of $res_p(\psi)$ are the models of ψ plus the interpretations which differ from them with respect to at most the truth-value of p.

The definition of res is extended to a set of symbols:

DEFINITION 46. If $\{p_1, p_2, \ldots, p_k\}$ is a set of symbols and ψ is a propositional formula, then

$$\begin{aligned} &res_{\emptyset}(\psi) = \psi \\ &res_{\{p_1,p_2,\ldots,p_k\}}(\psi) = res_{\{p_2,\ldots,p_k\}}(res_{p_1}(\psi)) \end{aligned}$$

Weber also proved that the order of the propositional variables chosen in the definition above is not important, so one can pick any desired enumeration.

THEOREM 47 ([Dalal, 1988b, Theorem 5.5]). *Let ψ be a formula and* $\mathrm{var}(\psi) = \{p_1, p_2, \ldots, p_k\}$.

$$G(\psi) = res_{p_1}(\psi) \vee \ldots \vee res_{p_k}(\psi)$$

We are now in a position to define $\psi \circ_d \varphi$ syntactically:

DEFINITION 48 ([Dalal, 1988a]). Let ψ and φ be two formulae of propositional logic. The revision of ψ by φ, $\psi \circ_d \varphi$, is defined as:

$$\psi \circ_d \varphi = G^k(\psi) \wedge \varphi$$

where k is the least value of i for which $\mathrm{mod}(G^i(\psi)) \cap \mathrm{mod}(\varphi) \neq \emptyset$.

Notice that the definitions above only work for non-contradictory formulae ψ.

PROPOSITION 49. *If ψ is contradictory, then so is $G^k(\psi)$, for any k.*

Proof. First notice that $g^k(\emptyset) = \emptyset$, for any k (by Definition 41). If ψ is contradictory, then $\mathrm{mod}(\psi) = \emptyset$. By Definitions 41 and 42, $\mathrm{mod}(G^k(\psi)) = g^k(\mathrm{mod}(\psi)) = \emptyset$. ■

As a result, $\circ_d$ as defined above cannot always verify the following postulate:[11]

(R3) If A is satisfiable, then $K \circ_a A$ is also satisfiable

By making a special provision for when ψ is contradictory, $\circ_d$ can indeed satisfy all of the AGM postulates for belief revision. It is sufficient to define

[11] In order to see this, consider the revision of any contradictory ψ by any non-contradictory φ.

$\psi \circ_d \varphi = \varphi$ if ψ is contradictory. A proof overlooking the extreme case above was given by Katsuno and Mendelzon. The reader is referred to [Katsuno and Mendelzon, 1991a; Katsuno and Mendelzon, 1992] for more details. A simpler and equivalent revision operation defined in terms of formulae written in disjunctive normal form (DNF) was proposed by Rodrigues in [Gabbay and Rodrigues, 1996a]. It is convenient to regard each disjunct in a DNF formula as an easy to manipulate syntactical representation of a class of models of that formula (the set of all models of the formula will correspond to the union of all such classes). Measuring distance between classes of models amounts to counting literals with different signs in a pair of disjuncts and conflict can be solved by superimposition of the stronger disjunct (coming from the new information) over the weaker one (coming from the old belief base).

4 ITERATION OF THE REVISION PROCESS

One of the main criticisms against the AGM framework is that it gives little guidance with respect to how future belief change operations should interact with the one currently taking place. Postulates (K°7) and (K°8) give some guidelines about revisions followed by expansions, but say nothing about revisions followed by revisions. Many formalisms for belief revision use extraneous mechanisms for deciding what beliefs to keep and this makes it harder to iterate the process. Such mechanisms include the selection functions used in partial meet revisions and partial meet contractions. The main criticism against these is that whereas one starts with a belief set and a selection function, the revision operation determines what the new belief set should be, but says nothing about how to update the selection function.

Postulates (K°7) and (K°8) only constrain the interaction between revisions and *expansions*:

(K°7) $K \circ_a (\varphi \wedge \psi) \subseteq \mathrm{Cn}((K \circ_a \varphi) \cup \{\psi\})$
(K°8) If $\neg\psi \notin K \circ_a \varphi$, then $\mathrm{Cn}(K \circ_a \varphi \cup \{\psi\}) \subseteq K \circ_a (\varphi \wedge \psi)$

The constraint about interaction between successive *revisions* is limited to the case of a revision $K \circ_a \varphi$ followed by a revision by ψ when ψ is consistent with $K \circ_a \varphi$. We have seen that in this case, (K°3) and (K°4) on the one hand and (K°7) and (K°8) on the other amount to the following conditions

($\mathrm{K}^\circ_{3,4}$) If φ is consistent with K, then $K \circ_a \varphi = \mathrm{Cn}(K \cup \{\varphi\})$.
($\mathrm{K}^\circ_{7,8}$) If ψ is consistent with $K \circ_a \varphi$, then $\mathrm{Cn}((K \circ_a \varphi) \cup \{\psi\}) = K \circ_a (\varphi \wedge \psi)$

Since ψ is consistent with $K \circ_a \varphi$, $(\mathrm{K}^\circ_{3,4})$ dictates that $(K \circ_a \varphi) \circ_a \psi = \mathrm{Cn}((K \circ_a \varphi) \cup \{\psi\})$. On the other hand, $(\mathrm{K}^\circ_{7,8})$ requires that $\mathrm{Cn}((K \circ_a \varphi) \cup \{\psi\}) = K \circ_a (\varphi \wedge \psi)$, and hence $(K \circ_a \varphi) \circ_a \psi = K \circ_a (\varphi \wedge \psi)$.[12]

Even though the postulates do capture the intuitions behind *rational changes of belief*, many authors seem to agree that seeing the agent's *corpus of beliefs* as a single coherent unit (a sentence in the finite case or a theory, otherwise) poses some problems and have thus sought to enrich the structure of a belief set in order to model the iteration of the operation [Nebel, 1991b; Nebel, 1992; Ryan, 1992; Rodrigues, 1998; Gabbay and Rodrigues, 1996b; Gabbay *et al.*, 2003]. We motivate the need for this richer structure next, before presenting a number of formalisms that deal specifically with the problem of iterated revision.

4.1 The problem of iteration and the need for extralogical information to guide the process

In the introduction to this chapter we have briefly mentioned the differences between the coherence and foundational approaches to belief revision. In this section we will motivate the importance of having a more structured representation of the beliefs of an agent to allow for truly rational changes of belief, and in particular when these changes occur over a period of time, which is the main concern of *iterated belief revision*.

Let us start with an example. Suppose the current belief state of an agent includes the belief that `tweety is a penguin`, and let us represent this belief by the formula

(1) pt

In addition, assume that the agent also believes that `penguins cannot fly`. One instance of this general rule is that `if tweety is a penguin, it does not fly`, which can be represented by the formula

(2) $pt \to \neg ft$

The two beliefs together imply the conjunction $pt \wedge \neg ft$:

$$\frac{pt \quad \wedge \quad (pt \to \neg ft)}{pt \wedge \neg ft}$$

Intuitively speaking, faced with the information that tweety can in fact fly (i.e., ft), there would be only two possibilities for the agent to reconciliate this information with her previous beliefs:

[12] This property is verified by the iterated revision mechanism proposed by Boutilier [Boutilier, 1996] (see (CB2) in Section 4.4).

① either by rejecting pt, that is,
`Since penguins cannot fly, tweety must not be a penguin`

② or by rejecting the fact that penguins cannot fly
`Tweety is a penguin and it does fly. Thus, some penguins` *`can`* `indeed fly`

However, remember that the AGM theory is *coherentist*. The beliefs (1) and (2) will reside side by side in the agent's belief set with an infinite number of other beliefs (all equally regarded) including, for instance, $pt \vee \neg ft$, $ft \to \neg pt$, $pt \wedge \neg ft$, $\neg ft$, etc. As it turns out, (1) and (2) and $\neg ft$ are all accepted, but the information that $\neg ft$ is there *only* as a consequence of (1) and (2) is lost. Notice that all models of pt and $pt \to \neg ft$ are also models of $pt \wedge \neg ft$, so any revision operation based on a simple evaluation of change such as the one provided by the distance d will keep the belief in pt and reject the belief in $\neg ft$, resulting in $pt \wedge ft$ [Gabbay and Rodrigues, 1996a; Dalal, 1988a].

The use of a belief base would improve matters, since it would allows us to distinguish between the beliefs $pt \to \neg ft$ and $\neg ft$ (the latter being in the belief set, but not in the belief base). However, it would not be enough to solve the problem of choosing between ① and ②. This secondary issue can be addressed by preference orderings, such as the epistemic entrenchment relations seen in Section 2.5, but this is again a mechanism *external* to the belief set and does not deal explicitly with iteration.

Not surprisingly, iterated revision is usually modelled by mechanisms that explicitly distinguish between an agent's belief state and his/her *epistemic state* [Darwiche and Pearl, 1994; Lehmann, 1995; Boutilier, 1996; Konieczny and Pérez, 2000]; the epistemic state containing the additional structural information needed to support the rational iteration of the revision process.

From the axiomatic point of view, the basic ideas consist of either augmenting the original AGM formulation with new postulates, such as (C1)–(C4) proposed by Darwiche and Pearl [Darwiche and Pearl, 1994; Darwiche and Pearl, 1996; Darwiche and Pearl, 1997] and presented here later, or by proposing entirely new formulations based on the original AGM ideas [Lehmann, 1995].

We shall start by presenting and discussing Darwiche and Pearl's approach.

4.2 Darwiche and Pearl's approach

Darwiche and Pearl's approach [Darwiche and Pearl, 1994] was based on the formulation of some postulates to constrain the expected behaviour of iterated revisions. They pointed out that a distinction between belief states and epistemic states was essential in the iteration process, although a reformulation of the AGM postulates in order to reflect this only appeared later in [Darwiche and Pearl, 1996] and [Darwiche and Pearl, 1997]. The main idea was to consider belief sets as being obtained from the agent's *epistemic states*, which possess a richer structure. An immediate consequence of this change in paradigm is that the equivalence of epistemic states cannot be derived from the equivalence between belief sets. Under this approach, two equivalent belief sets may be reached from two completely different epistemic states. This departure from coherentism has been supported by other authors [Friedman and Halpern, 1996; Lehmann, 1995; Rodrigues, 1998].

The starting point is again the assumption of a finite set of beliefs from which the belief set is derived. In Katsuno and Mendelzon's terms, an epistemic state would be associated with a belief base, whereas the belief set would be obtained by closing the base under the consequence relation. Since an epistemic state carries more information than a belief set, revisions must take into account differences arising from the distinction. The original postulates for revision for the finite case (R1)–(R6) were reformulated with this in mind and motivated by the observation that they were incompatible with the new set of proposed postulates to deal explicitly with the iteration of the revision process ((C1)–(C4)) (see [Freund and Lehmann, 1994]).

In the following presentation, Ψ will be used to represent an epistemic state and bel(Ψ) to represent the belief set obtained from Ψ. However, in order to lighten the notation and where the context is clear, we follow Darwiche and Pearl and use Ψ instead of bel(Ψ). By this we mean that, for instance, in (R*4) below, it is the epistemic states Ψ_1 and Ψ_2 that are meant in the first half of the postulate, but the belief sets bel($\Psi_1 \circ \varphi_1$) and bel($\Psi_2 \circ \varphi_2$) in the second one.

Darwiche and Pearl's postulates for belief revision of epistemic states

(R*1) $\Psi \circ \varphi$ implies φ
(R*2) If $\Psi \wedge \varphi$ is satisfiable, then $\Psi \circ \varphi \equiv \Psi \wedge \varphi$
(R*3) If φ is satisfiable, then $\Psi \circ \varphi$ is also satisfiable
(R*4) If $\Psi_1 = \Psi_2$ and $\varphi_1 \equiv \varphi_2$, then $\Psi_1 \circ \varphi_1 \equiv \Psi_2 \circ \varphi_2$
(R*5) $(\Psi \circ \varphi) \wedge \psi$ implies $\Psi \circ (\varphi \wedge \psi)$
(R*6) If $(\Psi \circ \varphi) \wedge \psi$ is satisfiable, then $\Psi \circ (\varphi \wedge \psi)$ implies $(\Psi \circ \varphi) \wedge \psi$

The above presentation is essentially the same as Katsuno and Mendelzon's, except for ($R^\star 4$) which is strictly weaker than (R4). In ($R^\star 4$), the condition for the equivalence of the resulting belief sets is that the original epistemic states are *identical*, instead of equivalent as in (R4). This is reflected immediately in the semantical characterisation of revision operators satisfying ($R^\star 1$)–($R^\star 6$) given in [Darwiche and Pearl, 1997].

Notation 2. The models of an epistemic state Ψ will be denoted by $\mathrm{Mod}(\Psi)$. By this we mean that $\mathrm{Mod}(\Psi) = \mathrm{mod}(\mathrm{bel}(\Psi))$.

In the formulation below, $I \preceq_\Psi I'$ represents the fact that I is at least as good at satisfying Ψ as I' is.

DEFINITION 50 (Faithful assignment for revision of epistemic states). A *faithful assignment for belief revision of epistemic states* is a function mapping each epistemic state Ψ to a total pre-order $\preceq_\Psi$ on $\mathcal{I}$, such that

1. If $I, I' \in \mathrm{Mod}(\Psi)$, then $I \equiv_\Psi I'$.
2. If $I \in \mathrm{Mod}(\Psi)$ and $I' \notin \mathrm{Mod}(\Psi)$, then $I \prec_\Psi I'$ holds.
3. If $\Psi = \Delta$, then $\preceq_\Psi = \preceq_\Delta$.

Obviously, $I \in \mathrm{Mod}(\Psi)$ if and only if $I \Vdash \mathrm{bel}(\Psi)$. Note that condition 3 above requires that the two epistemic states are *identical.* Darwiche and Pearl proved the following theorem, which is the counterpart of Theorem 21 for epistemic states:

THEOREM 51. *A revision operator* $\circ$ *satisfies postulates* ($R^\star 1$)—($R^\star 6$) *if there exists a faithful assignment that maps each epistemic state* Ψ *to a total pre-order* $\preceq_\Psi$*, such that*

$$\mathrm{Mod}(\Psi \circ \varphi) = \min_{\preceq_\Psi}(\mathrm{mod}(\varphi)).$$

($R^\star 1$)–($R^\star 6$) were then augmented with a new set of postulates dealing specifically with the iteration of the revision process.

(C1) If $\varphi \vDash \psi$, then $(\Psi \circ \psi) \circ \varphi \equiv \Psi \circ \varphi$
(C2) If $\varphi \vDash \neg\psi$, then $(\Psi \circ \psi) \circ \varphi \equiv \Psi \circ \varphi$
(C3) If $\Psi \circ \varphi \vDash \psi$, then $(\Psi \circ \psi) \circ \varphi \vDash \psi$
(C4) If $\Psi \circ \varphi \nvDash \neg\psi$, then $(\Psi \circ \psi) \circ \varphi \nvDash \neg\psi$

The meaning of the postulates above can be described as follows. (C1) says that revising an epistemic state Ψ by some information ψ and then revising it again by some more specific information φ is the same as revising Ψ by φ only. In [Lehmann, 1995], Lehmann showed that (C1) together with the AGM postulates imply (C3) and (C4). However, this is only the case when the distinction between epistemic states and belief sets is ignored. The importance in this distinction was shown by Darwiche and Pearl in [Darwiche and Pearl, 1997, Theorem 15], where they provided the reformulation of the AGM postulates in terms of epistemic states seen above, i.e., $(\mathrm{R}^{\star}1)$–$(\mathrm{R}^{\star}6)$.

Notice the subtle difference between (C1) and (C2): in (C1) the new information is just a specialisation of some information learned previously: φ gives more detail about some information ψ received previously. In (C2), however, the new information φ *contradicts* something that had just been learnt (ψ). Herzig et al. have discussed the plausibility of (C1), by showing that if an AGM revision operator satisfies (C1), then any non-trivial revision (i.e., a revision that does not correspond to a single expansion) can be described in terms of the new belief only, that is, without taking into account the previous belief set [Herzig *et al.*, 2003].

(C2) is one of the most controversial postulates. It says that if an agent learns ψ first and then she is given some information that contradicts this evidence (i.e., φ), then she should completely ignore the information conveyed by ψ. In [Freund and Lehmann, 1994], Freund and Lehmann showed that (C2) was incompatible with the original AGM postulates, but again this is only the case if one ignores the distinction between epistemic states and belief sets. However, (C2) and other variants seem to undermine the principle of minimal change. It does not seem reasonable to *always* discard the information conveyed by ψ completely simply because of the arrival of the contradictory information φ. For instance, suppose ψ is $p \wedge q$ and φ is $\neg p \vee \neg q$. Since $\neg p \vee \neg q \vDash \neg(p \wedge q)$ (C2) applies, but the question remains as to whether an agent should completely discard the information conveyed by $p \wedge q$ in face of $\neg p \vee \neg q$. After all, the belief in $p \leftrightarrow \neg q$ is compatible with $\neg p \vee \neg q$ and keeps some of the informational content of $p \wedge q$. Compliance with (C2) requires an agent to completely ignore the fact that she ever believed in p and in q. The example by which Darwiche and Pearl justify the plausibility of (C2) is arguable [Darwiche and Pearl, 1997, page 12]. In the example, an agent's current epistemic state includes the formula $\mathit{smart} \wedge \mathit{rich}$, representing the belief that `lady` X `is smart and she is also rich`. The epistemic state is then revised by $\neg \mathit{smart}$ and, subsequently, by smart. Since $\mathit{smart} \wedge \mathit{rich}$ is followed by some information that contradicts this observation, namely, $\neg \mathit{smart}$, (C2) applies and requires that the resulting epistemic state is equivalent to the initial state revised by the second observation only. In other words, the intermediate observation should be disregarded:

$$\left(\begin{array}{c}\left(\begin{array}{c}\text{smart}\wedge\text{rich}\\ \uparrow\\ \neg\text{smart}\end{array}\right)\\ \uparrow\\ \text{smart}\end{array}\right) \equiv \left(\begin{array}{c}\text{smart}\wedge\text{rich}\\ \uparrow\\ \text{smart}\end{array}\right)$$

However, the reason why (C2) seems reasonable in this example is because nothing of the informational content of the first revising sentence (i.e., $\neg smart$) can be kept after the second revision (by $smart$) is performed. After all, what could be consistently kept from "$\neg smart$" in the face of "$smart$"?

However, consider the following modified scenario. We start with an initially empty epistemic state which is to be revised by $smart \wedge rich$ and then by $\neg smart$:

$$\left(\begin{array}{c}\left(\begin{array}{c}\top\\ \uparrow\\ \text{smart}\wedge\text{rich}\end{array}\right)\\ \uparrow\\ \neg\text{smart}\end{array}\right) \equiv \left(\begin{array}{c}\top\\ \uparrow\\ \neg\text{smart}\end{array}\right)$$

The postulate applies here too, because $\neg smart \vDash \neg(smart \wedge rich)$. However, if one adopts the principle of minimal change, it seems counterintuitive to give up the belief in "$rich$" just because of the arrival of "$\neg smart$". It might have been the case that the observation with respect to lady X's being smart was wrong, but this does not necessarily mean that the belief in her being rich was inaccurate too. The acceptance of (C2) requires this though.

We now continue with the analysis of the remaining postulates. (C3) says that if after revising the epistemic state Ψ by φ, the agent holds the belief in ψ, then this belief should also hold in the epistemic state obtained after revising Ψ by ψ and then by φ.

(C4) is the negative counterpart of (C3). It says that if after revising Ψ by φ the agent does not believe that $\neg\psi$ holds, then it should not believe that $\neg\psi$ holds after revising Ψ by ψ and then by φ.

Taking into account a *temporal* perspective of the iterated revision process. Rodrigues [Rodrigues, 1998] advocated for the following stronger condition (where $M \prec_\Delta N$ means M is strictly better than N at satisfying Δ):

(CR$\star$) If $M \in \text{mod}(\beta)$ and $N \in \text{mod}(\neg\beta)$, then $M \prec_{\Gamma\circ\beta} N$.

The motivation for (CR$\star$) comes from the following observation: when Γ is revised by β, the new top priority with respect to the newly generated epistemic state $\Gamma\circ\beta$ is to satisfy β (if possible at all). If valuation M satisfies β, whereas N does not, it should not matter how M and N related before with respect to $\preceq_\Gamma$. In the new ordering $\preceq_{\Gamma\circ\beta}$, M should be preferred to N. Of course, this is biased by the AGM perception that the more recent some information is, the better it is regarded. A similarly flavoured behaviour of the iteration of the revision process was proposed by Boutilier [Boutilier, 1996] (see Section 4.4).

If Ψ is the current epistemic state and φ is used to revise Ψ, then the ordering $\preceq_\Psi$ must relate to $\preceq_{\Psi\circ\varphi}$ appropriately in order for (C1)–(C4) to hold. The following representation theorem by Darwiche and Pearl [Darwiche and Pearl, 1997] dictates the relationship between the faithful assignments of two consecutive epistemic states.

THEOREM 52. *If a given revision operator* $\circ$ *satisfies postulates* (R*1)—(R*6)*, then* $\circ$ *satisfies* (C1)—(C4) *if and only if the operator and its corresponding faithful assignment satisfy:*

(CR1) *If* $M, N \in \text{mod}(\varphi)$*, then* $M \preceq_\Psi N$ *if and only if* $M \preceq_{\Psi\circ\varphi} N$.

(CR2) *If* $M, N \in \text{mod}(\neg\varphi)$*, then* $M \preceq_\Psi N$ *if and only if* $M \preceq_{\Psi\circ\varphi} N$.

(CR3) *If* $M \in \text{mod}(\varphi)$ *and* $N \in \text{mod}(\neg\varphi)$*, then* $M \prec_\Psi N$ *implies* $M \prec_{\Psi\circ\varphi} N$.

(CR4) *If* $M \in \text{mod}(\varphi)$ *and* $N \in \text{mod}(\neg\varphi)$*, then* $M \preceq_\Psi N$ *implies* $M \preceq_{\Psi\circ\varphi} N$.

(CR1) states that the relationship between models of a formula φ should not change with respect to an epistemic state before and after a revision of this state by φ. (CR2) says that whenever two valuations M and N do not satisfy a formula φ, their relationship with respect to satisfying the previous epistemic state is transferred to the epistemic ordering of the new epistemic state after the revision by φ. (CR3) and (CR4) can be seen in the following way: (CR4) "preserves" the preference in favour of a model of a formula φ against a non-model of φ after a revision by φ, whereas (CR3) "preserves" the non-preference of non-models of φ against models of φ.

4.3 Lehmann's approach: belief revision, revised

Lehmann also argued that the AGM postulates in their original belief set interpretation were incompatible with some desired properties of the iteration process (including postulate (C1) above) and advocated for a complete reformulation of the revision postulates [Lehmann, 1995]. This led to the definition of the new set of postulates (I1)–(I7).

For Lehmann, an epistemic state corresponds to a finite sequence of revisions by non-contradictory formulae. The corresponding belief set is obtained by the successive application of a particular revision procedure, which is denoted by $\cdot$. For instance, $[\alpha]$ represents the initial epistemic state, obtained by the initial acceptance of α. Subsequent revisions are represented by the concatenation of formulae to the end of the list. Thus, $[\alpha \cdot \beta]$ represents the epistemic state obtained from the revision of $[\alpha]$ by β, etc.

The new set of postulates proposed by Lehmann is given below:

Lehmann's postulates for iterated belief revision

(I1) $[\alpha]$ is a consistent theory
(I2) $\alpha \in [\sigma \cdot \alpha]$
(I3) If $\beta \in [\sigma \cdot \alpha]$, then $\alpha \to \beta \in [\sigma]$
(I4) If $\alpha \in [\sigma]$, then $[\sigma \cdot \gamma] = [\sigma \cdot \alpha \cdot \gamma]$
(I5) If $\beta \models \alpha$, then $[\sigma \cdot \alpha \cdot \beta \cdot \gamma] = [\sigma \cdot \beta \cdot \gamma]$
(I6) If $\neg\beta \notin [\sigma \cdot \alpha]$, then $[\sigma \cdot \alpha \cdot \beta \cdot \gamma] = [\sigma \cdot \alpha \cdot \alpha \wedge \beta \cdot \gamma]$
(I7) $[\sigma \cdot \neg\beta \cdot \beta] \subseteq \mathrm{Cn}([\sigma], \beta)$

The only postulates that actually add new properties to the original AGM presentation are (I5) and (I7). The others are simply a reformulation of the original postulates in terms of the new representation. It is worth emphasising though, that as for Darwiche and Pearl, the notion of an epistemic state plays a fundamental role in the postulates. The condition below, which is a consequence of (K°6), does not follow from (I1)–(I7):

(IS) If $[\sigma] = [\gamma]$, then $[\sigma \cdot \alpha] = [\gamma \cdot \alpha]$

(I5) is Lehmann's counterpart for Darwiche and Pearl's (C1) presented before. It says that if α is a logical consequence of β, then revising by α and then revising by β is the same as revising by β only.

(I7) is a weaker version of (C2). It asserts that beliefs acquired from the revision by $\neg\beta$ should not be retained if the belief state is immediately revised by β. Notice that the precondition for (C2) is strictly weaker than that for (I7), whereas its postcondition is in turn at least as strong.

In order to discuss these postulates, let us call a *trivial revision procedure* the operation that assigns $\mathrm{Cn}(A)$ to $[\sigma \cdot A]$, whenever $\neg A \in [\sigma]$, and $\mathrm{Cn}([\sigma] \cup \{A\})$, otherwise. A negative result of Lehmann's framework is that any revision procedure satisfying postulates (I1)–(I7) becomes trivial after an arbitrarily long sequence of revisions. This is obviously against the principle of minimal change. Lehmann suggests that, because postulates (I1)–(I7) are closely related to AGM's original formulation, no reasonable revision satisfying these postulates can comply with the principle of minimal change. However, we argue that the problem may in fact be with postulate (I7) below.

(I7) $\quad [\sigma \cdot \neg\beta \cdot \beta] \subseteq \mathrm{Cn}([\sigma], \beta)$

This is because (I7) is closely related to Darwiche and Pearl's postulate (C2), whose incompatibility with the principle of minimal change we argued in Section 4.2.

4.4 Iterated revision according to Boutilier

In [Boutilier, 1996], Boutilier presented an influential work on belief revision that made use of a similarity notion and also distinguished belief sets from epistemic states. The terminology there was slightly different. He called the former *objective belief sets* and the latter *revision models*. The revision model contained structural information used to guide the revision process of a given belief set. We present the details next. As before, we consider a language $\mathcal{L}$ over a set of propositional variables $\mathcal{P}$.

DEFINITION 53 (Revision model). Given a belief set K, a revision model for K is a tuple $M = \langle W, \leq, v\rangle$, where W is a set of worlds; $\leq$ is a total pre-order on W; and $v : \mathcal{P} \longrightarrow 2^W$ is a valuation function.

In Boutilier's work, a valuation v identifies those worlds in W in which a given propositional variable holds. v is extended to complex formulae as usual. $M \vDash_w \varphi$ denotes the fact that $w \in v(\varphi)$ in M. The collection of all such worlds is denoted by $|\varphi|$. For worlds $v, w \in W$, $v \leq w$ is interpreted as world v being at least as *plausible* a state of affairs as world w is. The relation $\leq$ induces at total order on equivalence classes in W and can be seen as a ranking relation (much in the same way as in Grove's spheres).

Revision models are required to verify a number of rationality conditions. One such condition is that $|K|$ is constituted exactly by the minimum equivalence class in W. In other words,

$$w \leq v \text{ for all } v \in W \text{ if and only if } M \vDash_w K$$

It is also required that for all $p \in \mathcal{P}$, $v(p) \neq \emptyset$ in addition to a well foundedness condition on $\leq$.

Conditional assertions of the kind $\varphi \Box\!\!\rightarrow \psi$ are evaluated in a model M in the following way:

DEFINITION 54 (Evaluation of conditionals in a revision model). Let M be a revision model and φ and ψ formulae of propositional logic.

$$M \vDash \varphi \Box\!\!\rightarrow \psi \text{ if and only if } \min(M, \varphi) \subseteq |\psi|$$

where

$$\min(M, \varphi) = \{w \in W | M \vDash_w \varphi, \text{ and } M \vDash_v \varphi \text{ implies } m \leq v \text{ for all } v \in W\}$$

The above notion is used to define a revision function $\circ_b$ for a particular revision model M (we will drop the subscript M to lighten the notation):

DEFINITION 55. Let K be a belief set; M a revision model for K; and φ and ψ formulae of propositional logic.

$$K \circ_b \varphi = \{\psi \mid M \vDash \varphi \Box\!\!\rightarrow \psi\}$$

Note that $\circ_b$ is defined in terms of first-degree conditionals, since φ and ψ are restricted to propositional logic formulae only and cannot contain conditionals themselves. In [Boutilier, 1994], Boutilier proved the following correspondences:

THEOREM 56. *If M is a K-revision model and $\circ_b$ the revision function determined by M according to Definition 55, then $\circ_b$ verifies the AGM postulates for belief revision.*

THEOREM 57. *Let $\circ$ be any revision function satisfying the AGM postulates. For any belief set K, there is a K-revision model M such that $K \circ \varphi = K \circ_b \varphi$.*

Note that if φ is contradictory, then $\min(M, \varphi) = \emptyset$. Therefore, $M \vDash \varphi \Box\!\!\rightarrow \psi$ for all ψ and hence $K \circ_b \varphi = K_\perp$, as expected.

Boutilier rightly points out that a K-revision model M is only sufficient to determine the revision of the belief set K by an arbitrary belief φ. Additional iterations of the revision procedure will require new revision models of their own. In his own words:

> "Conditionals and judgements of entrenchment form an integral part of an agent's epistemic state."

The difficulty is in establishing how the subsequent $(K \circ \varphi)$-revision model $(M \circ \varphi)$ should relate to K's original revision model M. One of the basic conditions is that $M\circ\varphi$ satisfies the same requirements originally imposed on the revision model M. This is called the *basic requirement*:

DEFINITION 58 (Basic requirement). If M is a K-revision model and $\circ_b$ is used to revise K by φ, then the revision model $M\circ\varphi$ must be such that $\min(M\circ\varphi, \top) = \min(M, \varphi)$.

There are similiarities with Gärdenfors' *belief revision systems* [Gärdenfors, 1988], although Boutilier's approach is semantical in flavour.

Since the ordering $\leq$ in a revision model M is the underlying mechanism determining the revision of K by φ, Boutilier argues that the principle of minimal change should also be applied to $\leq$ to obtain the ordering $\leq'$ of the revision model $M\circ\varphi$. Because $\leq'$ is expected to verify the basic requirement given in Definition 58, all worlds in which φ holds (i.e., $|\varphi|$) must constitute the *minimum* equivalence class with respect to $\leq'$. The remaining worlds are free to relate to each other as one pleases, but it makes sense to *preserve* the way they used to relate in the old K-revision model (via $\leq$). This is formalised as follows:

DEFINITION 59. Let $M = \langle W, \leq, v\rangle$ be a revision model for K. The *minimal conditional revision operator* (MC-revision operator) $\circ_b$ maps M into $M \circ_b\varphi$ for any propositional formula φ, such that $M \circ_b\varphi = \langle W, \leq', v\rangle$ where

1. if $v \in \min(M, \varphi)$ then $v \leq' w$ for all $w \in W$ and $w \leq' v$ if and only if $w \in \min(M, \varphi)$; and
2. if $v, w \notin \min(M, \varphi)$, then $w \leq' v$ if and only if $w \leq v$.

Note the similarity of condition 1. above and (F1) and (F2) (see page 21). Condition 2. encapsulates the application of the principle of minimal change to $\leq$. This process can be applied iteratively and Boutilier showed in [Boutilier, 1996] that iterated revisions obtained in this way will have the following properties:

(CB1) If $(K \circ_b\varphi) \vdash \neg\psi$, then $(K \circ_b\varphi) \circ_b\psi = K \circ_b\psi$
(CB2) If $(K \circ_b\varphi) \not\vdash \neg\psi$, then $(K \circ_b\varphi) \circ_b\psi = K \circ_b(\varphi \wedge \psi)$

We saw in the beginning of this section that (CB2) is a natural consequence of $(\mathrm{K}^\circ_{3,4})$ and $(\mathrm{K}^\circ_{7,8})$. (CB1) implies Darwiche and Pearl's (C2):

(C2) If $\varphi \models \neg\psi$, then $(\Psi\circ\psi)\circ\varphi \equiv \Psi\circ\varphi$

For assume that $K \circ_b\varphi \vdash \neg\psi$. If $\varphi \models \neg\psi$, then $\psi \models \neg\varphi$. In this case, (C2) requires that $(K \circ_b\varphi) \circ_\psi \equiv K \circ_b\psi$, which is the consequent of (CB1). On the other hand, if $K \circ_b \varphi \not\vdash \neg\psi$, then $\exists I \in \mathrm{mod}(\varphi)$ such that $I \notin \mathrm{mod}(\neg\psi)$, in which case $\varphi \not\models \neg\psi$ and hence (C2) is vacuously true.

(C2) however does not imply (CB1), for it is possible that $\varphi \not\models \neg\psi$, but $K \circ_b \varphi \vdash \neg\psi$, in which case (CB1) constrains $(K \circ_b \varphi) \circ_b \psi$, whereas (C2) does not. This happens because it is not possible to know a priori what models of φ will be chosen as the models of $K \circ_b \varphi$, but it is possible that all such chosen models satisfy $\neg\psi$ without necessarily implying that all models of φ do.

Furthermore, we can see from Definition 59 that the worlds in $\min(M, \varphi)$ become the preferred class with respect to $\leq'$ and the relationships between all other worlds remain unchanged. This rearrangement of $\leq$ only works well if revisions are *kind*, i.e., *forward-compatible* in Boutilier's terminology.

DEFINITION 60 (Forward compatibility). Let M be a revision model determining a MC-revision function $*$. The revision sequence $\varphi_1, \ldots, \varphi_n$ is *forward-compatible* with respect to $*$ (or model M), if and only if $\neg\varphi_{i+1} \notin (\ldots((K \circ_b \varphi_1) \circ_b \varphi_2) \ldots) \circ_b \varphi_i$ for each $1 \leq i < n$.

One could argue that revision sequences that are always forward compatible are not interesting, since they can be accomplished by expansions only. Boutilier has noticed that:

THEOREM 61. *If $\varphi_1, \ldots, \varphi_n$ is forward compatible for K, then $((K \circ_b \varphi_1) \circ_b \varphi_2 \ldots) \circ_b \varphi_n = K \circ_b (\varphi_1 \wedge \ldots \wedge \varphi_n)$.*

Unfortunately, when an incompatible formula comes in, the model forgets all revisions occurring up to the last formula in the sequence that is compatible with the new information:

THEOREM 62 (Forgetfulness). *Let $\varphi_1, \ldots, \varphi_{n+1}$ be an incompatible sequence such that $\varphi_1, \ldots, \varphi_n$ is compatible. Let k be the maximal element of $\{i \leq n \mid \neg\varphi_{n+1} \notin ((K \circ_b \varphi_1) \circ_b \varphi_2 \ldots) \circ_b \varphi_i\}$. It follows that $((K \circ_b \varphi_1) \circ_b \varphi_2 \ldots) \circ_b \varphi_{n+1} = K \circ_b (\varphi_1 \wedge \ldots \wedge \varphi_k \wedge \varphi_{n+1})$.*

We will discuss this in some more detail in Section 4.6. For further considerations of this model of iterated revision, including special conditions under which sequences of revisions can be reduced to one step revisions, the reader is referred to [Boutilier, 1996].

4.5 Prioritised base revision

In [Nebel, 1991b; Nebel, 1992], Nebel introduced the concept of a *prioritised belief base*. A prioritised belief base is a pair $\Gamma = \langle K, \sqsubseteq \rangle$, where K is a set of formulae and $\sqsubseteq$ is a total pre-order on K. $\sqsubseteq$ is called an *epistemic relevance* *ordering* and represents priorities in K, where $x \sqsubseteq y$ denotes that y has at least the same priority as x. If $\sqsubseteq$ is also antisymmetric, Γ is called a *linear prioritised belief base.* $\sqsubseteq$ is assumed to always have maximal elements.

Since $\sqsubseteq$ is total, K can be partitioned into a set of equivalence classes induced by $\sqsubseteq$. That is, a family of subsets of K whose elements are at the same priority level modulo $\sqsubseteq$. Let us denote such a partitioning by $\overline{K}$. If K is finite, Γ can be represented as a list $\Gamma = [K_1, \ldots, K_n]$ where each K_i is a partition of K generated by the equivalence relation induced by $\sqsubseteq$ (K_n being the partition associated with the maximal elements of K in $\sqsubseteq$). Moreover, if Γ is linear, then each K_i is just a singleton.

A revision of H by φ, in symbols, $H \circledast \varphi$, is obtained via the Levi identity (see page 2). H is first contracted by $\neg\varphi$ and then expanded by φ. The contraction of H by $\neg\varphi$ uses the epistemic relevance ordering and is called the *prioritised removal of* $\neg\varphi$ *from* H, in symbols, $H \Downarrow \neg\varphi$. This is defined as follows

DEFINITION 63 (Prioritised removal of formulae). Let $\langle \Gamma, \sqsubseteq \rangle$ be a prioritised belief base and φ a formula. The prioritised removal of φ from Γ is the family of subsets of K defined as follows:

$$X = \bigcup_{K_i \in \overline{K}} \{H_i\}$$

where each H_i is a subset of K_i and $X \in \Gamma \Downarrow \neg\varphi$ if and only if $\forall X' \subseteq K$ and $\forall i$:

$$X \cap (K_i \cup \ldots \cup K_n) \subset X' \cap (K_i \cup \ldots \cup K_n) \text{ implies } \neg\varphi \in Cn(X')$$

That is, starting from the most prioritised equivalence class, we consider maximal subsets of each class that together with the set being constructed do not imply $\neg\varphi$. There will possibly be a number of such $X's$, as there are possibly many combinations of subsets of each class that do not imply $\neg\varphi$. Therefore, to obtain the belief state in the prioritised removal of $\neg\varphi$ from Γ it is necessary to take the set $\bigcap\{Cn(X) \mid X \in \Gamma \Downarrow \neg\varphi\}$.

The prioritised base revision of Γ by φ is then defined as follows

DEFINITION 64 (Prioritised base revision). Let $\langle \Gamma, \sqsubseteq \rangle$ be a prioritised belief base and φ a formula. The revision of Γ by φ ($\Gamma \circledast \varphi$) is defined as

$$\Gamma \circledast \varphi =^{\text{def}} Cn(\bigcap\{Cn(X) \mid X \in \Gamma \Downarrow \neg\varphi\} \cup \{\varphi\})$$

Thus, in the general case it is not possible to iterate the revision process. However, if K is finite, then $\Gamma \circledast \varphi$ can be at least finitely represented and the result is simply

$$Cn(\bigvee(\Gamma\Downarrow\neg\varphi) \wedge \varphi)$$

4.6 Prioritised databases

In the case of finite belief bases (representable as a formula) both arguments of the revision operation are of the same type. Therefore, it is possible to take $\circ$ as a right or left-associative operation. Furthermore, we argued in [Gabbay and Rodrigues, 1997] that $\circ$ is in general non-associative. It turns out that there are advantages to adopting the right-associative application of $\circ$. The right-associative interpretation of the revision operation yields a refined notion of similarity between worlds for a sequence $\Delta = \delta_1\circ(\delta_2\circ\ldots(\delta_{k-1}\circ\delta_k)\ldots)$. In this section, we analyse the properties of such a sequence and show that this interpretation reconciliates the coherentist approach with "memory".

For simplicity, we consider epistemic states that arise from a finite sequence of input formulae $\Delta = [\delta_1, \delta_2, \ldots, \delta_k]$, where for all $j > i$, δ_j is received after δ_i. In [Gabbay and Rodrigues, 1997], we called this a *prioritised database* (PDB) and required that each δ_i was a formula in disjunctive normal form (DNF). The DNF requirement was due to the fact that the revision operator used took advantage of the properties of a formula in DNF.[13] However, in general we can consider *any* revision operation $\circ$ satisfying the AGM postulates for finite bases. Notice that Lehmann's approach seen in Section 4.3 as well as the *linear* version of Nebel's prioritised belief bases (Section 4.5, [Nebel, 1991b]) are also based on sequences of revisions.

If we assume that the belief set associated with a PDB Δ is obtained by successively applying $\circ$ to the formulae in Δ and provided $\circ$ is not associative, there will be two natural ways of interpreting the sequence of revisions in Δ: either by considering the operation left-associative or by considering it right-associative. We will use ${}^*\Delta$ (read *left delta*) to denote the left associative interpretation of the sequence of revisions in Δ and Δ^* (read *right delta*) to denote the right-associative one. Formally,

DEFINITION 65. Let $\Delta = [\varphi_1, \varphi_2, \ldots, \varphi_k]$ be a PDB.

$${}^*\Delta = \begin{cases} \varphi_k & \Rightarrow \text{if } k = 1 \\ ((\varphi_1\circ\varphi_2)\circ\ldots)\circ\varphi_k & \Rightarrow \text{if } k > 1 \end{cases}$$

$$\Delta^* = \begin{cases} \varphi_k & \Rightarrow \text{if } k = 1 \\ \varphi_1\circ(\ldots\circ(\varphi_{k-1}\circ\varphi_k)) & \Rightarrow \text{if } k > 1 \end{cases}$$

[13] The operator was defined in [Gabbay and Rodrigues, 1996a].

In [Gabbay and Rodrigues, 1997], we argued that the right associative interpretation was the most interesting one, because the inevitable reapplication of the revision steps, although costly, provided an *opportunity* to revisit past decisions.[14] In the following, we assume $\circ$ is any revision operator complying with the AGM postulates for finite bases.

We start by defining an ordering $\preceq_\Delta$ that will help us to analyse how valuations of L compare with each other with respect to the epistemic state Δ (much in the sense of Lehmann's epistemic states and Darwiche and Pearl's faithful assignments). As before, $I \preceq_\Delta I'$ denotes the fact that I is at least as good at satisfying Δ as I' is.

DEFINITION 66. Let $\Delta = [\delta_1, \delta_2, \ldots, \delta_k]$ be a PDB and consider the belief set Δ^* where the revision operator $\circ$ satisfies the AGM postulates. Let $\leq_{\delta_i}$ be the faithful assignment for the operation for each formula δ_i in Δ as in Definition 20, and take $i, j \in \{1, \ldots, k\}$.

$$I \preceq_\Delta I' \text{ if and only if for all } i,\ I' <_{\delta_i} I \text{ implies } \exists j > i \text{ such that } I <_{\delta_j} I'.$$

The ordering above was motivated by the observations about associativity of $\circ$ made in [Gabbay and Rodrigues, 1997] and first appeared in [Rodrigues, 1998], where a comprehensive account of its properties can be found.

Note that if Δ is the empty sequence ε, then $I \equiv_\Delta I'$, for all $I, I' \in \mathcal{I}$ (vacuously). The greater the index of a formula in a PDB, the more recent it is the information it represents. Thus, what the definition above says is that the failure of a valuation I to be at least as good as another valuation I' at satisfying a formula received at time i can only be compensated by I being better than I' at satisfying some other formula received at a later time j. Similarity with respect to the formula at each point is determined by the faithful assignment of the operator used in the sequence of revisions at that point.

A number of properties of $\preceq_\Delta$ can be found in [Rodrigues, 2005]. We show some of interest below.

Notation 3. $\Delta; \beta$ will be used to denote the sequence obtained by appending β to the end of the sequence Δ, i.e., $\Delta :: [\beta]$.

Note that $\Delta; \beta$ is also a PDB.

REMARK 67. The revision of a PDB Δ by a formula β is $(\Delta; \beta)^*$.

DEFINITION 68. For each PDB Δ, $\varkappa$ is a function defined as follows:

$$\varkappa(\Delta) = \{\preceq_\Delta, \mathcal{I}\}.$$

[14] In day-to-day reasoning, we do this routinely.

where $\preceq_\Delta$ conforms to Definition 66.

THEOREM 69. *$\varkappa$ is a faithful assignment for revision of epistemic states.*

THEOREM 70. *The revision scheme obtained by* $(\Delta)^*$*, where Δ is a PDB, satisfies postulates* $(\mathrm{R}^\star 1)$–$(\mathrm{R}^\star 6)$*.*

Remember that $(\mathrm{R}^\star 1)$–$(\mathrm{R}^\star 6)$ are Darwiche and Pearl's basic rewriting of the AGM postulates for the case of epistemic states (see Section 4.2, page 40). Proofs of both theorems above can be found in [Rodrigues, 1998; Rodrigues, 2005].

We now discuss more specifically how the ordering $\preceq$ evolves as new formulae are added to a PDB. The evolution itself depends on the particular characteristics of the faithful assignment for the revision operator. For illustration purposes, we pick $\leq_\delta$ as determined by the distance function used by Dalal and others [Dalal, 1988a; Gabbay and Rodrigues, 1997], in which $I \leq_\delta J$ if and only if the minimum number of disagreements of truth-values of propositional letters between I and any model of δ is at most the minimum number of disagreements between J and any model of δ.

Suppose the initial PDB is composed solely by the formula p. The only ordering in this case is the faithful assignment $\leq_p$ and hence $\preceq_{[p]}=\leq_p$. Therefore, we have the ordering below. For simplicity, we consider $\mathcal{L}$ over $[p, q, r]$ only. A valuation I is represented as a sequence of binary digits PQR where $P = 1$ if and only if $I \Vdash p$, and $P = 0$ otherwise; $Q = 1$ if and only if $I \Vdash q$, and $Q = 0$ otherwise and $R = 1$ if and only if $I \Vdash r$, and $R = 0$ otherwise. A valuation appearing lower in the graph is preferred to a valuation appearing above it. Thus, in the diagram below, we have that for instance $100 \prec_{[p]} 000$. Note that the valuations displayed at the same level are all equivalent modulo $\preceq_{[p]}$.

$$\begin{array}{c} 000, 001, 010, 011 \\ \uparrow \\ 100, 101, 110, 111 \end{array}$$

There is not much information in the PDB when only p is present and the ordering above reflects that; it only makes a distinction between the valuations that satisfy p and those which do not. This is the rather weak requirement imposed by (F2) of Definition 31. The class with the models of p must be as represented above because of (F1) and the fact that the faithful assignments must be total. Also note that the models of $\neg p$ are not constrained in any way by Definition 31, but they are all equivalent in this case because of the definition of $\leq_p$.

By adding the formula $p \to q$ to the PDB above, the ordering $\preceq_{[p,p\to q]}$ for the PDB $[p, p \to q]$ changes to reflect the new priorities:

$$\begin{array}{c} 100, 101 \\ \uparrow \\ 000, 001, 010, 011 \\ \uparrow \\ 110, 111 \end{array}$$

Now we have three classes of valuations: in the minimal (preferred) one we have 110 and 111, which are exactly the only two models of p and $p \to q$. According to this similarity ordering, the next class of valuations that best satisfy $[p, p \to q]$ are $000, 001, 010, 011$. These correspond exactly to the valuations that failing to satisfy p, at least satisfy the most important formula in the PDB ($p \to q$). Finally, the last level contains the valuations that do not satisfy $p \to q$, but at least satisfy p.

The addition of r to the previous PDB results in the following rearrangement of the similarity ordering.

$$\begin{array}{c} 100 \\ \uparrow \\ 000, 010 \\ \uparrow \\ 110 \\ \uparrow \\ 101 \\ \uparrow \\ 001, 011 \\ \uparrow \\ 111 \end{array}$$

The minimal elements of this ordering are exactly the models of the three formulae. The next class contains the valuations that satisfy the two formulae with highest priority in the PDB, followed by the class with the valuations that satisfy the formula with highest priority and the one with least priority.

The other four classes of the ordering in the top half follow a similar pattern. They contain the valuations that fail to satisfy r. Whenever that is not possible, the next best thing is to satisfy the other two formulae. This is represented by valuation 110. The next two valuations satisfy the second formula and the least preferred one satisfies only the least important formula.

Perhaps a little more interesting is to see how the ordering above is obtained. Each class of valuations is recursively ordered in the reverse order of the sequence of formulae in the PDB. That is, firstly the set of all valuations is ordered according to the most recent information received: r. This will result in a number of equivalence classes totally ranked amongst themselves (since each faithful assignment is total). Next, each class is internally ordered according to the faithful assignment of the previous formula in the PDB: $p \to q$. Finally, the resulting classes are ordered according to the first formula in the sequence: p (see Figure 7).

$$\mathcal{I} \xrightarrow{\leq_r} \begin{array}{c} 000, 010,\\ 100, 110 \\ \uparrow \\ 001, 011,\\ 101, 111 \end{array} \xrightarrow{\leq_{p \to q}} \begin{array}{c} 100 \\ \uparrow \\ 000, 010, 110 \\ \uparrow \\ 101 \\ \uparrow \\ 001, 011, 111 \end{array} \xrightarrow{\leq_p} \begin{array}{c} 100 \\ \uparrow \\ 000, 010 \\ \uparrow \\ 110 \\ \uparrow \\ 101 \\ \uparrow \\ 001, 011 \\ \uparrow \\ 111 \end{array}$$

Figure 7. Embedded orderings in PDBs.

More formally, we can show the relationship between the ordering of a given PDB Δ and the ordering of Δ revised by a formula β, in the following way:

PROPOSITION 71. *Let $\Delta = [\delta_1, \ldots, \delta_k]$ be a PDB, β a formula, $\preceq_\Delta$ the faithful assignment for Δ and $\leq_\beta$ the faithful assignment for the formula β.*

$$I \preceq_{\Delta;\beta} I' \textit{ if and only if } I' \leq_\beta I \textit{ implies } I \leq_\beta I' \textit{ and } I \preceq_\Delta I'$$

where $\preceq_{\Delta;\beta}$ is the faithful assignment for the PDB $[\delta_1, \ldots, \delta_k, \beta]$.

The characterisation given in Proposition 71 above allows us to relate revisions achieved by PDBs with Darwiche and Pearl's formalisation given in Section 4.2 in the following way:

THEOREM 72. *The revision method achieved by PDBs satisfies postulates* (C1), (C3) *and* (C4).

The proofs of Proposition 71 and Theorem 72 can be found in [Rodrigues, 2005; Rodrigues, 1998]. Note that PDBs do not satisfy (C2). However, we have argued against this postulate in Section 4.2.

There are similarities between the revision scheme obtained by PDBs and Boutilier's approach (see Section 4.4). However, PDBs are more forgiving, as showed in the next example.

EXAMPLE 73. Consider the sequence of formulae p, $q \wedge r$ and $\neg q$ and suppose the revision model for p is $M = \langle W, \leq, v \rangle$, where $\leq = \leq_p$. According to Definition 59, the MC-revision operator will update $\leq$ successively as follows.

$$
\begin{array}{ccccc}
\leq_{[p]} & \longrightarrow & \leq_{[p,q\wedge r]} & \longrightarrow & \leq_{[p,q\wedge r,\neg q]}
\end{array}
$$

$$
\begin{array}{ccc}
000, 001, 010, 011 & 000, 001, 010, 011 & 000, 001, 010, 011 \\
\uparrow & \uparrow & \uparrow \\
100, 101, 110, 111 & 100, 101, 110 & 110 \\
 & \uparrow & \uparrow \\
 & 111 & 111 \\
 & & \uparrow \\
 & & 100, 101
\end{array}
$$

As a result, $(p \circ_b (q \wedge r)) \circ_b \neg q = \mathrm{Cn}(p \wedge \neg q)$. Notice that the revision by $q \wedge r$ is completely "forgotten" (see Theorem 62).

However, the ordering in a PDB would evolve as follows.

$$
\begin{array}{ccccc}
\leq_{[p]} & \longrightarrow & \leq_{[p,q\wedge r]} & \longrightarrow & \leq_{[p,q\wedge r,\neg q]} \\
 & & & & \\
000, 001, 010, 011 & & 000 & & 010 \\
\uparrow & & \uparrow & & \uparrow \\
100, 101, 110, 111 & & 100 & & 110 \\
 & & \uparrow & & \uparrow \\
 & & 001, 010 & & 011 \\
 & & \uparrow & & \uparrow \\
 & & 101, 110 & & 111 \\
 & & \uparrow & & \uparrow \\
 & & 011 & & 000 \\
 & & \uparrow & & \uparrow \\
 & & 111 & & 100 \\
 & & & & \uparrow \\
 & & & & 001 \\
 & & & & \uparrow \\
 & & & & 101
\end{array}
$$

and the final result of the sequence of revisions would be $[p, q \wedge r, \neg q]^* = \mathrm{Cn}(p \wedge \neg q \wedge r)$. In PDB's case, at least the compatible component r of $q \wedge r$ is retained.

There are similarities between the way PDBs construct the orderings for epistemic states and the way Konieczny and Pérez's *basic memory operator* operates [Konieczny and Pérez, 2000]. However, the basic memory operator satisfies (C2), whereas the revision process obtained by PDBs does not. We have argued against (C2) in page 42. A number of variations of the memory operator is also given in [Konieczny and Pérez, 2000].

4.7 Ordered theory presentations

Another formalism that can be used in the reasoning about epistemic changes is Ryan's *ordered theory presentations* (OTPs) [Ryan, 1992]. The main idea is again to have a finite belief base associated with some priority information for its formulae. For the case of belief revision, belief bases are presented as lists of formulae $[\varphi_1, \ldots, \varphi_k]$, where φ_i has priority over any φ_j with $j < i$ [Ryan, 1992, Section 4.4]. The priority information is used to solve possible inconsistencies arising from conflicting elements in the base. As before, revisions are performed by *appending* formulae to the list.

The main difference between OTPs and the other formalisms presented here is based on the way conflicts are solved. In an OTP, change is evaluated through a restricted notion of classical consequence called *natural consequence*. In order to introduce that, we need a few definitions first.

DEFINITION 74 (Positive and negative occurrences of propositional symbols). Let $p \in \mathcal{P}$ and φ a formula of propositional logic.

- p occurs positively in p.
- If p occurs positively (negatively) in φ, then it occurs negatively (positively) in $\neg\varphi$.
- If p occurs positively (negatively) in φ or in ψ, then it occurs positively (negatively) in $\varphi \wedge \psi$ and $\varphi \vee \psi$.
- If p occurs negatively (positively) in φ or positively (negatively) in ψ, then it occurs positively (negatively) in $\varphi \rightarrow \psi$.
- If p occurs at all in φ or ψ, then it occurs both positively and negatively in $\varphi \leftrightarrow \psi$.

Using the notion of positive (resp., negative) occurrences, Ryan then defines the sets of symbols $p \in \mathcal{P}$ in which a formula φ is monotonic (resp., anti-monotonic).

DEFINITION 75 (Monotonicity and anti-monotonicity of formulae with respect to propositional variables).

1. φ is monotonic in p, if it is equivalent to a formula in which all occurences of p (if any) are positive.
2. φ is anti-monotonic in p, if it is equivalent to a formula in which all occurences of p are negative.
3. φ^+ and φ^- are the sets of symbols in which φ is monotonic and anti-monotonic, respectively.

DEFINITION 76 (Natural consequence). Let φ and ψ be formulae. ψ is a natural consequence of φ, in symbols $\varphi \breve{\models} \psi$, if

1. $\varphi \models \psi$;
2. $\varphi^+ \subseteq \psi^+$; and
3. $\varphi^- \subseteq \psi^-$

Given the notion of natural consequence, it is now possible to define a partial pre-order $\sqsubseteq_\varphi$ on valuations of $\mathcal{L}$ that can be used to compare how good interpretations are at satisfying a formula φ.

DEFINITION 77 (Similarity via natural consequence). Let I and I' be two valuations of $\mathcal{L}$, $I \sqsubseteq_\varphi I'$ if, for each ψ

$$\varphi \breve{\models} \psi \text{ implies (if } M \Vdash \psi, \text{ then } N \Vdash \psi)$$

In the definition above, $I \sqsubseteq_\varphi I'$ is read as "I is at least as good at satisfying φ as I' is". For illustration purposes, the notion of similarity according to Definition 77 and that provided by the distance d for a belief base containing $p \wedge q$ are compared in Figure 8.

Analogously to what happened with the similarity ordering defined in terms of the distance *diff* presented before, the similarity measurement defined via natural consequence is associated with a partial pre-order on the set of valuations of L. This impairs the satisfaction of postulates (K°4) and (K°8).

For simplicity, the language $\mathcal{L}$ in the diagrams of Figure 8 is defined over the propositions p and q only and the valuations are represented as sequences of binary digits as before (the first digit for p and the second one for q). The lower a valuation appears in the diagram, the more preferred it is. As such, 10 and 01 are incomparable in ① because they satisfy different natural consequences of $p \wedge q$, namely p and q, respectively. In ②, they are considered equivalent, because they have the same distance to the only model of $p \wedge q$ (i.e., $d = 1$ to model 11). In this sense, OTPs offer a *qualitative* evaluation of change, whereas d offer a purely *quantitative* one. The distance function d yields a total ordering on the set of valuations $\mathcal{I}$, which ultimately leads to the satisfaction of all of the AGM postulates.

What remains to be shown is how to compare how good valuations are at satisfying a given belief base $\Gamma = [\varphi_1, \ldots, \varphi_k]$. As we said, the ordering $\sqsubseteq_\varphi$ is used subject to the priorities of the formulae in Γ.

DEFINITION 78 (Comparing valuations with respect to belief bases). Let I and I' be two valuations and $\Gamma = [\varphi_1, \ldots, \varphi_k]$ a belief base.

$$I \sqsubseteq^\Gamma I' \text{ if } \forall \varphi_i \exists \varphi_j \text{ such that } (I \not\sqsubseteq_{\varphi_i} I' \text{ implies } j > i \text{ and } I \sqsubset_{\varphi_j} I')$$

Intuitively, the definition above says that a valuation I is at least as good at satisfying a belief base Γ as a valuation I' is if for any formula φ_i, whenever I fails to be as good as I' at satisfying φ_i, then there exists a formula φ_j, which is more important than φ_i and I is strictly better at satisfying φ_j than I' is.

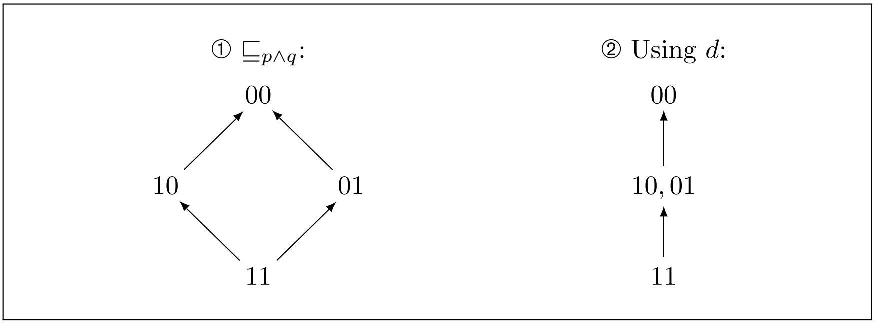


Figure 8. Comparative similarity for $p \wedge q$ using natural consequence and the function d.

Revisions are defined in a semi ad hoc manner by considering *linear* ordered theory presentations (i.e., a list of formulae). An overall order for the theory presentation is constructed by utilising the individual induced orders on interpretations obtained from the natural consequence relation over each formula on the list. In order to perform a revision, the revising formula is simply concatenated to the end of the list, giving it the highest priority in the base.

OTPs used for belief revision as defined above were shown to verify postulates (K°1), (K°3), (K°5), (K°6) and (K°7). (K°2) cannot be verified directly when the revising belief is contradictory, but this can be circumvented by the introduction of an inconsistent OTP $\perp$ and defining $\Gamma \circ \perp = \perp$. However (K°4) and (K°8) are not verified. Some arguments against (K°4) and (K°8) are presented in [Ryan, 1992]. An application of this belief revision approach is discussed in Section 7.

5 SPECIALISED BELIEF REVISION

Apart from the use of a richer structure to enable the distinction between epistemic states and belief sets, in all of the formalisms presented so far, some of the underlying assumptions remained the same: the logic was essentially classic (although the notion on consequence was restricted in the case of OTPs); the input was a single formula (although sequences of single inputs were considered in Section 4); and the revision process itself was mainly described in an axiomatic way. In this section, we present and discuss a number of formalisms for belief revision that depart from these assumptions. We start with resource-bounded revision.

5.1 Resource-bounded revision

Resource-based revision is a specialised belief revision approach proposed by Wassermann [Wassermann, 1999]. The shift from the AGM paradigm, where agents are idealised, is in recognising that in most practical applications agents are in fact entities with limited memory and capacity of inference, i.e., they have limited resources.

In the resource-bounded revision framework the beliefs of an agent are divided into *explicit*, *active* or *inferred* (or implicit) beliefs. Active beliefs are beliefs currently available for use. These can be beliefs recently acquired; intermediate conclusions in an argument; or beliefs related to a current topic. Beliefs have to become active first in order for them to be subsequently explicitly accepted, revised or rejected. However, explicit beliefs do not need to be necessarily active at all times and not all active beliefs are necessarily explicit.

In the formalism, a belief state is defined as a tuple $\langle E, \mathrm{Inf}, A\rangle$, where E is the set of explicit beliefs, Inf is the agent's inference relation (which can be used to determine the set of implicit beliefs), and A is the set of active beliefs.

Normally, when new information arrives it is firstly added to the set of active beliefs as a *provisional belief*, and then questioned by the agent. The depth of the enquiry on a provisional belief is determined by the agent and her interest in the subject [Wassermann, 1999]. If the new information "survives" the enquiry, it is incorporated into the set of explicit beliefs. Wassermann defines six main operations on a belief state $\langle E, \mathrm{Inf}, A\rangle$ that can be used to change the status of a belief. These operations enable beliefs

- which are explicit to become active ($+_r$);
- which are active to be rejected ($+_c$);
- to be observed as provisionally active ($+_o$);
- to change status from active to explicit ($+_a$);
- to be inferred from the active beliefs ($+_i$);
- to be put into question by being temporarily removed from the set of explicit beliefs ($+_d$).

The operations can be combined into complex operations to manipulate a belief state and were shown to be complete with respect to all possible changes that a belief state may undergo [Wassermann, 1997]. In other words, for any two belief states $K_1 = \langle E_1, \mathrm{Inf}, A_1 \rangle$ and $K_2 = \langle E_2, \mathrm{Inf}, A_2 \rangle$, there is a sequence of basic operations that transforms K_1 into K_2.

A revision is performed on the active beliefs. Given a belief state $K = \langle E, \mathrm{Inf}, A \rangle$, a new belief α is provisionally made active by taking $K +_o \alpha = \langle E, \mathrm{Inf}, A \cup \{\alpha\} \rangle$. At this point, the set of active beliefs may well be inconsistent. Consistency can be restored through revision. During the process, both explicit and active beliefs can be rejected, and explicit (but not active) beliefs retrieved and made active so that they can be taken into consideration during the reasoning process.

In [Wassermann, 1999], Wassermann also showed how to combine the operations given above to obtain Hansson's *local semi-revision operation* [Hansson, 1997b], according to which a belief base is revised in such a way that the incoming information is not necessarily given the highest priority. As a result, the new belief can either be accepted (i.e., made *explicit*) or rejected — a departure from the *principle of the primacy of the update*, imposed by (K°2).[15]

The following example illustrates how the revision process in this formalism takes place.

EXAMPLE 79 (Resource bounded revision). Let the current belief state be $\Psi = \langle B, \mathrm{Cn}, \emptyset \rangle$, where $B = \{\neg p, q, q \rightarrow p, r\}$ and assume the belief p is received as input. The first step is to *observe* it. This produces the new belief state $\Psi_1 = \Psi +_o p = \langle B, \mathrm{Cn}, \{p\} \rangle$ (i.e., p is added to the set of active beliefs). In order to incorporate the new belief as *explicit*, an enquiry process is performed on p. This involves retrieving explicit beliefs that are "relevant" to p. The amount of retrieval is bounded by the agent's resources. For this example, assume that the beliefs $\{\neg p, q \rightarrow p, q\}$ are retrieved (r is not relevant for this inference). An intermediate belief state $\Psi_2 = \Psi_1 +_r \{\neg p, q \rightarrow p, q\} = \langle B, \mathrm{Cn}, \{p, \neg p, q \rightarrow p, q\} \rangle$ is then obtained. At this point the set of active beliefs is inconsistent and thus needs to be revised. A possible revision would be one of the maximal consistent subsets that includes p, for instance, $\{p, q \rightarrow p, q\}$. Effectively, this corresponds to rejecting the active belief $\neg p$. The result would then be the new belief state $\Psi_3 = (\Psi_2 +_d \neg p) +_c \neg p = \langle B \backslash \{\neg p\}, \mathrm{Cn}, \{p, q \rightarrow p, q\} \rangle$. The resulting active beliefs would now be consistent and can be made explicit, finally giving the belief state $\Psi_4 = \Psi_3 +_a \{p, q \rightarrow p, q\} = \langle (B \backslash \{\neg p\}) \cup \{p\}, \mathrm{Cn}, \emptyset \rangle$.

Resource-based revision is general enough to encompass the standard AGM belief revision framework as well as Nebel's belief base approach

[15]This also happens in the controlled revision approach described in Section 5.2 and some authors have called it the *non-insistent policy*.

[Nebel, 1989]. Note, however, that the set of explicit beliefs is not necessarily closed under the consequence relation and that the active beliefs consist only of a limited part of the current set of beliefs of an agent. This means that it is quite possible that the agent has an inconsistent belief state without being aware of it (so long as the inconsistency is not within the set of active beliefs). This is an idea which seems to represent well how common-sense reasoning operates in practice.

In order to show how the traditional AGM approach relates to this approach, we make the following assumptions: for a belief state $\Psi = \langle K, \mathrm{Cn}, K\rangle$, K is a theory closed under the consequence relation Cn (the belief set of the agent) and the set of active beliefs of Ψ is *equal* to the set of explicit beliefs. Note that this is an idealised interpretation, since it assumes that an agent has unlimited memory and inference capability. However, it can be seen as an (AGM) idealisation of the process. Revision of a belief set K by a belief α (in the AGM sense) is obtained via the Levi identity (see page 5) in a sequence of $+_o$ (observation), $+_d$ (doubting) and $+_c$ (rejection) operations of the beliefs contradicting α followed by the acceptance of α itself. The operations would query K to check whether the active (= explicit) beliefs contradict α and, if they do, would perform the deletion of an appropriate subset in order to attain consistency with α — at which point α can be added via the acceptance operation. A formal account of the process loosely described above can be found in [Wassermann, 1997].

Resource-based revision can also be used to define Nebel's belief base revision [Nebel, 1989], in which there is a distinction between explicit beliefs (i.e., those included in a finite belief base) and implicit ones (i.e., those derivable from the base). However, as we mentioned before, Nebel's formalism does not distinguish between active and explicit beliefs. A belief base Φ would be represented by a belief state $\langle \Phi, \mathrm{Cn}, \mathrm{Cn}(\Phi)\rangle$ where the explicit beliefs consist of the belief base Φ, and the active beliefs contain all of its consequences (modulo Cn). Under this perspective, it is possible to show that Nebel's revision operation can also be expressed via an appropriate sequence of resource-bounded operations [Wassermann, 1997].

5.2 Controlled revision

Controlled revision is an algorithmic approach to belief revision that makes use of the history of updates in order to try and minimise the loss of information in the process and achieve more rational changes of belief. The basic idea is to activate/deactivate formulae as necessary in order to maintain consistency and also to keep record of the history of a formula's status through time. The history allows for previously deactivated formulae to be reinstated should they no longer cause inconsistency at a later stage. In addition, it is possible to analyse how *stable* beliefs have been through the evolution of an agent's belief state.

The approach builds upon the idea of a *labelled deductive system* [Gabbay, 1996], whereby belief bases are represented as sets of labelled formulae; the consequence relation defines a relation between sets of labelled formulae and individual labelled formulae; and inference rules specify how formulae and labels can be manipulated. The extra information in the labels may be used to constrain the application of the rules according to desired criteria. In the case of controlled revision, a label is a pair of terms (l, h), where l is a unique name for a formula and h is a list of constants $\pm i$ $(i \geq 1)$ indicating whether the formula was active $(+)$ or inactive $(-)$ at stage i. For instance, a labelled formula of the form $(l^i_j, +j) : A^i_j$ states that the formula A^i_j (uniquely associated to the term l^i_j) is *active* in the belief base at the j-th revision step (i.e., at time j).

With respect to the standard presentation, a belief base at stage n can be seen as the collection of all formulae active at step n, i.e., those formulae whose label's second term is a list $(\pm i, \ldots, \pm m)$, with $1 \leq i \leq m$ and $k \leq n \leq m$ and where n appears with a positive sign. The approach uses Johansson's minimal logic [Johansson, 1936] as its underlying logic, which includes the intuitionistic implication $\rightarrow$ and the constant symbol $\perp$ for *falsity*, with no special axiom. Within this logic, negation is defined as $\neg A =_{\text{def}} A \rightarrow \perp$, and the classical axiom $\perp \rightarrow A$, for every formula A, does not hold. A belief base is said to be *inconsistent* if it can derive $\perp$.

In the controlled revision framework the initial belief base K_0 is allowed to contain *integrity constraints* expressed in *denial* form $(A_1, A_2, \ldots, A_n) \rightarrow \perp$. Intuitively, this represents $A_1 \rightarrow (A_2 \rightarrow \ldots \rightarrow (A_n \rightarrow \perp) \ldots)$. Integrity constraints cannot be deactivated. An arbitrary belief base K_n is said to be *consistent*, if the set of its active formulae and integrity constraints is consistent.[16]

The mechanism for revision uses two main algorithms: a *selection algorithm*, which computes subsets of the base that are involved in an inconsistency; and a *reinstatement algorithm*, which reactivates previously deactivated formulae that no longer cause inconsistency. A new input A_{n+1} is *temporarily* included into a consistent belief base K_n and the set $K_n^{'} = K_n \cup \{(l_{n+1}, +(n+1)) : A_{n+1}\}$ checked for consistency. If $K_n^{'}$ is consistent, then K_{n+1} is defined by *i)* adding $(l_{n+1}, +(n+1)) : A_{n+1}$ to K_n

[16] The concepts of consistency and integrity constraints are subsumed by the more general concept of *acceptability*, presented in Section 5.4.

and *ii)* appropriately modifying the labels of the formulae in K_n so as to include the constant $+(n+1)$ into the history part of the active formulae's labels and the constant $-(n+1)$ into the history part of the inactive ones. No re-instantiation is needed because the logic behaves monotonically in this case, i.e., any formula causing inconsistency in K_n would still cause inconsistency in K_{n+1}.

The more interesting case is when $K_n^{'}$ turns out to be inconsistent and the selection algorithm is applied to compute one (or more) *selection* sets Γ_i^{n+1} of formulae from $K_n^{'}$. In effect, this algorithm determines which active formulae in $K_n^{'}$ should become inactive. The constant $-(n+1)$ is added to the history part of the labels of these formulae and the constant $+(n+1)$ added to the history part of the labels of the remaining formulae. The new base K_{n+1} is then computed using the re-instantiation algorithm, which identifies the inactive formulae in K_n that are no longer inconsistent with the newly determined active formulae. These formulae are then reactivated in K_{n+1}, i.e., they get the label $+(n+1)$. Notice that one would in general expect the sets Γ_i^{n+1} to be minimal and that some policy has to be applied to decide which of these to use. Furthermore, for any given Γ_i^{n+1}, there is possibly more than one way of applying the reinstantiation algorithm depending on the ordering on which the inactive formulae are considered for reactivation.

The following example illustrates the overall revision process.

EXAMPLE 80 (Controlled revision). Let $K_0 = \emptyset$ be the initial belief base and σ the following sequence of updates performed in the left to right order.

(3) $\quad \sigma = [l_1 : D \to C, l_2 : C \to \bot, l_3 : D]$

The first update amounts to a simple expansion of K_0, thus giving us the new base

$$K_1 = \{(l_1, (+1)) : D \to C\}$$

The second update is now applied to K_1. A temporarily expanded base $K_1^{'}$ is defined as $K_1 \cup \{(a_1, +1) : C \to \bot\}$, where a_1 is an arbitrary but unique temporary name for the input $C \to \bot$. $K_1^{'}$ is then checked for consistency. Since it is consistent, the new belief base K_2 is defined as

$$K_2 = \{(l_1, (+1, +2)) : D \to C, (l_2, (+2)) : C \to \bot\}$$

The third update is then applied to K_2. The temporarily expanded base $K_2^{'}$ is $K_2 \cup \{(a_2, +2) : D\}$ and checked for consistency. $K_2^{'}$ is inconsistent, since $\bot$ can be derived from D, $D \to C$ and $C \to \bot$ (the actual derivation uses the underlying labelled deductive system proof theory for Johansson's minimal logic, but this suffices for our discussion).[17]

Several policies can be applied to resolve the inconsistency. The *principle of the primacy of the update* would force the new input to be accepted in the revised base and some of the previous labelled formulae to be deactivated. In this case, $\{(a_2, (+2)) : D\}$ is kept, resulting in the following candidate subsets of $K_2^{'}$:[18]

$$\begin{array}{rcl} \Gamma_1^2 & = & \{(l_1, (+1, +2)) : D \to C\} \\ \Gamma_2^2 & = & \{(l_2, (+2)) : C \to \bot\} \end{array}$$

Deactivating all of the formulae in either Γ_1^2 or Γ_2^2 would prevent the inconsistency from being derived. The selection algorithm would choose one of them taking various parameters into account. One of these is *persistence* of the formulae in a given candidate subset. Persistence is calculated as the total number of times the formulae in the subset have been active. Those formulae with the lowest persistence value are then rejected. These correspond intuitively to the formulae that have been less stable in the base. If there is more than one candidate subset with the same minimum persistence value, then the smallest subset is chosen.[19] If the subsets with lowest persistence also have the same cardinality, then the history of the data in each subset is considered by computing the number of $\pm$ changes of each formula in the set. The subset whose formulae have lower *reliability priority* (i.e., higher number of changes) is selected. If none of these checks identify a singular candidate, a *tree-revision mechanism* is adopted. This mechanism structures the alternative ways of revising an inconsistent base into a tree and adds to each of them the input formula — this eventually helps to reinforce or reject some of the alternatives. The tree revision mechanism provides a criteria for choosing among candidate subsets for rejection in terms of newly acquired information instead of the random choice mechanism adopted in systems like TMS [Doyle, 1979].

Going back to the example above, from the two candidate subsets Γ_1^2 and Γ_2^2, the selection algorithm would choose Γ_2^2 since it has persistence value 1 which is lower than Γ_1^2's persistence value (2). The revised base would then be given by

$$K_3 = \{(l_1, (+1, +2, +3)) : D \to C, (l_2, (+2, -3)) : C \to \bot, (l_3, (+3)) : D\}$$

[17] The reader is referred to [Gabbay *et al.*, 2003] for a description of a labelled proof system for Johansson's minimal logic.

[18] If a policy does not verify the principle of the primacy of the update, then there could be a third candidate subset $\Gamma_3^2 = \{(a_2, (+2)) : D\}$.

[19] One can see this as minimal change *subject to* persistence.

The selection algorithm is followed by the reinstatement algorithm to reactivate formulae previously made inactive if they no longer generate inconsistency in the newly revised base. In the example above, the formulae in K_2 are all active, so the reinstatement algorithm does nothing.

One of the main characteristics of this approach is that it defines the revision process in terms of the history of the sentences in the belief base. The principle of primacy of the update is not necessarily satisfied by the selection algorithm, as it is also the case for the local semi-revision operation mentioned in Section 5.1 [Hansson, 1997b].

When the new belief is consistent with the belief base, the controlled revision algorithm behaves as other existing approaches for base revision. However, when the new belief is not consistent with the belief base, it is straightforward to show that postulate (R1) is not necessarily satisfied since the selection algorithm could well reject the update information (in fact, it will always do so if the update is itself inconsistent). For the same reason, postulate (R5) is also not satisfied. Consider any base K revised by a formula A and expanded by the formula $\neg A$. The result will be a base K' from which $l : \bot$ could be inferred with some label l. On the other hand, revising K by $A \wedge \neg A$ will result in a belief base K'' from which $\bot$ *cannot* be inferred. Postulates (R2), (R3), (R4) and (R6) are all satisfied.

The non-required enforcement of the principle of the primacy of the update makes the controlled revision approach in general different from the approaches proposed in [Darwiche and Pearl, 1997; Freund and Lehmann, 1994]. Controlled revision satisfies Lehmann's postulates (I1) and (I3)–(I6) and Darwiche and Pearl's ($R^\star 2$), ($R^\star 3$), ($R^\star 4$) and ($R^\star 6$). However, it fails to satisfy Darwiche and Pearl's ($R^\star 1$) and ($R^\star 5$) and Lehmann's (I2) and (I7), for the same underlying reason that prevented the satisfaction of (R1): contradicting information may not necessarily be accepted in the revision process.

For further details, the reader is referred to [Gabbay *et al.*, 2003].

5.3 Multiple belief revision

All belief revision approaches described so far are defined for the revision of a belief set (or belief base) by a single formula, although sometimes a sequence of such revisions is considered as part of an iterative revision process. We now consider formalisms that are defined for revisions of belief sets or bases by a *set* of formulae. We will refer to this class of formalisms as *multiple belief revision*.

Several approaches have been proposed to extend the AGM theory of belief revision so that it can handle revisions by sets of formulae. Some depart from the principle of primacy of the update and compute the revision based on the concept of an *explanation*, e.g., [Falappa *et al.*, 2002]; whereas others provide a general framework for multiple revision including the cases of revisions and contractions by infinite sets of formulae [Zhang, 1996; Zhang *et al.*, 1997; Zhang and Foo, 2001]. In this section, we will present these ideas in some detail and relate them to the original AGM theory described in the beginning of this chapter.

Explanation-based belief revision

As mentioned before, the principle of the primacy of the update requires that the new information is always accepted by a revision operation. However, we have seen that some approaches to belief revision have departed from this principle. Hansson called such operations *semi-revisions* [Hansson, 1997a]. Examples of semi-revisions include the controlled revision approach described in Section 5.2; the revision operators based on a *non-insistent policy* described in [Gabbay *et al.*, 2010] and the *explanation-based belief revision* operation proposed by Falappa [Falappa *et al.*, 2002], which we now present.

Explanation-based belief revision is based on the idea that new beliefs are supported by *explanations*. When new information that contradicts the current set of beliefs is received, the explanation for the information is considered and evaluated with respect to the current set of beliefs. If the explanation "resists the argument" then it is incorporated into the belief set and consequently the new belief is also accepted.

By definition, an *explanation* for a formula φ is a minimal consistent set of sentences Γ whose closure under the consequence relation properly includes all consequences of φ ($\mathrm{Cn}(\{\varphi\}) \subset \mathrm{Cn}(\{\Gamma\})$). Note that an explanation Γ for a sentence φ cannot be the singleton set $\{\varphi\}$ itself. To take into account the notion of explanation given above, a revision operation has to allow revision of a belief base by *sets* of sentences. In order to do this, Falappa proposes a set of postulates, some of which are direct generalisations of the original AGM ones, whereas others are stricter versions. We only present the key postulates below. In all postulates, K is a belief base; Γ, Γ_1 and Γ_2 are (finite) sets of sentences; φ is a formula and $K \circ_F \Gamma$ is the explanation-based revision of K by Γ. For a full account of the framework the reader is referred to [Falappa *et al.*, 2002].

(Vacuity) If $K \cup \Gamma \nvdash \bot$ then $K \circ_F \Gamma = K \cup \Gamma$

(Consistency)	If $\Gamma \not\vdash \bot$ then $K \circ_F \Gamma \not\vdash \bot$
(Congruence)	If $K \cup \Gamma_1 = K \cup \Gamma_2$ then $K \circ_F \Gamma_1 = K \circ_F \Gamma_2$
(Strong consistency)	$K \circ_F \Gamma \not\vdash \bot$
(Weak Success)	If $K \cup \Gamma \not\vdash \bot$ then $\Gamma \subseteq K \circ_F \Gamma$
(Core Retainment)	If $\varphi \in (K \cup \Gamma) \backslash (K \circ_F \Gamma)$ then there is a set H such that $H \subseteq (K \cup \Gamma)$, H is consistent but $H \cup \{\varphi\}$ is inconsistent
(Relevance)	If $\varphi \in (K \cup \Gamma) \backslash (K \circ_F \Gamma)$ then there is a set H such that $(K \circ_F \Gamma) \subseteq H \subseteq (K \cup \Gamma)$, H is consistent but $H \cup \{\varphi\}$ is inconsistent
(Inclusion)	$K \circ_F \Gamma \subseteq K \cup \Gamma$
(Reversion)	If $K \cup \Gamma_1$ and $K \cup \Gamma_2$ have the same minimally inconsistent subsets then $(K \cup \Gamma_1) \backslash (K \circ_F \Gamma_1) = (K \cup \Gamma_2) \backslash (K \circ_F \Gamma_2)$

One can easily see that by taking the conjunction of the sentences in K and Γ, the first three postulates above are essentially equivalent to Katsuno and Mendelzon's postulates (R2)–(R4) given in Section 2.4. The *weak success postulate* is new. It states that the explanation Γ for the new belief is accepted after the revision only when it is consistent with the initial base. This is a weakening of postulate (R1), which requires instead that the new belief is always accepted in the revised base. The *core retainment postulate* requires that nothing should be removed from the belief base or the set of sentences unless the removal is necessary to establish consistency in the revised belief base. The *relevance postulate* is a stronger version of core retainment. The *reversion postulate* states that if expanding a belief base by two different sets of formulae gives the same collection of minimal inconsistent subsets, then the sentences removed from the base in each respective revision should be the same.

The basic mechanism for revising a belief base K by a set of sentences Γ with *partial acceptance* consists of first expanding K by Γ, and thus provisionally accepting the new belief (entailed by Γ), and then eliminating from $K \cup \Gamma$ all possible inconsistencies. Note that the second phase may result in the new belief no longer being entailed.

Restoration of consistency to an inconsistent set $K \cup \Gamma$ is done by defining an *incision function* σ that returns a set containing formulae from each of the minimal inconsistent subsets of $K \cup \Gamma$ and then removing this set from $K \cup \Gamma$ to define a non-prioritised belief revision operation. Note that Fallapa only requires that an arbitrary formula is removed from *each* of these minimally inconsistent subsets. This will not necessarily correspond to the minimal contraction of $K \cup \Gamma$ needed to restore consistency, as there might be formulae in the intersection of some of these minimally inconsistent subsets — these would be the ideal candidates in order to achieve minimal loss of information. We will return to this point after the following example illustrating the process.

EXAMPLE 81. Consider the belief base $K = \{t, u, r, r \to s\}$ and the explanation $\Gamma = \{\neg t, p, p \to \neg s\}$ for a belief $\neg s$. Clearly, the new belief and its explanation are inconsistent with K. The first step is to consider the expanded base $K \cup \Gamma$ given below

$$K \cup \Gamma = \{t, u, r, r \to s, \neg t, p, p \to \neg s\}.$$

A revision function would attempt to remove from $K \cup \Gamma$ as little information as possible so as to re-establish its consistency (as required by the core retainment postulate). There are only two minimally inconsistent subsets of $K \cup \Gamma$ in this case, namely $\{t, \neg t\}$ and $\{p, p \to \neg s, r, r \to s\}$. Each of these sets is called a $\perp$-*kernel* set of $K \cup \Gamma$. The collection of all such sets, namely $\{\{t, \neg t\}, \{p, p \to \neg s, r, r \to s\}\}$, is denoted $(K \cup \Gamma)^{\amalg}\perp$.

Note that removing sentences from just some of the $\perp$-kernel sets of $K \cup \Gamma$ is not sufficient to re-establish consistency. On the other hand, removing from $K \cup \Gamma$ the union of all $\perp$-kernel sets removes more information than is necessary. In this example, for instance, it would result in the revised belief base $\{u\}$ — a revision that does not satisfy the relevance postulate. A more "economic" solution would be to pick a formula from each of the minimally inconsistent subsets, namely to "cut" from each $\perp$-kernel set enough information so as to establish consistency of $K \circ_F \Gamma$. As an example, a possible incision function σ on the set $(K \cup \Gamma)^{\amalg}\perp$ could return the set $\sigma((K \cup \Gamma)^{\amalg}\perp) = \{p, t\}$. Each element in $\sigma((K \cup \Gamma)^{\amalg}\perp)$ is taken from a different $\perp$-kernel set of $K \cup \Gamma$. More formally, for each $X \in (K \cup \Gamma)^{\amalg}\perp$, $\sigma((K \cup \Gamma)^{\amalg}\perp) \cap X \neq \emptyset$. In this case, the revised base would then be defined as $K \circ_F \Gamma = K \cup \Gamma \backslash \sigma((K \cup \Gamma)^{\amalg}\perp) = \{u, r, r \to s, \neg t, p \to \neg s\}$.

Note that in the above example the revised base does not explicitly include the new belief $\neg s$ and hence explanation-based revision does not guarantee the success of the revision operation (in AGM terms), but the revised set can in general include at least *part* of the new belief's explanation.

We can now return to our previous remark on the failure of this type of revision's satisfaction of the principle of informational economy. Consider the set $K \cup \Gamma = \{\neg p, p, p \to q, \neg q\}$. In this case, $(K \cup \Gamma)^{\amalg}\bot = \{\{\neg p, p\}, \{p, p \to q, \neg q\}\}$. An arbitrary incision function σ could return the set $\{\neg p\} \cup \{p\}$ (one formula from each $X \in (K \cup \Gamma)^{\amalg}\bot$). However, in this particular case, the removal of p alone would be sufficient.

The explanation-based revision illustrated above was formally called *kernel revision* by Falappa in [Falappa *et al.*, 2002]. It uses the following notion of a kernel set proposed by Hansson in [Hansson, 1994].

DEFINITION 82 (Kernel set). Let K be a set of formulae and φ a formula. The set of φ-kernel sets of K, in symbols $K^{\amalg}\varphi$, is defined as the set of all $K' \subseteq K$ such that

① $K' \vdash \varphi$

② if $K'' \subset K'$ then $K'' \nvdash \varphi$.

In order to ensure that each minimally inconsistent subset is sufficiently contracted so that $K \circ_F \Gamma$ is consistent, an *incision function* is defined by selecting *some* elements from each one of the $\bot$-kernel sets of $K \cup \Gamma$.

DEFINITION 83 (External incision function). Let K be a set of sentences. An *external incision function* for K is a function σ such that for any (finite) set of sentences Γ:

① $\sigma((K \cup \Gamma)^{\amalg}\bot) \subseteq \bigcup((K \cup \Gamma)^{\amalg}\bot)$

② If $X \in (K \cup \Gamma)^{\amalg}\bot$ and $X \neq \emptyset$ then $(X \cap \sigma((K \cup \Gamma)^{\amalg}\bot)) \neq \emptyset$

DEFINITION 84 (Kernel revision). Let K and Γ be finite sets of sentences and σ an external incision function for K. The *kernel revision* of K by Γ, in symbols, $K \circ_F \Gamma$ is defined as

$$K \circ_F \Gamma = (K \cup \Gamma) \backslash \sigma((K \cup \Gamma)^{\amalg}\bot)$$

Falappa has showed that an operator $\circ_F$ is a kernel revision operation if and only if it satisfies the postulates of inclusion, strong consistency, core retainment and reversion (see [Falappa *et al.*, 2002] for details).

Revision by finite sets of formulae

The idea of *multiple belief change* was initially proposed by Fuhrmann [Fuhrmann, 1988] and eventually evolved into a framework called *package contraction.* This defines the contraction of a belief set K by a set of formulae F as the operation of removing the set F as a whole from K [Fuhrmann and Hansson, 1994]. Later on, Zhang proposed a full generalisation and extension of the AGM revision postulates to deal with revisions by finite and infinite sets [Zhang, 1996]. His generalised contraction operator, called *set contraction*, contracts a belief set K by a set of formulae F, in symbols $K \ominus F$, by removing from K enough sentences so as to make the remaining subset $K \ominus F \subseteq K$ closed under the consequence relation and *consistent* with F. Obviously, these two operations produce different results. We illustrate this with an example.

EXAMPLE 85. Suppose that the belief set $K = \mathrm{Cn}(\{\neg\alpha, \neg\beta, \gamma\})$ is to be contracted by the set $F = \{\alpha, \neg\beta, \gamma\}$. A possible set contraction $K \ominus F$ in this case could be the set $K' = Cn(\{\neg\beta, \gamma\})$, since $K' \cup F$ is consistent. On the other hand, a package contraction of K by F would instead remove $\neg\beta$ and γ (and their logical consequences) from K, giving as a result the set $\mathrm{Cn}(\{\neg\alpha\})$.

Set contraction can be seen as an operation for selecting a consistent subset of a set $K \cup F$. If F contains a single non-contradictory formula A, this amounts to removing from K any information that proves $\neg A$. Informally, the set contraction of K by A in this particular case can be achieved by the contraction of K by $\neg A$ (in symbols, $K - \neg A$). As expected, if K is already consistent with F, the set contraction of K by F would simply return K itself. The motivation for set contraction is the rationale that beliefs ought to be contracted only when they conflict with new information.

Some postulates for set contraction were proposed in [Zhang, 1996] and subsequently shown to be insufficient for the characterisation of contraction by infinite sets [Zhang *et al.*, 1997].

Zhang also proposed a number of postulates for a second belief change operation to revise a belief set K by a (possibly infinite) set of formulae F in [Zhang, 1996]. The operation was called *set revision* and denoted by $K \oplus F$. The postulates are direct generalisations of the AGM revision postulates.

Revision postulates for set revision

(K$^\oplus$1) $K \oplus F = \mathrm{Cn}(K \oplus F)$
(K$^\oplus$2) $F \subseteq K \oplus F$
(K$^\oplus$3) $K \oplus F \subseteq K + F$
(K$^\oplus$4) If $K \cup F$ is consistent, then $K + F \subseteq K \oplus F$
(K$^\oplus$5) $K \oplus F = K_\perp$ iff F is inconsistent
(K$^\oplus$6) If $\mathrm{Cn}(F_1) = \mathrm{Cn}(F_2)$, then $K \oplus F_1 = K \oplus F_2$
(K$^\oplus$7) $K \oplus (F_1 \cup F_2) \subseteq K \oplus F_1 + F_2$
(K$^\oplus$8) If $F_2 \cup (K \oplus F_1)$ is consistent, then $(K \oplus F_1) + F_2$
$\subseteq K \oplus (F_1 \cup F_2)$

Zhang showed in [Zhang, 1996] that when F is a singleton set (K$^{\oplus}$1) – (K$^{\oplus}$8) are equivalent to (K$^{\circ}$1) – (K$^{\circ}$8).[20] One may wonder about the relationship between set revisions and set contractions. This can be investigated by extending the Levi and Harper identities to deal with sets as the second argument of the operations. More specifically:

$$K \oplus F = (K \ominus F) + F \qquad \text{(Extended Levi identity)}$$
$$K \ominus F = (K \oplus F) \cap K \qquad \text{(Extended Harper identity)}$$

According to Zhang, the revision of a belief set K by a set of sentences F can be performed by removing enough information from K so as to attain consistency with F and then subsequently expanding the result by F (this expansion is assumed to be closed under the consequence relation).

The full semantical framework for set revision and set contraction can be found in [Zhang and Foo, 2001]. The framework generalises the representation theorems of the AGM postulates (see beginning of this chapter or [Alchourrón *et al.*, 1985]). Whilst the semantical results were initially defined for the set contraction operation, it is their corresponding formulation for set revision that will be referred to here, which can be directly obtained via the extended Levi identity mentioned above [Zhang *et al.*, 1997].

We still need to define what set of formulae $K' \subseteq K$ need to be removed from K in order to make it consistent with an input set. The problem of constructing such a K' constitutes both the result of performing a set contraction of K by an input set F as well as the intermediate result necessary in the definition of a set revision of K by F. A possible starting point is the collection of all maximal subsets of K that are consistent with F. We will use the symbol $K \parallel F$ to denote this collection. Note that when F is a singleton set $\{A\}$, $K \parallel F$ corresponds to the familiar collection of maximally consistent subsets of K that fail to imply A, in symbols, $K_{\perp}\neg A$ (see Definition 2). Each element in $K \parallel F$ is therefore a potential candidate for our K'. However, given a belief set K and a set of formulae F, in general there will be many such maximal subsets. The usual notion of a *selection function* S is used in the process.

DEFINITION 86 (Selection function for multiple belief revision). Given a belief set K and a set of formulae F, a *selection function* S is such that:

① $S(K \parallel F) \subseteq K \parallel F$

[20] This can be done by defining $K \circ_a \phi = K \oplus F$.

② $K \parallel F \neq \emptyset$ implies $S(K \parallel F) \neq \emptyset$

Different selection functions can be defined by imposing additional constraints. For instance, one can require that S return a singleton set, i.e., $S(K \parallel F)$ would choose *one* of the maximal subsets of K that are consistent with F. Correspondingly, this would yield a *maxichoice set revision* function, defined in terms of the *maxichoice set contraction* given below.

DEFINITION 87. Let K be a belief set, F a set of sentences and S a selection function such that $S(K \parallel F)$ is a singleton set. A *maxichoice set contraction* of K by F ($K \ominus F$) is defined as:

$$K \ominus F = \begin{cases} K, & \text{if } F \text{ is inconsistent} \\ K' \in S(K \parallel F), & \text{otherwise} \end{cases}$$

The set revision function defined in terms of a maxichoice set contraction function satisfies postulates ($\mathrm{K}^{\oplus}1$) – ($\mathrm{K}^{\oplus}5$), a weaker version of postulate ($K^{\oplus}6$), and the following condition: (using the extended Levi identity and Theorem 3.14 of [Zhang and Foo, 2001]):

If $A \in K \backslash K \ominus F$, then there exists $B \in K$ such that $F \vdash \neg B$ and $A \rightarrow B \in K \ominus F$

However, as we have seen to be the case in the original AGM formulation, a maxichoice set contraction function produces belief sets that are *too large*, and thus the corresponding set revisions are also maximal (see page 7). An alternative approach is to make the selection function return the entire set $K \parallel F$, i.e., $S(K \parallel F) = K \parallel F$ and define the set contraction function as the intersection of all of its elements.

DEFINITION 88. Let K be a belief set, F a set of sentences and S a selection function such that $S(K \parallel F) = K \parallel F$. A *full meet set contraction* of K by F ($K \ominus F$) is defined as:

$$K \ominus F = \begin{cases} K, & \text{if } F \text{ is inconsistent} \\ \bigcap(K \parallel F) & \text{otherwise} \end{cases}$$

Zhang and Foo showed that set revisions defined in terms of full meet set contractions satisfy postulates ($\mathrm{K}^{\oplus}1$) – ($\mathrm{K}^{\oplus}5$); a weaker version of postulate ($\mathrm{K}^{\oplus}6$) and the following condition: (using the extended Levi identity and Theorem 3.14 of [Zhang and Foo, 2001]):

If $F_1 \subseteq F_2$ and $F_1 \cup K$ is inconsistent, then $K \ominus F_1 \subseteq K \ominus F_2$

Here again, similarly to what happens in the original AGM formulation, full meet set contractions yield belief sets that are *too small*, thus generating the smallest set revision of a belief set K by a set of formulae F. The result of the intersection would include only the sentences that belong to all of K's maximal subsets that are consistent with F. Maxichoice set contraction and full meet set contraction have therefore opposite drawbacks. However, they provide upper and lower bounds for any set contraction operation (and hence to a set revision operation too).

An intermediate solution, as expected, is the definition of a *partial meet set contraction* using the basic notion of a selection function given in Definition 86.

DEFINITION 89. Let K be a belief set, F a set of sentences and S a selection function as given in Definition 86. A *partial meet set contraction* of K by F ($K \ominus F$) is defined as:

$$K \ominus F = \begin{cases} K, & \text{if } F \text{ is inconsistent} \\ \bigcap S(K \parallel F), & \text{otherwise} \end{cases}$$

The selection function chooses *some* of the maximal subsets of K that are consistent with F and the partial meet contraction takes the intersection of such subsets. Set revisions defined in terms of partial meet contractions can be shown to satisfy postulates ($\mathrm{K}^{\oplus}1$) – ($\mathrm{K}^{\oplus}5$) and a weaker version of postulate ($K^{\oplus}6$). In order to provide a set contraction operation (and therefore also a set revision) that satisfies postulates ($\mathrm{K}^{\oplus}7$) and ($\mathrm{K}^{\oplus}8$) as well, it is necessary to consider the notion of entrenchment over the maximal subsets of a given belief set K. In this way, the selection function S will pick from $K \parallel F$ only those maximal subsets that are *epistemologically most entrenched.* This is captured by a relation $\leq$ on $\mathcal{U}_K = \bigcup\{K \parallel F \mid F \subseteq \mathsf{wff}(\mathcal{L})\}$, where $\mathsf{wff}(\mathcal{L})$ is the set of all well-formed formulae of $\mathcal{L}$ and $Y \leq X$, represents the fact that X is at least as epistemologically entrenched as Y.

DEFINITION 90. Let K be a belief set, F be a set of sentences and $\leq$ a relation on $\mathcal{U}_K$. A selection function S is said to be *relational* if:

$$S(K \parallel F) = \{X \in K \parallel F \mid \forall Y \in K \parallel F.\, Y \leq X\}$$

Zhang and Foo present some conditions under which a set revision operator $\oplus$ derived via the Extended Levi identity in terms of a relational partial meet contraction operator satisfies postulates ($K^{\oplus}1$) – ($\mathrm{K}^{\oplus}8$). For more details, the reader is referred to [Zhang and Foo, 2001].

Revision by infinite sets of formulae

As we mentioned above, the semantical characterisation just presented does not deal with infinite revision (or contraction). This is defined axiomatically by an additional postulate called the *limit postulate* expressing the relationship between the contraction of a belief set K by an infinite set of formulae F. The limit postulate ($K^{\oplus}LP$) is what in fact really *extends* the basic AGM theory to cover revisions by infinite sets. The full set of postulates for the infinite case is therefore given by ($\mathrm{K}^{\oplus}1$)–($\mathrm{K}^{\oplus}8$), where the sets F, F_1 and F_2 may be infinite, plus the limit postulate below.

$$(K^{\oplus}LP)\ K \oplus F = \bigcup_{\bar{F} \subseteq F}\ \bigcap_{\bar{F} \subseteq \bar{F}' \subseteq_f \mathrm{Cn}(F)} K \oplus \bar{F}'$$

$K^{\oplus}LP$ states that a revision of a belief set K by an infinite set of formulae F can be performed by revising K by all finite subsets $\bar{F}'$ of $\mathrm{Cn}(F)$ (in symbols, $\bar{F}' \subseteq_f \mathrm{Cn}(F)$).

A semantical characterisation of the limit postulate is based on the concept of a *nice-ordering partition*, in its turn defined in terms of a *total-ordering partition*. A total-ordering partition is a tuple $\langle \Gamma, \Pi, < \rangle$ where Γ is a set of formulae (e.g., a belief set); Π is a partition of Γ; and $<$ is a strict relation induced by a total order on Π. The *rank* of a formula $A \in \Gamma$ is defined as the set $\Delta \in \Pi$ such that $A \in \Delta$ and denoted by $b(A)$. Given a relation $<$, the higher the degree a belief has, the lower in the ordering its ranking will be. A nice-ordering partition is a total-ordering partition that satisfies the following property:

$$\text{If } A_1, \ldots, A_n \vdash B, \text{then} \max\{b(A_1), \ldots, b(A_n)\} \geq b(B)$$

In other words, a nice-ordering partition guarantees that the degree of belief of a formula is not higher than the maximum degree of belief of the formulae that entail it.[21]

Given a nice-ordering partition, it is possible to define a *nice-ordering partition contraction* function, which, together with the Levi identity, gives a set revision function that satisfies postulates $K^{\oplus}1$–$K^{\oplus}8$ [Zhang, 1996]. The remaining question is whether a nice-ordering partition is sufficient for the construction of a set contraction function (and therefore a set revision function as well) that also satisfies the limit postulate $K^{\oplus}LP$.

[21]Of course, this is related to AGM's original notion of epistemic entrenchment. In fact, Zhang showed the relationship between epistemic entrenchment relations $\leq_{EE}$ and nice-ordering partitions by considering beliefs with lower rank as being more entrenched [Zhang, 1996].

A possible way to define the contraction operation would be to consider the contraction of a belief set K by an infinite set F as the *limit* of set contractions of K by all *finite* subsets of F. Formally, this is defined as

$$lim_{\bar{F} \subseteq_f F} \, K \ominus \bar{F}$$

However, the limit of a set sequence exists only when the *limit superior* and the *limit inferior* of the same set sequence exist and are equal, which is not necessarily the case here. The lack of convergence is partly due to the fact that each individual operation using a finite subset $\bar{F}$ of F depends on the syntax of $\bar{F}$. The sequence over the sets $\bar{F} \subseteq \mathrm{Cn}(F)$ also does not converge [Zhang, 1996]. Alternatively, one could use either the limit inferior or the limit superior of the contractions of K by the finite subsets of F. To explain which of these two possibilities is most appropriate, let us briefly recap how these two limits are defined.

DEFINITION 91. Let $\{\Gamma_i : i \in \mathbb{N}\}$ be a infinite sequence of sets. The *limit superior* and *limit inferior* of this sequence are defined respectively by:

$$\overline{lim}_{i \to \infty} \Gamma_i = \bigcap_{i=0}^{\infty} \bigcup_{j=i}^{\infty} \Gamma_j \qquad \underline{lim}_{i \to \infty} \Gamma_i = \bigcup_{i=0}^{\infty} \bigcap_{j=i}^{\infty} \Gamma_j$$

It can be seen that the limit superior would not necessarily give a set closed under the consequence relation, whereas the limit inferior would. Contractions (resp., revisions) of a belief set K by an infinite set of formulae F are hence defined as the limit inferior of the sequence of contractions (resp., revisions) of K by all finite subsets of F. Formally,

$$K \ominus F = \bigcup_{\bar{F} \subseteq \mathrm{Cn}(F)} \; \bigcap_{\bar{F} \subseteq \bar{F}' \subseteq_f \mathrm{Cn}(F)} K \ominus \bar{F}'$$

Intuitively, given a belief set K and an infinite set of sentences F, a formula B will be included in the contraction of K by F, in symbols $B \in K \ominus F$, if there exists a finite subset $\bar{F}$ of F such that B is in the contraction of K by each finite superset of $\bar{F}$ included in $\mathrm{Cn}(F)$. The above is the actual definition of the limit postulate $K^{\ominus}LP$ for set contraction.

Analogously,

$$K \oplus F = \bigcup_{\bar{F} \subseteq F} \; \bigcap_{\bar{F} \subseteq \bar{F}' \subseteq_f \mathrm{Cn}(F)} K \oplus \bar{F}'$$

As expected, this is the limit postulate $K^{\oplus}LP$ for revisions given previously.

In terms of limit inferior these can be given as

$$K \ominus F = \underline{lim}_{\bar{F} \subseteq_f \mathrm{Cn}(F)} K \ominus \bar{F} \text{ and}$$

$$K \oplus F = \underline{lim}_{\bar{F} \subseteq_f \mathrm{Cn}(F)} K \oplus \bar{F}$$

In summary, the limit postulate essentially provides an approximation of the revision by an infinite set in terms of revisions by finite subsets. Using this postulate it is possible to show that infinite belief change operations can be reduced to finite belief change operations. In particular, if $\oplus$ is a belief revision operator satisfying the postulates $K^{\oplus}1$–$K^{\oplus}8$, then the limit postulate $K^{\oplus}LP$ and the following requirement are equivalent [Zhang, 1996]:

$$K \oplus F = \bigcap_{\bar{F} \subseteq_f \mathrm{Cn}(F)} K \oplus \bar{F}$$

Partial meet semantics for multiple revision

An alternative semantical approach to revision by sets of formulae can be given in terms of the notion of a *partial meet revision* [Falappa *et al.*, 2002]. The underlying mechanism of this revision is to add to a belief base K an input set of formulae Γ and then remove from the (possibly inconsistent) union as few formula as possible so as to restore consistency. This will result in a collection of maximal consistent subsets of $K \cup \Gamma$, which is again denoted by $(K \cup \Gamma)_{\perp}\perp$.

By using an appropriate selection function, a choice can be made between the maximal consistent subsets of $K \cup \Gamma$. The revised base is subsequently constructed from the intersection of the selected maximal consistent subsets.[22] The selection function is in principle defined according to the revision postulates that it needs to satisfy. In the case of the $\circ_F$ revision operator such a function is called *equitable* and defined as follows.

DEFINITION 92 (Equitable selection function). Let K be a belief base. A selection function γ for K is *equitable* if $(K \cup A)^{\amalg}\perp = (K \cup B)^{\amalg}\perp$ implies that $(K \cup A) \cap \gamma(K \cup A)_{\perp}\perp = (K \cup B) \cap \gamma(K \cup B)_{\perp}\perp$.

The intuition behind the above definition is that if the set of minimally inconsistent subsets of $K \cup A$ is equal to the set of minimally inconsistent subsets of $K \cup B$ then the same sentences can be eliminated from $K \cup A$ and $K \cup B$ in order to produce consistent revised bases with A and with B, respectively.

[22]The notion of maximal consistent subset here is related to the notion presented in Definition 2, except that K is not closed under the consequence relation and the input is a set.

Partial meet revisions are then defined as follows.

DEFINITION 93. Let K be a belief base, Γ a set of sentences and γ an equitable selection function for K. Then $K \circ_F \Gamma = \bigcap \gamma((K \cup A)_\perp \perp)$.

This type of partial meet revision can be related to the previously presented partial meet contraction and revision operations proposed by Zhang. However, note that $\circ_F$ does not necessarily satisfy the principle of the primacy of the update, whereas Zhang's operation $\oplus$ does. Consider the following example.

EXAMPLE 94. Let $K = \{a, b, \neg c\}$, $\Gamma = \{\neg b, c, d, e\}$ and consider what the partial meet revisions $K \oplus \Gamma$ and $K \circ_F \Gamma$ would give in this case.

$K \oplus \Gamma = \bigcap s(K \parallel \Gamma) + \Gamma$. The set $K \parallel \Gamma = \{\{a\}\}$ in this case. Hence, independently of the selection function s used, $\bigcap s(K \parallel \Gamma) = \{a\}$, which expanded by Γ would result in $K \oplus \Gamma = \{a, \neg b, c, d, e\}$.

In the case of $K \circ_F \Gamma$, the operation would consider first the set $K \cup \Gamma = \{a, b, \neg c, \neg b, c, d, e\}$, which is inconsistent. The remainder set with respect to $\perp$ would, in this case, be $(K \cup \Gamma)_\perp \perp = \{\{a, b, \neg c, d, e\}, \{a, b, c, d, e\}, \{a, \neg c, \neg b, d, e\}, \{a, c, \neg b, d, e\}\}$. Depending on the selection function used to pick some of these sets, the resulting revision may or may not include Γ. For instance, if the selection function γ picks just $\{a, c, \neg b, d, e\}$, then $K \circ_F \Gamma = \{a, c, \neg b, d, e\} = K \oplus \Gamma$. However, a different selection function could also select the two maximal consistent subsets $\{\{a, c, \neg b, d, e\}, \{a, c, b, d, e\}\}$, in which case $K \circ_F \Gamma = \{a, c, d, e\}$, and hence $K \circ_F \Gamma \neq K \oplus \Gamma$. Note that $\circ F$ would not verify the principle of the primacy of the update since $K \circ_F \Gamma \not\vdash \neg b$ and $\neg g \in \Gamma$.

The above example shows that only in some cases does Falappa's partial meet revision agree with Zhang's multiple belief revision. In general terms, Falappa's revision approach falls into the category of non-insistent revisions whereas Zhang's approach to multiple revision always satisfies the principle of the primacy of the update. In the cases where the two revision operations coincide, further correspondences could be drawn by appropriately constraining the equitable selection functions so as to capture the various partial meet contraction properties given before.

5.4 Revision by translation

An altogether different approach to belief revision was proposed in [Gabbay *et al.*, 2000; Rodrigues *et al.*, 2008]. It provided a uniform framework for revising theories of non-classical logics. The general idea is to define a belief revision operator for these logics in terms of a standard belief revision operator for classical logic. Typically, given a classical logic theory Δ and an input formula ψ, a revision process $\circ_a$ gives a new theory $\Gamma = \Delta \circ_a \psi$, corresponding to the revision of Δ by ψ and satisfying some desirable properties. In many belief revision approaches, these properties are often considered to be the AGM postulates [Gärdenfors, 1988], whereas in database updates the notion of *integrity constraints* often appears [Katsuno and Mendelzon, 1991a; Katsuno and Mendelzon, 1992]. Gabbay et. al. propose the notion of *acceptability* [Gabbay *et al.*, 2000] used to *export* the AGM revision process to non-classical logics whose semantical properties are axiomatisable in first-order logic.

The approach consists of translating the underlying semantics of a given object logic L into classical logic, performing the revision process in classical logic and then translating the results back into the original object logic. Let τ denote a translation function from L into classical logic, let $\mathcal{A}_L$ be a sound and complete axiomatisation of the semantical features of L in classical logic, and Acc a first-order characterisation of *acceptable* L-theories. A revision operator $*_L$ in the logic L can be defined as follows:

(4) $\Delta *_L \psi = \{\beta \mid \Delta^\tau \circ_a (\psi^\tau \wedge \mathcal{A}_L \wedge Acc) \vdash \beta^\tau\}$

where Δ^τ, and ψ^τ are the classical logic translations of Δ and ψ respectively.

The revision of Δ by ψ is therefore defined in terms of the revision (in classical logic) of Δ^τ by ψ^τ, $\mathcal{A}_L$ and Acc. The presence of $\mathcal{A}_L$ in the revising part ensures that the semantical properties of the object logic L are preserved by the revision operation and allows that the revised theory can be mapped back into the original logic L. Similarly, the formula (Acc) represents additional acceptability conditions of the revised theory. These conditions can be used to capture semantical differences between the object and the target logics as well as to express application driven constraints.

For non-classical logics that are extensions of classical logic, e.g., modal logics, the notion of inconsistency is identical to the notion of inconsistency of the target (classical) logic. If a given formula ψ is inconsistent with a given theory Δ, then ψ^τ would also be inconsistent (in classical logic) with Δ^τ. Hence the classical revision would be able to revise the (translated) theory and provide a revised theory that is also consistent in the object logic. For such logics, Acc could be simply $\top$ and the revision operator $*_L$ defined in a more straightforward manner as

$$\Delta *_L \psi = \{\beta \mid \Delta^\tau \circ_a (\psi^\tau \wedge \mathcal{A}_L) \vdash \beta^\tau\}.$$

This is the approach adopted and discussed in [Gabbay *et al.*, 2000]. However, even in these cases, domain-dependent notions of acceptability could still be provided to further refine the revision process, as it is done, for instance, with integrity constraints in database updates.

In the case of logics that are not extensions of classical logic, e.g., Belnap's four-valued logic [Belnap, 1977a; Belnap, 1977b], the notion of consistency may differ from that of classical logic. Classical inconsistency of a set of formulae is associated with the non-existence of a valuation that makes formulae in the set true — i.e., the non-existence of *models*. Object theories may not necessarily yield to translated theories that are inconsistent in the classical sense. For instance, the formula $p \wedge \neg p$, is consistent within the context of Belnap's four-valued logic, as it accepts semantic valuations (*set-ups* in Belnap's terminology) that assign the value *both* to the formula. A "model" of the classical-logic translation of this statement would therefore be a first-order interpretation that satisfies the *truthfulness* of the correspondence between Belnap's truth-values of p, $\neg p$ and $p \wedge \neg p$ with respect to the intended meaning of the $\wedge$ connective in the underlying object logic.

Such a translation in classical logic would bind the value "true" for p in a (Belnap) valuation, with the value "false" for $\neg p$ in the valuation and the value "both" for $p \wedge \neg p$ (i.e. "both true and false"). It is then not surprising the fact that, if the translation to classical logic of the underlying object-level semantics is sound and complete, then $\{p^{\tau}\}$ is consistent in classical logic with $(\neg p)^{\tau}$, and consequently, the revision of p by $\neg p$ via the classical logic translation would simply result in the conjunction $\{p \wedge \neg p\}$. This is because the AGM revision mechanism heavily relies upon the notion of (classical) consistency: AGM revision is triggered only when the new information is inconsistent with the current belief set (see postulate $(\mathrm{K}^{\circ}_{3,4})$).

As a result, for certain types of non-classical logics which do not possess a similar notion of consistency, revisions done via translation would be equivalent to simple expansions of the belief set, as it is the case in [Restall and Slaney, 1995]. For these logics, the use of the acceptability theory *Acc* enables the revision by translation mechanism to capture more refined revision processes without necessarily embedding the revision mechanism into the object logic itself, as is the case, for instance, in the system **R** [Martins and Shapiro, 1988; Shapiro, 1992]. The view adopted in [Gabbay *et al.*, 2000] is that a revision process for such logics (including paraconsistent logics) benefits from a shift towards the notion of *acceptable belief sets* — the imposition of specific restrictions on what an acceptable (revised) theory should be in the object logic L.

To illustrate this approach the case of modal logic is briefly summarised below. The reader is referred to [Gabbay *et al.*, 2000] where a detailed analysis for other logics is also given.

Different translation mechanisms can be defined from a given object logic to classical logic, depending on the underlying semantical structures used by the object logic. In the case of modal logic — an extension of classical logic — the translation mechanism essentially maps the notion of satisfiability of atomic formulae to first-order predicates with one or more (additional) arguments to encapsulate elements of the underlying semantical structure. For instance, the semantical notion $w \Vdash p$, for a possible world w and atomic formula p can be associated with the first-order formula $P(w)$, and analogously, $w \nVdash p$, with the formula $\neg P(w)$. In addition, appropriate axioms need to be defined to encode the correct behaviour of the connectives and indeed other semantical properties of the object logic, e.g., properties of the accessibility relation [Ohlbach, 1991]. These have to be sound and complete with respect to the notion of entailment of the object logic L. In general, the objective is to ensure that, given a translation function τ and a classical logic axiomatisation of the object level semantics $\mathcal{A}_L$, for any theory Δ and formula α in L the following correspondence holds

$$\Delta \vdash_L \alpha \text{ iff } \mathcal{A}_L \cup \Delta^\tau \vdash \alpha^\tau$$

where Δ^τ and α^τ are the translations to classical logic of Δ and α respectively.[23]

Translating the modal logic K into first-order logic requires a binary predicate R in classical logic to represent the accessibility relation and unary predicates $P_1, P_2, P_3, \ldots$, for each propositional symbol p_i of K. The idea is to encode the information of satisfiability of modal formulae by worlds into the variable of each unary predicate. In general, for a given world w and formula β the translation method can be defined as follows, where $\beta^\tau(w)$ represents $w \Vdash_k \beta$.

$$\begin{array}{rcl}
p_i^\tau(w) & = & P_i(w) \\
(\neg\beta)^\tau(w) & = & \neg(\beta^\tau(w)) \\
(\beta \wedge \gamma)^\tau(w) & = & \beta^\tau(w) \wedge \gamma^\tau(w) \\
(\beta \rightarrow \gamma)^\tau(w) & = & \beta^\tau(w) \rightarrow \gamma^\tau(w) \\
(\Box\beta)^\tau(w) & = & \forall y(wRy \rightarrow \beta^\tau(y))
\end{array}$$

The translation to classical logic of a modal theory Δ is given by the set

$$\Delta^\tau(w) = \{\beta^\tau(w) \mid \beta \in \Delta\}.$$

[23]We use the symbol $\vdash$ without subscript to denote entailment in classical logic and with subscript to denote entailment on a object logic.

The soundness and completeness of the translation function can be easily shown by proving the following property:

(Correspondence)

(5) $\Delta \vdash_k \beta$ iff $\mathcal{A}_L \cup \Delta^\tau(w_0) \vdash \beta^\tau(w_0)$

where for the modal logic K, $\mathcal{A}_L$ is empty (i.e., truth).[24] If Δ is finite, one can simply use the conjunction of its formulae instead ($\delta = \bigwedge \Delta$) and the correspondence between the object modal logic entailment and its translated first-order logic formalisation would be expressed as follows:

(Correspondence reformulated)

$$\delta \vdash_k \beta \ \text{ iff } \ \vdash \forall x(\delta^\tau(x) \to \beta^\tau(x))$$

A revision operator $*_k$ for K is then defined as follow.

DEFINITION 95. [Belief revision in K]

$$\Delta *_k \psi = \{\alpha \mid \Delta^\tau \circ_a (\psi^\tau \wedge T^\tau) \vdash \alpha^\tau\}$$

The following proposition shows that the above revision operator satisfies the AGM postulates. The proof relies on the fact that $*_k$ is defined in terms of $\circ_a$, which satisfies the AGM postulates for the target first-order logic, and that the first-order derivability relation captures the modal logic derivability relation.

PROPOSITION 96. *Let Δ be a belief set and ψ be a formula of the modal logic K. Let $*_k$ be as belief revision operator as specified in Definition 95. Then $*_k$ satisfies the AGM postulates* (K°1)–(K°8).

Proof. For a complete proof the reader is referred to [Gabbay *et al.*, 2000] ■

6 COMPLEXITY ISSUES

The computational complexity associated with the problem of belief revision has been investigated by many authors [Nebel, 1991a; Nebel, 1992; Nebel, 1998; Gärdenfors and Rott, 1995; Eiter and Gottlob, 1992b; Eiter

[24] In the case of a modal logic other than K, $\mathcal{A}_L$ would be non-empty. For instance, in the case of $S4$, $\mathcal{A}_L = \{\forall x(xRx) \wedge \forall x \forall y \forall z(xRy \wedge yRz \to xRz)\}$. Note that this first-order translation approach allows also for non-normal logics, e.g., if w_0 is the actual world, a first-order theory of the form $\mathcal{A}_L = \{w_0 R w_0\}$ would impose reflexivity only on the actual world w_0.

and Gottlob, 1992a]. Postulate (K°5) (or (R3) in the finite case) stipulates that the result of a revision operation is consistent provided that the revising sentence itself is not contradictory. Thus, regardless of whether one considers closed belief sets or belief bases, belief revision is at least as hard as deciding satisfiability of a set of formulae and deciding derivability from a set of formulae [Nebel, 1992].

In addition to that, the principle of minimal change requires that the result of the revision operation retains as much as possible of the information in the previous belief set. The notion of evaluation of change was briefly introduced in Section 3.1 and complexity analyses in general consider two main cases: the one where change is evaluated with respect to models of the belief set and the one where change is evaluated based on the syntactical representation of the set. The two results are in general similar except that in the model-based approaches, the size of the revising formula determines the number of models to analyse and therefore a restriction on it has an impact on the overall complexity of the method in question [Eiter and Gottlob, 1992a; Nebel, 1998].

Eiter and Gottlob's complexity analysis is based on what they call the *implication problem*, i.e., whether a given belief ψ is entailed by a belief set $K \circ \varphi$ (in symbols, $K \circ \varphi \vdash \psi$). This obviously involves two components (whose complexity interacts in the majority of the approaches to belief revision):

1. the inference process itself, i.e., whether a formula ψ is entailed by a belief set

2. the search for a maximal subset of a belief set that is consistent with a given formula

The second component is the one related to the revision operation itself and the principle of minimal change mentioned earlier. In [Eiter and Gottlob, 1992b; Eiter and Gottlob, 1992a], the implication problem is analysed for propositional logic under a number of different scenarios arising from the possible combinations of placing a restriction on the type of formulae allowed (e.g., full propositional formulae or horn clauses only) and on the size of the update formula itself (e.g., unlimited length or length bound by a constant).

Eiter and Gottlob's findings are that the complexity problem for the implication problem resides, in general, at the second level of the polynomial hierarchy (Π_2^P). These results confirm the initial investigations done by Nebel in [Nebel, 1991a], but go further by analysing scenarios under which the complexity can be reduced.[25] It turns out that tractability can only be attained by imposing a restriction on the size of the revising formulae and on the type of formulae allowed in the language — i.e., horn clauses only. Part of this is attributed to the fact that propositional inference for horn clauses is tractable.

Revision mechanisms that use priorities to help in the decision of what beliefs to give up, such as prioritised base revision [Fagin *et al.*, 1983; Nebel, 1991a] (see Section 4.5) in general share the same complexity of those who do not employ priorities, i.e., they are Π_2^p-complete. However, Nebel has proved that the implicational problem $K \circ \varphi \vdash \psi$, for the special case of horn languages and bases which are linearly ordered, can be decided in time $O(n^2)$, where $n = |K| + |\varphi| + |\psi|$.

On the other hand, the problem of generating a revised belief base $K \circ \varphi$ was investigated by Gogic et al. [Gogic *et al.*, 1994]. For a quite comprehensive analysis of the complexity of several satisfiability problems, the reader is referred to [Dalal and Etherington, 1992]. On the more specific topic of the complexity of iterated belief revision, see [Liberatore, 1997b; Eiter and Gottlob, 1993]. The complexity of *update* operations was also analysed by several authors and relevant references include [Liberatore, 1997a; Eiter and Gottlob, 1992b].

7 APPLICATIONS

Applications of belief revision have been proposed for the solution of many computer science problems. In this section, we provide a brief description of the applications that are more closely related to the belief revision approaches presented in this chapter.

7.1 Belief Revision in Requirements Engineering

A key problem in large systems engineering projects is the management of frequently changing requirements, both during initial system development and subsequent maintenance and evolution. It has been noted [Williams, 1998] that changes occurring during the development process are estimated to be responsible for 80% of the overall total costs of software development [Williams, 1993].

[25] For more information about computational complexity, the reader is referred to [Stockmeyer, 1987; Garey and Johnson, 1979; Johnson, 1990; Krentel, 1988; Papadimitriou and Steiglitz, 1982; Papadimitriou and Yannakakis, 1982].

The ability to record, track, analyse and control changes in requirements is crucial because modifications have a particularly significant impact on the consistency of specifications. Changes may introduce inconsistencies and conversely, handling existing inconsistencies requires changes to the requirements. Inconsistencies may arise for different reasons, for example, when multiple conflicting viewpoints are embodied in the specification, or when the specification itself is at a transient stage of evolution. These inconsistencies cannot always be resolved immediately — this fact is captured by Zowghi's and Rodrigues et. al.'s approaches [Zowghi, 1997; Rodrigues *et al.*, 2003]. In addition, manual assessment of the effects of these changes is very hard due to the complexity involved in modern large software systems. As a result, a number of techniques have been proposed to provide automated assistance to this task.

We describe four approaches based on belief revision principles for handling inconsistencies in the evolution of system requirements specifications: the first one uses AGM revision for default theories; the second one is based on the OTPs described in Section 4.7; a third one is based on epistemic entrenchment and finally the fourth one uses the clustered belief revision approach proposed in [Rodrigues, 2003].

Modelling and reasoning about the evolution of requirements

In [Zowghi, 1997], Zowghi formalised specifications as default theories where each requirement is defined as either *defeasible* or *non-defeasible*. In isolation, both defeasible and non-defeasible information are assumed to be consistent, but when put together, they can be inconsistent. Inconsistencies introduced by an evolutionary change can be resolved by performing a revision operation over the entire specification which changes the status of information from defeasible to non-defeasible or vice-versa.

Zowghi's approach has three main characteristics. It allows explicit representation of default information that can be used to construct complete requirements models from given (incomplete) sets of requirements; it permits the identification of consistent alternative models that resolve contradictions arising in an (incomplete) requirements specification; and finally it provides a constructive formal mechanism for computing the evolution of requirements models arising from the addition or retraction of requirements. The framework represents requirements specifications as *default theories* [Poole *et al.*, 1987; Poole, 1988b] and requirements evolution as the mapping between default theories through a process of rational belief revision. A non-monotonic derivability relation $\mid\!\sim$ is used to derive from a given (incomplete) set of requirements nonmonotonic consequences (or extensions) of the specifications using assumptions and defaults about the underlying problem domain. An AGM revision operator $\circ$ satisfying the AGM postulates is used to provide a rational revision of requirements specifications. This operator and its properties *confine*, but does not *uniquely determine* the new default theory obtained from the addition of information to a given partial requirements specifications.[26] The revision process used is based on a priority relation over the requirements expressing their relative importance — changes in the priorities affect the result of the revision.

The representation of requirements specifications is based on the nonmonotonic reasoning framework THEORIST proposed in [Poole *et al.*, 1987; Poole, 1988b]. Requirements specifications are therefore defined as a tuple $\langle F, H, C\rangle$, where F is a set of formulae that are *necessarily* true, called *facts*, H is a set of formulae that are *tentatively* true, called *hypotheses* and C is a set of *constraints*, such that $F \cup C$ is satisfiable. Requirements that are known to be true in a given domain are treated as facts, together with domain knowledge, whereas defaults or tentative requirements are treated as elements of H. Within this default logic representation framework, a requirements specification $\langle F, H, C\rangle$ is said to be *consistent* if $\langle F, H, C\rangle$ is a *consistent default theory*, namely if the set $F \cup C$ is satisfiable. Given a default theory Δ, the set of extensions of *maximal scenarios* of Δ, denoted with $E(\Delta)$ is the usual maximal default extension of Δ as defined by Reiter [Reiter, 1980].

Given a specification $S = \langle F, H, C\rangle$, a complete requirements specification (or maximal extension) constructed from S is the set $K \cup h$ such that $h \subseteq H$, $F \cup h \cup C$ is consistent and there is no h' such that $h \subset h' \subseteq H$ and $F \cup h' \cup C$ is consistent.

The evolution and change management of a given requirements specification is modelled by a revision operation that takes as input a given specification and the new information to be included and computes the revised specification. The operation satisfies the basic AGM postulates. The most important property in this context is the *principle of informational economy* (see page 6). In this setting, this amounts to retaining as much as possible of the old requirements. Obviously, if the new requirements are consistent with the current specification they are simply added and nothing is lost in the process. The difficulty arises when they are inconsistent with the current specification. In order to retain consistency some requirements in the current specification need to be retracted. In practice, what happens

[26] A contraction operation is also used on the same lines.

is the deactivation of these requirements. However, as the specification evolves, previously deactivated requirements may be reactivated provided they do not cause inconsistency. This is similar in spirit to the reinstatement algorithm of the controlled revision framework described in page 64. The revision mechanism defined in [Zowghi and Offen, 1997] provides the theoretical framework to this process and is based on the revision scheme for default theories proposed in [D. Zowghi and Peppas, 1996], which uses revision operators closely related to those defined in [Brewka, 1989a].

Given a default theory $\langle F, H, C\rangle$ representing a requirements specification and some new information φ, the revision of the specification by φ, denoted by $\langle F, H, C\rangle \circ \varphi$, is required to satisfy the following properties:

- the revised specification is a consistent default theory, i.e., the revised set of facts is consistent with the integrity constraints;
- the revision operation is successful, namely all the maximal scenarios of $\langle F, H, C\rangle \circ \varphi$ derive φ, and
- the outcome is independent from the syntactical form of φ. This means that for any formula ψ such that $\psi \equiv \varphi$, the revised specifications $\langle F, H, C\rangle \circ \varphi$ and $\langle F, H, C\rangle \circ \psi$ have the same maximal scenarios

A revision operator for default theories satisfying the above criteria was formally defined in [Zowghi and Offen, 1997] as follows.

DEFINITION 97. Let $\langle F, H, C\rangle$ be a default theory where F is a belief base, H is a set of default hypotheses and C is a set of constraints. Let φ be a formula. The revision of $\langle F, H, C\rangle$ by φ, denoted by $\langle F, H, C\rangle \circ \varphi$, is a default theory $\langle F', H', C'\rangle$ where:

- $F' = F \circ_a \varphi$
- $H' = H \cup (F \backslash F')$
- $C' = C - \neg F'$

The above operator satisfies the following properties (see [Zowghi and Offen, 1997]).

THEOREM 98. *Let $\langle F, H, C\rangle$ be a default theory and φ be a formula. The revised theory $\langle F, H, C\rangle \circ \varphi = \langle F', H', C'\rangle$ satisfies the following properties:*

- *$F' \cup C'$ is consistent*
- *for all e: $e \in E(\langle F', H', C'\rangle)$ implies $e \vdash \varphi$*

- *if* $\varphi \equiv \varphi'$*, then* $E(\langle F, H, C\rangle \circ \varphi) = E(\langle F, H, C\rangle \circ \varphi')$

Zowghi also defined a contraction operation for default theories in a similar way. The contraction is used to retract information from a requirements specification.

DEFINITION 99. Let $\langle F, H, C\rangle$ be a default theory and φ be a formula. The contraction of $\langle F, H, C\rangle$ by φ, denoted by $\langle F, H, C\rangle - \varphi$, is a default theory $\langle F', H', C'\rangle$ where:

- $F' = F - \neg C'$
- $H' = H \cup (F \backslash F')$
- $C' = C \circ_a \neg\varphi$

Zowghi showed that the above operator satisfies the following properties [Zowghi and Offen, 1997].

THEOREM 100. *Let* $\langle F, H, C\rangle$ *be a default theory and let* φ *be a formula. The default theory* $\langle F', H', C'\rangle = \langle F, H, C\rangle - \varphi$ *satisfies the following properties:*

- $F' \cup C'$ *is consistent*
- *there is no* $e \in E(\langle F', H', C'\rangle)$ *such that* $e \vdash \varphi$
- *if* $\varphi \equiv \varphi'$*, then* $E(\langle F, H, C\rangle - \varphi) = E(\langle F, H, C\rangle - \varphi')$

Informally, the contraction of a default theory by a formula φ consists of revising the set of constraints C with $\neg\varphi$ in order to guarantee that no default maximal extension of the theory will be consistent with φ; then contracting from the set of facts F the constraints that are consistent with φ; and finally adding the information removed from F into the set of default assumptions H. The latter ensures that no requirement is ever discarded from a given specification. If, for instance, a requirement φ is added to a specification that already contains $\neg\varphi$,[27] then $\neg\varphi$ is demoted to "default" status (i.e., moved from F to H) thus guaranteeing that no maximal scenario will contain it. If later on φ is retracted from the specification (e.g., by revising the constraints C with $\neg\varphi$), the set of facts F is contracted with all information that is inconsistent with the new set of constraints and therefore φ is contracted from F and added to H. In so doing the information $\neg\varphi$ is again included in the maximal scenarios of the new revised specification.

[27] φ is added to F.

In general, all requirements that have been considered during an evolution process are included in either F or H at all future times. Elements in F can be seen as resulting from prior revisions of a given specification whereas all the elements in C can be seen as resulting from prior contraction. The revision (resp., contraction) of F (resp., C) makes use of the notion of *epistemic entrenchment* to represent the relative importance of requirements. This ordering is clearly application-specific. For instance, in safety-critical systems, requirements related to safety factors may be considered most entrenched, whereas in real-time systems, requirements related to response time instead may be considered so. During the revision process inconsistencies between requirements are resolved by giving up those requirements that have lower epistemic entrenchment.

The following example, taken from [D. Zowghi and Peppas, 1996], illustrates how the revision and contraction operations defined above are used for evolving requirements.

EXAMPLE 101. Consider the set of requirements for a word processor

$$Spec = \{screencolour \leftrightarrow \neg screenmono, targetchildren \rightarrow screencolour\}$$

where *screencolour*/*screenmono* denote the requirement that the monitor screen be in colour or monochrome, and *targetchildren*/*targetadults* indicates whether the word processor is targetted at schoolchildren or adults.

With the addition of the facts *wordproc* and *targetadults* we get the default theory $\langle F_0, H_0, C_0\rangle$ given by

- $F_0 = \{screencolour \leftrightarrow \neg screenmono, targetchildren \rightarrow screencolour,$ $wordproc, targetadults\}$
- $H_0 = \emptyset$
- $C_0 = \emptyset$

Let us now consider the revision of this theory by *screenmono*. This formula is consistent with F_0 and therefore can be directly added to the set of requirements, thus generating the following requirements theory $\langle F_1, H_1, C_1\rangle$:

- $F_1 = F_0 \cup \{screemono\}$
- $H_1 = \emptyset$
- $C_1 = \emptyset$

A second revision of the specification by *targetchildren* would instead require a revision of the theory generating the new theory $\langle F_2, H_2, C_2\rangle$, where:

- $F_2 = F_1 \circ targetchildren = F_0 \cup \{targetchildren\}$
- $H_2 = H_1 \cup (F_1 \backslash F_2) = F_1 \backslash F_2 = \{screemono\}$
- $C_2 = \emptyset$

A subsequent evolution of the specification requiring the contraction of *screemono* would generate the new theory $\langle F_3, H_3, C_3 \rangle$, where:

- $F_3 = F_2 - \{screemono\} = F_2$
- $H_3 = H_2 \cup (F_2 \backslash F_3) = H_2 = \{screemono\}$
- $C_3 = C_2 \circ \neg screemono = \{\neg screemono\}$

Although this approach builds upon Poole's notion of default theories [Poole *et al.*, 1987; Poole, 1988b], the revision process of each component F and C is based on standard AGM revision with epistemic entrenchement. Subsequently, a set of operators closely related to the two operators given in Definition 97 and Definition 99 were shown to satisfy a reformulation of the AGM postulates based on the concept of maximal extensions of a given default theory. For further details the reader is referred to [Ghose, 1991].

Defaults in specifications

In [Ryan, 1993], the use *ordered theory presentations* (OTPs)[28] was proposed to handle the representation of default information and revision of software specifications.

The main motivation was to use default information to help narrowing the gap between formal specifications and the customers' initial requirements and to use belief revision to perform changes required by the realisation of undesirable consequences or inconsistency of a specification. A formal specification (or theory) is structured by splitting its *signature* into smaller components. For example, the signature of the specification of a lift system could be divided into three smaller signatures governing the behaviour of the lift's buttons; the behaviour of its door and the lift's current position.

Requirements are associated with each of the classes identified in the structure which then induces a partial ordering relation over a given system specification. The finite set of requirements specifications equipped with such a partial order (or priority relation) defines an ordered theory presentation. The ordering relation over the system requirements enables the use

[28]See Section 4.7 or [Ryan, 1991]).

of a *specificity principle* for the resolution of conflicts between specifications of different but dependent classes of objects. The more specific the class of a requirement is, the lower in the ordering the requirement is and the higher its priority is. In other words, more general classes have less priority than more specific ones. The specificity principle states that *defaults about a class of objects override those about a more general class* [Ryan, 1993]. In the lift example given above, a "lift" class would be considered to be more specific than a "button" class. Consider, for example, the following two statements, in which the first requirement refers to an object of class "lift" and the second one to an object of class "button".

- when the lift is on the i-th floor, the indicator light of the i-th floor is off
- pressing a button for a floor causes the corresponding light indicator to come on

In the particular situation when a lift is on the floor whose corresponding button has been pressed, the conflicting statements *indicator light of the i-th floor is off* and *indicator light of the i-th floor comes on* could be derived. However, in this case the specificity principle would block the consequences imposed by the second requirement since *indicator light of the i-th floor is off* would override the requirement associated the more general button class.

Logically speaking, a system specification is defined as a pair $\langle \Delta, \Gamma \rangle$, where Δ is the set of inviolable requirements (an ordinary theory presentation) and Γ is a set of norms (an *ordered* theory presentation). The models of a specification $\langle \Delta, \Gamma \rangle$ are then the $\sqsubseteq^{\Gamma}$-maximal elements in the set of interpretations that satisfy Δ. From the point of view of belief revision, Δ represents a belief base and the ordering of the OTP Γ comes from the process by which the specification is obtained, i.e., its revision history. The history can be thought of as showing the structure (through refinement) of a system component. The more general classes are introduced earlier during the development cycle and the more specific classes are introduced later as a refinement (or revision) of existing (more general) components. This process would generate a *linear* OTP, as discussed in Section 4.7. Revision of a specification would be obtained by the addition of sentences to the OTP.

Goal-structured analysis and design revision

One of the difficulties in automating some aspects of the process of software engineering is that although a formal mathematical language is needed in order to ensure a thorough analysis and verification of system properties, requirements from stakeholders are usually expressed in a natural language that is prone to ambiguities and other weaknesses.

In [Duffy *et al.*, 1995], Duffy et al. propose a formalism that attempts to bridge this gap by allowing information about requirements and other design decisions to be expressed both in natural language and in sentences of a formal logic language. The latter is used to automate the process of change evaluation. The framework was called *goal-structured analysis* (GSA), which is described next.

In GSA, logical representations of requirements and design decisions are embedded in a *goal structure* — a directed acyclic graph where the nodes, here called *frames*, contain information about the system. The information in a frame consists of natural language assertions as well as sentences in a logical language. Frames are divided into four main classes: *goals* — possibly structured statements about objectives to be achieved by the system; *effects* — similar to goals, but describing an outcome which is possibly undesirable; *facts* — statements that cannot be further decomposed (i.e., atomic); and *conditions* — statements that are also atomic, but which are possibly false in a given scenario. GSA does not distinguish between facts that support other facts. We will later see that this distinction is necessary during the process of goal structured revision.

Scenarios in GSA correspond to a set of conditions. A goal is said to be *globally supported* if it is satisfied by facts and conditions only. During the top-down decomposition, it is not usually possible to tell whether a goal will eventually be globally supported. However, the framework's objective is to identify conditions to ensure global support for the goal in the final structure. A goal which satisfies these conditions is said to be *locally supported.* Duffy et al. have shown in [Duffy *et al.*, 1995] that local support leads to global support.

In order to avoid certain difficulties, such as cycles in the support chain of a goal, the formalism adopts a partial order $\preceq_G$ on goals. A goal must be entailed by goals appearing lower in the order $\preceq_G$ for it to be locally supported.

The formalism presented in [Duffy *et al.*, 1995] does not contain a formalisation of how goal structures should change (for instance to accommodate changes in requirements). In [MacNish and Williams, 1998], Williams et al. put forward belief revision as the formal mechanism for revision of user specifications in top-down designs (which is the case of goal structures).

The idea is to associate the partial order on the requirements with an entrenchment ranking $0 < \mathcal{R}_c < \mathcal{R}_f < \mathcal{R}_r < \mathcal{R}_{max}$ where 0 is the least entrenched point; $\mathcal{R}_{max}$ is the most entrenched one and in which

- *conditions* are ranked in the interval $[0, \mathcal{R}_c]$;
- *facts* are ranked in the interval $(\mathcal{R}_c, \mathcal{R}_f]$;
- *goals* and *effects* are ranked in the interval $(\mathcal{R}_f, \mathcal{R}_r]$; and
- *causal rules* are ranked in the interval $(\mathcal{R}_r, \mathcal{R}_{max}]$

The only interval above that does not directly correspond to GSA's partition of assertions is $(\mathcal{R}_r, \mathcal{R}_{max}]$. Williams et al. note in [MacNish and Williams, 1998] that some facts actually represent causal relationships and consider them to play a more fundamental role. These are called *causal rules* and placed in the most entrenched interval. Also note that this particular way of ranking separates conditional requirements from mandatory ones (assumed to be true in any given scenario) by placing the former in the least entrenched rank $[0, \mathcal{R}_c]$. Intuitively, these will be the requirements that one is most willing to give up.

The entrenchment ranking is used to allow the development of goal (decomposition) structures without *loops*, and also to define a belief revision process over a given goal structure. Three types of operations on goal structures were proposed: test of alternative scenarios; revision of goals and effects; and contraction of causal rules. The first type of operation corresponds to the type of analysis usually associated with "what-if" questions. It tests different scenarios to see whether goals are (still) supported. The change in this case is in the set of conditional requirements — a new scenario might require removing existing conditions and adding new ones. As conditions are placed in the least entrenched interval of requirements, additions and deletions can be made on this part of the specification without affecting requirements in more entrenched ranks such as those associated with goals and causal rules. Only conditions that contradict newly added conditions are retracted leaving other existing unrelated conditions unaffected. The revision of goals and effects arises when a goal decomposition leads to disjunctive antecedents.

Given a partial entrenchment ranking, an assertion g is defined to be *globally supported* with respect to a scenario if and only if there exists a set S of assertions in the specification that entails g and such that each assertion s_i in S has a rank value $deg(s_i)$ either in the interval $(0, \mathcal{R}_f]$ or in the interval $[\mathcal{R}_r, \mathcal{R}_{max}]$.

For an assertion g to be *locally supported* in a scenario, there must exist a set S of assertions that entails g and such that the rank value $deg(s_i)$ of each assertion s_i in S is either in the interval $(0, deg(g))$ or in the interval $[\mathcal{R}_r, \mathcal{R}_{max}]$.

Williams et al. make other considerations with respect to improving efficiency by imposing syntactical constraints on the representation of rules (and hence simplifying theorem proving in the process). For more information, the reader is referred to [MacNish and Williams, 1998; Williams, 1998].

Clustered belief revision

As we have seen, conflicting viewpoints inevitably arise in the process of requirements analysis. Conflict resolution, however, may not necessarily happen until later in the development process. This highlights the need for requirements engineering tools that support the management of inconsistencies [Easterbrook and Nuseibeh, 1996; Spanoudakis and Zisman, 2001; Hunter and Nuseibeh, 1998].

Many formal methods of analysis and elicitation rely on classical logic as the underlying formalism. Model checking, for example, typically uses temporal operators on top of classical logic reasoning [Huth and Ryan, 2000] with well-behaved and established proof procedures. However, it is well known that classical logic theories trivialise in the presence of inconsistency and this is clearly undesirable in the context of requirements engineering.

Paraconsistent logics [da Costa, 1974] attempt to circumvent the problem of theory trivialisation by weakening some of the axioms of classical logic, often at the expense of reasoning power. For instance, Belnap's four valued logic [Belnap, 1977a] allows for non trivial logical representations where propositions can be both true and false, but does not verify basic inference rules such as modus ponens. While appropriate for concise modelling, logics of this kind are too weak to support practical reasoning and the analysis of inconsistent specifications.

In [Rodrigues *et al.*, 2003], Rodrigues et al. argue that a formal framework for the analysis of evolving specifications should be able to *tolerate* inconsistency without trivialising the reasoning process and enable impact analyses of potential changes to be carried out. They proposed the use of *clustered belief revision* [Rodrigues, 2003] to help in this process. The idea is to use priorities to obtain plausible (i.e., not trivial) conclusions from an inconsistent theory whilst exploiting the full power of classical logic reasoning. Requirements with similar functionality are grouped into "clusters" and the clusters themselves may have relative priorities between them. By analysing the result of a cluster, an engineer can either choose to rectify problems in the specification or to postpone the changes until more information becomes available. This allows for the anlaysis of the results of different possible prioritisations through reasoning in classical logic, and the evolution of specifications that contain conflicting viewpoints in a principled way.

The analysis of user-driven cluster prioritisations can also give stakeholders a better understanding of the impact of certain changes in the specification. The formalism does not provide any support for the creation of such prioritisations other than the reasoning mechanism itself, because prioritising requirements is subjective to the interpretation of the engineer and stakeholders and is in itself a very complex issue.

The starting point is a basic priority relation $\leq$ associated with a set of requirements Γ. This relation is extended to a relation $\preceq$ associated with the power set of Γ, i.e., 2^Γ. The idea is to consider only subsets of 2^Γ whose requirements satisfy certain properties — these are called *plausible*. One such property is obviously logical *consistency*. However, any constraint that can be expressed in classical logic can be used. The extension of the original $\leq$ to $\preceq$ is then used to determine which among the plausible elements of 2^Γ are *preferred*preferred subsets of requirements.

The tool translates requirements given in the form of "if-then-else" rules into the (more efficient) disjunctive normal form (DNF) for classical logic reasoning (this simplifies the consistency checking, for instance). A cut-down version of the light control case study [Heitmeyer and Bharadwaj, 2000] was given in [Rodrigues *et al.*, 2003] as an illustration of the use of the framework for requirements engineering.

As discussed above, the emphasis of the work is on the use of priorities for reasoning about potentially inconsistent specifications. The same technique can be used to check the consequences of a given specification and to reason about "what-if" questions arising during evolutionary changes as discussed for GSA above. This can be done by imposing different plausibility conditions, for instance.

The extension of $\leq$ to $\preceq$ can be optimised by the use of a number of heuristics about its behaviour with respect to $\leq$. The use of DNF greatly simplifies the reasoning process, but the conversion to DNF sometimes generates complex formulae — these can be minimised via Karnaugh maps. If conversion to DNF is not desired, a suitable theorem prover could be used instead in the reasoning process.

Clustered belief revision applied to software engineering relates to the GSA formalism proposed by Duffy et al. and described in the beginning of Section 7.1 in the sense that the latter also allows for an analysis of the consequences of alternative changes by checking the verification of goals after modifications to a specification. Rodrigues et. al.'s approach in addition supports the evaluation of consequences of (evolutionary) changes through the analysis of requirements retention after changes to a specification.

Many other techniques can be found in the literature for the management of inconsistency, but much of the work has focused on consistency checking; consistency analysis and actions to be taken based on pre-defined inconsistency handling rules rather than on belief revision itself. For example, Easterbrook et al. proposed to augment consistency checking rules with pre-defined lists of possible actions [Easterbrook and Nuseibeh, 1996], but with no policy or heuristics to guide in the choice among alternative actions. Their approach takes decisions based on an analysis of the history of the development process (e.g., past inconsistencies and past actions). Klein et al. also proposed the use of pre-defined hierarchies of conflicts and associated resolutions [Klein and Lu, 1990], while Robinson et al. suggested the use of negotiation for this purpose [Robinson and Volkov, 1998].

8 CONCLUSIONS

One can divide the evolution of the investigation of belief revision in two main fronts. The more philosophical/mathematical approach, in which the fundamental aspects of the process have been analysed; and the more applied one, in which the belief revision principles identified by the former are applied in the solution of specific computer science problems.

In [Hansson, 2003], Hansson presents a number of interesting philosophical problems for belief revision. One such problem is the representation and reasoning of conditional assertions. Suppose an agent's language for the representation of beliefs would allow for the representation of hypothetical reasoning with statements of the kind "it is possible that A" (in symbols $\diamond A$), whenever $\neg A$ is not held by the current belief set K. If $A \in K$ and K is not inconsistent, then it follows that $\diamond\neg A \notin K$. Now notice that $\diamond\neg A \in K - A$, contradicting one of the contraction postulates that asserts that $K - A \subseteq K$. Many issues of this nature are still open to investigation.

We discussed at some length in Section 4 some of the difficulties incurred by the intrinsic lack of a richer structure in a belief set. We argued that this was one of the main problems in the application of belief revision in practical reasoning systems because it makes it difficult to represent some meta-level constraints that are essential for the correct modelling of these systems. The lack of structure of the belief set was circumvented in a sense in the original AGM formulation with the proposal of epistemic entrenchment orderings and the like. However, no guidelines were provided as to how these orderings should evolve themselves. Many authors have highlighted this categorial matching problem preventing a more natural definition of revision mechanisms supporting iteration [Nayak, 1994b; Rott, 2001; Hansson, 2003].

As a result, all of the main proposals for the modelling of iteration of the revision process make use of an enriched notion of a belief set and provide some guidance on how this enriched structure (sometimes called an *epistemic state*) is affected by a belief change operation. In the formalisms for iterated revision, the meta-level information is used mainly to record the *history of revisions*, but the meta-level information has many other applications. It could be used for instance to represent priorities associated with beliefs. The question remains as to whether a more realistic approach can depart from principles that need to make reference to the richer structure explicitly. We will discuss this further later on in this section. In what now follows, we present a number of possible directions for the expansion of the field.

Revision for structured theories

In Section 2, the problem of belief revision was presented and formulated as follows:

(Def $*$) Let K be a set of formulae representing the beliefs of some agent in the language $\mathcal{L}$ of some logic L, with consequence relation Cn. Assume some additional information φ is received and that φ is consistent in L but $K \cup \{\varphi\}$ is inconsistent. Belief revision seeks to find a new consistent theory $K' = K \circ \varphi$ which contains φ.

The context and discussion above was initiated in [Alchourrón and Makinson, 1982; Alchourrón *et al.*, 1985] where the postulates the revision $\circ$ should satisfy were proposed. At that time, the main logic L was classical logic. The notion of a theory K was a set of well formed formulae and the notions of consistency and consequence were mainly those of classical logic or of similar logics.

Nowadays, as a result of deeper interaction of logic with computer science, artificial intelligence, common sense reasoning, legal reasoning, language analysis, etc., the notion of a logical system has changed and evolved. It is time to discuss the concepts of revision theory as applied to the new logics. We will do so without going into the character of the new logics, but giving just enough detail to put revision theory in perspective. For more details, the reader is invited to consult the book on *Labelled Deductive Systems* [Gabbay, 1996] and the chapter on labelled deduction in this series. The basic concepts used by $\circ$ are:

($\Pi°1$) A theory K is a set. φ is in K means set theoretical inclusion.
($\Pi°2$) Consistency is that of classical logic or similar.
($\Pi°3$) Consequence is monotonic, therefore maintaining consistency is done by deletion.
($\Pi°4$) Input φ is done by set theoretic inclusion.

The new concepts are at least as follows:

($\Pi°1'$) A theory K is a labelled structure of formulae. φ is in K probably means that φ appears somewhere in the structure
($\Pi°2'$) Consistency is no longer a central notion. A better notion is *acceptability*. Certain structured theories are not allowed.
($\Pi°3'$) Consequence is labelled deduction with quite a complex hierarchy of rules. Non-monotonicity is a natural side-effect of the proof system. Acceptability can be maintained by deletion, re-structuring or even addition.
($\Pi°4'$) The notion of input is part of the logic. Since a theory K is a complex structure, the input function must tell us how to integrate the input φ (which may come with a label) into the structure K.

The problem of revision now becomes as follows:

(Def $*'$) Given a theory K and an acceptable input φ, turn K into a new theory K', with φ appearing in the structure. If K' is not acceptable, find another theory K'' in which φ appears and which is acceptable.

Some examples are given to illustrate this idea:

EXAMPLE 102. Let a police inspector record information (expressed by formulae in the language of classical logic) by its source and reliability. Thus, he may have a database with the following structure:

1. (witness $a, 0.7$) : φ
2. (witness $b, 0.9$) : ψ
3. common sense : $\varphi \to \neg\psi$
4. common sense : $\varphi \to \gamma$

Assuming the evidence from witness a is acceptable (i.e., it was obtained by legal means), this knowledge base is problematic. How can one revise it?

Option 1: One can consider one of the witnesses less reliable, by for example, changing the label of item 1 to, say, '(witness $a, 0.3$)'. This will affect the value with which γ is derived.

Option 2: One can modify the view of the common-sense rule $\varphi \rightarrow \neg\psi$. Perhaps it has more exceptions than anticipated.

Option 3: Leave the data as it is and wait for more data to become available (or *request* for more data before making any conclusions).

The reader can see that the process of revision in such complex contexts is different and has more options than AGM and its variations.

EXAMPLE 103. This example arises naturally from iterated revision. Suppose one starts with a theory K and keep revising it to get $K_1 = K \circ \varphi_1, K_2 = K_1 \circ \varphi_2, \ldots$. If a record of the input stream and of how revisions of K into K_1, K_1 into K_2, etc., is kept, then this complex historical structure can be used to help in the decision of how to revise from K_i to K_{i+1}. This generalisation is the basis of *controlled revision* and other similar revision mechanisms using the history of the updates (see Section 4).

Connections with social choice theory

We also note connections with voting and Arrow's paradox [Arrow, 1963; Geanakoplos, 1996]. Consider a domain of alternatives $D = \{a_1, \ldots, a_n\}$ and assume there are voters $K_1, \ldots, K_m$ giving total orders of preferences on this domain. The problem is to resolve some compromise ordering C on the domain D. Under very reasonable assumptions on the compromise function $c = f(K_1, \ldots, K_m)$, one finds that the solution C will be (under these assumptions) one of the K_i. This is the essence of Arrow's paradox and is considered an undesirable outcome.

Now looking at the situation from the point of view of revision theory, the problem can be layed out as follows. Let $a_1, \ldots, a_n$ be n constants and consider a predicate language with equality $=$ and a binary order symbol $<$. Let φ be a formula containing as a conjunction all axioms of total order together with the axiom

$$\forall x(x = a_1 \vee x = a_2 \vee \ldots \vee x = a_n).$$

Let K be a consistent theory of order and consider $K \cup \{\varphi\}$. If this theory is not consistent, then one could invoke an AGM revision operator $\circ$ to find a theory $K' = K \circ \varphi$ such that $K' \vdash \varphi$. In fact, K' is some theory on how to order the set of elements $D = \{a_1, \ldots, a_n\}$. As we have seen, one way of finding K' is to look at all maximal consistent subsets $K_1^\circ, \ldots, K_m^\circ$ of K that fail to imply $\neg\varphi$ (i.e., $K_\perp \neg\varphi$). In other words, revision will be considering consistent theories $L_i^\circ = K_i^\circ \cup \{\varphi\}$ on D. These correspond to disjunctions L_i° of total orderings on the set D. The important point to bear in mind is that a maxichoice revision function will pick *one* of these, but this is exactly what social choice theory does not want. In voting terms a maxichoice revision operation is a process which, faced with a conflict between different voters, determines how to eliminate them all but one! In other words, this special kind of revision welcomes Arrow's paradox.

What would the voting compromise function look like from the point of view of revision theory?

Well, given an inconsistent theory $K \cup \{\varphi\}$, one seeks some K' such that $K' \cup \{\varphi\}$ is consistent and K' bears some relation to all acceptable possible AGM revision candidates K_i° of K. K' need not be a contraction of K but a different theory satisfying different requirements.

Note that since L_i° are disjunctions of possible total orderings, it may still be possible to find a total order consistent with the majority of the disjunctions (i.e., acceptable to the majority of voters). This does not contradict Arrow's paradox. What it says is that if the voters give several alternatives each for total preferences, maybe one can accommodate a majority of them.[29]

Another promising line of research flowing from this new point of view is the following:

"What would iterated revision be in the context of voting?"

"Would iterated changes in preferences be inconsistent with previous compromises?"

We should note in this context that revision for voting makes use of the ordering and the details of the theory involved. This suggests context revision, where the principles of how to revise a theory K makes use of the theory K itself and may be different for different theories K. So for example for a theory involving colours of objects (unary predicates $C_i(x)$), we may want to restore consistency without rendering any predicate $C_i(x)$ empty.

Object level revision

The class of AGM operators $K \times \varphi \longrightarrow K \circ \varphi$ is not defined in the object level, but in the meta-level, since the binary operation $\circ$ is not part of the language $\mathcal{L}$. However, there would be advantages in having the revision operator defined in the object level. This could lead to the design of systems which fix themselves axiomatically and would be good for applications where

[29] One can let the voters have preferences on the linear orders and try to accommodate them. From the revision theory point of view, this means each maximal subtheory of K (maximally consistent with the input φ) has its model preferentially ordered and looking at the models can help in making a choice.

inconsistencies can mean fault, undesirable behaviour, violation of integrity constraints and the like. Thus, moving into the object level may seem simple and straightforward.

One can immediately add $\circ$ to the object language and write (AGM?) axioms for it. This can indeed be done but it is not satisfactory. There are plenty of algorithms for actually finding $K \circ \varphi$, and the natural question is how to reflect them in the object level.

Obviously, the only way to do it is to present the object-level logic in a proof theoretical formulation (proof theory is algorithmic) and to add to it additional algorithmic features which can be used to do revision. Thus the particular algorithm for revision will be reflected and manifested as special proof rules for the object-level connective $\circ$. This exercise has yet to be done in detail, but one thing is already clear at this stage: since revision is connected with contraction, one must also be able to perform object-level deletion!

In fact, once one thinks of object-level deletion, one realises that perhaps deletion is a more fundamental notion than revision (in the spirit of the Levi identity — see page 2). Figure deletion out in the object level and many other notions can be defined, including revision, abductive revision, contraction and what else? Note that it would also be necessary to formalise the connections between the meta-level revision and the object-level ones. For more details, the reader is referred to [Gabbay *et al.*, 2002; Gabbay *et al.*, 2004].

Reactive systems

There is a more general way of looking at belief revision. One can regard the belief set as a system and the input as an interaction with the system. The system reacts to the attempt to interact with it. If the input is inconsistent, the system revises itself.

This idea can be applied in many contexts. Take for example the bridge from point a to point b seen in Figure 9.

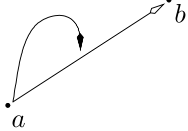


Figure 9. A bridge between points a and b.

As one passes through the bridge, if the load is too heavy, the bridge can collapse. The arrow with the white head indicates the connection, the arrow with the black head indicates the possible signal to disconnect.

Figure 10 depicts a more sophisticated scenario. In that figure, if one crosses from a to b without proper procedure, the connection between b and c would be automatically dropped and one would not be able to continue. Note that such system has object-level properties. It revises itself automatically.

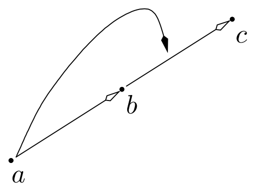


Figure 10. Bridges between points a, b and c.

Another example of interaction is revising consequence relations. Let **Cn** be a collection of consequence relations of the form $x|\!\!\rightsquigarrow y$ over some language $\mathcal{L}$. Assume $\mathcal{L}$ contains some form of implication $\rightarrow$. Define the revision of $|\!\!\rightsquigarrow$ by α ($|\!\!\rightsquigarrow_\alpha$) as $x|\!\!\rightsquigarrow_\alpha y$ iff $x|\!\!\rightsquigarrow \alpha \rightarrow y$. This notion turns out to be useful in quantum logic. In quantum logic, the elements are subspaces of the Hilbert space and revision is done by projection [Engesser and Gabbay, 2002]. For more details on reactive systems the reader is referred to [Gabbay, 2004].

First-order belief change

There is a vast body of literature on belief change, but almost all of it is propositional and mainly involves theories of the classical propositional logic. There is no study of belief change involving quantificational theories. This is not unique to belief change. Other areas of applied logic are also mainly propositional. For example, nonmonotonic logics are mainly propositional. There are some exceptions like circumscription and some predicate default logics but the main bulk of the research is propositional. We have a similar situation for probabilistic logics, and to a lesser extent with fuzzy logic. In restrospect, the entire area of non-classical possible-world logics is heavily slanted towards the propositional case. There are some studies of the quantifiers but not really enough. There is still the notion in people's minds that the main character of a logic is given by its propositional part and that the quantifiers are standard additions, acting like big conjunctions (for $\forall$) and big disjunctions (for $\exists$) over variable instantiations.

It is time, however, to turn our attention to predicate logics and apply our machinery to theories containing quantifiers, variables and individuals.

The received wisdom about passing from the propositional case to the predicate case was that all we need to do is add variables and basically add the quantifier rules:

1. $\forall x A(x) \Rightarrow A(y)$
2. $A(y) \Rightarrow \exists x A(x)$
3. $\dfrac{\vdash A(x) \Rightarrow B}{\vdash \exists x A(x) \Rightarrow B}$, x not in B
4. $\dfrac{\vdash B \Rightarrow A(x)}{\vdash B \Rightarrow \forall x A(x)}$, x not in B

There may be some variations depending on the meaning of $\Rightarrow$ (in e.g. substructural logics) but on the whole there is nothing special to the move of introducing the quantifiers into a logic. Of course one may study generalised quantifiers or linguistic quantifiers and so on but the ordinary $\forall$ and $\exists$ are considered as routinely introduced and require no special treatment and cause no special problems.

The first surprise was when this recipe did not exactly work for relevance logics (as Kit Fine has shown). Still, this was regarded as an aberration, as something to be fixed. There was no realisation that we need to look at 'reasoning at the presence of quantifiers' as something completely different from propositinal reasoning, having its own completely different set of problems and methods of solution.

Such a realisation is needed when we look at belief change, as we are going to discuss next.

We begin with an example.

EXAMPLE 104. Suppose we have two witnesses w_1 and w_2. w_1 says she saw Mr. Jones at Kings' Cross at 12.00 January 1st, 2010; whereas w_2 says he saw Mr. Jones at the Strand at 12.00, January 1st, 2010.

Obviously given a standard theory of people and time, these two statements are inconsistent. Standard AGM revision theory or any algorithmic refinements of it will seek to eliminate (delete) one of the statements. A more practical commonsense approach is to accept the honesty of the witnesses and seek to investigate the possibility of mistaken identity. Maybe w_1 saw Mr. Jones, but w_2 saw a Mr. Smith, and mistook him for Mr. Jones. This is an easier way out of the inconsistency, and with less side effects in the context of the application.

If we look at the previous example more formally, we have that we are given an inconsistent predicate theory talking about the individuals $a_1, \ldots, a_n$ involving atomic predicates $P_1(x), \ldots, P_k(x)$.

Question. Can we restore consistency by one of the following operations?

1. Swapping identities of individuals for predicate P_i: replace all or some occurrences of $P_i(a_j)$ by $P_i(a_r), r \neq j$.

2. Identifying elements: add $a_j = a_r, r \neq j$ and delete all assumptions $a_j \neq a_r$.

3. Postulate a new element $b \neq a_i, i = 1, \ldots, n$ and swap some occurrences of $P_i(a_j)$ by $P_i(b)$.

4. Split the predicate P into two predicates P' and P'', and replace each occurrence of P by either P' or P''. (For example, we have that the witness is partially colour blind and had identified two different shades of red!)

Technical challenges

1. For the method of reidentification and shifting of predicate extensions is belief revision the same as contraction and then expansion? (We expect the answer to be negative.)

2. Write reasonable postulates for this kind of first-order belief change.

3. Characterise belief change where only reidentification of elements is involved.

4. Do the same where postulating new elements is involved.

5. Can any inconsistent theory be made consistent by a reasonable shifting of predicate extensions? If not, characterise the boundaries.

6. Investigate inconsistencies for the monadic fragment and how it can be rectified by reidentification and shifting.

7. Investigate the role of equality in this area.

EXAMPLE 105. Consider the following assertions.

1. there are three elements a_1, a_2 and a_3

2. two of them are green

3. two of them are brown
4. an element cannot be both green and brown.

How would AGM deal with this? How would *we* deal with this?
If we turn this into a propositional problem, we would have

1. $\bigvee_{i \neq j} G(a_i) \wedge G(a_j)$
2. $\bigvee_{i \neq j} B(a_i) \wedge B(a_j)$
3. $\neg(G(a_i) \wedge B(a_i)), i = 1, 2, 3.$

Ordinary revision will take something out. Will first-order revision add a fourth element?

BIBLIOGRAPHY

[Adams, 1975] E. Adams. *The Logic of Conditionals.* D. Reidel, Dordrecht, Netherlands, 1975.

[Alchourrón and Makinson, 1982] C. A. Alchourrón and D. Makinson. On the logic of theory change: Contraction functions and their associated revision functions. *Theoria*, 48:14–37, 1982.

[Alchourrón and Makinson, 1985] C. A. Alchourrón and D. Makinson. The logic of theory change: Safe contraction. *Studia Logica*, 44:405–422, 1985.

[Alchourrón *et al.*, 1985] C. A. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50:510–530, 1985.

[Arrow, 1963] K. Arrow. *Social choice and individual values.* Cowles Foundation Monograph Series, second edition, 1963.

[Belnap, 1977a] N. Belnap. How a computer should think. In G. Ryle, editor, *Contemporary Aspects of Philosophy*, pages 30–56. Oriel Press, 1977.

[Belnap, 1977b] N. Belnap. A useful four-valued logic. In J. M. Dunn and G. Epstein, editors, *Modern Uses of Multiple-valued Logic*, pages 8–37. D. Reidel, 1977.

[Bench-Capon and Dunne, 2007] T. J. M. Bench-Capon and P. E. Dunne. Argumentationnext term in artificial intelligence. *Artificial Intelligence*, 171:619–641, 2007.

[Besnard and Hunter, 2008] Philippe Besnard and Anthony Hunter. *Elements of Argumentation.* MIT Press, 2008.

[Borgida, 1985] A. Borgida. Language features for flexible handling of exceptions in information systems. *ACM Transactions on Database Systems*, 10:563–603, 1985.

[Boutilier, 1994] C. Boutilier. Unifying default reasoning and belief revision in a modal framework. *Artificial Intelligence*, 68:33–85, 1994.

[Boutilier, 1996] C. Boutilier. Iterated revision and minimal change of conditional beliefs. *Journal of Philosophical Logic*, 25:262–305, 1996.

[Brewka and Hertzberg, 1993] G. Brewka and J. Hertzberg. How to do things with worlds: on formalizing actions and plans. *Journal of Logic and Computation*, 3(5):517–532, 1993.

[Brewka, 1989a] G. Brewka. Belief revision in a framework for default reasoning. In *Proceedings of the Konstanz Workshop on Belief Revision*, pages 206–222, 1989.

[Brewka, 1989b] G. Brewka. Preferred subtheories - an extended logical framework for default reasoning. In *Proceedings of the International Joint Conference on Artificial Intelligence*, Detroit, 1989.

[Brewka, 1990] G. Brewka. *Nonmonotonic Reasoning — Logical Foundations of Commonsense*, volume 12 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, October 1990.

[Brewka, 1991] G. Brewka. Cumulative default logic: In defense of nonmonotonic inference rules. *Artificial Intelligence*, 50:183–205, 1991.

[Chisholm, 1946] R. Chisholm. The contrary-to-fact conditional. *Mind*, 55:289–307, 1946. Reprinted in *Readings in Philosophical Analysis*, edited by H. Feigl and W. Sellars, Appleton-Century-Crofts, New York, 1949, pp. 482–497.

[D. Zowghi and Peppas, 1996] A. K. Ghose D. Zowghi and P. Peppas. A framework for reasoning about requirements evolution. In *Proceedings of the Forth Pacific Rim International Conference on Artificial Intelligence (PRICAI 96)*, 1996.

[da Costa, 1974] N. C. A. da Costa. On the theory of inconsistent formal systems. *Notre Dame Journal of Formal Logic*, 15(4):497–510, 1974.

[Dalal and Etherington, 1992] M. Dalal and D. W. Etherington. A hierarchy of tractable satisfiability problems. *Information Processing Letters*, 44:173–180, 1992.

[Dalal, 1988a] M. Dalal. Investigations into a theory of knowledge base revision: Preliminary report. In Paul Rosenbloom and Peter Szolovits, editors, *Proceedings of the Seventh National Conference on Artificial Intelligence*, volume 2, pages 475–479, Menlo Park, California, 1988. AAAI Press.

[Dalal, 1988b] M. Dalal. Updates in propositional databases. Technical report, Department of Computer Science, Rutgers University, February 1988. Technical Report DCS-TR-222.

[Dargam, 1996] F. C. C. Dargam. *On Reconciling Conflicting Updates (A Compromise Revision Approach)*. PhD thesis, Department of Computing, Imperial College, October 1996.

[Darwiche and Pearl, 1994] A. Darwiche and J. Pearl. On the logic of iterated belief revision. In Ronald Fagin, editor, *Proceedings of the 5th International Conference on Principles of Knowledge Representation and Reasoning*, pages 5–23. Morgan Kaufmann, Pacific Grove, CA, March 1994.

[Darwiche and Pearl, 1996] A. Darwiche and J. Pearl. On the logic of iterated belief revision. Technical Report R-202, Cognitive Science Laboratory, Computer Science Department, University of California, Los Angeles, CA 90024, November 1996.

[Darwiche and Pearl, 1997] A. Darwiche and J. Pearl. On the logic of iterated belief revision. *Artificial Intelligence*, 89:1–29, 1997.

[Doyle, 1979] J. Doyle. A Truth Maintenance System. *Artificial Intelligence*, 12:231–272, 1979.

[Doyle, 1992] J. Doyle. Reason maintenance and belief revision: foundations vs. coherence theories. In P. Gärdenfors, editor, *Belief Revision*, Cambridge Tracts in Theoretical Computer Science, pages 29–51. Cambridge University Press, 1992.

[Duffy *et al.*, 1995] D. A. Duffy, C. MacNish, J. A. McDermid, and P. Morris. A framework for requirements analysis using automated reasoning. In *Conference on Advanced Information Systems Engineering*, pages 68–81, 1995.

[Easterbrook and Nuseibeh, 1996] S. Easterbrook and B. Nuseibeh. Using viewpoints for inconsistency management. *Software engineering journal*, 11(1):31–43, January 1996.

[Eiter and Gottlob, 1992a] T. Eiter and G. Gottlob. The complexity of propositional knowledge base revision, updates and counterfactuals. *Artificial Intelligence*, 57:227–270, 1992.

[Eiter and Gottlob, 1992b] T. Eiter and G. Gottlob. On the complexity of propositional knowledge base revision, updates and counterfactuals. In *ACM Symposium on Principles of Database Systems*, 1992.

[Eiter and Gottlob, 1993] T. Eiter and G. Gottlob. The complexity of nested counterfactuals and iterated knowledge base revisions. In *Proc. Thirteenth International Joint Conference on Artificial Intelligence (IJCAI '93)*, pages 526–531, 1993.

[Engesser and Gabbay, 2002] K. Engesser and D. M. Gabbay. Quantum logic, Hilbert space, revision theory. *Artificial Intelligence*, 136:61–100, March 2002.

[Fagin *et al.*, 1983] R. Fagin, J. D. Ullman, and M. Vardi. On the semantics of updates in databases. In *Proceedings of the ACM SIGACT-SIGMOD Symposium on Principles of Database Systems*, pages 352–365, Atlanta, GA, 1983. ACM.

[Falappa *et al.*, 2002] M. A. Falappa, G. K.-I., and G. R. Simari. Explanations, belief revision and defeasible reasoning. *Artificial Intelligence*, 141(1/2):1–28, 2002.

[Freund and Lehmann, 1994] M. Freund and D. Lehmann. Belief revision and rational inference. Technical Report TR 94-16, The Leibniz Center for Research in Computer Science, Institute of Computer Science, Hebrew University, July 1994.

[Friedman and Halpern, 1996] N. Friedman and J. Y. Halpern. Belief revision: A critique. In *Proceedings of the 5th International Conference on Principles of Knowledge Representation and Reasoning*, pages 421–431, Cambridge, Massachusetts, 1996.

[Friedman and Halpern, 1999] N. Friedman and J. Y. Halpern. Belief revision: A critique. *Journal of Logic, Language and Information*, 8:401–420, 1999.

[Fuhrmann and Hansson, 1994] A. Fuhrmann and S.O. Hansson. A survey of multiple contraction. *Journal of Logic, Language and Information*, 3:39–76, 1994.

[Fuhrmann, 1988] A. Fuhrmann. *Relevant Logic, Modal Logics, and Theory Change.* PhD thesis, Canberra: Australian National University, 1988.

[Gabbay and Rodrigues, 1996a] D. M. Gabbay and O. Rodrigues. A methodology for iterated theory change. In D. M. Gabbay and Hans Jürgen Ohlbach, editors, *Practical Reasoning — First International Conference on Formal and Applied Practical Reasoning, FAPR'96*, Lecture Notes in Artificial Intelligence. Springer Verlag, 1996.

[Gabbay and Rodrigues, 1996b] D. M. Gabbay and O. Rodrigues. Structured databases: a framework to reason about belief change. In Guy McCusker and Sofia Jourdan, editors, *Proceedings of the Theory and Formal Methods 1996 workshop.* Imperial College - Department of Computing, Imperial College Press, March 1996.

[Gabbay and Rodrigues, 1997] D. M. Gabbay and O. Rodrigues. Structured belief bases: a practical approach to prioritised base revision. In D. M. Gabbay, Rudolf Kruse, Andreas Nonnengart, and Hans Jürgen Ohlbach, editors, *Proceedings of First Internation Joint Conference on Qualitative and Quantitative Practical Reasoning*, pages 267–281. Springer-Verlag, June 1997.

[Gabbay *et al.*, 2000] D. M. Gabbay, O. Rodrigues, and A. Russo. *Information, Uncertainty, Fusion*, chapter Revision by Translation, pages 3–31. Kluwer Academic Publishers, 2000. ISBN: 0-7923-8590-X.

[Gabbay *et al.*, 2002] D. M. Gabbay, O. Rodrigues, and J. Woods. Belief contraction, anti-formulae and resource overdraft: Part I — Deletion in resource bounded logics. *Logic Journal of the IGPL*, 10(6):601–652, November 2002.

[Gabbay *et al.*, 2003] D. M. Gabbay, G. Pigozzi, and J. Woods. Controlled revision – an algorithmic approach for belief revision. *Journal of Logic and Computation*, 13(1):15–35, 2003.

[Gabbay *et al.*, 2004] D. M. Gabbay, O. Rodrigues, and J. Woods. Deletion in resource unbounded logics — Belief contraction, anti-formulae and resource overdraft: Part II. In S. Rahman, J. Symons, D. M. Gabbay, and J. P. van Bendegem, editors, *Logic, Epistemology and the Unity of Science*, volume 1, chapter 16, pages 291–326. Kluwer Academic Publishers, 2004.

[Gabbay *et al.*, 2010] D. M. Gabbay, O. T. Rodrigues, and A. Russo. *Revision, Acceptability and Context*. Springer, 2010. To appear.

[Gabbay, 1996] D. M. Gabbay. *Labelled Deductive Systems*. Oxford University Press, 1996. ISBN 0198538332.

[Gabbay, 2004] D. M. Gabbay. Reactive kripke semantics and arc accessibility. In *Proceedings of CombLog'04, Workshop on Combination of Logics: Theory and Applications*, 2004.

[Gärdenfors and Rott, 1995] P. Gärdenfors and H. Rott. Belief revision. In D. M. Gabbay, C. J. Hogger, and J. A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 4, pages 35–132. Oxford University Press, 1995.

[Gärdenfors, 1978] P. Gärdenfors. Conditionals and changes of belief. *Acta Philosophica Fennica*, 30:381–404, 1978.

[Gärdenfors, 1982] P. Gärdenfors. Rules for rational changes of belief. In T. Pauli, editor, *Philosophical Essays Dedicated to Lannart Aqvist on His Fiftieth Birthday*, volume 34 of *Philosophical Studies*, pages 88–101. Philosophical Society and the Department of Philosophy, Uppsala University, 1982.

[Gärdenfors, 1986] P. Gärdenfors. Belief revision and the Ramsey test for conditionals. *The Philosophical Review*, 95:81–93, 1986.

[Gärdenfors, 1988] P. Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. A Bradford Book — The MIT Press, Cambridge, Massachusetts, and London, England, 1988.

[Gärdenfors, 1990] P. Gärdenfors. The dynamics of belief systems: Foundations $\times$ coherence theories. *Revue Internationale de Philosophie*, 44(172):24–46, 1990.

[Garey and Johnson, 1979] M. R. Garey and D. S. Johnson. *Computers and intractability; a guide to the theory of NP-completeness*. W.H. Freeman, 1979.

[Geanakoplos, 1996] J. Geanakoplos. Three brief proofs of arrow's impossibility theorem. Technical Report 1123R3, Cowles Foundation, Yale University, April 1996. available at http://ideas.repec.org/p/cwl/cwldpp/1123r3.html.

[Ghose, 1991] A. K. Ghose. *Practical belief change*. PhD thesis, PhD thesis, Department of Computing Science, University of Alberta, 1995, 1991.

[Ginsberg and Smith, 1988a] M. L. Ginsberg and David E. Smith. Reasoning about action I: A possible worlds approach. *Artificial Intelligence*, 35:165–195, 1988.

[Ginsberg and Smith, 1988b] M. L. Ginsberg and David E. Smith. Reasoning about action II: A possible worlds approach. *Artificial Intelligence*, 35:311, 1988.

[Gogic *et al.*, 1994] G. Gogic, C. H. Papadimitriou, and M. Sideri. Incremental recompilation of knowledge. In *National Conference on Artificial Intelligence*, pages 922–927, 1994.

[Grahne, 1991a] G. Grahne. Updates and counterfactuals. In *Proc. Second International Conference on Principles of Knowledge Representation and Reasoning (KR '91)*, pages 269–276. Morgan Kaufmann, San Francisco, CA, 1991.

[Grahne, 1991b] G. Grahne. Updates and counterfactuals. In J. A. Allen, R. Fikes, and E. Sandewell, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, pages 269–276, San Mateo, California, 1991. Morgan Kaufmann.

[Grove, 1988] A. Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17:157–170, 1988.

[Hansson, 1994] S. O. Hansson. Kernel contraction. *Journal of Symbolic Logic*, 59:845–859, 1994.

[Hansson, 1997a] S. O. Hansson. Semi-revision. *Journal of Applied Non-Classical Logic*, 7:151–175, 1997.

[Hansson, 1997b] S. O. Hansson. Semi-revision (invited paper). *Journal of Applied Non-Classical Logics*, 7(2), 1997.

[Hansson, 2003] S. O. Hansson. Ten philosophical problems in belief revision. *Journal of Logic and Computation*, 13:37–49, 2003.

[Harman, 1986] G. Harman. *Change in View: Principles of Reasoning*. MIT Press, Cambridge, Massachussetts, 1986.

[Heitmeyer and Bharadwaj, 2000] C. Heitmeyer and R. Bharadwaj. Applying the SCR requirements method to the light control case study. *Journal of Universal Computer Science*, 6(7), 2000. Special Issue on Requirements Engineering: the Light Control Case Study.

[Herzig *et al.*, 2003] A. Herzig, S. Konieczny, and L. Perussel. On iterated revision in the agm framework. In T. D. Nielsen and N. L. Zhang, editors, *Proceedings of the 7th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, number 2711 in Springer Lecture Notes in Artificial Intelligence, July 2003.

[Hunter and Nuseibeh, 1998] A. Hunter and B. Nuseibeh. Managing inconsistent specifications: Reasoning, analysis and action. *Transactions on Software Engineering and Methodology*, October 1998.

[Huth and Ryan, 2000] M. R. Huth and M. D. Ryan. *Logic in Computer Science: Modelling and Reasoning about Systems*. Cambridge University Press, 2000. 387.

[Johansson, 1936] I. Johansson. Der minimalkalkül, ein reduzierter intuitionistischer formalismus. *Compositio mathematica*, 4:119–136, 1936.

[Johnson, 1990] D. S. Johnson. A catalog of complexity classes. In *Handbook of Theoretical Computer Science*, volume A, chapter 2. Elsevier Science Publishers B. V., 1990.

[Katsuno and Mendelzon, 1991a] H. Katsuno and A. O. Mendelzon. Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52:263–294, 1991.

[Katsuno and Mendelzon, 1991b] H. Katsuno and A. O. Mendelzon. Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52(3):263–294, 1991.

[Katsuno and Mendelzon, 1992] H. Katsuno and A. O. Mendelzon. On the difference between updating a knowledge base and revising it. In P. Gärdenfors, editor, *Belief Revision*, pages 183–203. Cambridge University Press, 1992.

[Katsuno and Satoh, 1991] H. Katsuno and K. Satoh. A unified view of consequence relation, belief revision and conditional logic. In J. Mylopoulos and Ray Reiter, editors, *Proceedings of the Twelfth International Conference on Artificial Intelligence*, pages 406–412. Morgan Kaufmann, Sydney, Australia, August 1991.

[Klein and Lu, 1990] M. Klein and S. C.-Y. Lu. Conflict resolution in cooperative design. *The International Journal for Artificial Intelligence in Engineering*, 4(4):168–180, 1990.

[Konieczny and Pérez, 2000] S. Konieczny and R. P. Pérez. A framework for iterated revision, 2000.

[Krentel, 1988] M. W. Krentel. The complexity of optimization problems. *Journal of Computer and System Sciences*, 36(3):490–509, 1988.

[Lehmann, 1995] D. Lehmann. Belief revision, revised. In *Proceedings of the 14th International Joint Conference of Artificial Intelligence (IJCAI-95)*, pages 1534–1540, 1995.

[Lewis, 1973] D. K. Lewis. *Counterfactuals*. Harvard University Press, 1973.

[Liberatore, 1997a] P. Liberatore. The complexity of belief update. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence (IJCAI'97)*, pages 68–73, 1997.

[Liberatore, 1997b] P. Liberatore. The complexity of iterated belief revision. In Foto N. Afrati and Phokion G. Kolaitis, editors, *Database Theory - ICDT '97, 6th International Conference, Delphi, Greece, January 8-10, 1997, Proceedings*, volume 1186 of *Lecture Notes in Computer Science*, pages 276–290. Springer, 1997.

[Łukaszewicz, 1990] W. Łukaszewicz. *Non-monotonic reasoning: formalization of commonsense reasoning.* Ellis Horwood, 1990.

[MacNish and Williams, 1998] C. MacNish and M.-A. Williams. From belief revision to design revision: Applying theory change to changing requirements. In *PRICAI '96: Selected Papers from the Workshop on Reasoning with Incomplete and Changing Information and on Inducing Complex Representations*, pages 206–220. Springer-Verlag, 1998.

[Makinson and Gärdenfors, 1989] D. Makinson and P. Gärdenfors. Relations between the logic of theory change and nonmonotonic logic. In Fuhrmann A. and M. Morreau, editors, *The Logic of Theory Change, Workshop, Lecture Notes in Artificial Intelligence, Volume 465*, Konstanz, FRG, October 1989. Springer Verlag.

[Makinson, 1993] D. Makinson. Five faces of minimality. *Studia Logica*, 52:339–379, 1993.

[Makinson, 2003] D. Makinson. Ways of doing logic: what was different about agm 1985? *Journal of Logic and Computation*, 13(1):3–13, 2003.

[Martins and Shapiro, 1988] J. P. Martins and S. C. Shapiro. A model for belief revision. *Artificial Intelligence*, 35(1):25–79, May 1988.

[McCarthy, 1958] J. McCarthy. Programs with common sense. In *Mechanisation of Thought Processes, Proceedings of the Symposium of the National Physics Laboratory*, pages 77–84, London, UK, 1958. Her Majesty's Stationary Office.

[Modgil, 2009] S. Modgil. Reasoning about preferences in argumentation frameworks. *Artificial Intelligence*, 173:901–934, June 2009.

[Nayak, 1994a] A. Nayak. Iterated belief change based on epistemic entrenchment. *Erkentnis*, 41:353–390, 1994.

[Nayak, 1994b] A. C. Nayak. Iterated belief change based on epistemic entrenchment. *Erkenntnis*, 41:353–390, 1994.

[Nebel, 1989] B. Nebel. A knowlede level analysis of belief revision. In *First International Conference on Principles of Knowledge Representation and Reasoning (KR'89)*, pages 301–311, 1989. Toronto, May 1989, Morgan Kaufmann.

[Nebel, 1991a] B. Nebel. Belief revision and default reasoning: syntax-based approaches. In *Proc. Second International Conference on Principles of Knowledge Representation and Reasoning (KR '91)*, pages 417–428. Morgan Kaufmann, San Francisco, CA, 1991.

[Nebel, 1991b] B. Nebel. Belief revision and default reasoning: Syntax-based approaches. In J. Allen, R. Fikes, and E. Sandewall, editors, *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning*, pages 417–428. Morgan Kaufmann, 1991.

[Nebel, 1992] B. Nebel. Syntax-based approaches to belief revision. In P. Gärdenfors, editor, *Belief Revision*, number 29 in Cambridge Tracts in Theoretical Computer Science, pages 52–88. Cambridge University Press, 1992.

[Nebel, 1998] B. Nebel. How hard is it to revise a belief base. In D. Dubois and H. Prade, editors, *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, volume 3: Belief Change, pages 77–145. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

[Nute, 1984] D. Nute. Conditional logic. In D. M. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, chapter Chapter II.8, pages 387–439. D. Reidel, Dordrecht, 1984.

[Ohlbach, 1991] H. J. Ohlbach. Semantics–based translations methods for modal logics. *Journal of Logic and Computation*, 1(5):691–746, 1991.

[Papadimitriou and Steiglitz, 1982] C. H. Papadimitriou and K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity.* Prentice-Hall, Englewood Cliffs, NJ, 1982.

[Papadimitriou and Yannakakis, 1982] C. H. Papadimitriou and M. Yannakakis. The complexity of facets (and some facets of complexity). *Journal of Computer and System Sciences*, 28(2):244–259, 1982.

[Pollock, 1981] J. L. Pollock. A refined theory of counterfactuals. *Journal of Philosophical Logic*, 10:239–266, 1981.

[Poole *et al.*, 1987] D. Poole, R. Goebel, and R. Aleliunas. Theorist: a logical reasoning system for defaults and diagnosis. In *The Knowledge Frontier: Essays in the Representation of Knowledge*, pages 331–352. Springer-Verlag New York, 1987.

[Poole, 1988a] D. Poole. A logical framework for default reasoning. *Artificial Intelligence*, 36:27–47, 1988.

[Poole, 1988b] D. Poole. A logical framework for default reasoning. *Artificial Intelligence*, 36:27–47, 1988.

[Ramsey, 1931] F. P. Ramsey. Truth and probability. In R. B. Braithwaite, editor, *The Foundations of Probability and other Logical Essays.* Harcourt Brace, New York, 1931.

[Ramsey, 1990] Frank Plumpton Ramsey. *Philosophical Papers.* Cambridge University Press, 1990. Edited by D. H. Mellor.

[Reiter, 1980] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[Restall and Slaney, 1995] G. Restall and J. Slaney. Realistic belief revision. In Michel De Glas and Zdzislaw Pawlak, editors, *WOCFAI 95: Proceedings of the Second World Conference on the Fundamentals of Artificial Intelligence*, pages 67–378. Angkor, July 1995. ISBN 2-87892-009-0.

[Robinson and Volkov, 1998] W. N. Robinson and V. Volkov. Supporting the negotiation life cycle. *Commun. ACM*, 41(5):95–102, 1998.

[Rodrigues *et al.*, 1996] O. Rodrigues, M. Ryan, and P.-Y. Schobbens. Counterfactuals and updates as inverse modalities. In Yoav Shoham, editor, *6th Conference on Theoretical Aspects of Rationality and Knowledge*, pages 163–174, 1996.

[Rodrigues *et al.*, 2003] O. Rodrigues, A. d'Avila Garcez, and A. Russo. Reasoning about requirements evolution using clustered belief revision. In G. Spanoudakis A. S. d'Avila Garcez and A. Zisman, editors, *Proceedings of ACM ESEC/FSE International Workshop on Intelligent Technologies for Software Engineering WITSE03*, 2003.

[Rodrigues *et al.*, 2008] O. Rodrigues, D. M. Gabbay, and A. Russo. Belief revision in non-classical logics. *Review of Symbolic Logic*, 1:267–304, 2008.

[Rodrigues, 1998] O. Rodrigues. *A methodology for iterated information change.* PhD thesis, Department of Computing, Imperial College, 1998.

[Rodrigues, 2003] O. Rodrigues. Structured clusters: A framework to reason with contradictory interests. *Journal of Logic and Computation*, 13(1):69–97, 2003.

[Rodrigues, 2005] O. Rodrigues. *Iterated Revision and Automatic Similarity Generation*, volume 2, pages 591–613. College Publications, 2005.

[Rott, 1992] H. Rott. Preferential belief change using generalized epistemic entrenchment. *Journal of Logic, Language and Information*, 1:45–78, 1992.

[Rott, 2001] H. Rott. *Change, Choice and Inference.* Oxford University Press, 2001.

[Ryan and Schobbens, 1996] M. D. Ryan and P.-Y. Schobbens. Intertranslating counterfactuals and updates. In W. Wahlster, editor, *12th European Conference on Artificial Intelligence (ECAI)*, pages 100–104. J. Wiley & Sons, Ltd., 1996.

[Ryan, 1991] M. D. Ryan. Defaults and revisions in structured theories. In *IEEE Symposium on Logic in Computer Science (LICS)*, pages 362–373, 1991.

[Ryan, 1992] M. D. Ryan. *Ordered Presentation of Theories — Default Reasoning and Belief Revision.* PhD thesis, Department of Computing, Imperial College, U.K., 1992.

[Ryan, 1993] M. D. Ryan. Defaults in specifications. In *IEEE Proceedings of International Symposium on Requirements Engineering*, pages 142–149. IEEE Computer Society Press, 1993.

[Satoh, 1988] K. Satoh. Non-monotonic reasoning by minimal belief revision. In *Proceedings of the International Conference on 5th Generation Computer Systems*, pages 455–462, Tokyo, 1988.

[Shapiro, 1992] S. C. Shapiro. *Entailment*, volume 2, chapter 12, "Applications and Discussion", §83 "Relevance Logic in Computer Science", pages 553–563. Princeton University Press, 1992.

[Spanoudakis and Zisman, 2001] G. Spanoudakis and A. Zisman. Inconsistency management in software engineering: Survey and open research issues. In S. K. Chang, editor, *Handbook of Software Engineering and Knowledge Engineering*, volume 1, pages 329–380. World Scientific Pub. Co., 2001.

[Stalnaker and Thomason, 1970] R. C. Stalnaker and Richmond H. Thomason. A semantic analysis of conditional logic. *Theoria*, 36:23–42, 1970.

[Stalnaker, 1968a] R. Stalnaker. A theory of conditionals. *Studies in Logical Theory, American Philosophical Quarterly Monograph Series, No. 2*, 1968.

[Stalnaker, 1968b] R. C. Stalnaker. *A Theory of Conditionals*, volume 2 of *American Philosophical Quarterly Monograph Series (Nicholas Rescher, ed.)*, pages 98–112. Blackwell, Oxford, 1968.

[Stockmeyer, 1987] L. J. Stockmeyer. Classifying the computational complexity of problems. *Journal of Symbolic Logic*, 52:1–43, 1987.

[Wassermann, 1997] R. Wassermann. Towards a theory of resource-bounded belief revision. In *3rd Dutch-German Workshop in Non-Monotonic Reasoning*, February 26 1997.

[Wassermann, 1999] R. Wassermann. Resource bounded belief revision. *Erkenntnis*, 50:429 – 446, May 1999. ISSN: 0165-0106.

[Weber, 1987] A. Weber. Updating propositional formulas. In L. Kerschberg, editor, *Proceedings of the First Conference on Expert Database Systems*, pages 487–500. The Benjamin Cummings, 1987.

[Williams, 1993] E. Williams. 1st DTI/SERC proteus project workshop: Understanding changing requirements. Address from industrial participants, 1993.

[Williams, 1998] M.-A. Williams. Applications of belief revision. *Lecture Notes in Computer Science*, 1472:287–316, 1998.

[Winslett, 1988a] M. Winslett. Reasoning about action using a possible models approach. In *Proceedings of AAAI-88*, pages 89–93, San Mateo, CA, Saint Paul, MN, 1988. Morgan Kaufmann.

[Winslett, 1988b] M. Winslett. Reasoning about action using a possible models approach. In *Proc. National Conference on Artificial Intelligence (AAAI '88)*, pages 89–93, 1988.

[Wobcke, 1992] W. Wobcke. On the use of epistemic entrenchment in nonmonotonic reasoning. In *10th European Conference on Artificial Intelligence (ECAI'92)*, pages 324–328, 1992.

[Zhang and Foo, 2001] D. Zhang and N. Foo. Infinitary belief revision. *Journal of Philosophical Logic*, 6(30):525–574, 2001.

[Zhang *et al.*, 1997] D. Zhang, S. Chen, W. Zhu, and Z. Chen. Representation theorems for multiple belief changes. In *Proceedings of the 15th IJCAI Conference*, pages 89–94, 1997.

[Zhang, 1996] D. Zhang. Belief revision by sets of sentences. *Journal of Computer Science and Technology*, 11(2):1–19, 1996.

[Zowghi and Offen, 1997] D. Zowghi and R. Offen. A logical framework for modeling and reasoning about the evolution of requirements. In *IEEE Proceedings of Third International Symposium on Requirements Engineering (RE'97)*, pages 247–257. IEEE CS Press, 1997.

[Zowghi, 1997] D. Zowghi. A logical framework for modeling and reasoning about the evolution of requirements. In *Proceedings of the 3rd IEEE International Symposium on Requirements Engineering RE'97*, January 1997.

TOMASZ SKURA

REFUTATION SYSTEMS IN PROPOSITIONAL LOGIC

1 INTRODUCTION

1.1 Basic Concepts

By a *refutation system* **N** we mean an inference system consisting of refutation axioms and refutation rules. Refutation axioms are non-valid formulas, and refutation rules are rules preserving non-validity.

For example, consider the following rule (called *reverse modus ponens*).

$$\frac{B}{A} \qquad (\text{where } \vdash A \to B)$$

It was introduced by Łukasiewicz (who was inspired by Aristotle) [Łukasiewicz, 1951; Łukasiewicz, 1952]. It is a refutation rule for every logic closed under *modus ponens* ($\vdash B$, whenever $\vdash A$ and $\vdash A \to B$).

Another typical example is the rule

$$\frac{A \quad B}{A \vee B} \tag{R_d}$$

reversing the disjunction property ($\vdash A$ or $\vdash B$, whenever $\vdash A \vee B$), which was introduced by Gödel. It is a refutation rule for Intuitionistic Logic ($\mathcal{INT}$), but it is not a refutation rule for Classical Logic ($\mathcal{CL}$).

In Modal Logic we have the following variant of R_d.

$$\frac{A \quad B}{\Box A \vee \Box B} \tag{R_D}$$

It reverses the rule of disjunction (studied in [Hughes and Cresswell, 1984]). We remark that R_D is a refutation rule for several standard modal logics, including $S4$.

We say that a formula A is *refutable* in **N**, if A is derivable from refutation axioms by refutation rules.

For example, consider the following refutation system.

Refutation axioms: All formulas that are not in $\mathcal{CL}$.

D. Gabbay and F. Guenthner (eds.),
Handbook of Philosophical Logic, Volume 16, 115–157.
DOI 10.1007/978-94-007-0479-4_2, © Springer Science+Business Media B.V. 2011

Refutation rule: R_d.

Note that it is a refutation system for $\mathcal{INT}$. We can refute the formula $p \vee \neg p$ in this system as follows.

1. p (axiom)
2. $\neg p$ (axiom)
3. $p \vee \neg p$ (1,2,R_d)

Thus our refutation systems are just like traditional axiomatic proof systems, but they generate non-valid formulas rather than valid ones.

Syntactic refutations can be described as *attacks* in argumentation networks. An argumentation network is a pair (S, R), where S is a non-empty set of arguments and $R \subseteq S \times S$ is an attack relation (for more information see [Gabbay, 2008], [Barringer *et al.*, 2005]). In the above example, both 1 and 2 attack 3. As a result, $p \vee \neg p$ is refuted, which can be written as $-p \vee \neg p$, and the derivation $p, \neg p, p \vee \neg p$ justifies this. Thus a node $x \in S$ can be regarded as

$$(\Delta_x, -A_x)$$

where Δ_x is a derivation refuting A_x.

Let

$$\frac{A_1, ..., A_n}{A}$$

be a refutation rule. The corresponding attack can be presented as follows.

$$\frac{(\Delta_1, -A_1), ..., (\Delta_n, -A_n)}{(\Delta_1, ..., \Delta_n, A, -A)}$$

The *reverse modus ponens* rule involves provability, which can be viewed as *support*. This rule can be presented as the following attack.

$$\frac{(\Delta, +A \to B) \quad (\Delta', -B)}{(\Delta', A, -A)}$$

In the context $(\Delta, +A)$, Δ is a proof of A. A node can also have the form $(\Delta_x, +A_x)$.

Reverse modus ponens is a hybrid refutation rule. It involves provability as well as refutability. This feature may seem a disadvantage. However, it is in fact an advantage. This concept is more general than that of pure refutation rule. (Pure rules are special cases of hybrid rules.) Moreover, it makes the refutation machinery stronger by providing an additional derivation engine. And the proof system need not be an axiom system. It can be any convenient procedure generating valid formulas.

1.2 A Problem

Let $\mathbf{N}$ be a refutation system for a logic $\mathcal{L}$. We say that $\mathbf{N}$ is *characteristic* (or *complete*) for $\mathcal{L}$, if every formula $A \notin \mathcal{L}$ is refutable in $\mathbf{N}$.

In some logics, like $\mathcal{CL}$ as well as the modal logics $\mathcal{K}$ and $S5$, finding complete refutation systems is easy. In $\mathcal{CL}$ the conjunctive normal form procedure provides the following characteristic refutation system.

Refutation axioms: All formulas $\bigwedge \mathcal{X} \to \bigvee \mathcal{Y}$, where $\mathcal{X}, \mathcal{Y}$ are finite sets of propositional variables such that $\mathcal{X} \cap \mathcal{Y} = \emptyset$.

Refutation rule: reverse modus ponens

By adding the following refutation rule (introduced in [Lemmon and Scott, 1977]) to the refutation system for $\mathcal{CL}$, we obtain a characteristic refutation system for the modal logic $\mathcal{K}$.

$$\frac{A \to B_1, ..., A \to B_n, C}{\Box A \to C \vee \Box B_1 \vee ... \vee \Box B_n}$$

where $A, B_1, ..., B_n$ are formulas and C is a modal-free formula.

A similar characteristic refutation system for the modal logic $S5$ is provided by the reduction procedure described in [Hughes and Cresswell, 1968].

But in other logics, like the modal logic $S4$, the task is not easy.

One way of showing that a rule

$$\frac{A_1, ..., A_n}{A}$$

is a refutation rule for a logic $\mathcal{L}$ consists in assuming that the sequent $\Rightarrow A$ is provable in some cut-free sequent system for $\mathcal{L}$, analysing *all* possible proofs of $\Rightarrow A$, and concluding that some $\Rightarrow A_i$ is provable. This should also give a complete refutation procedure.

This method was introduced by Gentzen and it became a standard proof-search method using cut-free sequent systems (see e.g. [Kleene, 1952]). As a refutation-search method, it works in some logics, like the modal logic $\mathcal{K}$. But in other logics, like $S4$, the method is problematic and it may produce suspicious results, because it is an intuitively described procedure rather than an exact proof that no proof of $\Rightarrow A$ exists.

Consider the following sequent system for the modal logic $\mathcal{K}$.

Axioms: All sequents $\mathcal{U} \Rightarrow \mathcal{V}$, where $\mathcal{U}, \mathcal{V}$ are sets such that $\mathcal{U} \cap \mathcal{V} \neq \emptyset$ or $\perp \in \mathcal{U}$.

Rules:

$$\frac{\mathcal{U}, A \Rightarrow \mathcal{V}, B}{\mathcal{U} \Rightarrow \mathcal{V}, A \rightarrow B}$$

$$\frac{\mathcal{U} \Rightarrow \mathcal{V}, A \quad \mathcal{U}, B \Rightarrow \mathcal{V}}{\mathcal{U}, A \rightarrow B \Rightarrow \mathcal{V}}$$

$$\frac{\mathcal{U} \Rightarrow A}{\mathcal{U}', \Box\mathcal{U} \Rightarrow \Box A, \mathcal{V}'}$$

Take any sequent $\alpha = \mathcal{U} \Rightarrow \mathcal{V}$ that is not an axiom. Assume that α is provable. If $\mathcal{U} \cup \mathcal{V}$ contains a formula $A \rightarrow B$, then α can be obtained by one of the rules introducing $\rightarrow$, and so some simpler sequents must be provable. And if $\mathcal{U} \cup \mathcal{V}$ has no formula $A \rightarrow B$, then α can be obtained only by the rule introducing $\Box$, so that some simpler sequent must again be provable. This reasoning justifies the following refutation system for $\mathcal{K}$.

Refutation axioms: All sequents $\mathcal{U} \Rightarrow \mathcal{V}$, where $\mathcal{U}, \mathcal{V}$ are sets of atomic formulas such that $\mathcal{U} \cap \mathcal{V} = \emptyset$ and $\bot \notin \mathcal{U}$.

Refutation rules:

$$\frac{\mathcal{U}, A \Rightarrow \mathcal{V}, B}{\mathcal{U} \Rightarrow \mathcal{V}, A \rightarrow B}$$

$$\frac{\mathcal{U}, B \Rightarrow \mathcal{V}}{\mathcal{U}, A \rightarrow B \Rightarrow \mathcal{V}}$$

$$\frac{\mathcal{U} \Rightarrow \mathcal{V}, A}{\mathcal{U}, A \rightarrow B \Rightarrow \mathcal{V}}$$

$$\frac{\{\mathcal{U} \Rightarrow A : A \in \mathcal{V}\}}{\mathcal{U}', \Box\mathcal{U} \Rightarrow \Box\mathcal{V}, \mathcal{V}'}$$

In the last rule $\mathcal{U}', \mathcal{V}'$ are sets of atomic formulas such that $\mathcal{U}' \cap \mathcal{V}' = \emptyset$ and $\bot \notin \mathcal{U}'$.

It may seem that applying this method to other standard modal logics is a routine exercise, and very elegant refutation systems for these logics can be obtained.

But let us consider the modal logic $S4$ and the sequent system presented in [Fitting, 1983, Section 3.6]. The rule for introducing $\Box$ on the left of $\Rightarrow$ is the following.

$$\frac{\mathcal{U}, A \Rightarrow \mathcal{V}}{\mathcal{U}, \Box A \Rightarrow \mathcal{V}} \qquad (L_1)$$

Here a sequent is a pair $\mathcal{U} \Rightarrow \mathcal{V}$, where $\mathcal{U}, \mathcal{V}$ are finite sets of formulas. Note that $\mathcal{U}, A, A \Rightarrow \mathcal{V} = \mathcal{U}, A \Rightarrow \mathcal{V}$ because $\mathcal{U} \cup \{A, A\} = \mathcal{U} \cup \{A\}$.

Now consider the following rule.

$$\frac{B \to C}{\Box B \to C}$$

where B is a formula and C is a propositional variable. It seems that it is a refutation rule for $S4$. Indeed, assume that the sequent $\Rightarrow \Box B \to C$ is provable. Then the sequent $\Box B \Rightarrow C$ is provable, and it can be obtained only by the rule L_1. Hence $B \Rightarrow C$ is provable, and so $\Rightarrow B \to C$ is also provable.

However, let $B = \neg\Box(p \vee \Box\neg\Box p)$ and $C = p$. Then it is true that $\Box B \to C$ is provable, but it is not true that $B \to C$ is provable.

Here is a proof for $\Box B \to C$.

1. $p \Rightarrow p, \Box\neg\Box p$ (axiom)
2. $\Box p \Rightarrow p, \Box\neg\Box p$ $(1, L_1)$
3. $\Box p \Rightarrow p \vee \Box\neg\Box p$ $(2, D_1)$
4. $\Box p \Rightarrow \Box(p \vee \Box\neg\Box p)$ $(3, L_2)$
5. $B, \Box p \Rightarrow$ $(4, N_2)$
6. $\Box B, \Box p \Rightarrow$ $(5, L_1)$
7. $\Box B \Rightarrow \neg\Box p$ $(6, N_1)$
8. $\Box B \Rightarrow p, \Box\neg\Box p$ $(7, L_2)$
9. $\Box B \Rightarrow p \vee \Box\neg\Box p$ $(8, D_1)$
10. $\Box B \Rightarrow \Box(p \vee \Box\neg\Box p), p$ $(9, L_2)$
11. $\Box B, B \Rightarrow p$ $(10, N_2)$
12. $\Box B, \Box B \Rightarrow p$ $(11, L_1)$
13. $\Box B \Rightarrow p$ $(12, \{A, A\} = \{A\})$
14. $\Rightarrow \Box B \to p$ $(13, I_1)$

In order to refute $B \to C$, we use the following refutation rule, which is a generalization of R_D.

$$\frac{A, B, C}{C \vee \Box A \vee \Box B} \qquad (R'_D)$$

where A, B are formulas and C is modal-free. By using a standard semantic construction (see Section 3.4 of this paper), it is easy to check that R'_D is a refutation rule for $S4$. Note that the refutation rule

$$\frac{A, C}{C \vee \Box A} \qquad (R''_D)$$

is a special case of R'_D.

And here is a refutation for $B \to C$.

1. p (refutation axiom)
2. $\neg\Box p$ (refutation axiom)
3. $p \vee \Box\neg\Box p$ (1,2,R''_D)
4. $p \vee \Box(p \vee \Box\neg\Box p)$ (1,3,R''_D)
5. $\neg\Box(p \vee \Box\neg\Box p) \to p$ (4,*reverse mp*, $\vdash (\neg A \to B) \to (B \vee A)$)

A complete refutation procedure for $S4$ is a genuine problem.

1.3 Proving Syntactic Completeness

We solve this problem by modifying the method of proving syntactic completeness that was introduced by Scott in [Scott, 1957]. It can be described as follows.

1. For every formula A we can construct a normal form F with the property that A is valid if and only if F is valid.

2. In addition to *reverse modus ponens* we have a characteristic refutation rule R that is applicable to normal forms and consists in simple syntactic transformations.

3. We adopt some convenient proof system, and we prove, by an inductive argument, that every normal form is either provable or refutable. We emphasize that such a proof is exact, constructive, and simple. It describes a procedure for constructing for a given normal form either a proof or a refutation. (In this article, for the sake of uniformity, we adopt the traditional proof systems.)

4. We adopt some convenient semantic (or algebraic) system, and we check that R preserves non-validity.

This method can be applied to standard modal logics by modifying the normal forms and the characteristic refutation rule. It seems that it works quite well in transitive logics (see [Skura, 2002]).

We remark that one can also prove that a refutation system **N** is characteristic for a logic $\mathcal{L}$ in an indirect way, by obtaining characteristic formulas of some semantic (or algebraic) structures for $\mathcal{L}$ and refuting these formulas in **N**. This method is studied in [Skura, 1989; Skura, 1992; Skura, 1995] and [Goranko, 1994].

1.4 Reduction Procedures

Our refutation systems have *reverse modus ponens* as a refutation rule. This rule involves provability. However, all the applications of *reverse modus ponens* are exactly described in the completeness proof. What is more, the applications of *reverse modus ponens* (as well as *modus ponens*) can be deleted altogether. As a result, we obtain a reduction procedure, which can be regarded as a generalization of the classical conjunctive normal form procedure. In Classical Logic every formula A is reducible to normal forms $F_1, ..., F_n$ such that $A \in \mathcal{CL}$ if and only if every $F_i \in \mathcal{CL}$. In a non-classical logic $\mathcal{L}$, the procedure can be presented as a reduction tree with origin A and finite sets $\mathcal{Z}_1, ..., \mathcal{Z}_m$ of normal forms (whose validity is easy to check) as the end nodes having the following property. $A \in \mathcal{L}$ if and only if some $\mathcal{Z}_i \subseteq \mathcal{L}$. Hence if every formula in some $\mathcal{Z}_i$ is in $\mathcal{L}$ then $A \in \mathcal{L}$, and if some formula in every $\mathcal{Z}_i$ is not in $\mathcal{L}$ then $A \notin \mathcal{L}$.

1.5 General Remarks

Our refutation procedures resemble tableau procedures. But, in essence, these methods are different. The most important differences are the following.

1. Refutation rules preserve non-validity, whereas tableau rules preserve satisfiability. For example, the rule

$$\frac{A \vee B}{A | B}$$

is a tableau rule for Classical Logic, for if $A \vee B$ is true, then A is true or B is true. But its reverse

$$\frac{A \quad B}{A \vee B}$$

is not a refutation rule for $\mathcal{CL}$, because $p \vee \neg p \in \mathcal{CL}$, but $p \notin \mathcal{CL}$ and $\neg p \notin \mathcal{CL}$.

2. The tableau method is primarily a *proof-search* method. A closed tableau for $\neg A$ is a proof of A. And our reduction method is primarily a *refutation-search* method. A finite reduction tree for A is a refutation of A.

   Tableau rules can also be used for producing semantic refutations. It is by *failure* to close a complete tableau for $\neg A$ that a countermodel for A can be constructed. In order to make sure that no proof of A exists, a systematic tableau procedure must be used (see e.g. [Fitting, 1983]). However, in transitive logics, like $S4$, the procedure may produce an infinite tableau. Several techniques for dealing with loops and obtaining finite countermodels are known (for example [Hughes and Cresswell, 1968], [Fitting, 1983]). Our syntactic method is an alternative method of obtaining loop-free refutation procedures. A syntactic refutation is, by definition, a finite procedure consisting in simple transformations of plain formulas (with no semantic information). Thus our method is *complementary* to the standard refutation methods. It produces refutations by using the traditional method of axiomatic proof systems. However, in contrast with the traditional method, our method, in its refined form of reduction procedures, is *modus ponens*-free and it provides a refutation-search algorithm. The idea is that you write a normal form F and then you write simpler normal forms $F_1, ..., F_n$. How efficient this procedure is, is a purely computational problem and will not be considered here. Our goal is a deep understanding of the syntactic workings of propositional logics. In our opinion, the simplicity of the basic concept, which is that of *inference*, makes the subject worth investigating.

Moreover, the method of refutation systems also has the following interesting aspects.

1. A syntactic proof-refutation system expresses positive and negative conditions. This may be useful in studying non-classical logics that are proper subsets of $\mathcal{CL}$ (so some $\mathcal{CL}$ laws are rejected), for example in paraconsistent logics, which are defined by both positive conditions and negative ones. A system $\mathbf{S} = (\mathbf{POS}, \mathbf{NEG})$ (where $\mathbf{POS}, \mathbf{NEG}$ are inference systems generating valid formulas and non-valid formulas, respectively) can be viewed as a refutation device determining the

set of **S**-refutable formulas. If **S** is characteristic for a logic $\mathcal{L}$, then $\mathcal{L}$ is the greatest set satisfying the conditions expressed by **S**.

2. It is possible to construct a refutation system for every finite (modal or intuitionistic) algebra, and then use such systems for defining a refutation system for every (modal or intermediate) logic with the finite model property (that is, characterized by a class of finite models). However, there are logics without the finite model property (that is, they cannot be characterized by any class of finite models) that do have finite characteristic refutation systems (see [Skura, 1992; Skura, 1994]).

This article is an introduction to refutation systems. In Section 2 we offer a systematic exposition of the refutation procedure involving normal forms for Intuitionistic Logic. (Section 2.2 is elementary and can be skipped.) In Section 3, which is the most important part of the paper, we extend this method to the modal logic $S4$. In Section 4 refined reduction procedures are distilled from our proof-refutation procedures. Finally, symmetric inference systems are briefly discussed.

2 INTUITIONISTIC LOGIC

2.1 Preliminaries

Let $\mathcal{FOR}$ be the set of all formulas generated from the set AT consisting of the constant $\bot$ (the false) and the propositional variables $p, q, r, p_1, p_2, ...$ by means of the connectives

$\rightarrow$ (implication), $\wedge$ (conjunction), $\vee$ (disjunction).

(The members of AT are called atoms.) We define the connectives: $\neg$ (negation), $\equiv$ (equivalence), and the constant $\top$ (the true) as follows.

$$\begin{aligned}
&\neg A = A \rightarrow \bot \\
&\top = \neg\bot \\
&A \equiv B = (A \rightarrow B) \wedge (B \rightarrow A).
\end{aligned}$$

We assume that $\wedge, \vee$ bind stronger than $\rightarrow$, and so we also write

$$A \wedge B \rightarrow C \vee D \text{ instead of } (A \wedge B) \rightarrow (C \vee D).$$

By a *model* we mean a system $\mathbf{W} = (\mathcal{W}, \leq, V)$, where $(\mathcal{W}, \leq)$ is a reflexive transitive tree and V is a function assigning either 0 or 1 to every propositional variable a at a point $x \in \mathcal{W}$ and satisfying the following condition.

If $V(a,x) = 1$ and $x \leq y$, then $V(a,y) = 1$.

It follows (by straightforward induction on the complexity of a formula A) that if $V(A,x) = 1$ and $x \leq y$, then $V(A,y) = 1$.

Such a valuation V is extended to the other formulas as follows.

$V(\bot, x) = 0$.
$V(A \wedge B, x) = 1$ iff $V(A,x) = 1$ and $V(B,x) = 1$.
$V(A \vee B, x) = 1$ iff $V(A,x) = 1$ or $V(B,x) = 1$.
$V(A \to B, x) = 1$ iff for all y such that $x \leq y$,
we have $V(B,y) = 1$ whenever $V(A,y) = 1$.

We say that a formula A is *valid* in the model $\mathbf{W}$, if $V(A,x) = 1$ for every $x \in \mathcal{W}$. And we say that A is valid in the tree $(\mathcal{W}, \leq)$, if A is valid in every model on that tree.

It is convenient to define *Intuitionistic Logic* ($\mathcal{INT}$) as the set of formulas valid in all finite reflexive transitive trees.

2.2 Proof System

Axioms:

(I) $A \to (B \to A)$

(II) $(A \to (B \to C)) \to ((A \to B) \to (A \to C))$

(III) $A \wedge B \to A \qquad A \wedge B \to B$

(IV) $A \to (B \to A \wedge B)$

(V) $A \to A \vee B \qquad B \to A \vee B$

(VI) $(A \to C) \to ((B \to C) \to (A \vee B \to C))$

(VII) $\bot \to A$

Rule:

(*modus ponens*) $A, A \to B \;/\; B$

We say that a formula A is *derivable* from a finite set $\mathcal{Z} \subseteq \mathcal{FOR}$ (in symbols $\mathcal{Z} \vdash A$), if there is a finite sequence

$$\begin{array}{c} F_1 \\ \vdots \\ F_n \end{array}$$

of formulas such that $F_n = A$ and for each $1 \leq i \leq n$ either F_i is an axiom or $F_i \in \mathcal{Z}$ or F_i is obtained from preceding formulas by *modus ponens*. If $\mathcal{Z} = \emptyset$ then A is said to be *provable* (in symbols $\vdash A$). We remark that every provable formula is in $\mathcal{INT}$.

We say that formulas A, B are *equivalent*, if $\vdash A \equiv B$.

In this section some elementary facts about this proof system are established. For a start, we show that $A \to A$ is provable.

(VIII) $\vdash A \to A$.

Proof.

(1) $A \to ((A \to A) \to A)$ (I)

(2) $(A \to ((A \to A) \to A)) \to B$
where $B = (A \to (A \to A)) \to (A \to A)$ (II)

(3) $(A \to (A \to A)) \to (A \to A)$ (1, 2, *mp*)

(4) $A \to (A \to A)$ (I)

(5) $A \to A$ (3, 4, *mp*)

■

By a *substitution* we mean a function s from the set of propositional variables to $\mathcal{FOR}$ extended as follows.

$$\begin{aligned} &s(\bot) = \bot \\ &s(A \to B) = sA \to sB \\ &s(A \wedge B) = sA \wedge sB \\ &s(A \vee B) = sA \vee sB. \end{aligned}$$

PROPOSITION 1. The set of provable formulas is closed under the substitution rule.

Proof. Assume that $\vdash A$, that is, there is a proof $A_1, ..., A_n$ of A with the property that every A_i is an axiom or is obtained form preceding formulas by *mp*. By induction on n show that $sA_1, ..., sA_n$, where s is a substitution, is a proof as well (so $\vdash sA_n$). And this follows from the fact that every substitution instance of an axiom is also an axiom and C is provable whenever $B, B \to C$ are. ■

By Classical Logic ($\mathcal{CL}$) we mean the following subset of $\mathcal{FOR}$.

$$\mathcal{CL} = \{A \in \mathcal{FOR} : v(A) = 1 \text{ for every Boolean valuation } v\}$$

where a *Boolean valuation* is a function v from $\mathcal{FOR}$ to $\{0, 1\}$ satisfying the following conditions.

$$\begin{array}{l} v(\bot) = 0. \\ v(A \to B) = 0 \text{ iff } vA = 1 \text{ and } vB = 0. \\ v(A \wedge B) = 1 \text{ iff } vA = vB = 1. \\ v(A \vee B) = 0 \text{ iff } vA = vB = 0. \end{array}$$

PROPOSITION 2. If $\vdash A$ then $A \in \mathcal{CL}$.

Proof. It is not hard to verify that $\mathcal{CL}$ is closed under *modus ponens* and each axiom is in $\mathcal{CL}$. Hence every proof $A_1, ..., A_n$ has the property that $A_n \in \mathcal{CL}$, which gives the result. ■

In order to establish several useful provable formulas of the kind $A \to B$, we assume A (as an additional axiom) and derive B. This standard technique is justified by the following theorem (called Deduction Theorem). Here we write $\mathcal{X}, A \vdash B$ for $\mathcal{X} \cup \{A\} \vdash B$.

PROPOSITION 3. Let $\mathcal{X}$ be a finite set of formulas. Then
$\mathcal{X}, A \vdash B$ if and only if $\mathcal{X} \vdash A \to B$.

Proof. If $\mathcal{X} \vdash A \to B$, then $\mathcal{X}, A \vdash B$ by *mp*, so assume that B is derivable from $\mathcal{X} \cup \{A\}$. Then there is a derivation $A_1, ..., A_n$ such that $A_n = B$ and each A_i is an axiom (or an element of $\mathcal{X} \cup \{A\}$) or is obtained from preceding formulas by *mp*. Starting with A_1, show that each formula $A \to A_i$ is derivable from $\mathcal{X}$. This is easily accomplished by using *mp* and the theorems I,II,VIII. Indeed, if $A_1 \in \mathcal{X} \cup \{A\}$ then $A \to A_1$ is derivable from $\mathcal{X}$ by I and *mp* (or by VIII). And if say A_3 is obtained from A_1 and $A_2 = A_1 \to A_3$ by *mp*, and we know that $A \to A_1$ and $A \to A_2$ are derivable from $\mathcal{X}$, then so is $A \to A_3$ by II and *mp*. Therefore $\mathcal{X} \vdash A \to A_n$, that is, $\mathcal{X} \vdash A \to B$, as required. ■

COROLLARY 4. $\{A_1, ..., A_m\} \vdash B$ iff $\vdash A_1 \to (... \to (A_m \to B)...)$.

Proof. By straightforward induction on m. ■

It is not hard to prove the following.

(IX) $\quad A \to ((A \to B) \to B)$

(X) $(A \to B) \to ((B \to C) \to (A \to C))$

(XI) $(A \equiv B) \to ((B \equiv C) \to (A \equiv C))$

(XII) $(A \to (B \to C)) \equiv (B \to (A \to C))$

(XIII) $(A \wedge B) \equiv (B \wedge A)$

(XIV) $(A \vee B) \equiv (B \vee A)$

(XV) $A \wedge (B \wedge C) \equiv (A \wedge B) \wedge C$

(XVI) $A \vee (B \vee C) \equiv (A \vee B) \vee C$

(XVII) $(A \to (B \to C)) \equiv (A \wedge B \to C)$

(XVIII) $(A \to B \wedge C) \equiv (A \to B) \wedge (A \to C)$

(XIX) $(A \vee B \to C) \equiv (A \to C) \wedge (B \to C)$

For example, in order to prove XI we show that $A \equiv B, B \equiv C \vdash A \equiv C$.

Proof.

(1) $A \equiv B$ (A_1)

(2) $B \equiv C$ (A_2)

(3) $A \to B$ (1, III, *mp*)

(4) $B \to C$ (2, III, *mp*)

(5) $(B \to C) \to (A \to C)$ (3, X, *mp*)

(6) $A \to C$ (4, 5, *mp*)

(7) $C \to A$ (1, 2, III, X, *mp*)

(8) $A \equiv C$ (6, 7, IV, *mp*)

We finally establish the following replacement theorem.

(XX) $(B \equiv C) \to (A \equiv A(B/C))$

where $A(B/C)$ results from A by replacing *some* occurrences of B by C.

Proof. By induction on the complexity of A. If $A \in \mathsf{AT}$ then XX is $(B \equiv C) \to (A \equiv A)$ (or VIII), so that XX is provable. And if A is say

$A_1 \to A_2$ and XX(A_1), XX(A_2) are provable, then $A_2(B/C)$ can be derived from

$$B \equiv C, A_1 \to A_2, A_1(B/C)$$

and A_2 can be derived from

$$B \equiv C, A_1(B/C) \to A_2(B/C), A_1$$

so (by Proposition 3) XX(A) is provable.

COROLLARY 5. $\vdash (A \wedge B) \equiv (A \wedge B(A/\top))$.

Proof. Note that $\vdash A \to (\top \equiv A)$. Moreover by XX

$$(\top \equiv A) \to (B(A/\top) \equiv B)$$

Hence B is derivable from $A \wedge B(A/\top)$, and $B(A/\top)$ is derivable from $A \wedge B$, which gives the result.

COROLLARY 6. Let

$$\begin{aligned} A &= A_1 \wedge (A_2 \wedge ...(A_{n-1} \wedge A_n)...) \\ B &= B_1 \wedge (B_2 \wedge ...(B_{n-1} \wedge B_n)...) \end{aligned}$$

where $\{A_1, ..., A_n\} = \{B_1, ..., B_n\}$ with all $A_i, B_i \in \mathcal{FOR}$. Then A is equivalent to B.

Proof. By induction on n. If $n = 1$ then $A = B$, and so $\vdash A \equiv B$. Assume that $n \geq 2$.

(Case 1) $A_1 = B_1$. Then $A' = A_2 \wedge ... \wedge A_n$ is equivalent to $B' = B_2 \wedge ... \wedge B_n$ by the induction hypothesis. Hence A is equivalent to B by *replacement*.

(Case 2) $A_1 \neq B_1$. Then $A_1 \in \{B_2, ..., B_n\}$, and by the induction hypothesis $B' = B_2 \wedge ... \wedge B_n$ is equivalent to

$$C = A_1 \wedge ... \wedge B_n$$

(obtained from B' by replacing B_2 with A_1 and A_1 with B_2). Hence by *replacement* B is equivalent to $B_1 \wedge C$, which is equivalent (by XIII, XV, and *replacement*) to $A_1 \wedge D$, where D results from C by replacing A_1 by B_1. Now A is equivalent to $A_1 \wedge D$ by the induction hypothesis and *replacement*.

Therefore (by XI) A is equivalent to B, as required. ■

We may thus ignore the order of formulas in such contexts and introduce the symbol $\bigwedge \mathcal{Z}$ standing for a conjunction of the formulas in $\mathcal{Z} = \{A_1, ...A_n\}$. The symbol $\bigvee \mathcal{Z}$ is defined in an analogous way.

For any formulas $A_1, ..., A_n, B$, the symbol

$$A_1, ..., A_n \longrightarrow B$$

stands for $A_1 \to (... \to (A_n \to B)...)$.

PROPOSITION 7. Let $\mathcal{Z} = \{A_1, ..., A_n\}$. Then $A_1, ...A_n \longrightarrow B$ is equivalent to $\bigwedge \mathcal{Z} \to B$.

Proof. By induction on n. If $n = 1$ then this is true, so assume that $n \geq 2$. Then

$$A_1 \to (A_2, ..., A_n \longrightarrow B)$$

is equivalent (by the induction hypothesis and *replacement*) to

$$A_1 \to (\bigwedge\{A_2, ..., A_n\} \to B)$$

which is equivalent (by XVII) to $\bigwedge \mathcal{Z} \to B$, as required. ■

We also introduce the symbol

$$\mathcal{Z} \longrightarrow B$$

standing for any formula $A_1, ..., A_n \longrightarrow B$, where $\{A_1, ..., A_n\} = \mathcal{Z}$.

2.3 Normal Forms

By a *general form* we shall mean a formula

$$F = \mathcal{S} \longrightarrow a$$

where $\mathcal{S} = \mathcal{T} \cup \mathcal{U}_0 \cup \mathcal{U}$ and

$$\mathcal{U} = \{(a_i \to b_i) \equiv c_i : i \in \{1, ..., k\}\}$$

$\mathcal{U}_0$ is a finite set of formulas of the kind

$$b \to c \text{ or } b \to (c \to d) \text{ or } b \to c \vee d \text{ with } b, c, d \in \mathsf{AT},$$

$\mathcal{T}$ is finite set of atoms, and all $a, a_i, b_i, c_i \in \mathsf{AT}$.

The natural number k associated with $\mathcal{U}$ will be called the *rank* of F. (If $\mathcal{S} = \emptyset$ then F is a.)

Moreover we say that a general form is *normal*, if the following condition is satisfied.

If $b \to B \in \mathcal{U}_0$ then $b \notin \mathcal{T}$.

REMARK 8. Every general form $\mathcal{S} \longrightarrow a$ with $\mathcal{T} = \emptyset$ is normal.

And we say that a normal form F is *special*, if the formula

$$F_0 = \mathcal{T} \cup \mathcal{U}_0 \longrightarrow a \vee c_1 \vee ... \vee c_k$$

is not in $\mathcal{CL}$, that is, $v(F_0) = 0$ for some Boolean valuation v.

PROPOSITION 9. Let F be a normal form. Then $\vdash F_0$ if and only if $F_0 \in \mathcal{CL}$.

Proof. If $\vdash F_0$ then $F_0 \in \mathcal{CL}$ by Proposition 2, so let us assume that $F_0 \in \mathcal{CL}$. If $\{a, c_1, ..., c_k, \bot\} \cap \mathcal{T} \neq \emptyset$ then $\vdash F_0$. And if $\{a, c_1, ..., c_k, \bot\} \cap \mathcal{T} = \emptyset$ then $v(F_0) = 0$, where v is the Boolean valuation assigning 1 to the variables $b \in \mathcal{T}$ and 0 to the other ones, so that $F_0 \notin \mathcal{CL}$, which is impossible. Hence $\vdash F_0$, as required. ■

We now describe a procedure for constructing for a given formula A a normal form A'.

First of all, for any formula A we define the set $\mathsf{SUB}(A)$ of subformulas of A thus.

(1) $\mathsf{SUB}(a) = \{a\}$ if $a \in \mathsf{AT}$.

(2) If $A = B \otimes C$ and $\otimes \in \{\to, \wedge, \vee\}$,
then $\mathsf{SUB}(A) = \mathsf{SUB}(B) \cup \mathsf{SUB}(C) \cup \{A\}$.

Second, for each compound subformula B of A we introduce a new propositional variable p_B and we let $p_a = a$ for any atom a occurring in A, so that every subformula B of A has a unique corresponding atom p_B.

Third, we define the formula $\mathsf{N}(A)$ to be $\mathcal{S}_A \longrightarrow p_A$, where

$$\mathcal{S}_A = \{(p_B \otimes p_C) \equiv p_{B \otimes C} : B \otimes C \in \mathsf{SUB}(A), \otimes \in \{\to, \wedge, \vee\}\}$$

Fourth, we replace

$$a \wedge b \equiv c \text{ by } a \to (b \to c), (c \to a), (c \to b)$$
$$a \vee b \equiv c \text{ by } a \to c, b \to c, c \to a \vee b$$

The resulting formula $A' = \mathcal{S} \longrightarrow p_A$ is a general form equivalent to $\mathsf{N}(A)$. It is also a normal form because $\mathcal{S}$ contains no atoms.

REMARK 10. This technique was known to Wajsberg [Wajsberg, 1938], and it is now used in the resolution method (see e.g. [Mints, 1990]).

EXAMPLE 11. Let $A = (\neg p) \vee (\neg\neg p)$, $B = \neg\neg p$, $C = \neg p$. We define

$$p_C = p_1, p_B = p_2, p_A = p_3.$$

Then $\mathsf{N}(A) = \mathcal{S}_A \longrightarrow p_3$, where

$$\mathcal{S}_A = \{(p \to \bot) \equiv p_1, (p_1 \to \bot) \equiv p_2, (p_1 \vee p_2) \equiv p_3\},$$

and $A' = \mathcal{S} \longrightarrow p_3$, where $\mathcal{S} = \mathcal{T} \cup \mathcal{X}_0 \cup \mathcal{X}$ and

$$\begin{aligned}
&\mathcal{T} = \emptyset \\
&\mathcal{U}_0 = \{p_3 \to p_1 \vee p_2, p_1 \to p_3, p_2 \to p_3\} \\
&\mathcal{U} = \{(p \to \bot) \equiv p_1, (p_1 \to \bot) \equiv p_2\}.
\end{aligned}$$

The construction of $\mathsf{N}(A)$ from A has the following properties.

PROPOSITION 12. If $\vdash \mathsf{N}(A)$ then $\vdash A$.

Proof. Assume that $\vdash \mathsf{N}(A)$. Let s be a substitution such that

$$s(p_B) = B$$

for any $B \in \mathsf{SUB}(A) - \{\bot\}$. Since $s(\bot) = \bot$, we have $s(p_B) = p_B$ for every subformula B of A. In particular $sp_A = A$. Consider the formula $s\mathsf{N}(A) = s\mathcal{S}_A \longrightarrow sp_A$. The set $s\mathcal{S}_A$ consists of formulas of the kind $B \equiv B$, so $s\mathsf{N}(A)$ is equivalent to A, which gives the result by Proposition 1. ■

PROPOSITION 13. $\vdash A \to \mathsf{N}(A)$.

Proof. We first show that

$$(*) \qquad \vdash \mathcal{S}_A \longrightarrow (p_B \equiv B)$$

for every subformula B of A.

The proof is by induction on the complexity of B. If $B \in \mathsf{AT}$ then $p_B = B$, so $*$ holds. Assume now that $B = C \otimes D$ with $\otimes \in \{\to, \wedge, \vee\}$ and $*$ is true for C and D. By the definition of $\mathcal{S}_A$ we have

$$\vdash \mathcal{S}_A \longrightarrow ((p_C \otimes p_D) \equiv p_{C\otimes D})$$

and so $\vdash \mathcal{S}_A \longrightarrow (p_{C\otimes D} \equiv (p_C \otimes p_D))$. By the induction hypothesis

$$\vdash \mathcal{S}_A \longrightarrow (p_C \equiv C) \qquad \vdash \mathcal{S}_A \longrightarrow (p_D \equiv D)$$

By using *replacement* we can derive $p_{C\otimes D} \equiv (C \otimes D)$ from

$$p_{C\otimes D} \equiv (p_C \otimes p_D) \qquad p_C \equiv C \qquad p_D \equiv D$$

Hence

$$\bigwedge \mathcal{S}_A \vdash p_{C\otimes D} \equiv (C \otimes D)$$

so that (by Proposition 3)

$$\vdash \mathcal{S}_A \longrightarrow (p_{C\otimes D} \equiv (C \otimes D))$$

which is $*$. Therefore in particular

$$\vdash \mathcal{S}_A \longrightarrow (p_A \equiv A)$$

Hence $\vdash \mathcal{S}_A \longrightarrow (A \to p_A)$, and so $\vdash A \to (\mathcal{S}_A \longrightarrow p_A)$, which means that $\vdash A \to \mathsf{N}(A)$, as required. ■

Since $\mathsf{N}(A)$ is equivalent to A', we also have the following.

COROLLARY 14.

(i) If $\vdash A'$ then $\vdash A$.

(ii) $\vdash A \to A'$.

2.4 Refutation System

Refutation axioms: All special normal forms of rank 0.

Refutation rules:

(*reverse modus ponens*) B/A, where $\vdash A \to B$

$(\mathcal{N}_\mathcal{I})$ $$\frac{F_1, ..., F_k}{F}$$

where $F = \mathcal{S} \longrightarrow a$ is a special normal form and for each $1 \leq i \leq k$

$$F_i = a_i, \mathcal{S} \longrightarrow b_i.$$

(Here we write $A, \mathcal{X} \longrightarrow B$ for $\{A\} \cup \mathcal{X} \longrightarrow B$.) We say that A is *refutable* (in symbols $\dashv A$), if A is derivable in this system.

In order to verify that $\mathcal{N}_\mathcal{I}$ is a refutation rule for $\mathcal{INT}$, assume that $F_i \notin \mathcal{INT}$ $(1 \leq i \leq k)$. Then every F_i is not valid in some finite model $\mathbf{W}_i = (\mathcal{W}_i, \leq_i, V_i)$, so $V_i(F_i, x_i) = 0$ for some $x_i \in \mathcal{W}_i$. Since $(\mathcal{W}_i, \leq_i)$ is a finite tree, we can choose x_i in such a way that $V_i(F_i, x) = 1$ for every successor x of x_i. Consider the subtree $(\mathcal{W}_i', \leq_i')$ of $(\mathcal{W}_i, \leq_i)$ generated by x_i (consisting of x_i and its successors) together with the valuation V_i' defined by the condition: $V_i'(c,x) = V_i(c,x)$ for any $x \in \mathcal{W}_i'$ and any propositional variable c. It is not difficult to check that $V_i'(B,x) = V_i(B,x)$ for every formula B. Hence

$$V_i'(a_i, x_i) = V_i'(\bigwedge \mathcal{S}, x_i) = 1 \qquad V_i'(b_i, x_i) = 0$$

for every $1 \leq i \leq k$. Construct a new tree Υ with a new root w and $x_1, ..., x_k$ (with their trees) as the immediate successors of w. Define a new valuation V on Υ preserving the old valuations $V_1', ..., V_k'$ and such that for every propositional variable c we have: $V(c,w) = 1$ if $c \in \mathcal{T}$ and $V(c,w) = 0$ otherwise. Then the formulas in $\mathcal{T} \cup \mathcal{U}_0$ are true at w. (Note that the formulas in $\mathcal{S}$ are true at every x_i.) Since $\{a, c_1, ..., c_k\} \cap \mathcal{T} = \emptyset$ (for F is special), all a, c_i are false at w. Moreover each $a_i \rightarrow b_i$ is false at x_i, so that all $a_i \rightarrow b_i$ are also false at w (because $w \leq x_i$). Therefore $\mathcal{S}$ is true at w, and so $V(F,w) = 0$, which gives the result.

2.5 Syntactic Completeness

PROPOSITION 15. Let $A, B \in \mathcal{FOR}$.

(i) If If both A and B are provable, then $A \wedge B$ is provable.

(ii) If A is refutable or B is refutable, then $A \wedge B$ is refutable.

Proof. If $\vdash A$ and $\vdash B$, then $\vdash A \wedge B$ by *mp* and IV. And if $\dashv A$ or $\dashv B$, then $\dashv A \wedge B$ by *rmp* and III, as required. ■

By the *length* $l(A)$ of a formula A we mean the number of $\rightarrow$-occurrences in A. More precisely

(1) If $A \in \mathsf{AT}$ then $l(A) = 0$.

(2.1) If $A = B \otimes C$ and $\otimes \in \{\wedge, \vee\}$, then $l(A) = l(B) + l(C)$.

(2.2) If $A = B \rightarrow C$ then $l(A) = l(B) + l(C) + 1$.

We shall also write $A, B, \mathcal{X} \longrightarrow C$ instead of $\{A, B\} \cup \mathcal{X} \longrightarrow C$.

LEMMA 16. Every general form F of rank k is equivalent to a conjunction of normal forms of rank k.

Proof. By induction on the length l of F.

(1) $l = 0$. Then F is a normal form.

(2) $l \geq 1$ and the lemma holds for general forms of length $< l$. We assume that

$$F = b, b \to B, \mathcal{S}' \longrightarrow a$$

where $B \in \{c, c \to d, c \vee d\}$ with $c, d \in \mathsf{AT}$, and

$$\begin{aligned} \mathcal{S}' &= \mathcal{T}' \cup \mathcal{U}_0' \cup \mathcal{U} \\ \mathcal{T}' &= \mathcal{T} - \{b\} \\ \mathcal{U}_0' &= \mathcal{U}_0 - \{b \to B\} \end{aligned}$$

Then F is equivalent to

$$F' = b, B, \mathcal{S}' \longrightarrow a$$

because $b \wedge (b \to B)$ is equivalent to $b \wedge B$.

(Case 1) $B \in \{c, c \to d\}$. Then F' is a general form of length $< l$, so by the induction hypothesis it is equivalent to a conjunction of normal forms of rank k.

(Case 2) $B = c \vee d$. Then F' is equivalent (by XIX) to $G \wedge H$, where

$$\begin{aligned} G &= b, c, \mathcal{S}' \longrightarrow a \\ H &= b, d, \mathcal{S}' \longrightarrow a \end{aligned}$$

Since the induction hypothesis applies to G and H, this gives the result. ■

PROPOSITION 17. Let $A, B \in \mathcal{FOR}$.

(i) $\vdash (A \wedge ((A \to B) \equiv C) \equiv (A \wedge (B \equiv C))$

(ii) $\vdash (A \wedge (A \equiv B)) \equiv (A \wedge B)$.

Proof. By Corollary 5 and the fact that

$$\begin{aligned} &\vdash (\top \to B) \equiv B \\ &\vdash (B \to \top) \equiv \top \\ &\vdash (B \wedge \top) \equiv B. \end{aligned}$$

■

THEOREM 18. *Every normal form is either provable or refutable.*

Proof. By induction on the rank k of a normal form F.

(1) $k = 0$. Then $F = F_0$. If $F_0 \in \mathcal{CL}$ then $\vdash F_0$ by Proposition 9, and so $\vdash F$. And if $F_0 \notin \mathcal{CL}$ then F is a refutation axiom, so that $\dashv F$.

(2) $k \geq 1$ and we assume that the theorem is true for normal forms of rank $< k$. Consider the following formulas.

$$F_i = a_i, \mathcal{S} \longrightarrow b_i$$

$$G_i = a_i \to b_i, \mathcal{S} \longrightarrow a$$

where $1 \leq i \leq k$. Let F_i' result from F_i by replacing $(a_i \to b_i) \equiv c_i$ with $b_i \equiv c_i$, and let G_i' result from G_i by replacing $(a_i \to b_i) \equiv c_i$ with c_i. By Proposition 17 the formulas F_i', G_i' are equivalent to F_i, G_i, respectively. Moreover F_i', G_i' are general forms of rank $< k$. By Lemma 16 they are equivalent to conjunctions of normal forms of rank $< k$, which, by the induction hypothesis and Proposition 15, are provable or refutable. Hence, by using *mp* as well as *rmp*, we have

$$\vdash F_i \text{ or } \dashv F_i$$
$$\vdash G_i \text{ or } \dashv G_i$$

for every $1 \leq i \leq k$. Now if $\dashv G_i$ for some i then $\dashv F$ by *rmp*, so we may assume that $\vdash G_i$ (so that $\vdash \mathcal{S} \longrightarrow (c_i \to a)$) for all i. And if $F_0 \in \mathcal{CL}$ then

$$\vdash \mathcal{S} \longrightarrow a \vee c_1 \vee ... \vee c_k$$

by Proposition 9, so $\vdash \mathcal{S} \longrightarrow a$, and so we may also assume that $F_0 \notin \mathcal{CL}$.

(Case 1) Every F_i is refutable. Then F is refutable by $\mathcal{N}_{\mathcal{I}}$.

(Case 2) Some F_j is provable, so $\vdash \mathcal{S} \longrightarrow (a_j \to b_j)$. Since $\vdash G_j$, we get $\vdash F$.

Therefore either $\vdash F$ or $\dashv F$, which was to be shown. ■

COROLLARY 19. A is provable if and only if $A \in \mathcal{INT}$.

Proof. If A is provable then $A \in \mathcal{INT}$, so let us assume that A is not provable. Then its normal form A' is not provable. (Otherwise, $\mathsf{N}(A)$ is provable, and so $\vdash A$, by Proposition 12, which is impossible.) Hence A' is refutable, by Theorem 18, so that $A' \notin \mathcal{INT}$ (because no refutable formula is in $\mathcal{INT}$). Therefore $A \notin \mathcal{INT}$ (for $\vdash A \to A'$), as required. ■

2.6 Classical Logic

By modifying this procedure, we obtain a simple procedure for $\mathcal{CL}$. We consider only the set FOR of all formulas generated from AT by $\to$. The connectives $\wedge, \vee$ can be handled in a similar way, or they can be defined as follows.

$$A \wedge B = \neg(A \to \neg B) \text{ and } A \vee B = \neg A \to B.$$

By a general form of rank k we now mean a formula

$$F = \mathcal{S} \longrightarrow a$$

where $\mathcal{S} = \mathcal{T} \cup \mathcal{U}_0 \cup \mathcal{U}$ and

$$\mathcal{U} = \{(a_i \to b_i) \to c_i : i \in \{1, ..., k\}\}$$

$\mathcal{U}_0$ is a finite set of formulas of the kind

$$b \to c \text{ or } b \to (c \to d) \text{ with } b, c, d \in \mathsf{AT},$$

$\mathcal{T}$ is finite set of atoms, and all $a, a_i, b_i, c_i \in \mathsf{AT}$.

A general form is normal, if for every $b \to B \in \mathcal{U}_0$, we have $b \notin \mathcal{T}$.

And $\mathsf{N}(A)$ is now $\mathcal{S}_A \longrightarrow p_A$, where

$$\begin{aligned} \mathcal{S}_A = \quad & \{(p_B \to p_C) \to p_{B\to C}, \\ & p_{B\to C} \to (p_B \to p_C) : B \to C \in \mathsf{SUB}(A)\} \end{aligned}$$

Note that $\mathsf{N}(A)$ is a normal form. Moreover

PROPOSITION 20.

(i) If $\mathsf{N}(A) \in \mathcal{CL}$ then $A \in \mathcal{CL}$.

(ii) $A \to \mathsf{N}(A) \in \mathcal{CL}$.

Proof System

Axioms: All normal forms of rank 0 such that $a \in \mathcal{T}$ or $\perp \in \mathcal{T}$.

Rules:

$$(\mathcal{P}) \qquad \frac{A, \neg B, \mathcal{X} \longrightarrow D \qquad C, \mathcal{X} \longrightarrow D}{(A \to B) \to C, \mathcal{X} \longrightarrow D}$$

$$(\mathcal{PN}) \qquad \frac{A, B, \mathcal{X} \longrightarrow C}{A, A \to B, \mathcal{X} \longrightarrow C}$$

Refutation System

Refutation axioms: All normal forms of rank 0 such that $a \notin \mathcal{T}$ and $\bot \notin \mathcal{T}$.

Refutation rules: $\mathcal{PN}$ and

$$(\mathcal{N}') \qquad \frac{A, \neg B, \mathcal{X} \longrightarrow D}{(A \to B) \to C, \mathcal{X} \longrightarrow D}$$

$$(\mathcal{N}'') \qquad \frac{C, \mathcal{X} \longrightarrow D}{(A \to B) \to C, \mathcal{X} \longrightarrow D}$$

THEOREM 21. *Every normal form is either provable or refutable.*

Proof. By induction on the rank k of a normal form F.

(1) $k = 0$. If $a \in \mathcal{T}$ or $\bot \in \mathcal{T}$, then F is an axiom. And if $a \notin \mathcal{T}$ and $\bot \notin \mathcal{T}$, then F is a refutation axiom.

(2) $k \geq 1$ and this is true for normal forms of rank $< k$. Consider the following formulas.

$$F' = a_1, \neg b_1, \mathcal{S}' \longrightarrow a$$
$$F'' = c_1, \mathcal{S}' \longrightarrow a$$

where

$$\mathcal{S}' = \mathcal{T} \cup \mathcal{U}_0 \cup \mathcal{U}'$$
$$\mathcal{U}' = \mathcal{U} - \{(a_1 \to b_1) \to c_1\}$$

Note that F', F'' are general forms of rank $< k$. If F' (F'') is not a normal form, we replace all pairs $b, b \to B$ by b, B. As a result, we obtain normal forms G', G'' of rank $< k$. Note that F' (F'') is derivable from G' (G'') by $\mathcal{PN}$. By the induction hypothesis G', G'' are provable or refutable, so that F', F'' are also provable or refutable. If both F' and F'' are provable, then $\vdash F$ by $\mathcal{P}$. If F' is refutable then $\dashv F$ by $\mathcal{N}'$, and if F'' is refutable then $\dashv F$ by $\mathcal{N}''$. Therefore either $\vdash F$ or $\dashv F$, which was to be shown. ■

3 THE MODAL LOGIC $\mathcal{S}4$

3.1 Preliminaries

We now deal with the set $\mathcal{FORM}$ of all formulas generated from AT by the connectives: $\to$, $\wedge$, $\vee$, and $\Box$ (necessity). We define:

$$\neg A = A \to \bot \text{ and } A \equiv B = (A \to B) \wedge (B \to A).$$

The logic $\mathcal{S}4$ is here defined as the set of formulas valid in all finite cluster trees.

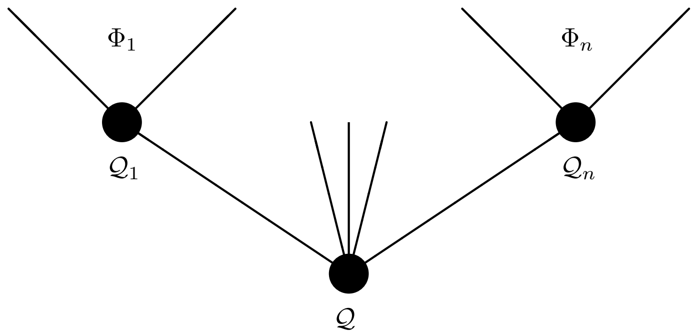
By a *cluster* we mean a system

$$\mathbf{Q} = (\mathcal{Q}, \mathcal{Q} \times \mathcal{Q})$$

where $\mathcal{Q}$ is a finite non-empty set. And by a *finite cluster tree* (with root $\mathcal{Q}$) we mean a system $\Phi = (\mathcal{Q}, \mathcal{W}, \lhd)$ defined as follows.

1. A cluster $\mathbf{Q}$ is a finite cluster tree.

2. If $\Phi_1 = (\mathcal{Q}_1, \mathcal{W}_1, \lhd_1), ..., \Phi_n = (\mathcal{Q}_n, \mathcal{W}_n, \lhd_n)$ are finite cluster trees and $\mathbf{Q}$ is a cluster, then the system Φ:



is a finite cluster tree. More precisely,

$\Phi = (\mathcal{Q}, \mathcal{W}, \lhd)$,
$\mathcal{W} = \mathcal{Q} \cup \mathcal{W}_1 \cup ... \cup \mathcal{W}_n$,
and the relation $\lhd \subseteq \mathcal{W} \times \mathcal{W}$ satisfies the condition: $x \lhd y$ if either $x \lhd_i y$ in some Φ_i or $x, y \in \mathcal{Q}$.

In other words, a finite cluster tree is a finite reflexive transitive tree with clusters instead of points.

3.2 Proof System

Axioms: Every $A \in \mathcal{FORM}$ that is a substitution instance of some $B \in \mathcal{CL}$, and

$$(K) \qquad \Box(A \to B) \to (\Box A \to \Box B)$$

$$(T) \qquad \Box A \to A$$

$$(4) \qquad \Box A \to \Box\Box A$$

Rules:

(*modus ponens*) $A, A \to B/B$
(*necessitation*) $A/\Box A$

A formula A is said to be *provable* (in symbols $\vdash A$), if A is derivable in this system. If $\vdash A \equiv B$, then we say that A is *equivalent* to B.

We now have the following replacement theorem.

$$\vdash \Box(B \equiv C) \to (A \equiv A(B/C)$$

3.3 Normal Forms

For any set $\mathcal{X}$ of formulas, the symbol $\Box\mathcal{X}$ stands for $\{\Box A : A \in \mathcal{X}\}$.

By a *normal form* we shall mean a formula

$$F = \Box\mathcal{S} \longrightarrow c_1$$

where $\mathcal{S} = \mathcal{T} \cup \mathcal{U}_0 \cup \mathcal{U}$,

$$\mathcal{U} = \{\Box a_i \equiv b_i : 1 \leq i \leq k\}$$
$$\mathcal{U}_0 = \{\Box c_j \equiv b_0 : 1 \leq j \leq l\}$$

$\mathcal{T}$ is a finite set of $\Box$-free formulas, and all a_i, b_0, b_i, c_j are atoms. By the *rank* of F we mean the number k. For any $A \in \mathcal{FORM}$ we construct such a form (with no $\mathcal{U}_0$) as follows.

First, we define the set $\mathsf{SUB}(A)$ of subformulas of A in the following way.

(1) $\mathsf{SUB}(A) = \{A\}$ if $A \in \mathsf{AT}$.

(2.1) If $A = B \otimes C$ with $\otimes \in \{\to, \wedge, \vee\}$, then $\mathsf{SUB}(A) = \mathsf{SUB}(B) \cup \mathsf{SUB}(C) \cup \{A\}$.

(2.2) If $A = \Box B$ then $\mathsf{SUB}(A) = \mathsf{SUB}(B) \cup \{A\}$.

Second, with every subformula B of A we associate a unique atom p_B in such a way that $p_b = b$ ($b \in \mathsf{AT}$).

Finally, we define the normal form $\mathsf{N}(A)$ of A thus.

$$\mathsf{N}(A) = \Box\mathcal{S}_A \longrightarrow p_A$$

where

$$\begin{array}{rcl}\mathcal{S}_A & = & \{(p_B \otimes p_C) \equiv p_{B\otimes C} : B \otimes C \in \mathsf{SUB}(A), \otimes \in \{\rightarrow, \wedge, \vee\}\} \cup \\ & & \{\Box p_B \equiv p_{\Box B} : \Box B \in \mathsf{SUB}(A)\}\end{array}$$

PROPOSITION 22.

(i) *If* $\vdash \mathsf{N}(A)$ *then* $\vdash A$.

(ii) $\vdash A \rightarrow \mathsf{N}(A)$.

Proof. See Section 2.3. ■

3.4 Refutation System

For any normal form F we define the following $\Box$-free formulas.

$$\begin{array}{l} F_0^1 = \mathcal{T} \longrightarrow c_1 \vee b_0 \vee b_1 \vee ... \vee b_k \\ F_0^2 = \mathcal{T} \longrightarrow c_2 \vee b_0 \vee b_1 \vee ... \vee b_k \\ \vdots \\ F_0^l = \mathcal{T} \longrightarrow c_l \vee b_0 \vee b_1 \vee ... \vee b_k. \end{array}$$

By a *special form* we mean a normal form F such that $F_0^j \notin \mathcal{CL}$ for every $1 \leq j \leq l$.

Refutation axioms:
Every special form F of rank 0.

Refutation rules:
reverse modus ponens (B/A with $\vdash A \rightarrow B$) and

$$(\mathcal{N}_4) \qquad \frac{F_1, ..., F_k}{F}$$

where F is a special form and for any $1 \leq i \leq k$

$$F_i = \Box\mathcal{S}, \Box c_1 \longrightarrow a_i$$

In order to show that $\mathcal{N}_4$ is a refutation rule for the logic $\mathcal{S}4$, let us assume that each $F_i \notin \mathcal{S}4$. Then there are finite cluster trees $\Phi_1, ..., \Phi_k$ with roots $\mathcal{Q}_1, ..., \mathcal{Q}_k$ and valuations $V_1, ..., V_k$, respectively, such that for every $1 \leq i \leq k$ there is $x \in \mathcal{Q}_i$ with $V_i(F_i, x) = 0$. Hence a_i is false there and the formulas in $\mathcal{S}$ are true everywhere in Φ_i. By the familiar construction, we now define a new finite cluster tree Φ with a new root $\mathcal{Q} = \{x_1, ..., x_l\}$ whose immediate successors are $\mathcal{Q}_1, ..., \mathcal{Q}_k$ (with their cluster trees) and with a new valuation V preserving the old valuations and satisfying the following conditions.

$$V(a, x_1) = v_1(a), ..., V(a, x_l) = v_l(a)$$

for any propositional variable a, where v_j is a Boolean valuation refuting F_0^j $(1 \leq j \leq l)$. Then c_j is false at x_j and the formulas in $\mathcal{T}$ are true at each $x \in \mathcal{Q}$. Also both $\Box a_i$ and $\Box c_j$ are false at every $x \in \mathcal{Q}$. Since $b_0, b_1, ..., b_k$ are false there, the formulas in $\mathcal{S}$ are true everywhere in Φ, so that F is false at x_1, which gives the result.

3.5 Syntactic Completeness

THEOREM 23. *Every normal form F is either provable or refutable.*

Proof. By induction on the rank k of $F = \Box\mathcal{S} \longrightarrow c_1$.

1. $k = 0$. Then $\mathcal{S} = \mathcal{T} \cup \mathcal{U}_0$. Consider the formulas $F_0^1, ..., F_0^l$. If no F_0^j is in $\mathcal{CL}$, then F is a refutation axiom, and so $\dashv F$. Assume that some $F_0^j \in \mathcal{CL}$. Then

$$\vdash \Box\mathcal{S} \longrightarrow c_j \vee b_0$$

   Since $\vdash \Box\mathcal{S} \longrightarrow (b_0 \to c_j)$ (by the definition of $\mathcal{U}_0$), we have $\vdash \Box\mathcal{S} \longrightarrow c_j$, so $\vdash \Box\mathcal{S} \longrightarrow \Box c_j$ (by *necessitation*), so $\vdash \Box\mathcal{S} \longrightarrow \Box c_1$, and so $\vdash F$. Hence $\vdash F$ or $\dashv F$.

2. $k \geq 1$ and the theorem holds for normal forms of rank $< k$. Consider the following formulas.

$$\begin{aligned} F_i &= \Box\mathcal{S}, \Box c_1 \longrightarrow a_i \\ A_i &= \Box\mathcal{S}, \Box a_i \longrightarrow c_1 \\ B_i &= \Box\mathcal{S}, \Box(\Box c_1 \equiv b_i) \longrightarrow c_1 \end{aligned}$$

   where $1 \leq i \leq k$. Let

$$F_i' = \Box\mathcal{T}_1, \Box\mathcal{U}_0', \Box\mathcal{U}' \longrightarrow a_i$$

   with $\mathcal{T}_1 = \mathcal{T} \cup \{b_0, c_1, ..., c_l\}, \mathcal{U}_0' = \{\Box a_i \equiv b_i\}, \mathcal{U}' = \mathcal{U} - \{\Box a_i \equiv b_i\}$,

$$A_i' = \Box\mathcal{T}_2, \Box\mathcal{U}_0, \Box\mathcal{U}' \longrightarrow c_1$$

with $\mathcal{T}_2 = \mathcal{T} \cup \{a_i, b_i\}$,

$$B_i' = \Box\mathcal{T}_3, \Box\mathcal{U}_0'', \Box\mathcal{U}' \longrightarrow c_1$$

with $\mathcal{T}_3 = \mathcal{T} \cup \{b_0 \equiv b_i\}, \mathcal{U}_0'' = \mathcal{U}_0 \cup \{\Box a_i \equiv b_0\}$.

By simple reductions all F_i, A_i, B_i are equivalent to F_i', A_i', B_i', respectively, which are normal forms of rank $< k$, which by the induction hypothesis are provable or refutable. Hence so are all F_i, A_i, B_i by *mp* and *rmp*.

(Case 1) All F_i are refutable and every $F_0^j \notin \mathcal{CL}$. Then $\dashv F$ by $\mathcal{N}_4$.

(Case 2) Some F_i is provable or some $F_0^j \in \mathcal{CL}$. We may assume that

$$\vdash A_i \qquad \vdash B_i$$

for each $1 \leq i \leq k$. (Otherwise $\dashv F$ by *rmp*.)

(Case 2.1) Some F_i is provable. Then $\vdash \Box\mathcal{S} \longrightarrow (\Box c_1 \equiv \Box a_i)$ (for $\vdash A_i$). Since $\vdash \Box\mathcal{S} \longrightarrow (\Box a_i \equiv b_i)$ (and $\vdash B_i$), this gives $\vdash F$

(Case 2.2) Some $F_0^j \in \mathcal{CL}$. Then

$$\vdash \Box\mathcal{S} \longrightarrow c_j \vee b_0 \vee ... \vee b_k$$

Also $\vdash \Box\mathcal{S} \longrightarrow (b_0 \to c_j)$ and $\vdash \Box\mathcal{S} \longrightarrow (b_i \to \Box a_i)$ $(1 \leq i \leq k)$. Since $\vdash \Box\mathcal{S} \longrightarrow (\Box a_i \to \Box c_1)$ (because $\vdash A_i$), we obtain $\vdash \Box\mathcal{S} \longrightarrow (\Box a_i \to c_j)$, and so $\vdash \Box\mathcal{S} \longrightarrow (b_i \to c_j)$ $(1 \leq i \leq k)$. Hence $\vdash \Box\mathcal{S} \longrightarrow c_j$, so $\vdash \Box\mathcal{S} \longrightarrow \Box c_j$, so $\vdash \Box\mathcal{S} \longrightarrow \Box c_1$, and so $\vdash F$.

Therefore either $\vdash F$ or $\dashv F$, as required. ■

COROLLARY 24. *A is provable if and only if $A \in \mathcal{S}4$.*

4 REDUCTION PROCEDURES

4.1 Reduction Rules

Roughly speaking, the idea behind reduction rules is to have a procedure that can transform a given formula A into a sequence

$$\begin{array}{c} A \\ A_1 \\ \vdots \\ A_z \end{array}$$

of simpler and simpler formulas such that every A_i is valid if and only if its immediate predecessor is valid. At the end we have a formula whose validity is easy to check. If the end formula A_z is valid, then A is valid. And if A_z is non-valid, A is also non-valid. (For some philosophical motivation of this concept (connected with Erotetic Logic) see [Wiśniewski, 2004].)

EXAMPLE 25. The following is a reduction rule for Classical Logic.

$$\frac{(A \to B) \to C}{(\neg A \to C) \land (B \to C)}$$

In such procedures sets of formulas (or sequences of formulas) are more useful than conjunctions of formulas, and so we modify this rule as follows.

$$\frac{(A \to B) \to C}{\neg A \to C \quad B \to C}$$

More precisely, by a (linear) *reduction rule* we mean a set H of pairs $\mathcal{Z}/\mathcal{Z}_1$, where $\mathcal{Z}, \mathcal{Z}_1$ are finite sets of formulas. We say that H is a reduction rule *for* a set $\mathcal{L} \subseteq \mathcal{FORM}$, if for every $\mathcal{Z}/\mathcal{Z}_1 \in \mathsf{H}$ the following condition is satisfied.

$\mathcal{Z} \subseteq \mathcal{L}$ if and only if $\mathcal{Z}_1 \subseteq \mathcal{L}$.

In non-classical logics it is useful to generalize this concept by introducing branching reduction rules. The linear rules are special cases of the branching ones. By a (branching) *reduction rule* we mean a set H of pairs $\mathcal{Z}/\Delta$, where $\Delta = \{\mathcal{Z}_1, ..., \mathcal{Z}_n\}$ and all $\mathcal{Z}, \mathcal{Z}_1, ..., \mathcal{Z}_n$ are finite sets of formulas. We say that H is a reduction rule *for* a set $\mathcal{L} \subseteq \mathcal{FORM}$, if for every $\mathcal{Z}/\Delta \in \mathsf{H}$ the following condition is satisfied.

$\mathcal{Z} \subseteq \mathcal{L}$ if and only if some $\mathcal{Z}_i \subseteq \mathcal{L}$.

In order to simplify the notation we also write

$$\frac{A^1; ...; A^{m_0}}{A_1^1; ...; A_1^{m_1} | ... | A_n^1; ...; A_n^{m_n}}$$

instead of

$$\frac{\{A^1, ..., A^{m_0}\}}{\{\{A_1^1, ..., A_1^{m_1}\}, ..., \{A_n^1, ..., A_n^{m_n}\}\}}$$

And if $n = 1$, we simply write

$$\frac{A^1; ...; A^{m_0}}{A_1^1; ...; A_1^{m_1}}$$

EXAMPLE 26. The following is *not* a reduction rule for $\mathcal{CL}$.

$$\frac{A \vee B}{A|B}$$

Indeed, $p \vee \neg p \in \mathcal{CL}$, but both $p \notin \mathcal{CL}$ and $\neg p \notin \mathcal{CL}$. However, it *is* a reduction rule for $\mathcal{INT}$.

REMARK 27. If H, H' are reduction rules, then $\mathsf{H} \cup \mathsf{H}'$ is also a reduction rule, so we may regard a number of reduction rules as a single one.

4.2 Reduction Systems

By a *reduction system* we shall mean a pair $\mathbf{H} = (\mathsf{H}_0, \mathsf{H}_1)$, where H_0 is a set of formulas (called *simple formulas*) and H_1 is a reduction rule. We say that $\mathbf{H}$ is a reduction system *for* a set $\mathcal{L} \subseteq \mathcal{FORM}$, if H_1 is a reduction rule for $\mathcal{L}$.

Let $\mathcal{Y}$ be a finite set of formulas, and let $\mathbf{H} = (\mathsf{H}_0, \mathsf{H}_1)$ be a reduction system. By an **H**-*reduction tree* Ψ *for* $\mathcal{Y}$ we mean a finite tree consisting of finite sets of formulas that satisfies the following conditions.

1. The origin of Ψ is $\mathcal{Y}$.

2. If $\mathcal{Z}_1, ..., \mathcal{Z}_n$ are the immediate successors of a node $\mathcal{Z}$, then $\{\mathcal{Z}_1, ..., \mathcal{Z}_n\}$ is obtained from $\mathcal{Z}$ by H_1.

3. If $\mathcal{Z}$ is an end node of Ψ, then $\mathcal{Z} \subseteq \mathsf{H}_0$.

We say that $\mathcal{Y}$ is **H**-*reducible*, if there is an **H**-reduction tree for $\mathcal{Y}$.

EXAMPLE 28. Let $\mathbf{H} = (\mathsf{H}_0, \mathsf{H}_1)$ be defined as follows.

(H_0) *Simple formulas*:

The formulas in $\mathcal{FOR} - \mathcal{CL}$.

(H_1) *Reduction rules*:

$(\mathsf{H}_\wedge)$ $\quad \dfrac{\{A \wedge B\}}{\{\{A, B\}\}}$

$(\mathsf{H}_\vee)$ $\quad \dfrac{\{A \vee B, C\}}{\{\{A, C\}, \{B, C\}\}}$

In the simplified notation these rules become

$(\mathsf{H}_\wedge)$ $\quad \dfrac{A \wedge B}{A; B}$

$$(\mathsf{H}_\vee) \qquad \frac{A \vee B; C}{A; C \mid B; C}$$

Then **H** is a reduction system for $\mathcal{INT}$.

EXAMPLE 29. Let **H** be the reduction system defined in Example 28, and let

$$\mathcal{Y} = \{p \wedge ((p \to q) \vee (q \to p))\}$$

Then the following is an **H**-reduction tree for $\mathcal{Y}$.

Reduction

$$\begin{array}{llll} \mathcal{Y} & & & \\ \{p, (p \to q) & \vee & (q \to p)\} & \mathsf{H}_\wedge \\ \{p, p \to q\} & & \{p, q \to p\} & \mathsf{H}_\vee \end{array}$$

In the simplified notation the above tree is this.

$$\begin{array}{lll} p \wedge ((p \to q) & \vee & (q \to p)) \\ p; (p \to q) & \vee & (q \to p) \\ p; p \to q & & p; q \to p \end{array}$$

Let **H** be a reduction system for a set $\mathcal{L} \subseteq \mathcal{FORM}$. We say that **H** is *characteristic for* $\mathcal{L}$ (or *complete for* $\mathcal{L}$), if every finite set $\mathcal{Y}$ is **H**-reducible.

4.3 Intuitionistic Logic

Recall that a general form of rank k is a formula $F = \mathcal{S} \longrightarrow a$, where

$\mathcal{S} = \mathcal{T} \cup \mathcal{U}_0 \cup \mathcal{U}$
$\mathcal{U} = \{(a_i \to b_i) \equiv c_i : i \in \{1, ..., k\}\}$
$\mathcal{U}_0$ is a finite set of formulas of the kind: $b \to c$ or $b \to (c \to d)$ or $b \to c \vee d$ with $b, c, d \in \mathsf{AT}$
$\mathcal{T}$ is finite set of atoms, and all $a, a_i, b_i, c_i \in \mathsf{AT}$.

A general form is normal, if for every $b \to B \in \mathcal{U}_0$ we have $b \notin \mathcal{T}$.

Simple formulas:
Every normal form of rank 0.

Reduction rules:

$$(\mathsf{H}^{\to}) \qquad \frac{A, A \to B, \mathcal{X} \longrightarrow C; \ldots}{A, B, \mathcal{X} \longrightarrow C; \ldots}$$

$$(\mathsf{H}^{\vee}) \qquad \frac{A \vee B, \mathcal{X} \longrightarrow C; \ldots}{A, \mathcal{X} \longrightarrow C; B, \mathcal{X} \longrightarrow C; \ldots}$$

$$(\mathsf{H}_{\mathcal{I}}) \qquad \frac{\{F\} \cup \mathcal{Y}}{\{\{F_1^1, F_2^1\} \cup \mathcal{Y}, \ldots, \{F_1^k, F_2^k\} \cup \mathcal{Y}, \mathcal{C}_0 \cup \mathcal{Y}\}}$$

where $F = \mathcal{S} \longrightarrow a$ is a normal form

$$\begin{array}{rcl} \mathcal{C}_0 & = & \{F_0, F_2^1, \ldots, F_2^k\} \\ F_0 & = & \mathcal{T}, \mathcal{U}_0 \longrightarrow a \vee c_1 \vee \ldots \vee c_k \end{array}$$

and for every $1 \leq i \leq k$

$$\begin{array}{rcl} F_1^i & = & a_i, b_i \to c_i, c_i \to b_i, \mathcal{S}_i \longrightarrow b_i \\ F_2^i & = & c_i, a_i \to b_i, \mathcal{S}_i \longrightarrow a \\ \mathcal{S}_i & = & \mathcal{S} - \{(a_i \to b_i) \equiv c_i\}. \end{array}$$

LEMMA 30. $\mathsf{H}_{\mathcal{I}}$ is a reduction rule for $\mathcal{INT}$.

Proof.

1. $F_1^i \notin \mathcal{INT}$ or $F_2^i \notin \mathcal{INT}$ for every $1 \leq i \leq k$, and $\mathcal{C}_0 \not\subseteq \mathcal{INT}$.

1.1. Some $F_2^i \notin \mathcal{INT}$. Then $a_i \to b_i, \mathcal{S} \longrightarrow a \notin \mathcal{INT}$ (because $A \wedge B$ is equivalent to $A \wedge (A \equiv B)$), so $F \notin \mathcal{INT}$.

1.2. Every $F_2^i \in \mathcal{INT}$. Then every $F_1^i \notin \mathcal{INT}$ and $F_0 \notin \mathcal{CL}$, so there are finite intuitionistic models

$$\mathbf{W}_1, \ldots, \mathbf{W}_k$$

such that each F_1^i is not valid in $\mathbf{W}_i$, and so $a_i, \mathcal{S} \longrightarrow b_i$ is not valid in $\mathbf{W}_i$ $(1 \leq i \leq k)$. By using the familiar construction we show that F is not valid in some finite intuitionistic model.

2.1. $F_1^i, F_2^i \in \mathcal{INT}$ for some $1 \leq i \leq k$. Then $\mathcal{S} \longrightarrow (a_i \to b_i) \in \mathcal{INT}$, and $a_i \to b_i, \mathcal{S} \longrightarrow a \in \mathcal{INT}$. Hence $F \in \mathcal{INT}$.

2.2. $\mathcal{C}_0 \subseteq \mathcal{INT}$. Then $F \in \mathcal{INT}$, which gives the result.

■

Let $\mathcal{Z}$ be a set of general forms. Then the rank of $\mathcal{Z}$ is the greatest rank of $F \in \mathcal{Z}$, and the length of $\mathcal{Z}$ is the greatest length of $F \in \mathcal{Z}$.

LEMMA 31. Every finite set $\mathcal{Z}$ (of rank k) of general forms is reducible to a finite set (of rank k) of normal forms.

Proof. By induction on the length l of $\mathcal{Z}$.

1. $l = 0$. Then every $F \in \mathcal{Z}$ is a normal form.

2. $l \geq 1$. We consider any general form F in $\mathcal{Z} = \{F\} \cup \mathcal{Y}$. We assume that
$$F = b, b \to B, \mathcal{S}' \longrightarrow a$$
with $B \in \{c, c \to d, c \vee d\}$. By applying the rule $\mathsf{H}^{\to}$ we reduce $\mathcal{Z}$ to
$$\{F'\} \cup \mathcal{Y}$$
(see the proof of Lemma 16). If $B \in \{c, c \to d\}$ then F' is a general form of length $< l$. And if $B = c \vee d$ then this set is reducible (by $\mathsf{H}^{\vee}$) to
$$\{G, H\} \cup \mathcal{Y}$$
where G, H are general forms of length $< l$. Note that these reductions do not affect the rank of F.

We now transform the other general forms of length l in the same way, and we obtain a finite set $\mathcal{Z}'$ of length $< l$ and of rank k. By the induction hypothesis $\mathcal{Z}'$ is reducible to a finite set (of rank k) of normal forms. Since $\mathcal{Z}$ is reducible to $\mathcal{Z}'$, this gives the result. ■

THEOREM 32. *Every finite set $\mathcal{Z}$ of normal forms is reducible.*

Proof. By induction on the rank k of $\mathcal{Z}$.

1. $k = 0$. Then every $F \in \mathcal{Z}$ is a simple formula.

2. $k \geq 1$. Let F be any normal form of rank k in $\mathcal{Z} = \{F\} \cup \mathcal{Y}$. By using the rule $\mathsf{H}_{\mathcal{I}}$ we reduce $\mathcal{Z}$ to
$$\{F_1^1, F_2^1\} \cup \mathcal{Y} \quad ... \quad \{F_1^k, F_2^k\} \cup \mathcal{Y} \quad \mathcal{C}_0 \cup \mathcal{Y}$$
Note that all F_1^i, F_2^i are general forms of rank $< k$. By Lemma 31 every set $\{F_1^i, F_2^i\}$ is reducible to a set (of rank $< k$) of normal forms. We transform the other normal forms of rank k in this way until we obtain
$$\mathcal{Z}_1 \quad ... \quad \mathcal{Z}_t$$

where each $\mathcal{Z}_i$ is a finite set (of rank $< k$) of normal forms. By the induction hypothesis every $\mathcal{Z}_i$ is reducible. Therefore $\mathcal{Z}$ is also reducible, which was to be shown.

■

EXAMPLE 33. Let $F = \mathcal{S} \longrightarrow p$, where $\mathcal{S} = \mathcal{U} \cup \mathcal{U}_0$,

$$\begin{array}{lcl} \mathcal{U} & = & \{(p_1 \to p_4) \equiv p_4, (p_2 \to p_5) \equiv p_5, (p_3 \to p_6) \equiv p_6\} \\ \mathcal{U}_0 & = & \{p_1 \to p, p_2 \to p, p_3 \to p, p_4 \to p_2 \vee p_3, p_5 \to p_1 \vee p_3, p_6 \to p_1 \vee p_2\}. \end{array}$$

Reduction

$$\begin{array}{lllll} \{F\} & & & & \\ \{F_1^1, F_2^1\} & \{F_1^2, F_2^2\} & \{F_1^3, F_2^3\} & \mathcal{C}_0 & \qquad (\mathsf{H}_{\mathcal{I}}) \end{array}$$

where

$$\begin{array}{lcl} F_1^i & = & p_i, p_{i+3} \to p_{i+3}, \mathcal{S}_i \longrightarrow p_{i+3} \\ \mathcal{S}_i & = & \mathcal{S} - \{(p_i \to p_{i+3}) \equiv p_{i+3}\} \end{array}$$

Note that every F_1^i is not in $\mathcal{CL}$, and so each $F_1^i \notin \mathcal{INT}$. Also $F_0 \notin \mathcal{CL}$. Thus some formula in every end set is not in $\mathcal{INT}$. Therefore $F \notin \mathcal{INT}$.

4.4 Classical Logic

Simple formulas:

Every normal form of rank 0.

Reduction rules: $\mathsf{H}^{\to}$ and

$$(\mathsf{H}_{\mathcal{CL}}) \qquad \frac{(A \to B) \to C, \mathcal{X} \longrightarrow D; \ldots}{A, \neg B, \mathcal{X} \longrightarrow D; C, \mathcal{X} \longrightarrow D; \ldots}$$

We say that a finite set $\mathcal{Y} \subseteq \mathsf{FOR}$ is *reducible*, if it is reducible in this system.

By the *rank* of a set $\mathcal{Y}$ of general forms we mean the greatest rank of $F \in \mathcal{Y}$.

THEOREM 34. *Every finite set $\mathcal{Y}$ of normal forms is reducible.*

Proof. By induction on the rank k of $\mathcal{Y}$.

1. $k = 0$. Then every $F \in \mathcal{Y}$ is a simple formula.

2. $k \geq 1$ and this holds for sets of rank $< k$. We assume that $\mathcal{Y} = \{F\} \cup \mathcal{U}$ and the rank of F is k. By using the rule $\mathsf{H}_{\mathcal{CL}}$ we reduce $\mathcal{Y}$ to the set $\mathcal{Y}' = \{F', F''\} \cup \mathcal{U}$ (see the proof of Theorem 21). Next, by using $\mathsf{H}^{\rightarrow}$, we reduce F', F'' to G', G'', if necessary. Note that G', G'' are normal forms of rank $< k$. We repeat such reductions with the normal forms of rank k that are in $\mathcal{U}$, until we obtain a finite set $\mathcal{Z}$ of normal forms of rank $< k$. By the induction hypothesis $\mathcal{Z}$ is reducible. Therefore $\mathcal{Y}$ is also reducible, which was to be shown.

■

EXAMPLE 35. Let $F = \mathcal{S} \longrightarrow p_4$, where

$$\begin{array}{lcl} \mathcal{S} & = & \mathcal{T} \cup \mathcal{U}_0 \cup \mathcal{U}, \\ \mathcal{T} & = & \{p, p_1\} \\ \mathcal{U}_0 & = & \{p_5 \rightarrow (p_2 \rightarrow p_5), p_6 \rightarrow (p_3 \rightarrow p_6), \\ & & p_2 \rightarrow p, p_3 \rightarrow p, p_4 \rightarrow p_3, p_5 \rightarrow p_3, p_6 \rightarrow p_2\} \\ \mathcal{U} & = & \{(p_2 \rightarrow p_5) \rightarrow p_5\} \end{array}$$

Reduction

$$\begin{array}{l} \{F\} \\ \{F', F''\}(\mathsf{H}_{\mathcal{CL}}) \\ \{G', F''\}(\mathsf{H}^{\rightarrow}) \end{array}$$

where

$$\begin{array}{lcl} F' & = & p_2, \neg p_5, \mathcal{S}' \longrightarrow p_4 \\ F'' & = & p_5, \mathcal{S}' \longrightarrow p_4 \\ \mathcal{S}' & = & \mathcal{T} \cup \mathcal{U}_0 \\ G' & = & p_2, \neg p_5, \mathcal{T}, \mathcal{U}_0' \longrightarrow p_4 \\ \mathcal{U}_0' & = & \mathcal{U}_0 - \{p_2 \rightarrow p\} \end{array}$$

Note that G' is a normal form of rank 0. Since $p_4 \notin \mathcal{T}$ and $\perp \notin \mathcal{T}$, we have $G' \notin \mathcal{CL}$, so that some formula in the end set is not in $\mathcal{CL}$. Hence $F \notin \mathcal{CL}$.

4.5 The Modal Logic $\mathcal{S}4$

Recall that a normal form of rank k is a formula $F = \Box\mathcal{S} \longrightarrow c_1$, where

$$\begin{array}{lcl} \mathcal{S} & = & \mathcal{T} \cup \mathcal{U}_0 \cup \mathcal{U} \\ \mathcal{U} & = & \{\Box a_i \equiv b_i : 1 \leq i \leq k\} \\ \mathcal{U}_0 & = & \{\Box c_j \equiv b_0 : 1 \leq j \leq l\} \end{array}$$

$\mathcal{T}$ is a finite set of $\Box$-free formulas, and all a_i, b_0, b_i, c_j are atoms.

For every $1 \le j \le l$ we define the following $\Box$-free formula.

$$F_0^j = \mathcal{T} \longrightarrow c_j \vee b_0 \vee ... \vee b_k$$

Simple formulas:

Every normal form of rank 0, and every $\Box$-free formula.

Reduction rule:

$$(\mathsf{H}_4) \qquad \frac{\{F\} \cup \mathcal{Y}}{\{\{F_1^1, F_2^1, F_3^1\} \cup \mathcal{Y}, ..., \{F_1^k, F_2^k, F_3^k\} \cup \mathcal{Y}, \mathcal{C}_1 \cup \mathcal{Y}, ..., \mathcal{C}_l \cup \mathcal{Y}\}}$$

where F is a normal form,

$$\mathcal{C}_j = \{F_0^j, F_2^1, F_2^2, ..., F_2^k\} \qquad (1 \le j \le l)$$

and for every $1 \le i \le k$,

$$F_1^i = \Box\mathcal{T}_1, \Box\mathcal{U}_0', \Box\mathcal{U}' \longrightarrow a_i$$

with $\mathcal{T}_1 = \mathcal{T} \cup \{b_0, c_1, ..., c_l\}, \mathcal{U}_0' = \{\Box a_i \equiv b_i\}, \mathcal{U}' = \mathcal{U} - \{\Box a_i \equiv b_i\}$,

$$F_2^i = \Box\mathcal{T}_2, \Box\mathcal{U}_0, \Box\mathcal{U}' \longrightarrow c_1$$

with $\mathcal{T}_2 = \mathcal{T} \cup \{a_i, b_i\}$,

$$F_3^i = \Box\mathcal{T}_3, \Box\mathcal{U}_0'', \Box\mathcal{U}' \longrightarrow c_1$$

with $\mathcal{T}_3 = \mathcal{T} \cup \{b_0 \equiv b_i\}, \mathcal{U}_0'' = \mathcal{U}_0 \cup \{\Box a_i \equiv b_0\}$.

LEMMA 36. Let F be a normal form of rank 0.

(i) If $F_0^j \notin \mathcal{CL}$ for all $1 \le j \le l$, then $F \notin \mathcal{S}4$.

(ii) If $F_0^j \in \mathcal{CL}$ for some $1 \le j \le l$, then $F \in \mathcal{S}4$.

Proof. See (1) in the proof of Theorem 23. ■

LEMMA 37. H_4 is a reduction rule for $\mathcal{S}4$.

Proof.

(1) $F_1^i, F_2^i, F_3^i \in \mathcal{S}4$ for some $1 \le i \le k$, or $\mathcal{C}_j \subseteq \mathcal{S}4$ for some $1 \le j \le l$.

Since F_1^i (that is F_i'), F_2^i (that is A_i'), F_3^i (that is B_i') are equivalent to F_i, A_i, B_i, respectively (see the proof of Theorem 23), it follows that $F \in \mathcal{S}4$ (see Case 2.1 and Case 2.2 in that proof).

(2) $F_1^i \notin \mathcal{S}4$ or $F_2^i \notin \mathcal{S}4$ or $F_3^i \notin \mathcal{S}4$ for every $1 \leq i \leq k$, and some $A \in \mathcal{C}_j$ is not in $\mathcal{S}4$ for every $1 \leq j \leq l$.

(2.1) Some $F_3^i \notin \mathcal{S}4$ or some $F_2^i \notin \mathcal{S}4$. Then $F \notin \mathcal{S}4$.

(2.2) All $F_2^i, F_3^i \in \mathcal{S}4$. Then every $F_1^i \notin \mathcal{S}4$ and every $F_0^j \notin \mathcal{CL}$ (so F is a special form). Since $\mathcal{N}_4$ is a refutation rule for $\mathcal{S}4$, we get $F \notin \mathcal{S}4$, as required.

As usual, the rank of a set $\mathcal{Z}$ of normal forms is the greatest rank of $F \in \mathcal{Z}$. ■

THEOREM 38. *Let $\mathcal{Z}$ be a finite set of normal forms, and let $\mathcal{X}$ be a finite set of $\Box$-free formulas. Then $\mathcal{Z} \cup \mathcal{X}$ is reducible.*

Proof. By induction on the rank k of $\mathcal{Z}$.

1. $k = 0$. Then every $F \in \mathcal{Z}$ is a simple formula.
2. $k \geq 1$. Let F be any normal form of rank k in $\mathcal{Z} \cup \mathcal{X} = \{F\} \cup \mathcal{Y}$. By using the rule H_4 we reduce this set to

   $$\{F_1^1, F_3^1\} \cup \mathcal{Y} \quad \dots \quad \{F_1^k, F_3^k\} \cup \mathcal{Y} \qquad \mathcal{C}_1 \cup \mathcal{Y} \quad \dots \quad \mathcal{C}_l \cup \mathcal{Y}$$

   Note that all F_1^i, F_2^i, F_3^i are normal forms of rank $< k$. (Of course, each F_0^j is $\Box$-free.) We transform the other normal forms of rank k in this way and we obtain

   $$\mathcal{Z}_1 \quad \dots \quad \mathcal{Z}_t$$

   where each $\mathcal{Z}_i$ is a finite set consisting of normal forms of rank $<$ k (and $\Box$-free formulas). By the induction hypothesis every $\mathcal{Z}_i$ is reducible. Therefore $\mathcal{Z} \cup \mathcal{X}$ is also reducible, which was to be shown.

■

EXAMPLE 39. Let $F = \Box(\Box p \equiv r), \Box(\Box q \equiv r), \Box(\Box s \equiv \bot) \longrightarrow p$.

Reduction

$$\begin{array}{l} \{F\} \\ \{F_1^1, F_2^1, F_3^1\} \quad \{F_0^1, F_2^1\} \quad \{F_0^2, F_2^1\} \qquad\qquad (\mathsf{H}_4) \end{array}$$

where

$$\begin{array}{rcl} F_1^1 & = & \Box(p \wedge q \wedge r), \Box(\Box s \equiv \bot) \longrightarrow s \\ F_0^1 & = & p \vee r \vee \bot \\ F_0^2 & = & q \vee r \vee \bot \end{array}$$

Note that F_1^1 is a normal form of rank 0 and $p \wedge q \wedge r \rightarrow s \vee \bot \notin \mathcal{CL}$, so $F_1^1 \notin S4$ (by Lemma 36). Also $F_0^1 \notin \mathcal{CL}$ and $F_0^2 \notin \mathcal{CL}$. Hence some formula in every end set is not in $\mathcal{S}4$, and so $F \notin \mathcal{S}4$.

5 SYMMETRIC INFERENCE SYSTEMS

5.1 Preliminaries

Let us consider a set $\mathcal{FM}$ consisting of all formulas generated from the atoms by standard connectives. An *inference* is a pair $\mathcal{Z}/A$, where $A \in \mathcal{FM}$ and $\mathcal{Z}$ is a finite set of formulas. Sets of inferences are called *rules.* Note that if $\mathcal{R}, \mathcal{R}'$ are rules, then $\mathcal{R} \cup R'$ is also a rule, so we may view a number of rules as a single one. An *inference system* is a pair $\mathbf{R} = (\mathcal{R}_0, \mathcal{R}_1)$, where $\mathcal{R}_1$ is a rule and $\mathcal{R}_0 \subseteq \mathcal{FM}$. A formula A is said to be **R**-*derivable* from a finite set $\mathcal{Z}$ of formulas (in symbols $\mathcal{Z} \vdash_{\mathbf{R}} A$), if there is a sequence $F_1, ..., F_n$ of formulas such that $F_n = A$ and every F_i is in $\mathcal{R}_0 \cup \mathcal{Z}$ or is obtained from preceding formulas by the rule $\mathcal{R}_1$. We say that a set $\mathcal{L}$ of formulas is **R**-*closed*, if $\mathcal{R}_0 \subseteq \mathcal{L}$ and $\mathcal{L}$ is closed under $\mathcal{R}_1$ (that is, $A \in \mathcal{L}$ whenever $\mathcal{Z} \subseteq \mathcal{L}$ for every inference $\mathcal{Z}/A \in \mathcal{R}_1$).

5.2 Syntactic Refutability

By a *symmetric inference system* we shall mean a pair

$$\mathbf{S} = (\mathbf{POS}, \mathbf{NEG})$$

where **POS** and **NEG** are inference systems. We say that **S** is *consistent*, if for no formula A we have both $\emptyset \vdash_{\mathbf{POS}} A$ and $\emptyset \vdash_{\mathbf{NEG}} A$.

A symmetric inference system can be regarded as a syntactic refutation device. If A is **NEG**-derivable from $\emptyset$, then we say that A is **S**-refutable. What is more, if $A \vdash_{\mathbf{POS}} B$ and B is **S**-refutable, then we also say that A is **S**-refutable.

More formally, we say that a formula A is **S**-*refutable* (in symbols $A \in \mathcal{REF}(\mathbf{S})$), if

$$\emptyset \vdash_{\mathbf{NEGP}} A$$

where

$$\mathbf{NEGP} = (\mathcal{NEG}_0, \mathcal{NEG}_1 \cup \mathcal{N_P}) \text{ and } \mathcal{N_P} = \{C/B : B \vdash_{\mathbf{POS}} C\}$$

Let $\mathcal{L} \subseteq \mathcal{FM}$. We say that $\mathcal{L}$ is **S**-closed, if $\mathcal{L}$ is **POS**-closed and $-\mathcal{L}$ (that is, $\mathcal{FM} - \mathcal{L}$) is **NEG**-closed.

We say that **S** is a symmetric inference system *for* $\mathcal{L}$, if $\mathcal{L}$ is **S**-closed.

PROPOSITION 40. *If* **S** *is a symmetric inference system for* $\mathcal{L}$*, then*

$$\mathcal{L} \subseteq -\mathcal{REF}(\mathbf{S})$$

Proof. It suffices to prove that every **S**-refutable formula A is not in $\mathcal{L}$. Let $F_1, ..., F_n$ be a **NEGP** derivation. We show by induction on n that $A_n \notin \mathcal{L}$.

1. $n = 1$. Then $F_1 \in \mathcal{NEG}_0$. Since $\mathcal{NEG}_0 \subseteq -\mathcal{L}$, we have $F_1 \notin \mathcal{L}$.

2. $n \geq 2$. (Case 1) F_n is obtained from say F_1 by $\mathcal{NEG}_1$. By the induction hypothesis $F_1 \notin \mathcal{L}$, so $F_n \notin \mathcal{L}$ because $-\mathcal{L}$ is $\mathcal{NEG}_1$-closed.

(Case 2) F_n is obtained from say F_1 by $\mathcal{N}_{\mathcal{P}}$. Then $F_n \vdash_{\mathbf{POS}} F_1$. Also $F_1 \notin \mathcal{L}$ by the induction hypothesis. Note that if $B \vdash_{\mathbf{POS}} C$ and $B \in \mathcal{L}$ then $C \in \mathcal{L}$ for $\mathcal{L}$ is **POS**-closed. Therefore $F_n \notin \mathcal{L}$, which gives the result. ∎

Let **S** be a symmetric inference system for $\mathcal{L}$. We say that **S** is *characteristic* for $\mathcal{L}$, if

$$\mathcal{T}(\mathbf{S}) \subseteq \mathcal{L}$$

where $\mathcal{T}(\mathbf{S}) = -\mathcal{REF}(\mathbf{S})$.

5.3 Syntactic Properties

Let $\mathbf{S} = (\mathbf{POS}, \mathbf{NEG})$ be a symmetric inference system. **S** expresses a syntactic property, which may be described as the class $\Phi(\mathbf{S})$ of all **S**-closed sets. This means that a set $\mathcal{L} \subseteq \mathcal{FM}$ has this property, if the following conditions are satisfied.

1. $\mathcal{L}$ contains $\mathcal{POS}_0$ and $\mathcal{L}$ is closed under $\mathcal{POS}_1$.

2. $-\mathcal{L}$ contains $\mathcal{NEG}_0$ and $-\mathcal{L}$ is closed under $\mathcal{NEG}_1$.

Proposition 40 enables us to establish the following interesting facts about the property $\Phi(\mathbf{S})$.

THEOREM 41. *If* **S** *is characteristic for* $\mathcal{L}$*, then* $\mathcal{L}$ *is the greatest* **S**-*closed set.*

Proof. Assume that **S** is characteristic for $\mathcal{L}$. Since **S** is an inference system for $\mathcal{L}$ (that is, $\mathcal{L}$ is **S**-closed), we have $\mathcal{L} \subseteq \mathcal{T}(\mathbf{S})$, so that $\mathcal{L} = \mathcal{T}(\mathbf{S})$. Now any **S**-closed set $\mathcal{L}' \subseteq \mathcal{T}(\mathbf{S})$ by Proposition 40. Hence $\mathcal{L}$ is the greatest **S**-closed set, as required. ∎

Let $\mathcal{X} \subseteq \mathcal{FM}$. By the *strengthening* of **POS** by $\mathcal{X}$ we mean the system

$$\mathbf{POS}^{\mathcal{X}} = (\mathcal{X} \cup \mathcal{POS}_0, \mathcal{POS}_1)$$

We define

$$\mathbf{S}^{\mathcal{X}} = (\mathbf{POS}^{\mathcal{X}}, \mathbf{NEG})$$

THEOREM 42. *If* $\mathbf{S}^{\mathcal{X}}$ *is characteristic for* $\mathcal{L}$*, then* $\mathcal{L}$ *is a maximal* $\mathbf{S}$*-closed set.*

Proof. Suppose that $\mathbf{S}^{\mathcal{X}}$ is characteristic for $\mathcal{L}$ but $\mathcal{L}$ is not a maximal $\mathbf{S}$-closed set. Then there is an $\mathbf{S}$-closed set $\mathcal{L}' \supset \mathcal{L}$, so $A \in \mathcal{L}'$ and $A \notin \mathcal{L}$ for some formula A. Since $\mathcal{T}(\mathbf{S}^{\mathcal{X}}) \subseteq \mathcal{L}$, we have $A \notin \mathcal{T}(\mathbf{S}^{\mathcal{X}})$. Also $\mathcal{L}$ is $\mathbf{S}^{\mathcal{X}}$-closed ($\mathbf{S}^{\mathcal{X}}$ being characteristic for $\mathcal{L}$), so $\mathcal{X} \subseteq \mathcal{L} \subseteq \mathcal{L}'$, and so $\mathcal{L}'$ is $\mathbf{S}^{\mathcal{X}}$-closed. Therefore $\mathcal{L}' \subseteq \mathcal{T}(\mathbf{S}^{\mathcal{X}})$ by Proposition 40, so that $A \notin \mathcal{L}'$, which is a contradiction. ■

We now give a few examples of syntactic properties of this kind that seem interesting. The relevant proofs can be found in the indicated papers.

Intermediate Logics

By *an intermediate logic* we mean a set $\mathcal{L} \subseteq \mathcal{FOR}$ such that

$$\mathcal{INT} \subseteq \mathcal{L} \subseteq \mathcal{CL}$$

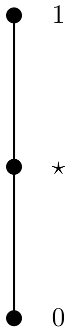
and $\mathcal{L}$ is closed under *substitution* and *modus ponens.*

EXAMPLE 43. [Skura, 1992] The intermediate logics without the law of excluded middle.

Let $\mathbf{S} = ((Int_0, Int_1), \mathbf{NEG})$, where

$$\begin{array}{lcl}
Int_0 & = & \{p \to (q \to p), \quad (p \to (q \to r)) \to ((p \to q) \to (p \to r)), \\
 & & p \wedge q \to p, \quad p \wedge q \to q, \quad p \to (q \to p \wedge q), \\
 & & p \to p \vee q, \quad q \to p \vee q, \quad (p \to r) \to ((q \to r) \to (p \vee q \to r)), \\
 & & \bot \to p\} \\
Int_1 & = & \{A, A \to B/B : A, B \in \mathcal{FOR}\} \cup \{A/s(A) : s \text{ is a substitution}\} \\
\mathcal{NEG}_0 & = & \{p \vee \neg p\} \\
\mathcal{NEG}_1 & = & \emptyset.
\end{array}$$

Then $\mathbf{S}$ is characteristic for the intermediate logic $\mathcal{VAL}(\mathbf{3})$, where $\mathcal{VAL}(\mathbf{3})$ is the set of formulas valid in the three-element intuitionistic algebra $\mathbf{3}$:



Therefore (by Theorem 41) $\mathcal{VAL}(\mathbf{3})$ is the greatest intermediate logic without the law of excluded middle.

We say that an intermediate logic $\mathcal{L}$ has the *disjunction property*, if for any formulas A, B we have

$$A \vee B \notin \mathcal{L} \text{ whenever both } A \notin \mathcal{L} \text{ and } B \notin \mathcal{L}.$$

In other words, $\mathcal{L}$ has the disjunction property, if the rule $A, B/A \vee B$ is a refutation rule for $\mathcal{L}$.

EXAMPLE 44. [Skura, 1992] The intermediate logics with the disjunction property.

Here $\mathbf{S} = ((Int_0, Int_1), \mathbf{NEG})$, where

$$\begin{array}{rcl} \mathcal{NEG}_0 & = & -\mathcal{CL} \\ \mathcal{NEG}_1 & = & \{A, B/A \vee B : A, B \in \mathcal{FOR}\} \end{array}$$

We remark that there are plenty of maximal intermediate logics with the disjunction property (see e.g. [Chagrov and Zakharyashchev, 1991]). For example, the logic $\mathcal{M}$ of finite problems, which is the set of formulas valid in all frames $(2^{\mathcal{X}} - \{\mathcal{X}\}, \subseteq)$, where $\mathcal{X}$ is a finite non-empty set.

Let $\mathbf{S}' = ((Int_0', Int_1), \mathbf{NEG})$, where

$$Int_0' = Int_0 \cup \{(\neg p \to q \vee r) \to (\neg p \to q) \vee (\neg p \to r)\}$$

Then $\mathbf{S}'$ is characteristic for $\mathcal{M}$. Therefore $\mathcal{M}$ is a maximal intermediate logic with the disjunction property (by Theorem 42).

We say that an intermediate logic $\mathcal{L}$ has the *generalized disjunction property*, if for any formulas A_i, B_i we have

$$A \to A_1 \vee ... \vee A_n \notin \mathcal{L} \text{ whenever } A \to A_1 \notin \mathcal{L}, ..., A \to A_n \notin \mathcal{L}$$

where $A = (A_1 \to B_1) \wedge ... \wedge (A_n \to B_n)$.

EXAMPLE 45. [Skura, 1989] The intermediate logics with the generalized disjunction property.

Let $\mathbf{S} = ((Int_0, Int_1), (-\mathcal{CL}, \mathcal{GD}))$, where

$$\begin{array}{rcl} \mathcal{GD} & = & \{A \to A_1, ..., A \to A_n / A \to A_1 \vee ... \vee A_n : \\ & & A = (A_1 \to B_1) \wedge ... \wedge (A_n \to B_n)\} \end{array}$$

Then $\mathbf{S}$ is characteristic for $\mathcal{INT}$. Therefore $\mathcal{INT}$ is the only intermediate logic with the generalized disjunction property.

Paraconsistent Logics

Let $\mathcal{FO}$ be the set of formulas generated form the propositional variables by the connectives:

$$\to, \wedge, \vee, \neg.$$

EXAMPLE 46. [Skura, 2004] The paraconsistent logics with the $\neg$-free $\mathcal{INT}$ and *contraposition.*

Let $\mathbf{S} = (\mathbf{POS}, \mathbf{NEG})$, where

$$\begin{array}{rcl} \mathcal{POS}_0 & = & \{p \to (q \to p), (p \to (q \to r)) \to ((p \to q) \to (p \to r)), \\ & & p \wedge q \to p, p \wedge q \to q, p \to (q \to p \wedge q), \\ & & p \to p \vee q, q \to p \vee q, (p \to r) \to ((q \to r) \to (p \vee q \to r)), \\ & & (p \to q) \to (\neg q \to \neg p)\} \\ \mathcal{POS}_1 & = & \{A, A \to B / B : A, B \in \mathcal{FO}\} \cup \{A/s(A) : s \text{ is a substitution}\} \\ \mathcal{NEG}_0 & = & \{p \to (\neg p \to q)\} \cup (-\mathcal{CL}) \\ \mathcal{NEG}_1 & = & \emptyset \end{array}$$

Then $\mathbf{S}$ is characteristic for the paraconsistent logic $\mathcal{CL} \cap \mathcal{PC}$, where $\mathcal{PC}$ is the set of formulas valid in the algebra $\mathbf{2}'$ resulting from the two-element Boolean algebra $\mathbf{2} = (\{0,1\}, -, \cap)$ by replacing the operation $-$ with the operation t defined by the condition: $t(0) = t(1) = 1$. (This logic was introduced by da Costa and Béziau [da Costa and Béziau, 1993].) Therefore $\mathcal{CL} \cap \mathcal{PC}$ is the greatest paraconsistent logic with the $\neg$-free $\mathcal{INT}$ and *contraposition.*

BIBLIOGRAPHY

[Barringer *et al.*, 2005] H. Barringer, D. Gabbay, and J. Woods. Temporal dynamics of argumentation network. In D. Hutter and W. Stephan, editors, *Mechanizing Mathematical Reasoning*, pages 59–98. Springer, 2005.

[Chagrov and Zakharyashchev, 1991] A. Chagrov and M. Zakharyashchev. The disjunction property of intermediate logics. *Studia Logica*, 50:189–216, 1991.

[da Costa and Béziau, 1993] N.C.A. da Costa and J.-Y. Béziau. Carnot's logic. *Bulletin of the Section of Logic*, 22:98–105, 1993.

[Fitting, 1983] M. Fitting. *Proof Methods for Modal and Intuitionistic Logics.* Reidel, Dordrecht, 1983.

[Gabbay, 2008] D.M. Gabbay. Logical modes of attack in argumentation networks. 2008.

[Goranko, 1994] V. Goranko. Refutation systems in modal logic. *Studia Logica*, 53:199–324, 1994.

[Hughes and Cresswell, 1968] G.G. Hughes and M.J. Cresswell. *An Introduction to Modal Logic.* Methuen, London, 1968.

[Hughes and Cresswell, 1984] G.G. Hughes and M.J. Cresswell. *A Companion to Modal Logic.* Methuen, London, 1984.

[Kleene, 1952] S.C. Kleene. *Introduction to Metamathematics.* North-Holland, Amsterdam, 1952.

[Lemmon and Scott, 1977] E.J. Lemmon and D. Scott. *An Introduction to Modal Logic, in collaboration with D. Scott.* Blackwell, Oxford, 1977.

[Łukasiewicz, 1951] J. Łukasiewicz. Aristotle's Syllogistic from the Standpoint of Modern Formal Logic. Clarendon, Oxford, 1951.

[Łukasiewicz, 1952] J. Łukasiewicz. On the intuitionistic theory of deduction. Indagationes Mathematicae, 14: 202-212, 1952.

[Mints, 1990] G. Mints. Gentzen-type systems and resolution rules. *Lecture Notes in Computer Science*, 417:198–231, 1990.

[Scott, 1957] D. Scott. Completeness proofs for the intuitionistic sentential calculus. In *Summaries of talks presented at the Summer Institute of Symbolic Logic.* Cornell University, 1957. second edition, Princeton, 1960, 231-241.

[Skura, 1989] T. Skura. A complete syntactical characterization of the intuitionistic logic. *Reports on Mathematical Logic*, 23:75–80, 1989.

[Skura, 1992] T. Skura. Refutation calculi for certain intermediate propositional logics. *Notre Dame Journal of Formal Logic*, 33:552–560, 1992.

[Skura, 1994] T. Skura. Syntactic refutations against finite models in modal logic. *Notre Dame Journal of Formal Logic*, 35:595–605, 1994.

[Skura, 1995] T. Skura. A Łukasiewicz-style refutation system for the modal logic S4. *Journal of Philosophical Logic*, 24:573–582, 1995.

[Skura, 2002] T. Skura. Refutations, proofs, and models in the modal logic K4. *Studia Logica*, 70:193–204, 2002.

[Skura, 2004] T. Skura. Maximality and refutability. *Notre Dame Journal of Formal Logic*, 45:65–72, 2004.

[Wajsberg, 1938] M. Wajsberg. Untersuchungen über den Aussagenkalkül von A. Heyting. *Wiadomości Matematyczne*, 46:429–435, 1938.

[Wiśniewski, 2004] A. Wiśniewski. Socratic proofs. *Journal of Philosophical Logic*, 33:299–326, 2004.

E.G. RUYS AND YOAD WINTER

QUANTIFIER SCOPE IN FORMAL LINGUISTICS

1 INTRODUCTION

The remarkable efficiency of language acquisition and linguistic communication must rely on some systematic mapping relating forms and meanings. As a result of this understanding, the study of the relations between syntactic and semantic descriptions has become a central element of all formal linguistic theories. Problems of *quantifier scope* constitute a perennial challenge for uncovering the relations between form and meaning. In some notorious examples, a linguistic element behaves semantically like a logical quantifier, but in a way that is not predicted from straightforward assumptions about its semantics or its position in the syntactic description. In many of these cases, a quantificational expression semantically behaves as if it appeared in a different position than its actual position in the sentence. Such effects are often referred to as *inverse scope effects*. Standard mechanisms that account for these phenomena often complicate the relations between the syntax and the semantics of natural language. As a result, much research has been devoted to the problem of quantifier scope, in an enduring attempt to reveal the status and the theoretical significance of scope shifting principles in formal linguistics.

In this article we give a broad overview of well-known empirical data about quantifier scope and about some proposals for its treatment in the linguistic-logical literature. The article is organized as follows. Section 2 introduces a toy grammar generating a simple fragment of English. This grammar will be used for illustrating and characterizing the problem of quantifier scope, and for discussing some methodological principles for assessing whether linguistic data support an analysis using scope shifting. In Section 3, we give a small inventory of other scope problems in natural language, and then concentrate on the problem of quantified NP scope that is illustrated by the fragment of Section 2. Section 4 discusses some syntactic and semantic theories that address the problem of quantifier scope. Section 5 looks beyond the scope phenomena present in the fragment, and considers some further evidence for the theories of scope shifting discussed in Section 4. Throughout, our empirical data will be drawn from English.

D. Gabbay and F. Guenthner (eds.),
Handbook of Philosophical Logic, Volume 16, 159–225.
DOI 10.1007/978-94-007-0479-4_3,

2 CHARACTERIZING INVERSE SCOPE EFFECTS

This section aims to characterize the problem of inverse scope effects with quantified NPs (QNPs). We start by introducing a small fragment of English with a semantics that illustrates the common notion of *direct scope.* In this semantics, the scope of semantic operators corresponds to simple structural relations in the syntax. However, we show that this simple conception of the syntax-semantics interface is insufficient for capturing some semantic intuitions, which are often referred to as *inverse scope* phenomena. The postulation of *scope ambiguity* is used for describing such cases. After the exposition of these basic notions, we move on to two confounding factors that are especially relevant for identifying inverse scope effects in natural language: the influence of pragmatic effects and of logical dependency between potential readings.

2.1 A "direct scope" grammar for a fragment of English

This section defines a toy grammar for an extremely simple fragment of English. The syntax and the lexicon define the set of *Structural Descriptions* (SDs) of well-formed expressions — the syntactic structures assigned to such expressions by the syntactic derivation, which in the given fragment are described using labeled bracketing notation. The semantics for the fragment is defined by means of a *translation function* (denoted by "⇒") into the simply typed lambda calculus. For details on these standard techniques see [Gamut, 1991].[1]

The following rules describe the syntax of our toy grammar.

Syntax

$$\begin{array}{lllllllll} S & \rightarrow & NPVP & N' & \rightarrow & N & S' & \rightarrow 7RelVP \\ VP & \rightarrow & V_{tr}NP & N' & \rightarrow & N_{tr}NP & & \\ VP & \rightarrow & V & N' & \rightarrow & N'S' & & \\ NP & \rightarrow & DN' & N' & \rightarrow & AN' & & \end{array}$$

These rules do not deal with number and person marking on nouns and verbs; we will silently amend the incorrect forms in our sample SDs. Note

[1] By using the translation to lambda-terms we do not take any position here regarding the necessity of this translation procedure. As many researchers (e.g. [Barker and Jacobson, 2007]) stress, it is possible that syntactic representations of natural language expressions are directly interpreted in a semantic model, with no translation to an intermediate logical language. A more complete discussion of this question and its relevance to the analysis of scope phenomena is beyond the limits of the present paper.

also that we use here a traditional noun phrase structure where modification occurs within an internal nominal labeled N', and the determiner appears at the NP level. For expository purposes, we will not use the more modern syntactic assumptions about *DP* structure (see [Abney, 1987]).[2]

For our exposition we will use the following lexicon over the above toy grammar, including corresponding logical types and translations to logical expressions of the higher order typed lambda calculus.

Lexicon

Cat	**Word**	**Translation**	**Type**
N	one, man, woman, city	$\Rightarrow$ PERSON, MAN, WOMAN, CITY	$\langle e,t\rangle$
$\mathbf{N}_{tr}$	inhabitant of	$\Rightarrow \lambda y\lambda x[\textsc{person}(x)\wedge \textsc{inhabit}(y)(x)]$	$\langle e,\langle e,t\rangle\rangle$
D	every	$\Rightarrow \lambda A\lambda B.\forall x[A(x)\rightarrow B(x)]$	$\langle\langle e,t\rangle,\langle\langle e,t\rangle,t\rangle\rangle$
	no	$\Rightarrow \lambda A\lambda B.\neg\exists x[A(x)\wedge B(x)]$	$\langle\langle e,t\rangle,\langle\langle e,t\rangle,t\rangle\rangle$
	some,*a*	$\Rightarrow \lambda A\lambda B.\exists x[A(x)\wedge B(x)]$	$\langle\langle e,t\rangle,\langle\langle e,t\rangle,t\rangle\rangle$
	$\emptyset$	$\Rightarrow \lambda A\lambda B.\exists 2x[A(x)\wedge B(x)]$	$\langle\langle e,t\rangle,\langle\langle e,t\rangle,t\rangle\rangle$
	three	$\Rightarrow \lambda A\lambda B.\exists 3x[A(x)\wedge B(x)]$	$\langle\langle e,t\rangle,\langle\langle e,t\rangle,t\rangle\rangle$
	five	$\Rightarrow \lambda A\lambda B.\exists x[A(x)\wedge B(x)]$	$\langle\langle e,t\rangle,\langle\langle e,t\rangle,t\rangle\rangle$
	exactly three	$\Rightarrow \lambda A\lambda B.\exists!3x[A(x)\wedge B(x)]$	$\langle\langle e,t\rangle,\langle\langle e,t\rangle,t\rangle\rangle$
	exactly five	$\Rightarrow \lambda A\lambda B.\exists!5x[A(x)\wedge B(x)]$	$\langle\langle e,t\rangle,\langle\langle e,t\rangle,t\rangle\rangle$
A	midwestern	$\Rightarrow \lambda A\lambda x[\textsc{midwestern}(x)\wedge A(x)]$	$\langle\langle e,t\rangle,\langle e,t\rangle\rangle$
NP	John, Mary	$\Rightarrow \lambda A.A(\textsc{john}_e), \lambda A.A(\textsc{mary}_e)$	$\langle\langle e,t\rangle,t\rangle$
V	participated	$\Rightarrow$ PARTICIPATED	$\langle e,t\rangle$
$\mathbf{V}_{tr}$	inhabit	$\Rightarrow$ INHABIT	$\langle e,\langle e,t\rangle\rangle$
	admire	$\Rightarrow$ ADMIRE	$\langle e,\langle e,t\rangle\rangle$
	meet	$\Rightarrow$ MEET	$\langle e,\langle e,t\rangle\rangle$
Rel	who	$\Rightarrow \lambda A\lambda B\lambda x.[A(x)\wedge B(x)]$	$\langle\langle e,t\rangle,\langle\langle e,t\rangle,\langle e,t\rangle\rangle\rangle$

Abbreviations

We use SOME to abbreviate $\lambda A\lambda B.\exists x[A(x)\wedge B(x)]$ (the translation of *some*, *a*);
we use EVERY to abbreviate $\lambda A\lambda B.\forall x[A(x)\rightarrow B(x)]$ (the translation of *every*).

Translation

1. Lexical items translate as stated in the lexicon.

[2]The grammar also uses the traditional designation S for sentence, rather than some theoretically more up-to-date notation, and its variant S' in the (wholly unrealistic) rule for relative clauses.

2. For all $\gamma \in$ SD s.t. $\gamma = [_X \beta]$:
   if $\beta \Rightarrow \beta$' then $\gamma \Rightarrow \beta$'.

3. For all $\gamma \in$ SD s.t. $\gamma = [_X \alpha\beta]$ or $\gamma = [_X \beta\alpha]$:
   if $\alpha \Rightarrow \alpha'_a$ and $\beta \Rightarrow \beta'_{\langle a,c\rangle}$, then $\gamma \Rightarrow \beta'(\alpha')$.

4. For all $\gamma \in$ SD s.t. $\gamma = [_{VP} \alpha\beta]$ (or $\gamma = [_{N'} \alpha\beta]$) where α is a Vtr (Ntr respectively):
   if $\alpha \Rightarrow \alpha'_{\langle e,\langle e,t\rangle\rangle}$ and $\beta \Rightarrow \beta'_{\langle\langle e,t\rangle,t\rangle}$, then $\gamma \Rightarrow \lambda x.\beta'(\lambda y.\alpha'(y)(x))$.

2.2 Incompleteness of the grammar's "direct scope" strategy

The reader may verify that for sentences (1) and (2), the toy grammar above derives the (simplified) SDs in (1a) and (2a), with the accompanying translations (1b) and (2b). The latter translations contain some obvious abbreviations and appear with their reductions to first order predicate calculus.

1. some woman admires every man
   (a) $[_{NP}$ some woman$]$ $[_{VP}$ admires $[_{NP}$ every man$]]$
   (b) $\textsc{some}(\textsc{woman})(\lambda x.(\textsc{every}(\textsc{man}))(\lambda y.\textsc{admire}(y)(x))$
   (c) $\equiv \exists x[\textsc{woman}(x) \wedge \forall y[\textsc{man}(y) \rightarrow \textsc{admire}(y)(x)]]$

2. some inhabitant of every midwestern city participated
   (a) $[_{NP}$ some inhabitant of $[_{NP}$ every midwestern city$]]$ participated
   (b) $\textsc{some}(\lambda x.(\textsc{every}(\textsc{midwest\_city}))(\lambda y.\textsc{inhabitant\_of}(y)(x)))$ $(\textsc{participated})$
   (c) $\equiv \exists x[[\textsc{person}(x) \wedge \forall y[[\textsc{midwestern}(y) \wedge \textsc{city}(y)] \rightarrow \textsc{inhabit}(y)(x)]] \wedge \textsc{participated}(x)]]$

Although highly simplified, the SDs in (1a) and (2a) display the commonly supposed constituent structures for the English sentences in (1) and (2). In particular, these SDs capture the commonly assumed syntactic asymmetry in (1) between VP-external subject and VP-internal object, and the part-whole relation in (2) between the NP-modifier *in every midwestern city* and the subject NP that contains it. The meanings of lexical items in this grammar are standard in natural language semantics. The four translation rules provide for each sentence in the fragment a translation in the simply typed lambda-calculus, which for (1b) and (2b) reduce to formulas of the first-order predicate calculus (1c) and (2c).[3] In the translations (1b) and (2b), scope relations between translations of quantificational expressions match the constituent structures assumed by the syntax. Also in more comprehensive grammars, keeping to this matching and to simple lexical semantics and interpretative strategies normally leads to the propositions in (1c) and (2c) for (1a) and (2a).

Of the four translation rules of the toy grammar, rules 1 and 2 are trivial. Translation rule 3 embodies a very simple assumption about meaning composition in natural language, under which two lambda terms (or their denotations) can only compose by way of *function application.* Translation rule 4, however, is rather *ad hoc* in the way it composes binary relations of type $\langle e, \langle e, t\rangle\rangle$ with QNP meanings of the type $\langle\langle e, t\rangle, t\rangle$ of *generalized quantifiers* (see section 4.1). The problem of composing meanings of relational predicates with quantifier meanings is conceptually distinct from the problem of scope ambiguity. However, as we shall see in section 4.2, most theories of QNP scope establish a connection between the two problems in one way or another. Thus, translation rule 4 in the above toy grammar should be considered as a provisional assumption for expository purposes, and not as a necessary part of theories of QNP interpretation.

The simple architecture that is assumed in our toy grammar is empirically inadequate, however, and its inadequacy illustrates the problem of quantifier scope. Consider first sentence (1). Many English speakers judge (1) true in case every man is admired by a different woman. Therefore, it is reasonable to assume that (1) has not only the reading in (1b), but also the one in (3) below.

3. $\forall y[\ \textsc{man}(y) \rightarrow \exists x[\textsc{woman}(x) \wedge\ \textsc{admire}(y)(x)]]$

A similar problem is even clearer in sentence (2). This sentence is unlikely to be interpreted using the statement (2b) that the grammar generates, which would entail the unlikely existence of a person who inhabits every midwestern city. Many English speakers judge (2) unambiguous, with a meaning that is expressible using the following formula.

4. $\forall y[[\textsc{midwestern}(y) \wedge\ \textsc{city}(y)] \rightarrow \exists x[\textsc{person}(x) \wedge\ \textsc{inhabit}(y)(x) \wedge\ \textsc{participated}(x)]]$

Here again, we see that the scope relations that the grammar generates in (2b) are different than what semantic intuitions require.

[3]Similar conversions of lambda terms are henceforth performed without mention.

As mentioned above, in the analyses (1b) and (2b), the scope relations between the logical operators are in agreement with the scope relations between the constituents that they correspond to. We will henceforth refer to such analyses as *direct scope*. But the semantics of sentences like (1) and (2) demonstrate that English may also exhibit opposite scope relations as in (3) and (4). Such interpretations will be referred to as *inverse scope*. By extension, when treating examples outside the fragment, we will also speak of *inverse scope* interpretations: statements whose representations in Predicate Calculus or the typed lambda calculus show reversed scope relations with respect to the scope relations between constituents in the commonly assumed syntactic structure. Since many scope relations in the syntactic structure are often obvious or taken for granted by syntactic theories, we will at times sloppily talk about the scope of a *constituent* (e.g. a QNP), where in fact we should properly speak of the scope of the corresponding operator in a translation of the sentence.

2.3 Methodological and empirical principles in the study of quantifier scope

Whether a given English sentence is ambiguous, and if so, whether the relevant ambiguity is one of scope, is a theoretical question that often relies on intricate syntactic and semantic intuitions. Various methodological issues arise when addressing this question, which we would like to discuss at the outset.

First, we will not assume that native speakers have direct knowledge of ambiguity. That is, we do not rely on speakers' intuitions as to whether a sentence is ambiguous; nor do we rely on speakers' ability to report reliably on semantic properties of selected readings of ambiguous sentences, which would require them to consciously differentiate between and select these readings. While the possibility of such "direct access" to different readings of a given sentence is ultimately an empirical question, we will prefer to err on the side of caution. Consequently, our primary data will be native speakers' intuitions on truth and inference as they relate to "raw" utterances. In the present section we discuss some of the difficulties that arise in drawing conclusions from such data. We will introduce some commonly accepted guidelines for evaluating the reliability of native speakers' intuitions, and deciding whether these intuitions support an analysis of the relevant sentences as scopally ambiguous. Section 2.3.1 briefly discusses how pragmatic preferences for particular readings may interfere with the reliability of judgments; Section 2.3.2 discusses the repercussions of logical dependence between readings for the analysis of scope ambiguity. We end this section on methodological issues with a brief note on cross-linguistic variation, section 2.3.3.

2.3.1 Pragmatic effects

Particular interpretations may prove more or less accessible to speakers depending on their plausibility in the given context; such effects may interfere with the semantic judgments we seek. Crucially, a reading may appear to be absent merely because it is implausible. For instance, consider the contrast between the following examples.

5. John saw the man with the telescope.

6. John saw the man with the dog.

Most syntactic theories assume that both (5) and (6) are structurally ambiguous. However, for obvious pragmatic reasons the ambiguity is much clearer in (5) than in (6). Thus, we should be wary of trusting the judgment that a sentence lacks a particular reading, if that reading is an implausible one. The safest course is to accept that a reading is absent only if we have found it absent despite its being plausible — or better, despite its being the only plausible reading of the sentence. Consider for instance the following sentence.[4]

7. $[_{NP}$ someone $[_{S'}$ who inhabits every midwestern city]] participated

   (a) $\exists x[[\textsc{person}(x) \wedge \forall y\ [[\textsc{city}(y) \wedge\ \textsc{midwestern}(y)] \rightarrow \textsc{inhabit}(y)(x)]]\ \textsc{participated}(x)]]$

   (b) $\forall y[[\textsc{city}(y) \wedge\ \textsc{midwestern}(y)] \rightarrow \exists x[\textsc{person}(x) \wedge \textsc{inhabit}(y)(x) \wedge\ \textsc{participated}(x)]]$

Sentence (7) is judged by most speakers to be pragmatically strange, as it asserts that one and the same person can inhabit every midwestern city. Thus, it is safe to assume that sentence (7), by contrast to (2), lacks reading (7b) and allows only reading (7a). As (7a) is an implausible proposition, and (7b) is a plausible one. Thus, the fact that we perceive sentence (7) as stating the implausible statement (7a) and not as stating (7b), is reliable evidence that sentence (7) does not have (7b) as one of its readings. This contrast between sentences (2) and (7) will play an important role in Section 3.2.

[4]To save space, we here and henceforth identify sentences with their SDs unless this may lead to confusion.

2.3.2 Logical dependence between readings

The need to postulate scope ambiguities is particularly hard to demonstrate (or disprove) in case one purported reading entails another one. Consider example (8):

8. $[_S[_{NP}$every man$]$ $[_{VP}$ admires $[_{NP}$ some woman $]]]$

   (a) $\forall x[\ \textsc{man}(x) \rightarrow \exists y[\textsc{woman}(y) \wedge\ \textsc{admire}(y)(x)\ ]]$

   (b) $\exists y[\textsc{woman}(y) \wedge \forall x\ [\textsc{man}(x) \rightarrow\ \textsc{admire}(y)(x)\ ]]$

Cases like (8) have often been cited as allowing the inverse scope reading (8b). However, considering that whenever (8b) is true, (8a) is true as well, how could we determine whether (8) allows the inverse scope reading in addition to its direct scope reading (8a)? Some speakers may indicate that they can interpret (8) as entailing (8b). However, as implied from our assumptions above, we do not want to rely on such judgments, which would presuppose that speakers have the ability to access intuitions on the properties of particular readings of sentences to the exclusion of other readings; the safer course is to assume that speakers cannot separate different readings. It should be stressed that this does not prevent sentences like (8) from being judged scopally ambiguous, we assume that but the decision on this question can only rely on theoretical considerations or on the behavior of similar sentences, and not on direct intuitions concerning the sentence in question. For relevant discussions of this point see [Reinhart, 1976, pp. 190-196; Cooper, 1979, p. 142; Kempson, 1977, ch.8; Kempson and Cormack, 1981; Ruys, 1992, pp. 6-20; Altman *et al.*, 2005].

We might choose to ignore examples such as (8) altogether, and concentrate on examples such as (1), for which the inverse scope reading (3) does not entail the direct reading (1b). Strictly speaking, of course, the argument against an ambiguity analysis of (8) also applies to (1), but an unambiguous analysis for (1) would have to state that it has only the inverse scope reading (3). This is highly unlikely, given the constituent structure of the sentence, and would still support the main point, namely the existence of inverse scope readings. On the other hand, admittedly, the intuitions on the existence of an inverse scope reading for (1) are not solid for all speakers (Kurtzman and MacDonald 1993). In addition, indefinite NPs such as *some woman* in (8) sometimes exhibit exceptional (inverse) scope options that are theoretically interesting, as will be discussed in section 3.3. For these reasons, let us illustrate some methods that are employed in order to test the ability of QNPs to take inverse scope.

2.3.2.1 A. We can construct examples for which the direct scope and inverse scope readings are independent. For instance, in the following example (9), the direct scope analysis in (9a) is logically independent from the inverse scope analysis in (9b).

9. exactly three men admire some woman

   (a) $\exists!3x[\textsc{man}(x) \wedge \exists y[\ \textsc{woman}(y) \wedge \textsc{Admire}(y)(x)\ ]]$

   (b) $\exists y[\ \textsc{woman}(y) \wedge \exists!3x\ [\textsc{man}(x) \wedge\ \textsc{admire}(y)(x)\ ]]$

It can be demonstrated that in cases such as (9), the inverse scope analysis in (9b) captures situations for which many speakers judge sentence (9) to be true, but which are not captured by the direct scope analysis (9a). Hence it is rather reasonable to conclude that indefinite object NPs allow a wide scope interpretation over the subject in cases like (8) as well.

2.3.2.2 B. Another potentially relevant piece of evidence for scope ambiguity in cases like (8) is obtained once they are embedded in negative entailing contexts, as in the following example.

10. it is not the case that every man admires some woman

   (a) $\neg\forall x\ [\ \textsc{man}(x) \rightarrow \exists y[\textsc{woman}(y) \wedge\ \textsc{admire}(y)(x)\ ]]$

   (b) $\neg\exists y[\textsc{woman}(y) \wedge \forall x[\textsc{man}(x) \rightarrow\ \textsc{admire}(y)(x)\ ]]$

The two relevant analyses of (10) in (10a) and (10b) are the negations of (8a) and (8b). Due to the negation, the inverse scope reading (10b) is not logically stronger than the direct scope reading (10a). Hence, we could demonstrate the existence of the (10b) reading by showing that (10) is true in a model in which (10b) is true but (10a) false. Despite the fact that negation as in (10) may facilitate the decision whether the inverse scope analysis reflects a reading of (10), actual judgments have proven rather insecure.[5] This "experimental" difficulty makes it harder to use sentences like (10) as even indirect evidence for deciding whether an inverse scope analysis is justified for (8) as well.

[5]This is further complicated by the possibility that in addition to (10a) and (10b), sentence (10) may also have four potential readings where one of the QNPs, or both of them, takes scope over the negation.

2.3.2.3 C. We can attempt to construct grammatical tests for inverse scope. For example, (11) shows that the indefinite object in (8) introduces a "discourse referent" for a pronoun to pick up:

11. Every man admires some woman. She is really smart.

Evidence that the anaphoric relation in (11) is available only in case the antecedent takes wide scope comes from (12), where anaphora is blocked in the second sentence in case the pronoun in the first sentence is interpreted as "bound" by the subject (i.e. the value picked for *he* in (12) may vary from man to man).

11. ?? Every man admires some woman he knows. She is really smart.

Tests like the ones surveyed in A–C above may be used to justify a scope mechanism for QNPs also in cases where direct semantic intuitions do not necessarily support a n account that is based on scope ambiguity.

2.3.3 A note on cross-linguistic variation

As the reader may have observed, all our empirical data so far have been drawn from one language, English, and we will continue to confine discussion in this article to this language. We feel that this limitation is serious, but defensible, for several reasons. Firstly, scope phenomena in English have been studied much more extensively and in more detail than in any other language. Secondly, our primary interest in this article is not so much in the description of all manner of scope phenomena as it is in the presentation of the various theoretical approaches to scope phenomena that have been proposed. Since English data suffice to illustrate the workings of the various theories of quantifier scope that we will discuss, and since we are not aware of scope phenomena in other languages that cannot in principle be described by means of the theoretical devices we will be presenting, our discussion is not seriously hampered by the limitation to English. Nonetheless, cross-linguistic variation in quantifier scope phenomena is an important topic, not only from a descriptive stand-point, but also because the existence and extent of such variation has important theoretical implications. There are numerous important works describing scope phenomena in other languages than English, including [Huang, 1982; Liu, 1990] for Chinese, [Hoji, 1985] for Japanese, as well as works that focus specifically on cross-linguistic variantion: e.g. [Gil, 1982); Aoun and Li, 1989] various papers in [Szabolcsi, 1997]; see [Pafel, 1994] for an overview.

3 SOME PROBLEMS OF QNP SCOPE

This section gives an overview of some of the major empirical generalizations concerning problems of scope ambiguity in natural language, especially scope ambiguities with QNPs. We first give a short catalogue of phenomena that have been analyzed as scope ambiguities, and then turn to some of the special effects with QNP scope: restrictions on their scope, unexpected wide scope/narrow scope of QNPs, and "mixed scope" effects.

3.1 Overview of some scope phenomena

3.1.0.1 A. QNP-QNP. The following two examples from the fragment of Section 2 have illustrated that sentences with multiple QNPs can show scope ambiguities:

13. some woman admires every man

14. some inhabitant of every city participated

The inverse wide scope for the embedded NP in (14) is known as "inverse linking" (see Gabbay and Moravcsik 1974 and May 1977 for early discussion). This inverse scope reading is prominent, often more prominent than the direct scope reading. Inverse wide scope for the object in simple transitive sentences like (13) is usually less prominent than the direct reading of such sentences, but nonetheless available.

3.1.0.2 B. Negation and QNPs. Sentences containing negation and a quantified NP may show scope ambiguity:

15. John doesn't speak exactly three languages

16. all that glitters is not gold

Sentence (15) can be understood as asserting the falsity of the claim that John speaks exactly three languages. In this case we say that negation takes scope over the QNP. But (15) can also be interpreted as stating that there are exactly three languages that John doesn't speak. In this case the QNP takes scope over the negation. Similarly, in (16) the sentence may either mean that nothing that glitters is gold or that there are glittering things that are not gold. For work on the scope of negation, also with respect to topic-focus structure, see [Jackendoff, 1972, pp. 352-362; Horn, 1989, p. 226; Beghelli and Stowell, 1997; Büring, 1997].

3.1.0.3 C. Intensionality. *De re/de dicto* ambiguities are also commonly analysed as scope ambiguities (cf. [Quine, 1956; Montague, 1973; Ben-Avi and Winter, 2007]). For instance:

17. John is looking for a book

18. an American runner is likely to win the race

In (17), whether the sentence means that John is looking for a specific book or for any book, is often analyzed as a scope ambiguity of the indefinite *a book* with respect to the predicate *look for*. The former, *de re*, reading is often analyzed as a case where the indefinite takes scope over the predicate, whereas the latter, *de dicto*, reading is often analyzed by giving the predicate scope over the indefinite. A similar distinction is made for (18).

3.1.0.4 D. QNPs inside questions. Questions containing quantified NPs and wh-phrases may show a scope ambiguity, as in (19):

19. which woman does every man love?

(19) permits an individual answer ("every man loves Mary"), or a "pair-list" answer ("John loves Sue, Peter loves Mary, ..."). The pair-list reading of the question can be treated as involving *every man* quantifying into the question, taking wider scope than *which woman*; for the individual answer, *every man* scopes below *which woman.* See [Karttunen and Peters, 1980; Engdahl, 1980; Jacobson, 1999; Groenendijk and Stokhof; 1984; 1997; May, 1985].

3.1.0.5 E. Adverbs. Scope relations between adverbs of different types and QNPs may also vary:

20. (a) John has never met a friend of mine
    (b) someone always wins

21. (a) John probably saw an article in this morning's *Times*
    (b) someone probably spiked the punch

For instance, (20a) can either mean that John has met no friend of mine, or that there is a friend of mine that John has never met. When adverbs are analyzed as quantifiers (over times, events, possible worlds etc.), this kind of ambiguity is often analyzed as similar to the QNP-QNP kind of scope ambiguity. For two studies of the scope of adverbs and relevant further references see [Larson, 2003; Schäfer (2004].

3.1.0.6 F. Coordination. Sentences like (22) have been analyzed (e.g. in [Bergamann, 1982]) as involving scope ambiguity of coordination:

22. (Exactly) four teachers and authors smiled.

Under one interpretation, where *and* is often assumed to take scope below *four*, the sentence makes a claim about (exactly) four people, each of them a teacher and an author.[6] Under another interpretation, where (22) makes a claim about four teachers and four authors, *and* can be analyzed as taking scope over *four*.

Another kind of sentence that was analyzed in terms of scope ambiguity of coordination is the following:

23. John is looking for a maid or a cook.

The interpretation under which it is either a maid or a cook that John is looking for, but not necessarily both, was analyzed by Rooth and Partee [1982] as involving wide scope for the disjunction over the intensional verb *look for*. The other interpretation, where John would be both satisfied by finding a maid and by finding a cook, is considered as a case where *or* takes scope below *look for*. For more analyses of scopal effects with coordination see [Hendriks, 1993; Larson, 1985a; Schwarz, 1999; Winter, 2000b], among others.

After this review of some scope ambiguity phenomena in English, the remainder of his section formulates some generalizations that govern the distribution of direct and inverse scope readings that appear with QNPs. The facts in this domain can roughly be summarised as follows. In almost all cases (as far as our fragment in section 2.1 goes: in all cases) direct scope is an option. Whether a structure containing quantified noun phrases A and B allows inverse scope of B over A depends on two factors: the syntactic configuration relating A and B, and the choice of B. The following subsections elaborate on some circumstances that empirically affect the availability of inverse scope.

[6]We do not discuss here a possible interpretation, where the sentence refers to four people, some of whom are teachers and the rest are authors. An analysis of this interpretation involves the complicated question of collective readings of *and* conjunctions with nominals and predicates [Heycock and Zamparelli, 2005]. Yet another possible interrepretation of (22), which is however irrelevant for our purposes, is the one where the constituency of the subject is *[exactly four teachers] and [authors]*.

3.2 Restrictions on scope

As is well known, not every sentence containing two QNPs allows scope inversion. The availability of inverse scope depends partly on the syntactic configuration that relates the two QNPs. Consider again the minimal pair (2) (=(14)) and (7), which are restated below.

24. some inhabitant of every midwestern city participated

25. someone who inhabits every midwestern city participated

As we have seen, the QNP *every midwestern city* in (24) can take scope over *some inhabitant*, which yields a pragmatically acceptable inverse scope reading. By contrast, sentence (25) allows only an unacceptable direct scope reading, as predicted by the direct scope strategy of the grammar in Section 2.

At this point we would not like to prejudge the issue whether the explanation of the inverse scope reading of (24) and its absence in (25) is to be found in syntax or semantics; we discuss this issue in some detail in Section 4. For convenience, however, we describe a generalization that roughly governs these facts in syntactic terms. A hypothesis that has often been pursued since the late 1960s is that those syntactic domains that QNPs cannot scope out of are exactly the ones that are opaque to syntactic "movement" (e.g. formation of *wh* questions). To understand this hypothesis, consider first the examples in (26):

26. (a) which city$_i$ did you meet inhabitants of t_i ?
    (b) * which city$_i$ did you meet people who inhabit t_i ?
    (c) did you meet inhabitants of this city ?
    (d) did you meet people who inhabit this city ?

The phrase *which city* in (26a) performs the same function as argument of *inhabitants of* that is performed by *this city* in (26c). In this sense, *which city* in (26a) is related to the position following *of*; we indicate the relation by marking that position with a symbol t (for "trace") coindexed with *which city* (see section 4.2.1 below for further explanation of the syntactic mechanisms involved). (26b) shows that the relation is disturbed when *which city* sits outside a relative clause, while the position t it is related to sits inside the relative clause (compare (26b) to (26d)). The restriction that (26b) illustrates is referred to as the *Complex NP Constraint* (CNPC); the NP containing the relative clause is said to function syntactically as an *island* for wh-extraction.

Returning now to the examples (24) and (25), we find a similar pattern: the quantified NP *every midwestern city* in (24), in the same position as t in (26a), can take sentential scope, as though, like *which city* in (26a), it occupied the sentence-initial position. But *every midwestern city* in (25), which is inside a relative clause, cannot take scope over the sentence as a whole, similarly to the unacceptability of (26b). In this sense, the NP with the relative clause in (25) functions as a scope island, and (25) suggests that the CNPC is not only an extraction island as in (26b), but also an island that holds of QNP scope. In syntactic terminology, we say that scope islands and extraction islands coincide.

As we shall see in section 4.2.1, this generalization about the similarity between scope islands and islands for extraction led to the hypothesis that inverse scope results from (covert) movement of the QNP. Below we provide some more examples of islands, not illustrated by the fragment.[7]

27. (a) * which man$_i$ will you inherit a fortune if t_i dies
    (b) you will inherit a fortune if every man dies

28. (a) * what$_i$ did John hiss that Smith liked t_i
    (b) John hissed that Smith liked every painting

Both the *if*-clause in (27) and the complement clause to a verb *hiss* in (28) are islands for *wh*-extraction, and disallow matrix scope.[8] See Section 4.2.1 for further discussion of the similarity of scope taking and wh-movement, and for further examples of islands.

3.3 Unexpected wide scope: simple indefinites

Given the constraints on inverse scope illustrated in (24)–(25), the absence of a similar contrast in (29)–(30) is unexpected.

29. every inhabitant of a/some midwestern city participated

30. everyone who inhabits a/some midwestern city participated

31. (a) $\exists x[\textsc{city}(x) \wedge \textsc{midwestern}(x) \wedge \forall y[[\textsc{person}(y) \wedge \textsc{inhabit}(x)(y)] \rightarrow \textsc{participated}(y)]]$

[7] Example (28) is from [May, 1977, p. 94,120].

[8] In many syntactic frameworks, constructions like the *if* clauses as in (27) are classified as sentential *adjuncts*, and accordingly illicit sentences like (27a) are classified as violating an *adjunct constraint*. Verbs like *hiss (that)* in (28) that prevent grammaticality in cases like (28a) are often referred to as *non-bridge* verbs (unlike other verbs like *think* and *say*); hence one might classify (28) as exemplifying a non-bridge verb island for extraction and scope.

(b) $\forall y[[\textsc{person}(y) \wedge \exists x\ [\textsc{city}(x) \wedge\ \textsc{midwestern}(x) \wedge$ $\textsc{inhabit}(x)(y)]] \rightarrow \textsc{participated}(y)]$

Many English speakers agree that both (29) and (30) allow the inverse scope reading as well as the direct scope reading, as stated in (31a) and (31b) respectively. This is clearly the case with the determiner *some*, but also (marginally) with the article *a*. A classical example, outside our fragment, that more clearly demonstrates the same effect is the following.

32. If a friend of mine from Texas had died in the fire, I would have inherited a fortune. [Fodor and Sag, 1982]

Sentence (32) can be interpreted as true if I have a certain friend whose death would make me rich, even if I have other friends for whom this does not hold. Again it seems, as in (30), that the indefinite can take scope over an island, in this case the adjunct island contributed by the conditional.

The generalization seems to be that simple indefinite NPs can scope out of relative clause islands, and this pattern also persists with other scope islands types. The contrast between the ill-formedness of (26b) and the availability of the inverse scope reading (31b) for (30) is illuminating: it suggests that it would be problematic to derive (31b) for (30) by the same rule that "fronts" wh-elements as in (26a). Using such a rule for both (30) and (26a) (or the inverse scope reading of (24)) would make the contrast between (26b)/(25) and (24) completely mysterious. See Section 4.3.3 for alternative theories about the behavior of indefinite NPs as in (30).

In our fragment, the exceptional wide scope behavior seen in (30) is displayed by those NPs that are headed by *a* and *some*, as well as the numeral *three*. For instance:

33. John met everyone who admires three midwestern cities

Sentence (33) has a reading, fitting in a context in which John is researching the popularity of three particular cities. This suggests a wide scope option for *three cities*. This description of the facts is a bit simplistic, as we shall see in See Section 3.5. We also postpone to sections 3.4 and 3.5 a discussion of the class of NPs that support this exceptional wide scope behavior.

3.4 Absence of inverse scope

There are several types of QNP that show an unexpected absence of inverse scope, even in syntactic contexts where other QNPs do show inverse scope. One prominent and fairly uncontroversial example is the bare plural. Bare plural NPs do not take inverse scope, as the following example illustrates [Carlson, 1977]:[9]

34. no man met women

    (a) $\neg\exists x[\text{MAN}(x) \wedge \exists 2y[\text{WOMAN}(y) \wedge \text{ MEET}(y)(x)]]$
    (b) $\exists 2y[\text{WOMAN}(y) \wedge \neg\exists x[\text{MAN}(x) \wedge \text{ MEET}(y)(x)]]$

Sentence (34) only has the reading in (34a), not the reading in (34b). The same holds for (35) and (36):

35. John met every man who inhabits midwestern cities

36. John met every inhabitant of midwestern cities

Intuitions are particularly clear for these examples, which are pragmatically infelicitous. Inverse scope readings would be a pragmatically acceptable, but the sentences are not, which means that these sentences do not allow an inverse scope analysis.

Unfortunately, for some other classes of QNPs the relevant semantic intuitions are not as clear-cut, and their scope properties have not been studied as extensively in the available literature as e.g. the scope properties of QNPs of the *every N* variety. Nonetheless, some tentative generalizations have been proposed in the literature that deserve to be mentioned.

Consider first NPs with modified numeral determiners. Given the inverse scope option for (13), repeated as (37), a similar option for (38) is expected. It is however claimed [Liu, 1990; Beghelli, 1993; 1995] that this option is not available.

37. some woman admires every man

38. some woman inhabits exactly three cities

    (a) $\exists x[\text{WOMAN}(x) \wedge \exists!3y\ [\text{CITY}(y) \wedge \text{ INHABIT}(y)(x)]]$
    (b) $\exists 3!y[\text{CITY}(y) \wedge \exists x[\text{WOMAN}(x) \wedge \text{INHABIT}(y)(x)]]$

According to these authors, a sentence such as (38) is intuitively judged to mean only (38a): it is true if (38a) is true, and false if (38a) is false. Specifically, (38) is false if exactly three cities are each inhabited by a different woman, as allowed by (38b). This means that for the case of (38), an inverse scope analysis would be incorrect (but see [Reinhart, 2006b] for a different view on these data). Absence of inverse scope can also be observed in (39) below.

[9]For ease of exposition, we translate the bare plural with the quantifier $\exists 2$, while noting that this is an extreme simplification.

39. every man admires less than three women

    (a) there are less than three women that every man admires

Sentence (39) does not allow the inverse wide scope reading for *less than three women*, as expressed by (39a).

In the second syntactic configuration our fragment contains, a PP-modified NP, it has been claimed that we find a similar effect [Beghelli, 1993]:

40. an inhabitant of exactly three cities participated

Although intuitions are less secure here, (40) shows a clear preference for the (pragmatically implausible) direct scope interpretation over the (more plausible) inverse scope analysis.

The exceptional wide scope found with simple indefinites in (30) and (33) is also not found with modified-numeral NPs:

41. John met every man who admires exactly three midwestern cities

This example is judged false if there is any man admiring any choice of three midwestern cities whom John did not meet.

Sentences (38)–(41) show NPs with a compound determiner that do not take inverse scope. We find the same behaviour with other monotone decreasing QNPs, even non-modified ones, as shown in (42):

42. some man admires few woman

A third category of NPs that, in some respects, belongs to the class discussed here are simple numeral NPs. We saw in the previous section (see (33)) that there are indications that these NPs can take exceptional wide scope. The example in (43), on the other hand, suggests that in other cases they must take direct scope:

43. some man admires three women

    (a) $\exists x[\textsc{man}(x) \wedge \exists 3y[\textsc{woman}(y) \wedge \textsc{admire}(y)(x)]]$

    (b) $\exists 3y[\textsc{woman}(y) \wedge \exists x[\textsc{man}(x) \wedge \textsc{admire}(y)(x)]]$

This sentence is judged to entail the proposition that there is at least one man who admires three women; the inverse scope reading is judged to be much harder to obtain than in (13). The seemingly contradictory behaviour of the *three N* class is the subject of the next section.

3.5 Mixed scope

Below we repeat examples (33) and (43), and add examples (46) and (47):

44. John met everyone who admires three midwestern cities
    (a) $\forall x[[\text{PERSON}(x) \wedge \exists 3y[\text{MIDWESTERN}(y) \wedge \text{CITY}(y) \wedge \text{ADMIRE}(y)(x)]] \rightarrow \text{MEET}(\text{JOHN}, x)\ ]$
    (b) $\exists 3y[\text{MIDWESTERN}(y) \wedge \text{CITY}(y) \wedge \forall x[[\text{PERSON}(x) \wedge \text{ADMIRE}(y)(x)] \rightarrow \text{MEET}(\text{JOHN}, x)\ ]]$
45. some man admires three women
    (a) $\exists x[\text{MAN}(\mathrm{x}) \wedge \exists 3y[\text{WOMAN}(y) \wedge \text{ADMIRE}(y)(x)]]$
    (b) $\exists 3y[\text{WOMAN}(y) \wedge \exists x[\text{MAN}(x) \wedge \text{ADMIRE}(y)(x)]]$
46. no man admires three midwestern cities
47. (a) John met someone who inhabits three midwestern cities
    (b) John met some inhabitant of three midwestern cities
    (c) $\exists x[[\text{PERSON}(x) \wedge \exists 3y[\text{MIDWESTERN}(y) \wedge \text{CITY}(y) \wedge \text{INHABIT}(y)(x)]] \wedge \text{MEET}(\text{JOHN}, x)\ ]$
    (d) $\exists 3y[\text{MIDWESTERN}(y) \wedge \text{CITY}(y) \wedge \exists x[[\text{PERSON}(x) \wedge \text{INHABIT}(y)(x)]] \wedge \text{MEET}(\text{JOHN}, x)\ ]$

Unmodified numeral plurals display a seemingly contradictory scope behaviour. On the one hand, their scope is not limited to their surface position: from object position, they can "escape" the scope of the subject in (46), and even from relative-clause embedded position, they seem to escape the scope of the containing quantifier in (44). Thus, (44) is not felt to entail that John met everyone who admired any choice of three midwestern cities; rather, this sentence can apparently be about three particular midwestern cities. Likewise, sentence (46) allows more than just the direct scope reading; this is clear from the fact that (46) is true even if some men do admire some choice of three midwestern cities. In view of these facts, these bare numeral QNPs appear to behave much like simple indefinites (aN), which allow both "normal" inverse scope and exceptional wide scope (see section 3.3, as well as section 4.3.3 below).

On the other hand, a description of these non-narrow scope readings for the plural indefinites in (44) and (46) as "wide scope readings" would be too simplistic, as already mentioned in Section 3.4. Bare numeral QNPs do not simply take inverse wide scope out of islands: (44) does not allow the reading (44b), and (47a) does not allow the wide scope reading (47d). Furthermore, the inability of such QNPs to take inverse wide scope extends to non-island contexts: (45) and (47b) show that these QNPs behave much like the exceptional narrow scope modified numeral QNPs of section 3.4: the inverse scope readings given in (45b) and (47d) are highly marked [Ioup, 1975; Liu, 1990; Beghelli 1993].

3.6 Summary of QNP scope problems

In Section 2 we showed readings of sentences which suggest that the simple "direct scope" interpretative strategy of our toy grammar does not cover the interpretations of sentences in the fragment. We saw that in many cases this incompleteness can be a result of an "inverse scope" strategy for interpreting syntactic structures. In this section we saw some central challenges for the inverse scope strategy. First, inverse scope readings are often constrained by syntactic restrictions that seem parallel (at least partially) to the restrictions on overt extraction. However, simple indefinite NPs seem exceptionally free, and can often display an inverse scope behavior that does not seem to obey these syntactic restrictions. Conversely, some other NPs seem exceptionally restricted, and hardly show any inverse scope phenomena. Further, some plural indefinite NPs seem both exceptionally free and exceptionally restricted in their inverse scope potential, in a way that may lead to "mixed" scope behavior. The next section is an overview of some theories that attempt to account for (parts) of this complex array of linguistic scope phenomena.

4 LOGICAL AND LINGUISTIC THEORIES OF QUANTIFIER SCOPE

4.1 Preliminaries on quantifier scope

A common approach to quantification in natural language, which was clearly manifested in [Montague, 1973], (henceforth PTQ) and substantiated and popularized in [Barwise and Cooper, 1981; Keenan and Stavi, 1986], is to consider all QNPs as denoting *generalized quantifiers.* This assumption means that all QNPs denote sets of sets of entities, or, isomorphically, predicates over predicates over entities. For instance, in this approach a QNP like *every man* denotes the predicate that holds of the predicates that hold of all individual men in the model. In grammars with an extensional semantics such as the one of Section 2, using typed lambda terms, such a generalized quantifier receives the type $\langle\langle e,t\rangle,t\rangle$ and is represented as follows:

48. $\lambda B.\forall x[\textsc{man}(x) \rightarrow B(x)]$

Compositionally, this treatment entails that the determiner *every* receives the denotation of a function from predicates over entities to generalized quantifiers. This means that the type of such determiners, as in the grammar of Section 2, is $\langle\langle e,t\rangle,\langle\langle e,t\rangle,t\rangle\rangle$. The standard lambda term assumed for representing the extensional meaning of the determiner *every* is also as in the grammar of Section 2:

49. $\lambda A\lambda B.\forall x[A(x) \rightarrow B(x)]$

Most semantic frameworks assume that transitive predicates like *admire* denote two-place relations, of type $\langle e,\langle e,t\rangle\rangle$. This is also the typing strategy assumed in the grammar of Section 2. In order to derive a meaning for sentences like *some/every woman admires every man*, the semantic mechanism should be able to compose the binary relation for *admire* with the two generalized quantifiers for the subject and the object. Many works assume, as we did in translation rule 4 of the grammar in Section 2, that the way to reach an interpretation for such transitive sentences involves lambda abstraction over variables that take the argument positions in the predicate. The two linear orders of composition with the binary predicate (object first or subject first) lead to the following two propositions, with Q1 and Q2 as the lambda terms for the subject and object quantifiers, respectively, and R the binary predicate.

50. (a) $Q1(\lambda x.Q2(\lambda y.R(x,y)))$

(b) $Q2(\lambda y.Q1(\lambda x.R(x,y)))$

For our expository purposes here it is important to note that representations like (50) involve what we call a “standard” approach to quantifier scope. Two basic principles underlie this approach: (i) QNPs denote generalized quantifiers; (ii) the denotations of QNPs and relational predicates are amalgamated using “linear” composition as in (50a) and (50b), equivalent to a formula where one of the quantifiers takes scope over the other. As we will see in the sequel, these assumptions are not necessarily sufficient for describing all scope phenomena. However, as a baseline approach these “standard scope” principles allow one to capture many basic facts about inverse scope phenomena. We now move on to some theories that adopt and substantiate these principles.

4.2 "Standard scope" mechanisms

Familiar mechanisms that adopt the standard approach to scope can roughly be divided into two categories: syntactic and semantic ones. We call an approach *syntactic* if it requires a modification of the rules of syntax in order to derive inverse scope. In case of scope ambiguity, syntactic approaches normally postulate multiple distinct syntactic representations (structures, derivations) underlying the same string, with a different meaning assigned to each of them. We call an approach *semantic* if it keeps to the most straightforward syntactic account of the constituent structure of the language and only postulates a modification of the semantics so as to derive inverse scope readings. In case of ambiguity, a semantic approach postulates a single syntactic representation, to which rules of semantic interpretation can apply in different ways.

It should be remarked, however, that the distinction between *syntactic* and *semantic* approaches is a rather crude one. In many cases a syntactic scope mechanism has semantic repercussions and vice versa. In the following sections, we shall outline a selection of syntactic and semantic scope mechanisms from the literature. We will first introduce two well-known scope mechanisms: the methods of *Quantifier Raising* and *Quantifying-in*. The first approach is mostly followed in works in generative linguistics following May [1977], whereas Quantifying-in was introduced in PTQ and followed by much work in the tradition of Montague Grammar. Then we move on to a brief overview of the semantic scope mechanism of Cooper [1975; 1983], known as *Cooper Storage*, and the flexible type mechanism of Hendriks [1993]. Lastly, we will outline two categorial approaches to quantifier scope, which show a tight interaction between syntax and semantics.

4.2.1 Quantifier Raising

The "Quantifier Raising" (QR) theory of quantifier scope ambiguity was first proposed by Chomsky [1976] and May [1977] as a revision of the dominant generative theory of the time, known as the Extended Standard Theory [Chomsky, 1972]. The QR theory persisted into the subsequent Principles & Parameters framework (the basis of the well-known Government Binding theory following [Chomsky, 1981], see [Chomsky and Lasnik, 1993], and it still plays a role in the currently dominant Minimalist Program of Chomsky [1993; 1995].

In these syntactic models, an expression is associated with multiple phrase structure representations, which are related by rules of movement (and other transformations). This is illustrated in (51):

51. I wonder who John thinks Peter likes

(a) I wonder $[_{CP}$ John thinks Peter $[_{VP}$ likes who $]]$

(b) I wonder $[_{CP}$ who$_i$ John thinks Peter $[_{VP}$ likes t_i $]]$

(51a) is the (simplified) Deep Structure (or: *D-Structure*) representation of (51). This representation is generated by a set of Base Rules (for instance, rewrite rules as used for our fragment in section 2.1 above). The constituent structure at this level of representation captures the fact that *who* in (51) functions as the object of *likes*. (51b), the Surface Structure (or: *S-Structure*) representation, is derived from (51a) via a *movement rule*, which displaces the wh-element *who* to the front of the embedded question. The movement operation leaves a "trace" denoted t, which functions like a phonetically empty constituent coindexed with the moved constituent. The presence of indexed traces makes it possible to recover the D-Structure role of moved elements from the S-Structure representation. In the S-Structure (51b), the coindexing between the wh-element *who* and the trace keeps track of the fact that *who* is related to the object position of the verb *likes*. The S-Structure in turn is further input to rules of the phonological component of the grammar, which yields the phonetic form of the sentence.

Chomsky [1976] proposed that the S-Structure representation is input to a further set of rules, QR among them, which derive the *Logical Form* of the sentence. Representations at the level of Logical Form are interpreted by a semantic mechanism.[10] A simple configuration of this setting is one in which each LF is mapped to one, and only one, semantic analysis of the sentence.[11] Since the rules deriving LF do not feed into the phonological component (hence do not affect phonetic form), they are known as "covert"

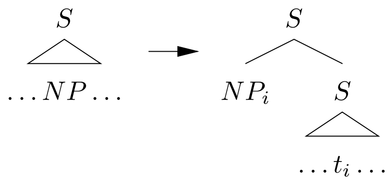
[10]The term *Logical Form* has of course been chosen to suggest similarity with the logician's notion of the logical form of a proposition which underlies its inference properties, as distinguished from the grammatical form. However, it has repeatedly been stressed, in particular by Chomsky (see e.g. [Chomsky, 1980]), that the representation of a sentence at the grammatical level of LF is not to be equated with its "logical form." The contribution of LF is in structural aspects of meaning determined by syntax, potentially leaving other aspects unspecified. Representations at the level of LF are (almost always) taken as phrase structure representations, derived by syntactic rules and subject to syntactic well-formedness conditions, for which independent evidence is sought in other grammatical phenomena (such as conditions on *wh*-phrases left in situ at S-Structure, and on the coreference behavior of pronouns and anaphors), not necessarily semantic ones.

[11]It has been proposed (esp. by [May, 1985; Aoun and Li, 1989]) that the level of LF is not disambiguated: these authors combine a syntactic account of scope ambiguities (QR deriving LF) with a non-syntactic approach to deriving various interpretations from a given LF-representation.

operations, as opposed to "overt" movement operations such as the *wh*-movement illustrated in (51).

On this approach, scope ambiguities arise through optionality in the application of a movement rule called *Quantifier Raising* (QR). This rule derives from one given S-Structure several different LFs with different scope relations between elements in the sentence. On its earliest formulation [May 1977], this movement rule operates as shown in (52):

52.



This version of the QR rule takes the S-structure representation of a sentence S containing a quantified noun phrase NP, moves NP, and attaches it to the node S by "splitting" S into two nodes and attaching the NP under the highest of these.[12] Since the rule is a standard movement rule, by convention it leaves a trace t coindexed with the moved NP, as shown in (52).[13]

Adding such a rule to the grammar of section 2.1 will yield at least the following LFs for examples (13) and (24), respectively:

53. (a) $[_S[_{NP}$ some woman$]_1[_S[_{NP}$ every man$]_2[_S t_1[_{VP}$ admires $t_2]]]]$

(b) $[_S[_{NP}$ every man$]_2[_S[_{NP}$ some woman$]_1[_S t_1[_{VP}$ admires $t_2]]]]$

54. $[_S[_{NP}$ every city$]_2[_S[_{NP}$ some inhabitant of $t_2]_1[_S t_1[_{VP}$ participated$]]]]$

The interpretive rules in the grammar should now apply to these LF representations. In order to modify our toy grammar so that structures derived by QR are properly interpreted, we add the following translation rule:

[12] The manner in which the moved NP in (52) is attached is known as "(Chomsky-)Adjunction"; hence May's formulation of the rule: "Chomsky-adjoin a QNP to S" [May, 1977, p. 18]. For a formal definition of Chomsky-adjunction, see [Lasnik and Kupin, 1977].

[13] The extensive literature on the topic contains a variety of other definitions, differing e.g. in what NP types are subject to QR [May, 1985; Reinhart, 1991; Ruys, 1992], what nodes besides S it may target (Williams 1977, May 1985), whether it may also move material downward [May, 1977; 1985; Fox, 1995], under what conditions it may or must apply [May, 1977; Fox, 1995; 2000; Reinhart, 2006a], whether it should be equated with other supposed covert operations [Hornstein, 1995], whether or not it involves adjunction [Hornstein, 1995; Beghelli and Stowell, 1997; Bruening, 2001] and even in whether or not it feeds into phonetic form [Kiss, 1991; Fox and Nissenbaum, 1999, and references cited there]. See [Kiss, 2006] for a recent overview.

5. For all $\gamma \in$ SD s.t. $\gamma = [_S\beta_i\ \alpha\ ]$, β is an NP and α is an S, for all $i \in N$: if $\alpha \Rightarrow \alpha'$ and $\beta \Rightarrow \beta'$, then$\gamma\ \Rightarrow \beta'(\lambda x_i.\alpha')$.

This rule relies on the afore-mentioned assumption that when a movement rule adjoins a QNP to a sentence node S, that sentence contains a trace that is coindexed with the QNP. Furthermore, Translation Rule 5 also relies on the assumption that a trace with an index i is translated using a free variable x_i.[14] This assumption about the translation of traces is satisfied by the following scheme for traces in the lexicon, which is added on top of the lexicon of section 2.1.

Cat	**Word**	**Translation**	**Type**
NP	t_i	$\Rightarrow \lambda P.P(x_i)$	$\langle\langle e,t\rangle,t\rangle$

This scheme follows Montague [1973] in that traces, like proper names, are translated into generalized quantifier terms, with the variable filling the role of the constant JOHN_e in terms like $\lambda \text{A}.\text{A}(\text{JOHN}_e)$, which appear in the lexicon of section 2.

The LFs in (53) and (54) are now interpreted similarly to the general representations we gave in (50) to simple transitive sentences. For instance, the verb phrase [*admires* t_2] in (53b) now receives the following translation using translation rule 4 in the grammar of section 2.1.

$$\lambda x.(\lambda P.P(x_2))(\lambda y.\text{ADMIRE}(y)(x)),$$

which is equivalent to:

$$\lambda x.\text{ADMIRE}(x_2)(x).$$

Using translation rule 5 above we get the following translation for the LF in (53b):

$$(\lambda A.\forall y[\text{MAN}(y) \rightarrow A(y)])(\lambda x_2.(\lambda B.\exists z[\text{WOMAN}(z) \wedge B(z)])(\lambda x.\text{ADMIRE}(x_2)(x))),$$

which is equivalent to the inverse scope reading of sentence (13):

$$\forall y[\text{MAN}(y) \rightarrow \exists x[\text{WOMAN}(x) \wedge \text{ADMIRE}(y)(x)]]$$

In most treatments, such rules as translation rule 5 above are left implicit (but see e.g. [May, 1989]). Keenan and Faltz [1985] and Heim and Kratzer [1998] manage without the extra translation rules by effectively adding the lambda operator for variable binding as an extra node to the syntactic representation.

[14] It is assumed further that every QNP receives a “fresh” index, to prevent accidental coindexing of variables.

Evaluating the LF / QR approach

The QR theory, which holds that quantifier scope is mediated through a syntactic movement rule deriving a syntactic level of representation, has several types of empirical consequences. First and foremost, it leads one to expect that conditions on quantifier scope can be stated as conditions on rules of movement; this implication is our primary topic in this section.

We want to stress, however, that the QR/LF theory has further implications, some of which are only indirectly related to quantifier scope phenomena. It has been argued, for instance, that additional (movement) operations besides QR apply in deriving LF from S-Structure (e.g. a rule fronting *wh*-phrases left in situ at S-Structure). These operations are held responsible for other semantic effects besides relative scope, and even for the well-formedness of certain constructions and aspects of cross-linguistic variation. Since the QR theory presupposes the existence of LF as a grammatical level of representation, any independent evidence for such LF operations affects the status of the QR theory. In addition, since QR derives a syntactic level of representation, one expects that there might be further rules of syntax that take the output of QR as their input: either further derivation rules, or syntactic constraints that apply to LF representations derived by QR; the treatment of "Antecedent Contained Deletion" (ACD) phenomena discussed in Section 5.2 may be a case in point. A full evaluation of the QR approach to scope phenomena must take these various types of indirect evidence into account.

Primarily, however, evidence for QR exists to the extent that generalizations on quantifier scope can be stated in terms of syntactic properties of the relevant constructions, and to the extent that these generalizations apply to other purported movement operations as well. Ultimately, on the QR approach, a unified theory explaining properties of both overt and covert movement should be possible.

As far as our fragment goes, there are two factors affecting quantifier scope, and the occurrence of inverse scope: choice of QNP, and syntactic context. Some of the effects of the choice of the QNP on scope are mentioned in section 3.4. However, by far the most widely discussed prediction associated with QR theory is that limitations on scope and limitations on (overt) movement which arise from the syntactic context should coincide. As mentioned in section 3.2 above, many of the islands for extraction that were discovered in Ross [1967] have also been identified as scope islands. We repeat some earlier examples and add some further types:

55. (a) $*$ which city$_i$ did you meet $[_{NP}$ people who inhabit t_i $]$

(b) [$_{NP}$ someone who inhabits every midwestern city] participated

56. (a) $*$ which man$_i$ will you inherit a fortune [$_{CP}$ if t_i dies]

(b) you will inherit a fortune [$_{CP}$ if every man dies]

57. (a) $*$ which student does Prof Jones [$_{VP}$ despise t_i] and [$_{VP}$admire the dean]

(b) some professor [$_{VP}$ despises every student] and [$_{VP}$ admires the dean]

58. (a) $*$ who$_i$ did you see [John's picture of t_i]

(b) I saw [John's picture of everyone]

(55) illustrates the effect of the Complex NP Constraint (CNPC), introduced in section 3.2: an NP with a relative clause does not allow overt *wh*-extraction of *which city* in (55a); it also does not allow wide scope for *every Midwestern city* in (55b). (56) shows the effect of an *Adjunct Island*: the *if*-clause serves as an island for both extraction (cf. the ungrammaticality of (56b)) and scope, witness the fact that sentence (56) does not have a reading where the noun phrase *every man* takes scope over the conditional (see more on this observation in section 4.3.3). (57) shows the effect of the *Coordinate Structure Constraint* (CSC): an element may not be extracted out of one conjunct in a coordination structure in (57a), and a quantifier does not scope out of such a construction in (57b) (*every student* does not scope over *some professor*). (58) illustrates the *Specificity Constraint*: a definite NP, especially one with an overt subject (*John's*) is an island for both extraction (58a) and scope: (58b) has only a narrow scope reading for *everyone* (it's a single group photograph).[15]

The parallelism between the a.-examples and the b.-examples in (55)–(58) provides strong prima facie evidence that the rule responsible for quantifier scope ambiguities is indeed a movement rule, of the same type, hence largely subject to the same conditions, as the rule responsible for overt (wh-)movement. This provides the primary motivation for the QR theory.[16]

---

[15] For more recent discussion of the CSC effect on quantifier scope, see [Ruys, 1992; Fox, 1995 (who discusses certain classes of exceptions)]. For the Specificity Constraint, see [Chomsky, 1973; Fiengo and Higginbotham, 1981; Davies and Dubinsky, 2003].

[16] In fact, the observation that islands for overt movement coincide with scope islands, and the account of quantifier scope in terms of a quantifier movement rule, precedes the QR theory. Lakoff [1970], working in the framework of Generative Semantics, presents many of the basic observations and proposes a similar account (see also [Postal, 1974]): in the deep structure underlying a sentence with a quantified NP, the quantificational

Prima facie evidence, of course, need not be conclusive. An alternative for describing scope islands in terms of restrictions on overt movement might be that quantifier scope is *clause-bounded*; this would trivially prevent scoping out of clauses that happen to be syntactic islands. For instance, CNPC islands consist of an NP containing a relative clause; if scoping out of the clause is blocked, scoping out of the island as a whole is blocked as well. This was the argument raised by Chomsky [1975] against Rodman's [1976] syntactic approach to quantifier scope based on Montague's Quantifying-in operation (see section 4.2.2 below). Rodman demonstrated that quantifier scope is sensitive to CNPC islands; Chomsky countered that quantifier scope simply cannot escape finite clauses, as shown by the non-ambiguity of (59):

59. John said that everyone had left [Chomsky, 1975]: ($\alpha 3 \alpha$)

One might thus argue that the similarity of syntactic islands to scope islands is an illusion, and that the former happen to be a subset of the latter — although it remains to be seen, of course, whether this would support any other approach to quantifier scope. That would depend on whether clause-boundedness could be implemented insightfully in, say, a semantic theory of quantifier scope.

One empirical answer to Chomsky's challenge can be based on the observation that clause-boundedness may both be too restrictive and too permissive as an account of quantifier scope. Sometimes, quantifier scope is more restricted than the minimal clause, namely when a QNP is embedded in an island smaller than the minimal clause: this is illustrated in (57b) and (58b) above. At the same time, a quantifier may sometimes scope out of the minimal clause, especially when the minimal clause in non-finite (see e.g. (60), from Hornstein 1995), and for some speakers also when the clause is finite, as in (59) (see [May, 1977, p. 217]; see also [Reinhart, 1997].

60. someone expected [$_S$ every Republican to win]

While Chomsky's alternative description may thus be rejected as oversimplified, the contrast between (59), where a QNP cannot scope out of a finite clause for most speakers, and (51) above, where a wh-phrase does move out of such a clause, is indicative of a more fundamental challenge to the QR theory: there is no perfect parallelism between scope and overt movement. In itself, this observation does not falsify the movement approach to scope phenomena: given that QR and *wh*-movement differ in the type of object

determiner (e.g., *many*) occupies the scope position, from which it is lowered to its surface position by a lowering rule which is sensitive to the same island conditions that block overt wh-movement.

being moved, the landing site for the movement, and the (c)overtness of the movement, it is possible that conditions on syntactic movement, properly formulated, predict some observational divergence of wh-movement and scope. In the case of Chomsky's example (59), May [1977] showed that his own account correctly predicts the distinction between (51) and (59) (see below). In general, whether the syntactic approach to quantifier scope is correct is not decided on the basis of superficial (dis-)similarities of *wh*-movement and quantifier scope. What matters for a critical evaluation of QR theory is whether we can construct a successful theory of movement which provides an insightful account of both *wh*-movement and QR.

In order to provide a concrete illustration of these points, we conclude this section by briefly returning to the fragment of section 2, and illustrating some of the problems that have arisen in providing a syntactic, QR account of the scope ambiguities it contains. Our purpose here is emphatically not to provide an overview, either historic or systematic, of constraints on movement developed in the generative framework and their applicability to QR; it is merely to provide some sense of the type of problems and solutions that arise.

May [1977] proposed that both (overt) *wh*-movement and QR obey Chomsky's [1973] *Subjacency* condition. This syntactic condition states that no movement may cross two bounding nodes, where S and NP are considered bounding nodes. This correctly predicts that quantifier scope obeys the CNPC. Given the subjacency restriction on QR, sentence (25) above does not allow the LF below:

61. $[_S$ every city$_i$ $[_S$ $[_{NP}$ someone $[_S$ who inhabits t$_i$]] participated]]

In (61) there are three bounding nodes (two Ss and one NP) that separate the trace t$_2$ and the landing site for *every city.* Likewise, the CNPC effect with overt wh-movement in (55a), repeated below, is explained by Subjacency: three bounding nodes separate *which city* from its trace.

62. * which city$_i$ $[_S$ did you meet $[_{NP}$ people who $[_S$ inhabit t$_i$]]]

Crucially, May [1977] argues that the Subjacency condition also explains the observation that QNPs cannot, but wh-phrases can, escape finite clauses. The LF for (59) that would give wide scope to *everyone*, given in (63), violates Subjacency:

63. $[_S$ everyone$_i$ $[_S$ John said that $[_S$ t$_i$ had left]]]

(51), however, does not violate Subjacency, since a *wh*-phrase moving out of a finite clause, unlike a quantifier NP, can make an intermediate landing at the left edge of the finite clause, as indicated in (64) by the trace $t{*}_i$ in this position:

64. I wonder $[_{S'}$ who$_i[_S$ John thinks $[_{S'} t *_i$ $[_S$ Peter $[_{VP}$ likes $t_i]]]]]$

While the empirical implications of this analysis are satisfactory, two problems arise. First, the supposed LF for (13) which yields the available inverse scope reading, given below, also appears to violate Subjacency.

65. $[_{S2}[_{NP}$ every man$]_i[_{S1}[_{NP}$ some woman$]_j[_{S0} t_j [_{VP}$ admires t_i $]]]]$

Assuming that the noun phrase *some woman* also undergoes QR, the noun phrase *every man* has to cross two S-nodes in order to take scope above it. This led May to a revision of the Subjacency condition for which no independent evidence from overt movement was available at the time (although the problem was resolved in the framework of [Chomsky, 1986a]): multiple S-nodes in a relation of immediate domination (S0 and S1 in (65)) count as one bounding node.

The problem illustrated in (65) was perhaps purely technical. However, example (66) below drives a more substantive wedge between QR and overt movement, and the problem it illustrates remains to this day:

66. $*$ who $[_S$ did $[_{NP}$ pictures of t $]$ please you$]$

This ill-formed example shows that *wh*-movement out of a subject-NP is disallowed, an effect that has also been attributed to Subjacency [Chomsky, 1977, p. 112]: the element *who* in (66) crosses both the NP and S nodes. On this count, we would expect that our third syntactic context disallows extraction as well: the LF that yields the inverse scope reading for (24) should also violate Subjacency according to its definition above. This LF is given below.

67. $[_S[_{NP}$ every city$]_i[_S[_{NP}$ some inhabitant of $t_i]_j[_S t_j[_{VP}$ participated$]]]]$

May's [1977, p. 214] solution to this puzzle added a clause to the Subjacency condition: NP does not count as a bounding node in case the relevant movement is QR. Clearly, this does not resolve the conflict, but rather codifies the divergence of wh-movement and QR seen in (66)—(67); a proliferation of such divergences would render the "unification" of QR with an overt movement rule vacuous.

One solution to this puzzle was offered by May [1985], who suggested that QR does not extract a QNP from another QNP in inverse linking constructions, but rather adjoins the embedded QNP to the containing one:

68. $[_S [_{NP} [_{NP}$ every city$]_i [_{NP}$ some inhabitant of $t_i]]_j [_S t_j [_{VP}$ participated$]]]$

In (68), Subjacency is not violated. For discussion of the semantic interpretation of structures like (68), see May (1989) and Larson (1985b). We will not trace the history of the treatment of these problems any further; the reader is referred to May and Bale (2006) for a recent overview.

A considerable amount of further work has been done on the QR theory of quantifier scope ambiguity in various stages of the generative framework; space does not allow us to discuss, or even outline, this body of literature. We refer to Reinhart (1997) for an overview of many of the issues involved in determining the conditions on covert movement, and arguments that, on balance, the Subjacency condition remains the preferred account of QR-determined quantifier scope phenomena. We conclude by observing that, at the time of writing, although progress has been made in several areas, a complete theory of movement restrictions as they apply to scope does not appear within reach.

4.2.2 Quantifying-in

Montague's [1973] PTQ introduced a grammar for a fragment of English that, among other phenomena, treats quantifier scope ambiguities. The syntactic formalism that is assumed in PTQ is somewhat non-standard, and it is therefore hard to illustrate its treatment of scope ambiguities using the phrase structure grammar of section 2. We will here illustrate only the general idea behind Montague's operation of *Quantifying-in* — a syntactic treatment of NPs that generates quantifier scope ambiguities in the PTQ fragment. For full details see PTQ itself, or the more friendly introductions in [Dowty *et al.*, 1981, ch. 7] and [Gamut, 1991, ch. 6].

A syntactic rule in PTQ takes a sequence of expressions $\alpha_1, \alpha_2, \ldots, \alpha_n$ and their syntactic categories $C_1, C_2, \ldots, C_n$, and generates an expression α and a category C. Crucially, syntactic rules in PTQ can generate α using non-concatenative operations on the expressions $\alpha_1, \alpha)2, \ldots, \alpha_n$. This is unlike ordinary phrase structure rules, which only concatenate the input expressions. To see how this allows PTQ to capture quantifier scope ambiguities, consider our example (13) from section 2.1, restated below:

69. some woman admires every man

To generate the object wide scope reading of this sentence, PTQ generates the following two expressions, with the respective categories:

α_1,	category t:	*some woman admires him$_n$*
α_2,	category T:	*every man*

The expression α_1 is of category t — PTQ's category for sentences — and it contains a pronoun *him*, derived with an arbitrary index n. The expression α_2 is of category T — PTQ's category for noun phrases. Both α_1 and α_2 are generated using rules that are mostly similar to standard phrase structure rules in using concatenation of lexical expressions. However, PTQ's quantification rule (S14) uses a syntactic operation of *substitution* to replace the pronoun in α_1 by the noun phrase α_2. This derives the output expression $\alpha =$ *some woman admires every man* in (69). The rule determines that this expression, like the input expression α_1, is of category t (=sentence).

On the semantic side, PTQ assigns each derived expression in the grammar a translation in Montague's *intensional logic* (IL — an intensional variant of the typed lambda calculus, see [Dowty *et al.*, 1981; Gamut, 1991]). Each syntactic rule has a corresponding translation rule responsible for deriving the translation of the output expression. In the case of the syntactic quantification rule, the corresponding translation rule (rule T14 in PTQ) is responsible for the wide scope interpretation of the object in (69). The sentential expression α_1, with the pronoun *him*$_n$, is derived with the following translation β_1, containing a free variable x_n. The translation β_1 is given here after some simplifications, and ignoring the intensional aspects of PTQ translations:

$$\beta_1 = \exists x[\textsc{woman}(x) \wedge \ \text{admire}(x_{)}(x)]$$

The noun phrase expression α_2 is derived with the following generalized quantifier translation β_2 (again, with some simplifications):

$$\beta_2 = [\lambda B.\forall y[\text{man}(y) \rightarrow B(y)]$$

The translation rule T14 for quantification lets the quantifier bind the free variable x_n in β_1 using lambda abstraction over this variable:

$$\beta = \beta_2(\lambda x_n.\beta_1)$$

The resulting translation β of the sentential output expression α (= *some woman admires every man*) is equivalent to the object wide scope reading of the sentence:

$$\forall y[\textsc{man}(y) \rightarrow \exists x[\text{woman}(x) \wedge \textsc{admire}(y)(x)]]$$

Montague's method of Quantifying-in does not on its own account for island constraints on quantifier scope. Considering the complex NP constraint (CNPC), Rodman [1976] addresses this problem for an extension of PTQ that also treats ordinary relative clauses with *who* and *that* relative pronouns, in addition to the rather artificial *such that* construction of PTQ. To see the problem for the PTQ grammar, reconsider sentence (25) from section 2.3.1, restated below as (70), or its equivalent in the PTQ grammar that is given in (71).

70. someone who inhabits every midwestern city participated

71. someone such that he inhabits every midwestern city participated

Without further assumptions, PTQ and its straightforward extension in Rodman [1976] allow such sentences to be interpreted with a sentential scope for the noun phrase *every midwestern city.* This is because rules S14 and T14 of quantification allow such NPs to compose with the sentential expression in (72), which contains the free pronoun it_n.

72. someone who inhabits it_n participated

Further, the translation rule T14 allows the quantifier translation of the noun phrase *every midwestern city* to bind the free variable corresponding to the pronoun it_n. This leads to the counterintuitive analysis of (70) (=(25)) in (7b). Rodman proposes to block such illicit interpretations by marking the indices on pronouns within relative clauses with a special sign that blocks application of the Quantifying-in rule from outside the relative construction. Rodman uses the superscript 'R' to mark such pronouns, disallowing sentential expressions like (72) that do not contain 'R' on pronouns within their relative clauses. As a result, the expression in (72) is not derived by Rodman's grammar, and instead the following expression is derived, with the index n^Ron the pronoun.

73. someone who inhabits it_{n^R} participated

The syntactic rule S14 in Rodman's extension of the PTQ fragment can only substitute a noun phrase for occurrences of pronouns with indices *unmarked* by 'R'. Consequently, sentence (70) cannot be derived using (73) and the noun phrase *every midwestern city.* The net outcome of this treatment is that sentences like (70) and (71) are derived by Rodman's fragment, but without an analysis that gives sentential scope to the NP within the relative clause.

Rodman further notes that because of the way in which his relative clause rules are construed, pronouns with an 'R-ed' index cannot be used for forming more complex relative clauses. For instance, from (73) Rodman's fragment cannot generate the following ungrammatical sentence:

74. * I admire the city which someone who inhabits participated

As Rodman argues, this proposal captures the parallelism between island restrictions on overt wh-"movement" (to wit, the movement of the relative clause operator *which* in (74)) and scope islands (witness the absence of a wide scope reading for (70)).

4.2.3 Cooper Storage

One of the first alternatives to Quantifying-in, and the first semantic approach to QNP scope, was the proposal of Cooper [1975; 1979; 1983] known as *Cooper Storage.* For Cooper, a central motivation for this technique [Cooper, 1975, ch. 4] was to show that syntactic representations for natural language sentences need not be "disambiguated" in the sense of PTQ. Each syntactic representation in Cooper's system may have more than one semantic analysis. The reason that examples showing quantifier scope ambiguity, such as (69), are seen to justify such an approach, is that straightforward syntactic considerations (relating e.g. to well-formedness, or to syntactic constituency tests) do not appear to support assumptions about syntactic ambiguity in the relevant examples, or about the existence of multiple syntactic derivations. For instance, sentences like (69) receive only one syntactic analysis in the grammar of Section 2. If one considers that inverse scope readings of such sentences are not by themselves sufficient reason for postulating a syntactic ambiguity, the conclusion that their syntactic representation must receive more than one semantic analysis may seem inevitable.

Cooper's account of "purely semantic" ambiguity is obtained by generalizing meaning representations.[17] First, meanings in Cooper's account are represented using ordered *pairs*, where the core lambda term representing the expression's meaning is coupled with a *store.* A store is a sequence of pairs of quantifiers and variables they bind. Such representations need to be processed in order to interpret the expression. For instance, one of the representations for sentence (69) above is the following one:

$$\{hi_1 = \langle \textsc{admire}(x)(y), \langle x/Q1, y/Q2\rangle\rangle,$$

where

$$\begin{array}{rcl} Q1 & = & \lambda A.\forall z[\textsc{man}(z) \rightarrow A(z)] \\ Q2 & = & \lambda B.\exists u[\textsc{woman}(u)) \wedge B(u)] \end{array}$$

The first element in such a representation as $\Phi 1$ is a lambda-term (in this case $\textsc{admire}(x)(y)$), possibly with free variables, which can be bound by

[17] Our account of Cooper Storage here essentially follows Carpenter [1997, ch. 7]. For another overview of Cooper Storage see Hendriks [1993, ch. 1].

one of the quantifiers in the store. Cooper essentially assumes that each quantifier in the store can bind the respective variable at any point in the process of meaning derivation. NPs are the syntactic elements that contribute quantifiers to meaning representation. As a result, each NP occurring in an expression may either combine with the core meaning directly, or be stored and, at a later stage of the semantic interpretation process, taken out of storage and combined with the core meaning at that point. Hence, each expression containing one or more NPs may in principle have more than one representation using Cooper storage. For instance, in addition to Φ1, sentence (69) also have the following representations:

$$\begin{array}{lcl} \Phi_2 & = & \langle Q1(\lambda x.\textsc{admire}(x)(y)), \langle y/Q2\rangle\rangle \\ \Phi_3 & = & \langle Q2(\lambda y.\textsc{admire}(x)(y)), \langle x/Q1\rangle\rangle \\ \Phi_4 & = & \langle Q2(\lambda y.Q1(\lambda x.\textsc{admire}(x)(y))), \langle -\rangle\rangle \\ \Phi_5 & = & \langle Q1(\lambda x.Q2(\lambda y\textsc{admire}(x)(y))), \langle -\rangle\rangle \end{array}$$

In representation Φ2, the quantifier $Q1$ binds the variable x associated with it, whereas $Q2$ remains on the store. In representation Φ_3 it is the opposite situation, whereas in representation Φ_4 and Φ_5, both quantifiers bind their variables and the store is empty, with the different scope construals of the quantifiers with respect to one another. Φ_5 is equivalent to the object wide scope reading of the sentence; it is obtained when the translation of the object is stored, and taken out of storage after the translation of the subject has been combined with the core meaning. Only the last two representations are fully interpretable, and lead to the actual two meanings of the sentence.

In [Cooper, 1979, 157-158] and [Cooper, 1983, 61] a preliminary account of CNPC restrictions on scope is proposed using Cooper's method of quantifier storage.

4.2.4 Type Flexibility

An approach to QNP scope that is also purely semantic, yet quite different from Cooper Storage, was proposed by Hendriks [1993]. The three approaches to standard QNP scope that were reviewed above all capture scope ambiguity using operations on noun phrases or the quantifiers they denote. Unlike these approaches, Hendriks' proposed method takes predicates to be the locus of scope ambiguity. Following Partee and Rooth [1983], Hendriks assumes that predicates in natural language have multiple semantic types. On the one hand, as in our toy grammar of section 2 and in many traditional accounts, it is assumed that natural language predicates can take *entities* as their arguments. On the other hand, Hendriks follows

Montague and Partee/Rooth, and allows predicates in natural language to have generalized quantifiers as their direct arguments.

One of Montague's motivations for allowing quantifiers to be direct arguments of predicates comes from the semantics of *intensional predication* in natural language. Consider the following classical example.

75. John is looking for a unicorn.

Sentences like (75) cannot be fully treated by analyzing the verb *look for* as a relation between e-type entities. Sentence (75) can of course be true if no unicorns exist. No plausible interpretation for the object *a unicorn* would capture this fact if the object argument of verb *look for* were to be treated as a simple entity. Montague's conclusion from this difficulty was to allow intensional predicates like *look for* to take an *intensional quantifier* as their direct object argument. In extensional terms, this analysis can be presented as treating the transitive verb *look for* with type $\langle\langle\langle e,t\rangle,t\rangle,\langle e,t\rangle\rangle$. For a discussion of this analysis and intensional verbs see [Gamut, 1991, pp. 168-9; Zimmermann, 1993], among others.

Following his general assumptions about uniform type assignment to syntactic categories, Montague also assigned intensional types to non-intensional predicates like *participate* above, or *admire* or *inhabit* in (69) and (70). In extensional terms, this means that also transitive verbs like *admire* or *inhabit* receive the analysis above of the predicate *look for* as allowing a quantificational direct argument. Partee and Rooth diverged from Montague's type uniformity, and proposed that extensional verbs like *participate*, *admire* or *inhabit* may have *multiple types*. In Partee and Rooth's account, extensional transitive predicates lexically denote relations between entities, but their arguments can be adjusted to fit the quantifier type $\langle\langle$e,t$\rangle$,t$\rangle$ when the meaning derivation process requires it.

Hendriks exploits the type ambiguity in Partee and Rooth's proposal in order to derive QNP scope ambiguity in his semantic system. Simplifying Hendriks' mechanism, we introduce the following *type shifting operators*, both of which map a relation of type $\langle e,\langle e,t\rangle\rangle$ between entities to a relation between quantifiers.

$$\begin{array}{lcl} \sum_{ONS} & = & \lambda R_{\langle e,\langle e,t\rangle\rangle}\lambda Q_{\langle\langle e,t\rangle,t\rangle}\lambda P_{\langle\langle e,t\rangle,t\rangle}.P(\lambda y.Q(\lambda x.R(x)(y))) \\ \sum_{OWS} & = & \lambda R_{\langle e,\langle e,t\rangle\rangle}\lambda Q_{\langle\langle e,t\rangle,t\rangle}\quad\lambda P_{\langle\langle e,t\rangle,t\rangle}.Q(\lambda x.P(\lambda y.R(x)(y))) \end{array}$$

Under the assumption that these operators can apply to the transitive predicate *admire* in (69), we derive the object-narrow-scope (direct scope) reading using the first operator and the object-wide-scope (inverse scope) reading using the second one. This is illustrated below.

$$\begin{array}{ll} \sum_{ONS} & (\text{ADMIRE})(\text{EVERY}(\text{MAN}))(\text{SOME}(\text{WOMAN})) \\ & \Leftrightarrow \exists x[\text{WOMAN}(x) \wedge \forall y[\text{MAN}(y) \rightarrow \text{ADMIRE}(y)(x)]] \\ \sum_{OWS} & (\text{ADMIR})(\text{EVERY}(\text{MAN}))(\text{SOME}(\text{WOMAN})) \\ & \Leftrightarrow \forall y[\text{MAN}(y) \rightarrow \exists x[\text{WOMAN}(x) \wedge \text{ADMIRE}(y)(x))]] \end{array}$$

In addition to these direct/inverse scope readings of simple transitive sentence, Hendriks shows that his system also allows deriving wide scope readings of QNPs beyond embedded structures, like the ones that were illustrated by example (60) in section 4.2.1. See [Hendriks, 1993, p.85-88] for further details.

Hendriks suggests that (island) constraints on quantifier scope may be captured in his system by only allowing type shifting operations for a subclass of lexical items (e.g., for transitive verbs, but not for relative clause operators).

4.2.5 Categorial approaches

A purely semantic proposal to quantifier scope was tentatively suggested by van Benthem [1986, 130-131; 1991, 61, 113-114], based on a *non-directional* version of Categorial Grammar, known as the *Lambek Calculus* with permutation (LP). Van Benthem's LP system can be conceived of as an extension of the core semantic calculus for meaning composition. Early categorical approaches (e.g. [Ajdukiewicz, 1935]) only allow *function application*, which is used as the core principle for composing types (or categories). In a general format, function application allows one to "eliminate" the functional relation between types A and B in a complex type $\langle A, B\rangle$ by providing an $\langle A, B\rangle$-type function $f$with its A-type argument x. Officially, this simple type/meaning change is written as follows in natural deduction format:

$$\frac{\langle A, B\rangle : f \qquad A : x}{b : f(x)}$$

This rule of function application corresponds to translation rule 3 of the grammar in section 2. For instance, when composing a quantifier EVERY(MAN) of type $\langle\langle e, t\rangle, t\rangle$ with a one place predicate PARTICIPATE of type $\langle e, td\rangle$, Ajdukiewicz Calculus allows function application with the appropriate outcome of type t. In natural deduction format, this simple inference of types and meanings is written as follows:[18]

$$\frac{\langle\langle e,t\rangle,t\rangle : \textsc{every}(\textsc{man}) \qquad \langle e,t\rangle : \textsc{participate}}{t : (\textsc{every}(\textsc{man}))(\textsc{participate})}$$

In addition to function application, Van Benthem's LP Calculus, like other calculi following Lambek [1958], also contains a rule of *hypothetical reasoning.* In natural deduction format, this rule allows one to introduce an element into the meaning derivation as a variable, which is later eliminated using *abstraction*, and which derives a function type. In natural deduction format, hypothetical reasoning looks as follows:

$$\begin{array}{ll} [A : x]1 & \text{– introduction of assumption 1} \\ \vdots & \text{– using assumption 1 for derivation} \\ \dfrac{\overline{\quad B : y \quad}}{\langle A,B\rangle : \lambda x.yg}\,E1 & \text{– eliminating assumption 1 using hypothetical reasoning} \end{array}$$

In this scheme, LP's hypothetical reasoning rule "pretends as if" a type A with meaning x is present in the derivation (assumption 1), uses it for deriving a type B with meaning y, and then "discharges" assumption 1 by creating a type $\langle A,B\rangle$ with meaning $\lambda x.y$.

Hypothetical reasoning allows LP to do away with the fairly artificial translation rule 4 of section 2, while at the same time deriving QNP scope ambiguity. Consider the following meaning derivation for sentence (69).

76. some woman admires every man

$$\dfrac{\langle\langle e,t\rangle,t\rangle : \textsc{some}(\textsc{woman}) \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\langle e,\langle e,t\rangle\rangle : \textsc{admire} \quad [e : y]1u}{\langle e,t\rangle : \textsc{admire}(y)} \qquad [e : x]2}{t : \textsc{admire}(y)(x)}}{\langle e,t\rangle : \lambda y.\textsc{admire}(y)(x)}\,E1 \qquad \langle\langle e,t\rangle,t\rangle : \textsc{every}(\textsc{man})}{t : (\textsc{every}(\textsc{man}))(\lambda y.\textsc{admire}(y(x))}}{\langle e,t\rangle : \lambda x.(\textsc{every}(\textsc{man})))(\lambda y.\textsc{admire}(y)(x))}\,E2}{t : (\textsc{some}(\textsc{woman}))(\lambda x.(\textsc{every}(\textsc{man}))(\lambda y.\textsc{admire}(y)(x)))}$$

This direct scope reading is derived because hypothetical reasoning allows LP to "feed" the transitive predicate with its entity arguments by postulating them in the derivation, and then to "discharge" these assumptions

[18] Also the meaning and type of the quantifier $\textsc{every}(\textsc{man})$ are derived using a similar application from the standard types of the determiner and the noun as assumed in the grammar.

at the steps preceding the application of the object/subject quantifier. A similar use of hypothetical reasoning also derives the inverse scope reading, as in the derivation below.

77. some woman admires every man

$$\dfrac{\langle\langle e,t\rangle,t\rangle : \textsc{some}(\textsc{woman}) \quad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\langle e,\langle e,t\rangle\rangle : \textsc{admire} \quad [e:y]1}{\langle e,t\rangle : \textsc{admire}(y} \quad [e:x]2}{t : \textsc{admire}(y)(x)}}{\langle e,t\rangle : \lambda x.\textsc{admre}(y)(x)}\,E2 \quad [\langle\langle e,t\rangle,t\rangle : Q]3}{t : Q(\lambda x.\textsc{admire}(y)(x))}}{\langle e,t\rangle : \lambda y.Q(\lambda x.\textsc{admire}(y)(x))}\,E1 \quad \langle\langle e,t\rangle,t\rangle : \textsc{every}(\textsc{man})}{t : (\textsc{every}(\textsc{man}))(\lambda y.Q(\lambda x.\textsc{admire}(y)(x)))}}{\langle\langle\langle e,t\rangle,t\rangle,t\rangle : \lambda Q.(\textsc{every}(\textsc{man}))(\lambda y.Q(\lambda x.\textsc{admire}(y)(x)))}\,E3}{t : (\textsc{every}(\textsc{man}))(\lambda y.(\textsc{some}(\textsc{woman}))(\lambda x.\textsc{admire}(y(x)))}$$

This inverse scope reading of (69) is here derived because of the possibility to introduce a subject quantifier Q by assumption 3, which takes narrow scope below the object *every man*, and which is later discharged before the quantifier denoted by the subject *some woman* is composed in the derivation.[19]

Unfortunately, as pointed out by Hendriks [1993, 69], this derivation of QNP scope ambiguity is accompanied by massive overgeneration. The principle of hypothetical reasoning allows the introduction of "traces" of arguments before they actually appear in the derivation. A too simplistic usage of this principle may also allow binding of such traces by the "wrong" quantifier in the sentence. For instance, if assumption 2 in derivation (76) were to be discharged immediately prior to the composition with the object quantifier, and similarly for assumption 1 and the subject quantifier, the result would have been the following one.

78. some woman admires every man

[19]Note that there is an apparently simpler way of deriving the inverse scope reading of (69) than the one in (77), using a derivation similar to (76) where assumption 1 is discharged after assumption 2, and the subject quantifier composes with the transitive verb before the object quantifier. This analysis would be completely symmetrical to the one in (76), and it is therefore often assumed in the categorical literature. We show here the more complicated derivation (77) of the inverse scope reading, in order to show that the problem demonstrated in (78) below persists even with the standard [Subject [Verb Object]] constituency, which we adopt throughout this paper.

$$
\dfrac{\langle\langle e,t\rangle,t\rangle : \textsc{some}(\textsc{woman}) \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\langle e,\langle e,t\rangle\rangle : \textsc{admire} \quad [e:y]1}{\langle e,t\rangle : \textsc{admire}(y)} \qquad [e:x]2}{t : \textsc{admire}(y)(x)}}{\langle e,t\rangle : \lambda x.\textsc{admire}(y)(x)}\,E2 \qquad \langle\langle e,t\rangle,t\rangle : \textsc{every}(\textsc{man})}{t : (\textsc{every}(\textsc{man}))\lambda x.\textsc{admire}(y)(x))}}{\langle e,t\rangle\lambda y, (\textsc{every}(\textsc{man}))(\lambda x.\textsc{admire}(y)(x))}\,E1}{}}{t : (\textsc{some}(\textsc{woman}))(\lambda y.(\text{every}(\text{man}))(\lambda x.\textsc{admire}(y)(x)))}
$$

In the resulting proposition, this analysis states that there is a woman who is admired by every man, which is not consistent with any interpretation of sentence (69).

In the categorial grammar literature on scope ambiguity there have been two major attempts to overcome this kind of overgeneration. Moortgat [1997] proposes a *multimodal* version of categorical grammar, which uses a special scoping type constructor different from the functional constructor in standard functional type $\langle A, B\rangle$. In this way hypothetical reasoning in the semantics is properly coupled with the syntax of the sentence without generating illicit derivations like (78). A more recent strategy, first proposed in De Groote [2001] and Muskens [2003], who attribute the original approach to Oehrle [1994], is that of *Abstract Categorial Grammar* (sometimes also referred to as *Lambda Grammar*), where the relations between syntax and semantics allow a more sophisticated separation between word order and semantic composition than in LP and traditional categorical grammars. We will not try to discuss the technical details of these works here, and refer the reader to the overviews in [Carpenter, 1997, ch. 7] and [Muskens, 2003]). Importantly, these categorical approaches keep the treatment of QNP scope phenomena rather close to the treatment of "overt movement" phenomena. Thus, it is conceivable that parallelisms between scope and movement can be captured in categorical approaches similarly to QR theory. For some remarks on this point in relation to the Coordinate Structure Constraint see [Carpenter, 1997, 241].

Yet another approach to quantifier scope, not entirely "categorial", but quite in the spirit of the categorial approaches surveyed above, was proposed by Barker [2002]. Barker uses the notion of *continuation* from computer science, as an account of the apparent mismatch between quantifier types and their function in syntax. This move allows an elegant account of quantifier scope as well.

A general survey and evaluation of various approaches to quantifier scope appears in Jacobson [2002]. Jacobson distinguishes among four types of theories: direct compositional approaches (e.g. Cooper Storage, Hendriks' type shifting); weaker compositional approaches that include some enrichments of the syntax (e.g. Quantifying-in); generative semantic approaches (modeling scope relations at Deep Structure); and "modern" syntactic approaches to scope (modeling scope relations at LF). We believe that it is worthwhile to consider Jacobson's classification as a basis for discussion on the merits and disadvantages of various techniques, also in light of categorial approaches (e.g. Moortgat) and more recent proposals like the ones by Barker, de Groote and Muskens. However, we will not attempt this analysis here.

4.2.6 Discussion — different emphases by different approaches to QNP scope

The approaches to QNP scope that were surveyed above are rather heterogeneous in terms of their empirical coverage and methodological standpoints. The QR theory, being a syntactic theory, is most concerned about characterizing different syntactic configurations for quantifier scope, and motivating the derivation of LF using the QR movement rule. Cooper Storage and Hendriks' type shifting mechanism are purely semantic theories which aim at avoiding syntactic representations of QNP scope. In the Quantifying-in technique and the categorial approaches surveyed, syntactic or derivational processes are still used for describing QNP scope, but the main focus is on securing a tight connection between these operations and the semantic component.

These different methodological and technical emphases complicate the comparison between the different approaches to QNP scope. In terms of empirical content the QR theory is by far the most comprehensive among these proposals. Problems like the nature of the restrictions on QNP scope have been much better studied and described in the QR literature. By contrast, the other approaches have studied more extensively the methodological and technical questions surrounding the notion of *compositionality*, and in general — the matching between syntax and semantics, as revealed by QNP scope phenomena. We cannot address here the question of compositionality in detail, and refer the reader to some of the many works on this topic: [Montague, 1970; Janssen, 1983; 1996; Hendriks, 1993, ch.2; Jacobson, 2002; Barker and Jacobson, 2007].

To compare specific theories in this situation is a rather difficult task. Explicit comparisons among the "Montagovian" theories have been at times carried out in the literature: see for instance Carpenter's [1997, ch. 7] comparison of Moortgat's scoping type constructor and the methods of Quantifying-in and Cooper Storage. However, while these comparisons are beneficial for choosing between the non-QR theories, they only lightly touch on the empirical concerns of most QR-theorists. Conversely: in the QR literature there is considerably less emphasis on foundational questions regarding the mathematical properties of the relations between syntax and semantics.

We believe that further developments in the theory of QNP scope may ultimately depend on the general understanding of "movement phenomena" (cf. discussion at the end of section 4.2.1). Perhaps only such a comprehensive theory could settle the current discrepancies between rival approaches to standard QNP scope. Once the more general problem is resolved, current technical differences between some alternative theories of scope may appear less central they currently do.

4.3 Non-Standard Scope Mechanisms

The direct scope and inverse scope readings that we have discussed so far can all be treated using standard scope mechanisms. Despite the many technical differences between these mechanisms, they all produce *linear* relations between QNPs as exemplified in (50). In most contemporary theories, such linear quantification — or "Fregean" quantification (cf. [Keenan, 1992]) — technically means that the QNPs in the sentence are interpreted as a sequence of standard $\langle\langle e,t\rangle,t\rangle$ generalized quantifiers, which are composed using standard translation rules or compositional principles. However, as mentioned above, there are semantic phenomena that involve more complicated mismatches between syntactic structure and the scopal semantics of QNPs. This section gives a brief overview of some of these challenges and attempts that have been made to address them.

4.3.1 Branching quantification

The assumption that scopal relations between quantifiers in natural language are essentially linear draws to a large extent on the tradition of first order Predicate Calculus. In the Predicate Calculus, quantifiers can only take scope (i.e. be prefixed to formulas) in a linear order, as in the following formula.

79. $\forall x \exists z \forall y \exists u \Phi(x,y,z,u)$

Following Henkin [1961], logicians have also explored other possible ordering relations between quantifiers, especially the following *branching* scheme.

80.

$$\begin{array}{l}\forall x \exists z \\ \\ \forall y \exists u\end{array} \!\!\!> \Phi(x, y, z, u)$$

Henkin proposed a semantics for branching schemes as in (80) using the notion of *Skolem functions*, which is defined below.

81. A n-ary *Skolem function* over a domain E is a function that sends any non-empty subset A of E and a tuple of n elements in E to an element of A.

For instance, a 2-ary Skolem function f over E sends every non-empty set $A \subseteq E$ and any two elements x and y in E to an element $f(x, y, A)$ in A. A 0-ary Skolem function f is a function that sends any non-empty subset A of E to one of its elements $f(A)$. Such 0-ary Skolem functions, which are discussed in more detail in section 4.3.3 below, are also known as *choice functions*.

In Henkin's proposal, linear quantification using Skolem functions is used for interpreting formulas with branching first-order quantifiers. For instance, the formula in (80) is interpreted as the non-first-order formula below, using linear existential quantification over Skolem functions f and g of arity 1, where the set E is the whole domain of individuals in the model.[20]

82. $\exists f \exists g \forall x \forall y \Phi(x, y, f(x, E), g(y, E))$

By the semantics in (82), the branching formula in (80) is not equivalent to any formula with a linear ordering of the quantifiers.

Let us now concentrate on possible linguistic manifestations of differences between branching interpretations and standard linear schemes of first order quantifiers. The claim that natural language sentences can exhibit branching quantification that should be interpreted similarly to Henkin's scheme was first made in Hintikka [1973] and Gabbay and Moravcsik [1974]. One of Hintikka's well-known examples is the following.

[20] In the literature on Skolem functions, the set argument in definition (81) of Skolem functions is sometimes suppressed when this set argument is the whole domain of individuals E. For linguistic purposes, however, quantification is often *restricted* and the set argument in (82) is replaced by a proper subset of E, as in formula (85) below representing the meaning of the restricted branching quantification in (84).

83. Some book by every author is referred to in some essay by every critic.

Hintikka suggests that sentence (83) should have an analysis equivalent to the scheme of branching quantification given in (84) below. Following Schlenker [2006], we adopt in (84) a format of *restricted quantification* that is more convenient than Henkin's scheme for displaying the parallelism between the branching formula and the sentence.[21]

84.

$$\left.\begin{array}{l} [\forall x : \textsc{author}(x)][\exists z : \textsc{book-by}(z,x)] \\ \\ \quad [\forall y : \textsc{critic}(y)][\exists u : \textsc{essay-b}(u,y)] \end{array}\right\rangle \ \textsc{referred-to-in}(z,u)$$

Using a proper adjustment of Henkin's strategy in (80)–(82), formula (84) can be interpreted as follows using Skolem functions of arity 1, similar to (82).

85. $\exists f \exists g [\forall x : \textsc{author}(x)][\forall y : \textsc{critic}(y)]$
$\textsc{referred-to-in}(f(x, \lambda z \text{sc book-by}(z,x)), g(y, \lambda u.\textsc{essay-by}(y,y)))$

Assuming that every author wrote at least one book and every critic wrote at least one essay, the proposition expressed by (85) can roughly be paraphrased as follows:

> "There is a way to map each author x and his books $B(x)$ to a particular book $b(x)$, and there is a way to map each critic y and his essays $E(y)$ to a particular essay $e(y)$ s.t. for each author x and critic $y : b(x)$ is referred to in $e(y)$."

What this paraphrase entails, in the terms of Sher [1991], is that there is a "massive nucleus" N of books and essays, such that each book in N is referred to by each essay in N, and the writers of the books and essays in N cover the set of all authors and critics.

Whether such a reading that involves a "massive nucleus" exists for sentences like (83) has been debated in the literature [Fauconnier, 1975; Beghelli *et al.*, 1997; Landman, 2000, ch. 9.5; Schlenker, 2006]. One of the problems for deciding on this question is similar to the problem discussed in relation to sentence (8) in section 2.3.2 with respect to inverse scope

[21] In a restricted quantifier notation, the formula $[\forall x : P(x)]\Phi$ is equivalent to the standard predicate calculus formula $\forall x[P(x) \rightarrow \Phi]$, whereas the formula $[\exists x : P(x)]\Phi$ is equivalent to $\exists x[P(x) \wedge \Phi]$.

readings. As Fauconnier pointed out, the branching scope analysis in (85) is logically stronger than some of the linear readings for (83). For instance, if a "massive nucleus" of books and essays exists as required by (85), then the following, linear scope reading of (83) is automatically satisfied as well.

86. $[\forall x : \text{AUTHOR}(x)][\forall y : \text{CRITIC}(y)][\exists z : \text{BOOK-BY}(z, x)][\exists u : \text{ESSAY-BY}(u, y)]$ $\text{REFERRED-TO-IN}(z, u)$

Thus, using truth-conditional evidence alone, it is hard to determine if sentence (83) should have an interpretation as formalized in (84) and (85).

Independently of this empirical debate, other works [Barwise, 1979; Westerstähl, 1987; van Benthem, 1989; Sher, 1991] suggested an extension of Henkin's definition of branching quantification to generalized quantifiers beyond the existential and universal quantifiers of first order logic. This makes it possible to construct branching schemes without a linear equivalent, using only two (generalized) quantifiers. For instance, Sher suggested the definition in (88) below for interpreting the branching formula (87) with the generalized quantifiers Q1 and Q2.

87.

$$\left.\begin{array}{l} Q1\ x \\ \\ Q2\ y \end{array}\right> \Phi(x, y)$$

88. Formula (87) is true iff there are sets X and Y such that the following conditions hold:

   (1) Q1 holds of X and Q2 holds of Y;

   (2) each element of X is in the relation Φ to each element of Y;

   (3) X and Y are maximal sets satisfying condition 2.[22]

Under this definition, the branching analysis of sentence (89) below, with non-monotone numeral quantifiers, should be interpreted as paraphrased in (90).

89. Exactly four critics read exactly ten books.

[22] That is, no element can be added to X or Y such that condition 2 remains satisfied.

90. There is a "massive nucleus" N of four critics and ten books, such that each critic in N read each book in N, and this nucleus is a *maximal* one: no critic c outside N read all the books in N, and no book b outside N was read by all the critics in N.

As in the case of Hintikka's original example, whether sentences like (89) require an analysis along the lines of (90) was debated in the literature [Beghelli *et al.*, 1997].

Here we will not try to settle the empirical debates that surround the linguistic status of branching analyses. Rather, we will now move on to other problems of non-linear scope, where there are fewer empirical doubts surrounding the validity of the core factual judgments challenging standard theories of linear QNP scope. However, as we will see, accounts of other non-linear scope phenomena have been proposed that bear a strong resemblance to the mechanisms that were used to characterize "branching" quantification.

4.3.2 Cumulative quantification

A non-linear scopal interaction between quantifiers, which is somewhat similar to "branching" but more solidly supported by empirical evidence, is *cumulative quantification.* The phenomenon was illustrated in [Scha, 1981] using the following example, which Scha paraphrased as in (92).

91. (exactly) 600 Dutch firms use (exactly) 5000 American computers.

92. The total number of Dutch firms that use an American computer is 600, and the total number of computers that are used by a Dutch firm is 5000.

Similarly to the "branching" analysis (90) of (89), the analysis of (91) in (92) does not give priority to the scope of any of the two QNPs over the other. Like branching analyses, also cumulative interpretations cannot be expressed using any linear combination of unary generalized quantifiers like the ones generated in section 4.1 above.[23] Empirically, the situation is clearer with such "cumulative" effects than with the "branching" effects discussed above. Even if Scha's strategy of paraphrasing sentence (91) in (92) is not completely accurate, it is rather clear that (92) comes close to

[23]For a proof of this fact, as well as more examples of such cases of *inherently polyadic* quantification, which is not reducible to linear composition of unary quantifiers, see [van Benthem, 1989; Keenan, 1992].

capturing a scope effect in (91) that does not involve simple linear composition of $\langle\langle e,t\rangle,t\rangle$ generalized quantifiers. Intuitively, speakers agree that sentences like (91) can be true in situations that render the linear scope analysis (or analyses) of the sentence false, but where the proposition expressed by (92) is true.[24] Unlike the branching paraphrase of (89) in (90), however, Scha's paraphrase of (91) in (92) makes no requirement of a "massive nucleus", which is the part of the branching semantics that is most debated by researchers who deny the relevance of this semantics for natural language [Beghelli *et al.*, 1997].

There are quite a few mechanisms that have been proposed in the literature in order to deal with cumulative effects. Scha proposed to compose the standard meanings of determiners like *exactly three* and *exactly five* into complex determiners, which can combine with the two nouns (e.g. *men* and *women* in (91)) and derive a cumulative reading. Another proposal, by Schein [1993, ch. 9], is to use a mechanism that combines *event semantics* with the *linear* scope mechanism of QR and anaphoric analysis for deriving a cumulative reading of sentences like (91) or (93). Landman [2000, pp. 222-280] addresses the problem of cumulative readings for such sentences using another mechanism in event semantics, involving *maximality principles* of the sort used for implicatures of numeral expressions [Krifka, 1989].

We will not embark here upon a critical evaluation of these proposals. One of the complicating factors in such an evaluation is the status of possible interactions between cumulativity and collective readings. For instance, Scha considers examples with two definites like *the soldiers hit the targets*, and contends that the prominent reading of such sentences is to be paraphrased using vague predication over collective entities, roughly: *there is a hitting relation between the group of soldiers and the group of targets.* This kind of interpretation is sometimes also referred to as *cumulative.* Whether such effects with "referential" plural NPs are to be distinguished from cases like (91) or (93) is an open question (see [Sternefeld, 1997; Winter, 2000a; Beck and Sauerland, 2001] among others). However, it should be noted that cumulative quantification in the sense of Scha is also observable with singular NPs, and not only with plurals. For instance, consider the following cases, classified as "resumptive" by May [1989]:

93. Exactly one man admires exactly one woman.

[24]This conclusion also holds when considering the object-wide-scope analysis of sentences like (91), since the cumulative analysis is also independent of this analysis. However, as mentioned in section 3.4, this inverse scope construal is unlikely to reflect a true reading of sentences like (91) with numeral indefinites.

94. No man admires no woman.

Cases like (93) and (94) also admit readings that are paraphrased using Scha's cumulative strategy in (92), which takes into account the total numbers of men admiring women and women admired by men. Thus, to say the least, the relations between cumulative quantification and plurality are not obvious.

4.3.3 Wide-scope indefinites and quantification over Skolem functions

As mentioned in sections 3.3 and 3.5, one of the long-standing challenges for theories of QNP semantics is the scopal behavior of indefinite NPs. So far, we have assumed that like other noun phrases, indefinites should denote generalized quantifiers, possibly augmented with branching/cumulative schemes of interpretation as discussed above. However, it is a well-established fact in the extensive literature about the restrictions on QNP scope (see section 3.3) that some indefinites do not seem to obey the same restrictions as other QNPs. We repeat example (32), from the *locus classicus*, [Fodor and Sag, 1982]:

95. If a friend of mine from Texas had died in the fire, I would have inherited a fortune.

Fodor and Sag's intuition, widely agreed on in the literature, is that sentence (95) can be interpreted as true if I have a certain friend whose death would make me rich, even if I have other friends for whom this does not hold. Under the standard treatment of indefinites as quantifiers, this behavior looks quite exceptional. This is because, as was pointed out in section 3.2, the scope of most other QNPs is restricted (at least) by island constraints. In (95) the indefinite is within an adjunct island (the *if* clause, see section 3.2). As a result, standard scope mechanisms are expected to be restricted so that if the indefinite denotes an existential quantifier, this quantifier would not take scope over the conditional. The only reading expected for (95) using island restricted standard scope mechanisms is the following one (where the conditional is treated as material implication).

96. $[\exists x[\text{FRIEND}(x) \wedge \text{DIE}(x)]] \rightarrow \text{INHERIT\_FORTUNE}(I)$

The proposition in (96) entails that in *any* event in which a friend of mine dies I inherit a fortune. The interpretation of (95) that Fodor and Sag point out is however more similar to the following analysis, where the existential quantifier takes sentential scope, over the material implication.

97. $\exists x[\textsc{friend}(x) \wedge [\textsc{die}(x) \rightarrow \textsc{inherit\_fortune}(\mathbf{I})]]$

The contrast between (95) and its variation (98) with a universal quantifier is instructive. For (98), the analysis in (99) is the *only* plausible reading available for the sentence.

98. If every friend of mine from Texas had died in the fire, I would have inherited a fortune.

99. $[\forall x[\textsc{friend}(x) \rightarrow \textsc{die}(x)]] \rightarrow \textsc{inherit\_fortune}(\mathrm{I})$

Indeed, sentence (98) unequivocally claims that I inherit a fortune if *all* my friends die, which is the statement that (99) models.

The following example includes more of the indefinite NPs that have been reported to show the same effect illustrated by (95) above.

100. If *a certain friend/some friend/some friends/three friends* of mine from Texas had died in the fire, I would have inherited a fortune.

Work on the scope of indefinites has shown a wide range of syntactic contexts where indefinites show a freer scopal behavior than other QNPs. Specifically, the singular and plural indefinite NPs in (95) and (100) seem able to violate all island constraints, not only adjunct islands, and including the CNPC island present in our fragment: see section 3.3, example (30). Since the standard scope mechanisms (such as QR, Storage, Quantifying in, Type Shifting) must be made subject to these island constraints (or we no longer account for the usual island effects with other QNPs illustrated in (98)), the exceptional scope of indefinites must be due to some non-standard scope mechanism, or due to some other peculiarity of their interpretation.

There have been various attempts to address the challenge that the behavior of indefinites raises for the theory of QNP scope. We can roughly identify two extremes in the approaches that have proposed. One approach has been to derive the "wide scope" behavior of indefinites from their traditional treatment as existential quantifiers. Another approach analyses the "wide scope" impression with indefinites as an effect resulting from their exceptional descriptive (or "referential") properties. According to the first approach, sentence (95) should have an analysis equivalent (or logically similar) to the one given in (97). According to the second, sentence (95) has no such reading, and the "wide scope" effect is a result of analyzing the indefinite *a friend of mine* in (95) as close in meaning to a definite description or a demonstrative (i.e. *the/this friend of mine*).

We will not review or analyze in detail these two approaches and the various ways in which they are combined in actual proposals. Instead, we refer the reader to some of the many works on this problem (Egli & Von Heusinger 1995, Schwarzschild 2002, Farkas 1997, Ruys 1992, Abusch 1994 etc.). In the context of the current discussion of non-standard scope mechanisms, however, it is worthwhile to mention one logical semantic mechanism that has been proposed for treating "wide-scope" indefinites: the mechanism of *choice functions*, or more generally, Skolem functions. Skolem functions as defined in (81) above were used in Henkin's treatment of branching schemes with first order quantifiers. Assuming that f is a variable over choice functions (0-ary skolem functions), the following formula can be used to model the "wide-scope" effect in (95).

101. $\exists f[\text{die}(f(\text{friend})) \rightarrow \text{inherit\_fortune}(I)]$

The proposition in (101) claims that there is a value for a choice function f that satisfies the following formula:

$$\text{DIE}(f(\text{FRIEND})) \rightarrow \text{INHERIT\_FORTUNE(I)}$$

Assuming that the set of my friends is non-empty, let us denote the element that f assigns this set by r. By definition of f as a choice function, r is a friend of mine. Hence, the following proposition now holds:

$$\text{DIE}(r) \quad \rightarrow \text{INHERIT\_FORTUNE(I)}$$

This means that using the choice function representation in (101) is logically close to the predicate calculus representation in (97).[25]

Semantic mechanisms using choice functions for treating "scopal" phenomena were used in [Reinhart, 1992; 1997; Kratzer, 1998; Winter, 1997] and many more recent works, with notable variations in the details of their usage. Importantly, Kratzer proposed to use choice functions as a "referential" (or *deictic*) semantic mechanism, without existential quantifiers like the one in (101). The reason that many works have found representations as in (101) attractive for treating the "wide-scope" of indefinites is that, unlike standard existential quantification (e.g. (97)), the representation using choice functions does not require that the restrictive predicate of the indefinite is "pulled out" of its surface position. In (101), the predicate FRIEND that is denoted by the indefinite's restriction *friend of mine* remains within the scope of the conditional, in accordance with the surface

[25]The difference between (101) and (97) is in the case where the argument of the choice function f, i.e. the predicate FRIEND, is empty (which may occur if I happen to have no friends). The implications of this point for the usages of choice functions in formal semantics were extensively discussed in [Winter, 1997; 2001]. See [Ruys, 2006] for a somewhat different view.

constituent structure of the sentence, and is not "pulled out" of the adjunct island. The felicitous consequence is that no scope mechanism is required that violates island conditions.

One advantage of not having to pull out the restriction of the indefinite out of the island was pointed out in [Winter, 1997], based on observations in [Ruys, 1992] discussed in section 3.5. These works show that despite the fact that some plural indefinites show wide scope effects beyond islands, the scope of their *distributivity* is restricted to remain within the island. Relevant examples were given in sections 3.3 and 3.5, and two more examples are the following ones.

102. If three friends of mine from Texas had died in the fire, I would have inherited a fortune.

103. If three workers in our staff have a baby soon we will have to face some hard organizational problems.

In both (102) and (103), the sentence can be interpreted as a statement on three people (relatives or works), and possible scenarios that would occur under certain events happening to these people (death, having a baby). However, in both cases the events would have to happen to *all* three people in order for the conditional to take effect. Thus, for instance, sentence (102) can be interpreted as in (104) below, but not as in (105).

104. $\exists A[|A| = 3 \wedge [\forall x \in A\ \textsc{friend}(x)] \wedge [[\forall y \in A\ \textsc{die}(y)] \rightarrow \textsc{inherit\_fortune}(I)]]$

105. $\exists A[|A| = 3 \wedge \forall x \in A\ [\textsc{friend}(x) \wedge [\textsc{die}(x) \rightarrow \textsc{inherit\_fortune}(I)]]$

In (104) distribution over elements in the set A is independent for each of the predicates FRIEND and DIE. By contrast, in (105) distribution takes scope over the conditional. Winter further discusses cases like (103) of mixed scope, where distribution over different workers (and different babies!) is pragmatically prominent due to world knowledge. Such cases strengthen the conclusion that distribution cannot violate syntactic islands, and must, if existent, remain constrained within the island. This fact cannot be easily captured if restrictions on indefinites are free to violate islands, but it directly follows from the choice function mechanism.

Other works on the scope of indefinites [Kratzer, 1998; Chierchia, 2001; Winter, 2004; Schlenker, 2006] have shown various reasons to adopt the more general Skolem function mechanism for treating not only branching quantifiers as in (83) above, but also for some cases of more ordinary scope taking indefinites. Schlenker [2006] points out that this decision has interesting implications for the debate surrounding branching quantification. Reconsider sentence (83), repeated below.

106. Some book by every author is referred to in some essay by every critic.

With previous works, Schlenker assumes a Skolem function mechanism for interpreting the *scopal* behavior of indefinites in sentences like (96) and (100). Schlenker then argues that using the same mechanism, we expect the indefinites *some book* or *some essay* in (106) to lead to a branching reading of this sentences as formalized in (82) following Henkin's use of unary Skolem functions. If this is the case, the origins of branching quantification in such cases may be explained on independent considerations about the scope of indefinites. Further, the same Skolem mechanism would be unlikely to derive any "branching" reading for sentence (89) and similar ones, with indefinites like *exactly four critics* or *exactly ten books*. The reason is that this kind of modified numeral indefinites was argued [Liu, 1990] not to show any exceptional "wide scope" behavior. Therefore, it was concluded (e.g. [Winter, 2001, ch.3-4]) that modified numeral indefinites like these ones should not be treated using Skolem functions. Non-linear quantificational effects in cases like (89) exist, but independently of whether we classify them as "cumulative" or "branching", they are not likely to be captured by a linguistic mechanism that employs Skolem functions for interpreting indefinites.

5 TWO EMPIRICAL EXTENSIONS

Our discussion so far of scope inversion phenomena and their theoretical implications has focused exclusively on one type of empirical data: examples in which a scope bearing element takes wider scope than would be expected given its standard semantics and its position in the syntactic structure. The present section discusses two additional types of data that provide evidence that some scope shifting rule, of which we discussed various implementations in section 4.2, is operative in natural languages. Section 5.1 deals with examples in which a scope bearing element takes narrower scope than expected given its syntactic position ("scope reconstruction"). Section 5.2 presents data in which ellipsis resolution data, rather than intuitions arising from relative scope, provide evidence for the operation of a scope shifting rule. Because most literature on these topics is in the QR/LF tradition surveyed in section 4.2.1, our discussion will also mostly take this perspective.

It is not our intention, however, to suggest that these phenomena necessarily constitute an argument in favor of the QR approach: other plausible analyses have been proposed, and we will briefly mention a few of them.

5.1 Quantifier scope reconstruction — syntactic and semantic accounts

Let us repeat some examples of quantifier scope ambiguity from section 3.1.

107. all that glitters is not gold

108. an American runner is likely to win the race

109. someone always wins

110. someone probably spiked the punch

The scope problems in our fragment that have been treated so far all involve a QNP which is optionally assigned a wider scope than would be expected given its position in the overt syntactic structure. The examples repeated above, on the other hand, allow the QNP to take *narrower* scope than its surface position would lead one to expect. Thus, (108) allows a reading which can be paraphrased by ‘it is likely that there exists some American runner who wins the race’. Under this reading, presumably, the noun phrase *an American runner* is interpreted with narrow scope relative to *likely*: a *de dicto* reading. Similarly, in the other examples the subject can scope below negation, the adverb of quantification, and the modal adverb, respectively.

May [1977] discussed examples like (108) and proposed that his QR rule (see section 4.2.1 above) can sometimes move a QNP downward. This instance of applying QR is referred to as ‘Quantifier Lowering’ (QL)[26]. For (108) QL results in the following derivation.

111. a. DS: is $[_{AP}$ likely $[_{S}$ [an American runner]$_i$ to win the race]]
   b. SS: [an American runner]$_i$ is $[_{AP}$ likely $[_{S}$ t$_i$ to win the race]]
   c. LF: *t* is $[_{AP}$ likely $[_{S}$ [an American runner]$_i$ $[_{S}$ t$_i$ to win the race]]]
   d. $\textsc{likely}(^\wedge\mathrm{kx}[\textsc{american}(x) \wedge \textsc{runner}(x) \wedge \textsc{win\_the\_race}(x)])$

(111a) is the D-Structure, where *an American runner* occupies its base position as logical subject of *win the race*. (111b) is the S-Structure, derived via

[26]The term Quantifier *Raising* is therefore somewhat of a misnomer in theories that assume QL.

NP-movement of *an American runner* to the grammatical subject position of *be likely.* QL then results in the LF given in (111c), where the lowered NP still binds its original trace, with scope relations giving rise to the *de dicto* reading formalized in (111d).[27] In each of the examples given above, a similar derivation can be proposed: the QNP moves across the negation or adverb at S-Structure, a movement that can optionally be 'undone' at LF via QL.

Motivation for a syntactic movement analysis is weaker in these examples than in examples that motivated QR, for which sensitivity to syntactic island effects can be demonstrated. Indeed, the QL operation is syntactically suspect, as it is believed that movement operations in general do not move material downward. For instance, the derivation in (112) is ill-formed:

112. a. DS you asked who$_i$ [$_{CP}$ [John loved Mary]]

b. SS * you asked t$_i$ [$_{CP}$ who$_i$ [John loved Mary]]

There is assumed to be no operation that derives the illicit SS in (112b) by "lowering" *who* from its position in the main clause, in the DS representation (112a), to a position in the embedded clause, as in the SS representation (112b). This is not decisive evidence against the QL hypothesis, however. The reason (112) is ruled out might be that the wh-operator at S-Structure does not bind a trace (violating a ban on vacuous quantification) and its trace is unbound (violating e.g. Chomsky's [1986b, 85] Strong Binding condition). The putative examples of QL in (1512)–(1517) are different. The QL operation in these examples is assumed to be preceded by an overt Raising operation. For instance, in (111), the SS (111b) is already assumed to be derived from the DS in (111a) by a movement operation. As a result, the lowered QNP in the LF representation (111c) still binds its original trace after QL; hence, (111c) does not violate a ban on vacuous quantification. Indeed, QL is not allowed if not preceded by a Raising operation, even if the ban on vacuous quantification is respected:

113. [an American runner]$_i$ is eager [PRO$_i$ to win the race]

a. t_i is eager [[an American runner]$_i$ [PRO$_i$ to win the race]]

In (113), the S-Structure of which is not derived by movement of *an american runner* out of the embedded clause, a *de dicto* reading for the noun

[27]The '^' operator from [Montague, 1973] is meant to guarantee that the argument of the operator LIKELY is the *intension,* rather than the extension, of the complement proposition (equivalent to the clause "an American runner will win the race"). For details see [Gamut, 1991], or, in a more perspicacious format, [Gallin, 1975].

phrase *an American runner* is not available, presumably because QL would result in the LF (113a) in which the trace of *an American runner* in argument position is unbound (see May 1977, 1985).

Because QL is assumed to (partly) undo a preceding movement, the data are often considered to reveal a "reconstruction" effect: the effect found when an element moved out of its DS position functions at LF as though it occupies its pre-movement position (these effects are also found, e.g., with Binding Theoretic phenomena; see [Sportiche, 2006] for a recent overview). This has led to the hypothesis that, instead of a QL operation, the trace of the raised QNP might be playing a role in determining its scope.[28] One possible implementation makes use of the supposition in Chomsky's [1995] Minimalist Program that a movement trace is in fact a copy of the moved element. If so, the surface syntactic form (111b) is actually (114):

114. [an american runner] is likely $[_{IP}$[an american runner] to win the race]

The phonological component is assumed to delete the downstairs copy of *an American runner*. By contrast, the semantic component is assumed to have an interpretative strategy that ignores the upstairs copy. This results in the *de dicto* reading.

A semantic alternative to the syntactic approaches to scope reconstruction phenomena outlined above appears to be readily available. Consider, for instance, the ambiguous example (115), which allows the readings paraphrased in (115a) and (115b):

115. [How many people]$_i$ should John talk to t_i

   a. for how many people x, John should talk to x

   b. for which number n, John should talk to n-many people

Roughly speaking, reading (115b) involves "reconstruction" of *n-many people* into the scope of *should*. The solution proposed by e.g. [Cresti, 1995] (see also references cited there) involves the assumption that the trace of *how many people* can be translated as a variable of different types. *How many people* (or rather, *n-many people*, after *how* is separated out) is composed with its sister after abstraction over this variable. If it is type e, the result is (115a). If it is type $\langle s, \langle\langle s, \langle e, t\rangle\rangle, t\rangle\rangle$ (the type of the intension of a Montagovian generalized quantifier) then the translation of *n-many people* is converted into the scope of *should*, resulting in (115b).

[28]As e.g. in [Aoun and Li, 1989]. In some languages, quantifier scope ambiguities between subject and object appear to arise *only* in situations that are analyzed as reconstruction.

A similar strategy can be employed to deal with the examples (107)–(110). This is illustrated in (116) below for (108).

116. $[_S$ [an american runner]$_i$ [is likely $[_S t_i$ to win the race]]

a. $\lambda P_{\langle s,\langle e,t\rangle\rangle}.\mathrm{k}y[\textsc{american}(y) \wedge \textsc{runner}(y) \wedge^{\vee} P(y)]$

b. $\lambda x_e.\textsc{likely}(^{\wedge}\textsc{win\_the\_race}(x))$

c. $\lambda X_{\langle s,\langle\langle s,\langle e,t\rangle\rangle,t\rangle\rangle}.\textsc{likely}(^{\wedge}([^{\vee}X](^{\wedge}\textsc{win\_the\_race})))$

Assume, with Cresti, that the composition rule can freely apply to the intension or extension of an expression (depending on type requirements); that the trace of NP movement can be translated as a variable of type e or $\langle s,\langle\langle s,\langle e,t\rangle\rangle,t\rangle\rangle$; and that in the translation of a structure $[_S$ NP$_i$ VP] the relevant x_i variable in the translation of VP is abstracted over. Then in (116), the translation of the subject (116a) can be combined either with (116b) or (116c), giving the two readings discussed above.

We will not discuss the relative merits of a syntactic or semantic approach to scope reconstruction here. See [Cresti, 1995] for arguments that her semantic approach can deal with the island effects observed with scope reconstruction after wh-movement. See [Fox, 1995; 1999], on the other hand, for arguments that scope reconstruction is subject to Binding Theory and economy constraints on movement.

5.2 Antecedent Contained Deletion (ACD)

The phenomenon of Antecedent Contained Deletion (ACD) has been used to argue for the existence of a scope shifting rule.[29] Consider (117) and (118) (from [May, 1985]):

117. Dulles $[_{VP1}$ suspected Philby], and Angleton did $[_{VP2}$ e] too

118. Dulles $[_{VP1}$ suspected $[_{NP}$ everyone who Angleton did $[_{VP2}$ e]]]

The VP in the second conjunct of (117) has been *elided*, where such *ellipsis* is an operation that is allowed under identity with the VP of the first conjunct (the *antecedent* of the ellipsis). Although the exact nature of the relevant

[29]We would like to stress that this section only scratches the surface of the large and expanding body of literature on ellipsis and ACD; our purpose here is merely to point out the connection between scope and ellipsis resolution phenomena. We repeat that, for convenience, our discussion is phrased mostly from the perspective of a QR theory of quantifier scope, and an LF VP-copying theory of VP-ellipsis, but the problems raised by ACD exist independently of this approach and reoccur in various forms in other approaches.

identity constraint and the nature of the ellipsis operation are subject to debate, cases as in (118) create a problem for almost every approach.[30] As noted by [May, 1985] (see also [Bouton, 1970; Sag, 1976], the elided VP2 is contained in its antecedent VP1.[31] If ellipsis resolution involves copying the antecedent into the empty VP, the copying procedure never terminates, as illustrated in (119) (the infinite regress problem). If ellipsis involves deletion of an underlying full form, (118) would require an infinite underlying structure.

119. a. Dulles $[_{VP1}$ suspected $[_{NP}$ everyone who Angleton did $[_{VP2}$ e]]]

b. Dulles $[_{VP1}$ suspected $[_{NP}$ everyone who Angleton did $[_{VP1}$ suspected $[_{NP}$ everyone who Angleton did $[_{VP2}$ e]]]]]

c. Dulles $[_{VP1}$ suspected $[_{NP}$ everyone who Angleton did $[_{VP1}$ suspected $[_{NP}$ everyone who Angleton did $[_{VP1}$ suspected $[_{NP}$ everyone who Angleton did $[_{VP2}$ e]]]]]]]

May [1985] proposed that QR moves the NP containing VP2 out of VP1 (step 2); the resulting LF allows copying without regress (step 3):

120. a. Dulles $[_{VP1}$ suspected $[_{NP}$ everyone who Angleton did $[_{VP2}$ e]]]

b. $[_{NP}$ everyone who Angleton did $[_{VP2}$ e $]]_i$ $[_S$ Dulles $[_{VP1}$ suspected t_i]]

c. $[_{NP}$ everyone who Angleton did $[_{VP1}$ suspected $t_i]]_i$ $[_S$ Dulles $[_{VP1}$suspected t_i]]

This approach predicts, correctly, that the QNP containing the elided VP must scope out of the antecedent VP (Sag 1976; examples from Bruening 2001):

121. a. Ozzy wanted every book that Kate wrote

b. Ozzy wanted every book that Kate did $[_{VP}$ e]

While (121a) allows a *de dicto* reading (see section 3.1) for the object, (121b) does not.

[30]See e.g. [Williams, 1977; Sag, 1976; Vanden Wyngaerd and Zwart, 1991; Lasnik, 1993; Hornstein, 1994; Rooth, 1992; Fiengo and May, 1994; Heim, 1997; Merchant, 2001; Wilder, 2003]. See [Jacobson, 1992; 1996; 1998; Jäger, 2001; 2005] for views of ACD and VP ellipsis in a categorial approach.

[31]For the simple case (118), the problem might be circumvented by assuming it is only V, not VP, that has been elided; then the antecedent does not contain the ellipsis site. But this simple expedient does not resolve the ACD in more complicated cases such as (122) or (123a); a scope shifting rule does. For a sophisticated version of a V-ellipsis approach, see [Cormack, 1985; Jacobson, 1992].

The hypothesis that the operation that resolves ACD is also the one responsible for inverse scope predicts that ACD will be allowed just where scope inversion is allowed. Thus, for instance, ACD is allowed in inverse linking structures (Kennedy 1997):

122. John $[_{VP1}$ wrote $[_{NP}$ a report on $[_{NP}$ every student Peter did $[_{VP2}$ e $]]]]$

ACD resolution is blocked by a CNPC island (section 3.2), but this may be due to a corresponding CNPC island violation inside the ellipsis site. Better evidence that scope islands affect ACD comes from the following paradigms (from [Larson and May, 1990]; see also [Wilder, 2003]):

123. a. John $[_{VP1}$ believed $[_{S}$ $[_{NP}$ everyone you did $[_{VP2}$ e $]]$ to be a genius $]]$

b. John $[_{VP1}$ believed $[_{S}$ $[_{NP}$ everyone you *believed to be a genius* $]$ to be a genius $]]$

c. * John $[_{VP1}$ believed $[_{CP}$ (that) $[_{NP}$ everyone you did $[_{VP2}$ e $]]$ was a genius $]]$

When the elided VP is contained in the subject of a non-tensed subclause, as in (123a), the matrix VP can antecede the ellipsis: (123a) allows the paraphrase (123b). But ACD cannot be resolved in this manner when the ellipsis site is contained in the subject of a tensed subclause; hence the ill-formedness of (123c). This corresponds to the scope options for quantified NPs in these positions: the subject of a non-tensed clause easily scopes into the matrix clause, even higher than the matrix subject; but the subject of a tensed subclause does not, as illustrated in (124) (although intuitions differ).

124. a. someone believes $[_{S}$ everyone to be a genius $]$

b. someone believes $[_{CP}$ (that) everyone is a genius $]$

Observe, incidentally, that (122) and (123a) are examples where the antecedent for the ellipsis contains more material than a single V, in a way that renders more secure the diagnosis that, barring a scope shifting operation, the ellipsis is antecedent-contained (cf. footnote 31).

It is further predicted that NPs that are not subject to QR (or other covert movement operations) do not allow ACD; this is confirmed by (125), from Lasnik (1993) (the indicated NP is not quantificational, and receives Case in situ):

125. * Mary stood near [$_{NP}$ Susan, who Emily did [$_{VP}$ e] as well]

Finally, we expect that the correlation between scope and ACD resolution breaks down with NPs that are subject to a non-standard scope mechanism which does not "displace" the NP. This is confirmed by (126) (from [Kennedy, 1997]):

126. John [$_{VP1}$ believed that Bill [$_{VP2}$ had seen a certain film that I did [$_{VP3}$ e]]]

(126) has a *de re* reading with *a certain film that I did* taking wide scope relative to *believed*, but even on this matrix scope reading, only VP2 may antecede the empty VP. The example does not have a reading 'there is a certain film I believed that Bill had seen, that John believed that Bill had seen'; the absence of this reading follows if ACD resolution indeed requires LF movement of the NP, but exceptional wide scope for indefinite NPs is due to a different mechanism (see section 4.3.3).

6 CONCLUSIONS

In this paper we have tried to give a broad overview of QNP scope phenomena and some prominent approaches to their treatment. By way of conclusion, we would like to highlight three topics that have reoccurred in our review at various places and seem to us especially central.

Syntax vs. Semantics

We believe that there is little reason to prejudge scope phenomena as belonging to either syntax or semantics. Even though the primary data of inverse and non-linear scope readings are always semantic, the mechanisms that account for them may reasonably involve syntactic considerations and principles. The real challenge, we think, is to provide a theory of scope effects that makes the optimal division of labor between syntax and semantics, in terms of empirical coverage, conceptual clarity and technical soundness and elegance. As our overview above has clarified, this is by no means an easy challenge: more comprehensive solutions to this challenge are still to be found.

Scope effects as "movement"?

One of the major theoretical decisions that any theory of scope has to make is whether to treat inverse scope effects as a "movement" phenomenon. Are the mechanisms that are responsible for inverse scope relations also responsible for phenomena that involve "overt" extraction? A positive answer to this question, as most clearly given in QR theory, does not yet determine completely the description of a phenomenon like QNP scope in the grammar. However, any answer to this question has direct implications for the overall organization of the grammar. Specifically, a positive answer to this question leads to far-reaching challenges that emerge from the many discrepancies between scope effects and "movement" effects, as surveyed in section 4.2.1. Conversely, a decision to clearly dissociate scope effects from "movement" phenomena may require substantial justification for any proposed account of the former.

"QNP Scope" as an epiphenomenon

Pre-theoretically, and in the face of such a simple set of examples as we initially employed in section 2.2 in order to illustrate the incompleteness of the direct scope strategy in our toy grammar, one might have expected that the available scope options for quantified NPs might be described as a simple permutation of quantifiers, as illustrated in (50), and might be explained by postulating one syntactic or semantic rule (such as QR, or Storage), or one set of rules of a given type, which would be enough to derive the available options. Instead, as has become clear from several decades of research on this topic, and as we have attempted to illustrate in this article, the scope of quantified NPs is not a unified phenomenon, and it is unlikely that it is mediated by one component of the grammar, a dedicated "scope module". Various NP types "take scope" in different ways: sometimes their scope is mediated through a "movement-like" rule, sometimes through one of the non-standard mechanisms described in section 4.3.3. Unforeseen factors have often been found to influence the available scope options. This makes the study of scope phenomena all the more challenging, as it requires a non-trivial balance between descriptive accuracy and theoretical frugality and elegance.

The many questions surrounding the notion of QNP scope, and scope effects in general, leave much room for further research. We do believe, however, that more than forty years of extensive linguistic-logical research of scope phenomena also leave room for hope. The important theoretical and empirical advances that have been made and the unique collaboration that they have prompted between logicians and formal linguists promise to keep the study of scope phenomena an active area of research for years to come.

BIBLIOGRAPHY

[Abney, 1987] S. P. Abney. *The English Noun Phrase in its Sentential Aspect*. PhD thesis, Massachusetts Institute of Technology, 1987.

[Aoun and Li, 1989] J. Aoun and Y. A. Li. Scope and Constituency. *Linguistic Inquiry* 20-2:141-172, 1989.

[Abusch, 1994] D. Abusch. The scope of indefinites. *Natural Language Semantics*, 3:88.135, 1994.

[Ajdukiewicz, 1935] K. Ajdukiewicz. Die syntaktische konnexität. *Studia Philosophia*, 1:1.27, 1935.

[Altman *et al.*, 2005] A. Altman, Y. Peterzil, and Y. Winter. Scope dominance with upward monotone quantifiers. *Journal of Logic, Language and Information*, 14:445-455, 2005.

[Barker, 2002] C. Barker. Continuations and the nature of quantification. *Natural Language Semantics* 10:211-242, 2002.

[Barker and Jacobson, 2007] C. Barker and P. Jacobson. Introduction: Direct compositionality. In Barker, C. and Jacobson, P., eds., *Direct Compositionality*. Oxford University Press, Oxford, 2007.

[Barwise, 1979] J. Barwise. On branching quantfiers in English. *Journal of Philosophical Logic*, 8:47-80, 1979.

[Barwise and Cooper, 1981] J. Barwise and R. Cooper. Generalized Quantifiers and Natural Language. *Linguistics and Philosophy* 4, 159-219, 1981.

[Beck and Sauerland, 2001] S. Beck and U. Sauerland. Cumulation is needed: a reply to Winter (2000a). *Natural Language Semantics*, 8:349–371, 2001.

[Beghelli, 1993] F. Beghelli. A minimalist approach to Quantifier Scope. NELS 23, Vol 1, 65-80, 1993.

[Beghelli, 1995] F. Beghelli. The Phrase Structure of Quantifier Scope. PhD diss, UCLA, 1995.

[Beghelli *et al.*, 1997] F. Beghelli, D. Ben-Shalom, and A. Szabolcsi. Variation, distributivity, and the illusion of branching. In Szabolcsi, A., editor, *Ways of Scope Taking*. Kluwer, Dordrecht, 1997.

[Beghelli and Stowell, 1997] F. Beghelli and T. Stowell. Distributivity and Negation: the Syntax of *each* and *every*. In Szabolcsi, A., editor, *Ways of Scope Taking*. Kluwer, Dordrecht. 71-107, 1997.

[Ben-Avi and Winter, 2007] G. Ben-Avi and Y. Winter. The semantics of intensionalization. In *Proceedings of the Workshop on New Directions in Type-theoretic Grammars*, 2007.

[van Benthem, 1986] J. van Benthem. *Essays in Logical Semantics*. D. Reidel, Dordrecht, 1986.

[van Benthem, 1989] J. van Benthem. Polyadic quantifiers. *Linguistics and Philosophy*, 12:437-464, 1989.

[van Benthem, 1991] J. van Benthem. *Language in Action: categories, lambdas and dynamic logic*. North-Holland, Amsterdam, 1991.

[Bergmann, 1982] M. Bergmann. Cross-categorial semantics for conjoined common nouns. *Linguistics and Philosophy* 5:399-401, 1982.

[Bouton, 1970] L. Bouton. Antecedent-contained pro-forms. In: *Proceedings of CLS* 6, 154-167, 1970.

[Bruening, 2001] B. Bruening. QR obeys Superiority: Frozen Scope and ACD. LI 32-2 p. 233-273, 2001.

[Büring, 1997] D. Büring. The great scope inversion conspiracy. *Linguistics and Philosophy* 20, 175-194, 1997.

[Crlson, 1977] G. N. Carlson. *Reference to Kinds in English*. PhD thesis, University of Massachusetts at Amherst, 1977.

[Carpenter, 1997] B. Carpenter. *Type-Logical Semantics.* MIT Press, Cambridge, Massachusetts, 1997.

[Chierchia, 2001] G. Chierchia. A puzzle about indefinites. In C. Cecchetto, G. C. and Guasti, M. T., eds., *Semantic Interfaces: reference, anaphora and aspect.* CSLI Publications, Stanford, 2001.

[Chomsky, 1972] N. Chomsky. *Studies on semantics in generative grammar.* The Hague, Mouton, 1972.

[Chomsky, 1973] N. Chomsky. Conditions On Transformations. In *Festschrift for Morris Halle*, ed. Stephen Anderson & Paul R. Kiparsky, 232-286. New York: Holt, Rinehart & Winston 1973.

[Chomsky, 1975] N. Chomsky. Questions of Form and Interpretation. *Linguistic Analysis* 1-1, 75-109, 1975.

[Chomsky, 1976] N. Chomsky. Conditions on Rules of Grammar. *Linguistic Analysis* 2, 303–351, 1976.

[Chomsky, 1977] N. Chomsky. On Wh-Movement. In *Formal Syntax*, ed. Peter W. Culicover, Thomas Wasow and Adrian Akmajian, 71-132. San Francisco, London: Academic Press, 1977.

[Chomsky, 1980] N. Chomsky. *Rules and representations*. Oxford : Blackwell, 1980.

[Chomsky, 1981] N. Chomsky. *Lectures on Government and Binding.* Foris, Dordrecht, 1981.

[Chomsky, 1986a] N. Chomsky. *Barriers*. MIT Press, Cambridge, Mass, 1986.

[Chomsky, 1986b] N. Chomsky. *Knowledge of Language: Its Nature, Origin and Use.* New York: Praeger, 1986.

[Chomsky, 1993] N. Chomsky. A minimalist Program for Linguistic Theory. In *The view from Building 20: Essays in honor of Sylvain Bromberger*, ed. Kenneth Hale and Samuel Jay Keyser, 1-52. Cambridge, Mass.: MIT Press 1993.

[Chomsky, 1995] N. Chomsky. *The Minimalist Program.* Cambridge, Mass: MIT Press, 1995.

[Chomsky and Lasnik, 1993] N. Chomsky and H. Lasnik. The theory of principles and parameters. In J. Jacobs et al. (eds.) Syntax: An International Handbook of Contemporary Research, Vol. 1. Walter de Gruyter, pp. 506-569, 1993. (Reprinted in Chomsky 1995).

[Cooper, 1975] R. Cooper. *Montague's Semantic Theory and Transformational Syntax.* PhD thesis, University of Massachusetts at Amherst, 1975.

[Cooper, 1979] R. Cooper. Variable binding and Relative Clauses. In F. Guenthner & S.J. Schmidt (eds.), *Formal Semantics and Pragmatics for Natural Languages*, Reidel, Dordrecht, pp. 131-169, 1979.

[Cooper, 1983] R. Cooper. *Quantification and Syntactic Theory.* Reidel, Dordrecht, 1983.

[Cormack, 1985] A. Cormack. VP Anaphora: Variables and Scope. In F. Landman and F. Veltman (eds.), *Varieties of Formal Semantics.* Dordrecht: Foris, 1985.

[Cresti, 1995] D. Cresti. Extraction and Reconstruction. *Natural Language Semantics* 3: 79-122, 1995.

[Davies and Dubinsky, 2003] W. D. Davies and S. Dubinsky. On Extraction from NPs. *Natural Language & Linguistic Theory* 21: 1–37, 2003.

[de Groote, 2001] Ph. de Groote. Towards Abstract Categorial Grammars. In: *Association for Computational Linguistics, 39th Annual Meeting and 10th Conference of the European Chapter, Proceedings of the Conference*, pages 148–155, Toulouse, France, 2001.

[Dowty *et al.*, 1981] D. Dowty, R. Wall, and S. Peters. *Introduction to Montague Semantics.* D. Reidel, Dordrecht, 1981.

[Egli and von Heusinger, 1995] U. Egli and K. von Heusinger. The epsilon operator and e-type pronouns. In Egli, U., Pause, P. E., Schwarze, C., von Stechow, A., and Wienold, G., editors, *Lexical Knowledge in the Organization of Language*. John Benjamins, Amsterdam-Philadelphia, 1995.

[Engdahl, 1980] E. Engdahl. *The Syntax and Semantics of Questions in Swedish.* PhD Dissertation, University of Massachusetts at Amherst, 1980.

[Farkas, 1997] D. Farkas. Evaluation indices and scope. In Szabolcsi, A., editor, *Ways of Scope Taking*. Kluwer, Dordrecht, 1997.

[Fauconnier, 1975] G. Fauconnier. Do quantifiers branch? *Linguistic Inquiry*, 6:555-567, 195.

[Fiengo and Higginbotham, 1981] R. Fiengo and J. Higginbotham. Opacity in NP. *Linguistic Analysis* 7.4. pp. 395-421, 1981.

[Fiengo and May, 1994] R. Fiengo and R. May. *Indices and Identity.* Cambridge, Mass.: MIT Press, 1994.

[Fodor and Sag, 1982] J. D. Fodor and I. Sag. Referential and quantificational indefinites. *Linguistics and Philosophy*, 5:355-398, 1982.

[Fox, 1995] D. Fox. Economy and Scope. *Natural Language Semantics* 3:283-341, 1995.

[Fox, 1999] D. Fox. Reconstruction, Binding Theory, and the Interpretation of Chains, LI 30, p. 157-196, 1999.

[Fox, 2000] D. Fox. *Economy and semantic interpretation.* Cambridge, Massachusetts : MIT Press, 2000.

[Fox and Nissenbaum, 1999] D. Fox and J. Nissenbaum. Extraposition and Scope: A case for overt QR, Proceedings of the West Coast Conference on Formal Linguistics 18, 1999.

[Gabbay and Moravcsik, 1974] D. Gabbay and J. Moravcsik. Branching quantifiers, English and Montague Grammar. *Theoretical Linguistics*, 1.1:139-157 1974.

[Gallin, 1975] D. Gallin. *Intensional and Higher-Order Logic.* North-Holland Publishing Company, Amsterdam 1975.

[Gamut, 1991] L. T. F. Gamut. *Logic, Language, and Meaning*; Volume 2: Intensional Logic and Logical Grammar, 1991.

[Gil, 1982] D. Gil. Quantifier Scope, Linguistic Variation, and natural Language Semantics. *L&P* 5:421-472, 1982.

[Groenendijk and Stokhof, 1984] J. Groenendijk and M. Stokhof. *Studies on the Semantics of Questions and the Pragmatics of Answers.* PhD Thesis, Universiteit van Amsterdam, 1984.

[Groenendijk and Stokhof, 1997] J. Groenendijk and M. Stokhof. Questions. In: J. van Benthem & A. ter Meulen (eds), *Handbook of Logic and Language*, Amsterdam/Cambridge, Mass., Elsevier/MIT Pess, pp. 1055-1124. 1997.

[Heim, 1997] I. Heim. Predicates or Formulas? Evidence from Ellipsis. In A Lawson & E. Cho (eds), *Proceedings of Semantics and Linguistic Theory VII*, p. 197–221. Ithaca, NY: CLC publications, 1997.

[Heim and Kratzer, 1998] I. Heim and A. Kratzer. *Semantics in generative grammar.* Oxford, Blackwell, 1998.

[Hendriks, 1993] H. Hendriks. *Studied Flexibility: categories and types in syntax and semantics.* PhD thesis, University of Amsterdam. 1993.

[Henkin, 1961] L. Henkin. Some remarks on infinitely long formulas. In *Infinitistic Methods, Proceedings of the Symposium on Foundations of Mathematics*, 1959. Pergamon Press, Warsaw, 1961.

[Heycock and Zamparelli, 2005] C. Heycock and R. Zamparelli. Friends and colleagues: plurality, coordintion, and the structure of DP. *Natural Language Semantics*, 13:201-270, 2005.

[Hintikka, 1973] J. Hintikka. Quantifiers vs. quantification theory. *Dialectica*, 27:329-358. Reprinted in *Linguistic Inquiry* 5 (1974):153-177, 1973.

[Hoji, 1985] H. Hoji. *Logical form constraints and configurational structures in Japanese.* Diss, U. of Washington, 1985.

[Hornstein, 1995] N. Hornstein. *Logical Form, from GB to Minimalism.* Blackwell 1995.

[Huang, 1982] C.-T. J. Huang. *Logical Relations in Chinese and the Theory of Grammar.* Diss. MIT, 1982.

[Ioup, 1975] G. L. Ioup. *The Treatment of Quantifier Scope in a Transformational Grammar.* Diss. City U. of New York, 1975.

[Jackendoff, 1972] R. Jackendoff. *Semantic Interpretation in Generative Grammar.* Cambridge, Mass.: MIT Press, 1972.

[Jacobson, 1992] P. Jacobson. Antecedent Contained Deletion in a Variable Free Semantics, in: *Proceedings of the Second Conference on Semantics and Linguistic Theory*, Ohio State University. pp. 193-213, 1992.

[Jacobson, 1996] P. Jacobson. The Syntax/Semantics Interface in Categorial Grammar, In *The Handbook Of Contemporary Semantic Theory*, Lappin, Shalom (Ed), Oxford, England: Blackwell, 1996, pp 89-116, 1996.

[Jacobson, 1998] P. Jacobson. Antecedent contained deletion and pied-piping: Evidence for a variable free semantics. In *Proceedings of SALT 8*, ed. by Devon Strolovitch and Aaron Lawson, 74–91. Ithaca, N.Y.: Cornell University, CLC Publications, 1998.

[Jacobson, 1999] P. Jacobson. Towards a variable-free semantics. *Linguistics and Philosophy*, 22:117-185, 1999.

[Jacobson, 2002] P. Jacobson. The (dis)organization of grammar: 25 years. *Linguistics and Philosophy*, 22:601-626, 2002.

[Jäger, 2001] G. Jäger. Anaphora and quantification in Categorial Grammar. In: M. Moortgat (ed.), *Logical Aspects of Computational Linguistics*, Springer Lecture Notes in Artificial Intelligence 2014, Springer, pp. 70-90, 2001.

[Jäger, 2005] G. Jäger. *Anaphora and Type Logical Grammar*, Springer, Dordrecht, 2005.

[Janssen, 1983] T. M. V. Janssen. *Foundations and Applications of Montague Grammar.* PhD thesis, Mathematisch Centrum, Amsterdam, 1983.

[Janssen, 1996] T. M. V. Janssen. Compositionality. In van Benthem, J. and ter Meulen, A., editors, *Handbook of Logic and Language.* Elsevier, Amsterdam, 1996.

[Horn, 1989] L. R. Horn. *A Natural History of Negation.* University of Chicago Press, Chicago, 1989.

[Karttunen and Peters, 1980] L. Karttunen and S. Peters. Interrogative Quantifiers. In: C. Rohrer (ed.), *Time, Tense and Quantifiers.* Tübingen: Max Niemeyer Verlag.181-205, 1980.

[Kempson, 1977] R. M. Kempson. *Semantic Theory.* Cambridge University Press: Cambridge, 1977.

[Kempson and Cormack, 1981] R. M. Kempson and A. Cormack. Ambiguity and quantification. *Linguistics and Philosophy*, 4:259-309, 1981.

[Keenan, 1992] E. Keenan. Beyond the Frege boundary. *Linguistics and Philosophy*, 15:199-221, 1992.

[Keenan and Faltz, 1985] E. Keenan and L. Faltz. *Boolean Semantics for Natural Language.* Reidel: Dordrecht, 1985.

[Keenan and Stavi, 1986] E. Keenan and J. Stavi. A semantic characterization of natural language determiners. *Linguistics and Philosophy*, 9:253-326 1986.

[Kennedy, 1997] C. Kennedy. Antecedent-contained deletion and the syntax of quantification. *Linguistic Inquiry* 28: 662-688, 1997.

[Kiss, 1991] K. É. Kiss. Logical Structure in Syntactic Structure: The Case of Hungarian. In: *Logical Structure and Linguistic Structure: Cross-Linguistic Perspectives.* C.-T. J. Huang and R. May (eds.), 123–148. Dordrecht: Reidel, 1991.

[Kiss, 2006] K. É. Kiss. Quantifier Scope Ambiguities. In: *The Blackwell Companion to Syntax.* Everaert, M. and H. van Riemsdijk (eds). Blackwell, 2006.

[Kratzer, 1998] A. Kratzer. Scope or pseudoscope? Are there wide scope indefinites? In Rothstein, S., editor, *Events and Grammar*. Kluwer, Dordrecht, 1998.
[Krifka, 1989] M. Krifka. Nominal reference, temporal constitution and quantification in event semantics. In Bartsch et al. (editors), *Semantics and Contextual Expression.* Foris: Dordrecht. 1989.
[Kurtzman and MacDonald, 1993] H. S. Kurtzman and M. C. MacDonald. Resolution of quantifier scope ambiguities. *Cognition*, 48, 243-279, 1993.
[Lakoff, 1970] G. Lakoff. Repartee. *Foundations of Language* 6, 389-422 1970.
[Lambek, 1958] J. Lambek. The mathematics of sentence structure. *American Mathematical Monthly*, 65:154-169, 1958.
[Landman, 2000] F. Landman. *Events and Plurality: the Jerusalem lectures.* Kluwer, Dordrecht, 2000.
[Larson, 1985a] R. K. Larson. On the syntax of disjunction scope. *Natural Language and Linguistic Theory*, 3:217-265, 1985.
[Larson, 1985b] R. K. Larson. *Quantifying into NP*. ms, 1985.
[Larson and May, 1990] R. K. Larson and R. May. Antecedent containment or vacuous movement: Reply to Baltin. *Linguistic Inquiry* 21:103-122, 1990.
[Larson, 2003] R. K. Larson. Sentence-final adverbs and "scope". In M. Wolf and K. Moulton (eds.) *Proceedings of NELS 34*. UMass: GLSA, 2003
[Lasnik, 1993] H. Lasnik. *Lectures on Minimalist Syntax.* University of Connecticut Working Papers in Linguistics, Occasional Papers Issue 1. Reprinted in: H. Lasnik. 1999. *Minimalist Analysis.* Malden, Ma: Blackwell, 1993.
[Lasnik and Kupin, 1977] H. Lasnik and J. Kupin. A restrictive theory of transformational grammar. *Theoretical Linguistics* 4:173-196, 1977.
[Liu, 1990] F.-H. Liu. *Scope Dependency in English and Chinese.* Diss, UCLA, 1990.
[Longobardi, 1987] G. Longobardi. Extraction from NP and the Proper Notion of Head Government. In: Alessandra Giorgi and Giuseppe Longobardi (eds.), *The Syntax of Noun Phrases*, Cambridge University Press, Cambridge, pp. 57-112, 1987.
[May, 1977] R. May. *The Grammar of Quantification.* PhD Diss., MIT, Cambridge, Mass. Repr. by IULC, 1977.
[May, 1985] R. May. *Logical Form, Its Structure and Derivation.* Cambridge, Mass.: MIT Press, 1985.
[May, 1989] R. May. Interpreting Logical Form. *Linguistics and Philosophy* 12, 387-435, 1989.
[May and Bale, 2006] R. May and A. Bale. Inverse Linking. In: M. Everaert & H. van Riemsdijk (eds), The Blackwell Companion to Syntax, Vol. II, 2006.
[Merchant, 2001] J. Merchant. *The syntax of silence: Sluicing, islands, and the theory of ellipsis.* Oxford University Press: Oxford, 2001.
[Milsark, 1974] G. L. Milsark. *Existential Sentences in English.* PhD Diss MIT, 1974.
[Montague, 1970] R. Montague. Universal grammar. *Theoria*, 36:373-398. Reprinted in R. Thomason, editor (1974), *Formal Philosophy: selected papers of Richard Montague*, Yale, New Haven, 1970.
[Montague, 1973] R. Montague. The Proper Treatment of Quantification in Ordinary English. In: *Approaches to Natural Language*, eds. J. Hintikka, J. Moravcsik & P. Suppes. Dordrecht: Reidel, 1973.
[Moortgat, 1997] M. Moortgat. Categorial type logics. In van Benthem, J. and ter Meulen, A., editors, *Handbook of Logic and Language*. Elsevier Science, Amsterdam, 1997.
[Muskines, 2003] R. Muskens. Language, Lambdas, and Logic. In Geert-Jan Kruijff and Richard Oehrle, editors, *Resource Sensitivity in Binding and Anaphora*, Studies in Linguistics and Philosophy, pages 23-54. Kluwer, 2003.
[Oehrle, 1994] R. T. Oehrle. Term-labeled categorial type systems. *Linguistics & Philosophy*, 17:633-678, 1994.

[Pafel, 1994] J. Pafel. Scope and word order. In J. Jacobs et. Al, (eds) *Syntax: an International Handbook of Contemporary Research.* De Gruyter, Berlin. 867-888, 1994.

[Partee and Rooth, 1983] B. Partee and M. Rooth. Generalized conjunction and type ambiguity. In Bauerle, R., Schwarze, C., and von Stechow, A., editors, *Meaning, Use and Interpretation of Language.* De Gruyter, Berlin, 1983.

[Postal, 1974] M. Postal. On Certain Ambiguities. *Linguistic Inquiry* 5-3, 367-424 1974.

[Quine, 1956] W. Quine. Quantifiers and propositional attitudes. *Journal of Philosophy* 53, 177-187, 1956.

[Reinhart, 1976] T. Reinhart. *The Syntactic Domain of Anaphora.* PhD. Diss., MIT, Cambridge, Mass, 1976.

[Reinhart, 1983] T. Reinhart. *Anaphora and Semantic Interpretation.* Croom Helm, London & Sydney, 1983.

[Reinhart, 1991] T. Reinhart. Non-Quantificational LF. In *The Chomskyan Turn*, ed. A. Kasher. Oxford: Blackwell, 1991.

[Reinhart, 1992] T. Reinhart. Wh-in-situ: an apparent paradox. In *Proceedings of the 8th Amsterdam Colloquium.* University of Amsterdam, Institute for Logic, Language and Computation, 1992.

[Reinhart, 1997] T. Reinhart. Quantifier Scope: How Labour is Divided between QR and Choice Functions, *Linguistics and Philosophy* 20, 335-397, 1997.

[Reinhart, 2006a] T. Reinhart. *Interface strategies: optimal and costly computations.* Cambridge, Massachusetts, MIT Press, 2006.

[Reinhart, 2006n] T. Reinhart. Scope shift with numeral indefinites — syntax or processing? In L. Tasmowski and S. Vogeleer (eds), *Non-definiteness and Plurality*, John Benjamins, Amsterdam/Philadelphia, pp. 291-311, 2006.

[Rodman, 1976] R. Rodman. Scope Phenomena, "Movement Transformations," and Relative Clauses. In: Barbara H. Partee (ed.), *Montague Grammar.* Academic Press, New York, 165-176, 1976.

[Rooth, 1992] M. Rooth. Ellipsis Redundancy and Reduction Redundancy, in: Steve Berman and Arild Hestvik (eds.) Proceedings of the Stuttgart Ellipsis Workshop, Sprachtheoretische Grundlagen für die Computerlinguistik, no. 29, SFB 340, IBM Heidelberg, 1992.

[Ross, 1967] J. R. Ross. *Constraints on Variables in Syntax.* Diss. MIT. Distr. by IULC. Reprinted as: *Infinite Syntax!.* Ablex, Norwood NJ, 1986.

[Ruys, 1992] E. G. Ruys. *The Scope of Indefinites.* Diss. Utrecht University. Utrecht: LEd, 1992.

[Ruys, 2006] E. G. Ruys. Unexpected Wide-Scope Phenomena. In Martin Everaert & Henk van Riemsdijk (eds), *The Blackwell Companion to Syntax; Volume V.* p. 175-228. Blackwell, Malden, 2006.

[Sag, 1976] I. Sag. *Deletion and Logical Form.* PhD. Diss., MIT, Cambridge, Mass, 1976.

[Scha, 1981] R. Scha. Distributive, collective and cumulative quantification. In Groenendijk, J., Stokhof, M., and Janssen, T. M. V., editors, *Formal Methods in the Study of Language.* Mathematisch Centrum, Amsterdam, 1981.

[Schäfer, 2004] M. Schäfer. Manner Adverbs and Scope. In: Steube, A. (Hrsg.). *Grammatik und Kontext: Zur Interaktion von Syntax, Semantik und Prosodie bei der Informationsstrukturierung.* Linguistische Arbeitsberichte 81, Universität Leipzig:33-51, 2004.

[Schein, 1993] B. Schein. *Plurals and Events.* MIT Press, Cambridge, Massachusetts 1993.

[Schlenker, 2006] P. Schlenker. Scopal independence: a note on branching and island-escaping readings of indefinites and disjunctions. *Journal of Semantics*, 23:281-314, 2006.

[Schwarz, 1999] B. Schwarz. On the syntax of either...or. *Natural Language and Linguistic Theory*, 17:339-370 1999.

[Schwarzschild, 2002] R. Schwarzschild. Singleton indefinites. *Journal of Semantics*, 19:289-314, 2002.

[Sher, 1991] G. Sher. *The Bounds of Logic: a generalized viewpoint.* MIT Press, Cambridge, Massachusetts, 1991.

[Sportiche, 2006] D. Sportiche. Reconstruction, Binding, and Scope. In: M. Everaert & H. van Riemsdijk (eds), *The Blackwell Companion to Syntax*, Vol. IV, 2006.

[Sternefeld, 1997] W. Sternefeld. Reciprocity and cumulative predication. In Hamm, F. and Hinrichs, E., editors, *Plurality and Quantification.* Kluwer, Dordrecht, 1997.

[Szabolcsi, 1997] A. Szabolcsi, ed. *Ways of Scope Taking.* Dordrecht, Kluwer, 1997.

[Vande Wyngaerd and Zwart, 1991] G. Vanden Wyngaerd and C. J.-W. Zwart. Reconstruction and vehicle change. In *Linguistics in the Netherlands 1991*, ed. Frank Drijkoningen and Ans van Kemenade, 151-160. Amsterdam: John Benjamins, 1991.

[Wilder, 2003] C. Wilder. Antecedent-Containment and Ellipsis, In *The Interfaces: Deriving And Interpreting Omitted Structures*, Schwabe, Kerstin, & Winkler, Susanne (Eds), Amsterdam: John Benjamins, 2003, pp 79-119, 2003.

[Westerstähl, 1987] D. Westerstähl. Branching generalized quantifiers and natural language. In Gärdenfors, P., editor, *Generalized Quantifiers.* D. Reidel, Dordrecht, 1987.

[Williams, 1977] E. Williams. Discourse and Logical Form. *Linguistic Inquiry* 8-1, 101-139, 1977.

[Winter, 1997] Y. Winter. Choice functions and the scopal semantics of indefinites. *Linguistics and Philosophy*, 20:399-467. 1997.

[Winter, 2000a] Y. Winter. Distributivity and dependency. *Natural Language Semantics*, 8:27-69, 2000.

[Winter, 2000b] Y. Winter. On some scopal asymmetries of coordination. In *Proceedings of the KNAW conference on Interface Strategies.* 2000.

[Winter, 2001] Y. Winter. *Flexibility Principles in Boolean Semantics: coordination, plurality and scope in natural language.* MIT Press, Cambridge, Massachusetts 2001.

[Winter, 2004] Y. Winter. Functional quantification. *Research on Language and Computation*, 2:331-363, 2004.

[Zimmerman, 1993] T. E. Zimmermann. On the proper treatment of opacity in certain verbs. *Natural Language Semantics*, 1:149-179.

ARNON AVRON AND ANNA ZAMANSKY

NON-DETERMINISTIC SEMANTICS FOR LOGICAL SYSTEMS

1 INTRODUCTION

1.1 The Key Idea

The principle of truth-functionality (or compositionality) is a basic principle in many-valued logic in general, and in classical logic in particular. According to this principle, the truth-value of a complex formula is uniquely determined by the truth-values of its subformulas. However, real-world information is inescapably incomplete, uncertain, vague, imprecise or inconsistent, and these phenomena are in an obvious conflict with the principle of truth-functionality. One possible solution to this problem is to relax this principle by borrowing from automata and computability theory the idea of non-deterministic computations, and apply it in evaluations of truth-values of formulas. This leads to the introduction of *non-deterministic matrices* (Nmatrices) — a natural generalization of ordinary multi-valued matrices, in which the truth-value of a complex formula can be chosen *non-deterministically* out of some non-empty set of options. There are many natural motivations for introducing non-determinism into the truth-tables of logical connectives. We discuss some of them below. They give rise to two different ways in which non-determinism can be incorporated: the *dynamic* and the *static*[1]. In both the value $v(\diamond(\psi_1, \dots, \psi_n))$ assigned to the formula $\diamond(\psi_1, ..., \psi_n)$ is selected from a set $\widetilde{\diamond}(v(\psi_1), \dots, v(\psi_n))$ (where $\widetilde{\diamond}$ is the interpretation of $\diamond$). In the dynamic approach this selection is made separately and independently for each tuple $\langle\psi_1, \dots, \psi_n\rangle$. Thus the choice of one of the possible values is made at the lowest possible (local) level of computation, or on-line, and $v(\psi_1), \dots, v(\psi_n)$ do not uniquely determine $v(\diamond(\psi_1, \dots, \psi_n))$. In contrast, in the static semantics this choice is made globally, system-wide, and the interpretation of $\diamond$ is a function, which is selected before any computation begins. This function is a "determinisation" of the non-deterministic interpretation $\widetilde{\diamond}$, to be applied in computing the value of any formula under the given valuation. This limits non-determinism, but still leaves the freedom of choosing the above function among all those that are compatible with the non-deterministic interpretation $\widetilde{\diamond}$ of $\diamond$.

[1]The dynamic approach was introduced together with the concept of Nmatrices. The static approach was later introduced in [Avron and Konikowska, 2005]

D. Gabbay and F. Guenthner (eds.),
Handbook of Philosophical Logic, Volume 16, 227–304.
DOI 10.1007/978-94-007-0479-4_4,

1.2 Some Intuitive Motivations

We start by presenting some cases in which the need for non-deterministic semantics naturally arises.

Syntactic "underspecification":
Consider the standard Gentzen-type system *LK* for propositional classical logic (see e.g. [Troelstra and Schwichtenberg, 2000]). Its introduction rules for $\neg$ and $\vee$ are usually formulated as follows:

$$\frac{\Gamma \Rightarrow \Delta, \psi}{\Gamma, \neg\psi \Rightarrow \Delta}\ (\neg \Rightarrow) \qquad\qquad \frac{\Gamma, \psi \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \neg\psi}\ (\Rightarrow \neg)$$

$$\frac{\Gamma, \psi \Rightarrow \Delta \quad \Gamma, \varphi \Rightarrow \Delta}{\Gamma, \psi \vee \varphi \Rightarrow \Delta}\ (\vee \Rightarrow) \quad \frac{\Gamma \Rightarrow \Delta, \psi, \varphi}{\Gamma \Rightarrow \Delta, \psi \vee \varphi}\ (\Rightarrow \vee)$$

The corresponding semantics is given by the following classical truth-tables:

	$\neg$
t	**f**
f	**t**

		$\vee$
t	**t**	**t**
t	**f**	**t**
f	**f**	**t**
f	**f**	**f**

Note that each syntactic rule of LK dictates some semantic condition on the connective it introduces: $(\neg \Rightarrow)$ corresponds to the condition $\tilde{\neg}(t) = f$, while $(\Rightarrow \neg)$ corresponds to the condition $\tilde{\neg}(f) = t$, thus completely determining the truth-table for negation. Similarly, $(\vee \Rightarrow)$ dictates the last line of the truth-table for $\vee$, i.e $\tilde{\vee}(f, f) = f$, while $(\Rightarrow \vee)$ dictates the other three lines. Now suppose we want to reject the *law of excluded middle* (LEM), in the spirit of intuitionistic logic. This can most simply be done by discarding the rule $(\Rightarrow \neg)$, which corresponds to LEM, while keeping the rest of the rules unchanged. What is the semantics of the resulting system? Intuitively, by discarding $(\Rightarrow \neg)$, we lose the information concerning the second line of the truth-table for $\neg$. Accordingly, we are left with a problem of *underspecification.* This can be modelled using Nmatrices in a very natural way: in case of underspecification, all possible truth-values are allowed. The corresponding semantics in the case we consider would be as follows (we use sets of possible truth-values instead of truth-values):

	$\neg$
t	{**f**}
f	{**t**,**f**}

		$\vee$
t	**t**	{**t**}
t	**f**	{**t**}
f	**f**	{**t**}
f	**f**	{**f**}

Linguistic ambiguity:
In many natural languages the meaning of the words "either ... or" is ambiguous. Thus the Oxford English Dictionary explains the meaning of this phrase as follows:

> The primary function of either, etc., is to emphasize the indifference of the two (or more) things or courses, ..., but a secondary function is to emphasize the mutual exclusiveness (i.e. either of the two, but not both).

Following this kind of common-sense intuition about "or", it follows that in many natural languages the word "or" has both an "inclusive" and an "exclusive" sense. For instance, when some mathematician promises: *"I shall either attack problem A or attack problem B"*, then in many cases he might at the end solve the two problems, but there are certainly situations in which what he means is "but do not expect me to attack them both". In the first case the meaning of "or" is inclusive, while in the latter case it is exclusive. Now in many cases one is uncertain whether the meaning of a speaker's "or" is inclusive or exclusive. However, even in cases like this one would still like to be able to make some certain inferences from what has been said. This situation can be captured by dynamic semantics based on the following non-deterministic truth-table for $\vee$:

		$\vee$
t	**t**	{**t**, **f**}
t	**f**	{**t**}
f	**t**	{**t**}
f	**f**	{**f**}

Note that the static semantics is less appropriate here, since the meaning of a speaker's "or" is not predetermined, and he might use both meanings of "or" in two different sentences within the same discourse.

Inherent non-deterministic behavior of circuits:
Nmatrices can be applied to model non-deterministic behavior of various

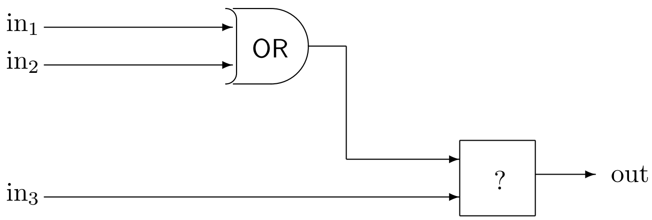


Figure 1. The circuit C

elements of electrical circuits. An ideal logic gate performing operations on boolean variables is an abstraction of a physical gate operating with a continuous range of electrical quantity. This electrical quantity is turned into a discrete variable by associating a whole range of electrical voltages with the logical values 1 and 0 (see [Rabaey et. al, 2003] for further details). There are a number of reasons, due to which the measured behavior of a circuit may deviate from the expected behavior. One reason can be the variations in the manufacturing process: the dimension and device parameters may vary, affecting the electrical behavior of the circuit. The presence of disturbing noise sources, temperature and other conditions are another source of deviations in the circuit response. The exact mathematical form of the relation between input and output in a given logical gate is not always known, and so it can be approximated by a non-deterministic truth-table. For instance, suppose that the circuit C given in Figure 1 consists of a standard OR gate and a faulty AND gate, which responds correctly if the inputs are similar, and unpredictably otherwise. The behavior of the gate can be described by the following truth-table, equipped with the dynamic semantics:

		AND
$\mathbf{t}$	$\mathbf{t}$	$\{\mathbf{t}\}$
$\mathbf{t}$	$\mathbf{f}$	$\{\mathbf{f}, \mathbf{t}\}$
$\mathbf{f}$	$\mathbf{t}$	$\{\mathbf{f}, \mathbf{t}\}$
$\mathbf{f}$	$\mathbf{f}$	$\{\mathbf{f}\}$

Computation with unknown functions:
Let us return to Figure 1, and suppose that this time it represents a circuit about which only some partial information is known. Namely, it is known that the gate labelled with "?" is either an XOR gate or an OR gate, but it is not known which one. Thus the function describing the second gate

is deterministic, but unknown to us. This situation can be represented by using the non-deterministic truth-table for $\vee$ given in the "linguistic ambiguity" example, equipped with the static semantics.

Verification with unknown evaluation models:
There are two well-known three-valued logics for describing different types of computational models. The first, which captures parallel evaluation, was described in the context of computational mathematics by Kleene ([Kleene, 1938]); the second, programming oriented method, in which evaluation proceeds sequentially, was proposed by McCarthy ([McCarthy, 1963]). Below are the corresponding truth-tables for $\vee$:

(Kleene)

$\tilde{\vee}$	$\mathbf{f}$	$\mathbf{e}$	$\mathbf{t}$
$\mathbf{f}$	$\mathbf{f}$	$\mathbf{e}$	$\mathbf{t}$
$\mathbf{e}$	$\mathbf{e}$	$\mathbf{e}$	$\mathbf{t}$
$\mathbf{t}$	$\mathbf{t}$	$\mathbf{t}$	$\mathbf{t}$

(McCarthy)

$\tilde{\vee}$	$\mathbf{f}$	$\mathbf{e}$	$\mathbf{t}$
$\mathbf{f}$	$\mathbf{f}$	$\mathbf{e}$	$\mathbf{t}$
$\mathbf{e}$	$\mathbf{e}$	$\mathbf{e}$	$\mathbf{e}$
$\mathbf{t}$	$\mathbf{t}$	$\mathbf{t}$	$\mathbf{t}$

Now suppose we are sending an expression $\psi \vee \varphi$ for evaluation to some distant computer, for which it is not known whether it performs parallel or sequential computations. Hence we know that $\psi \vee \varphi$ will be evaluated using a deterministic function $\tilde{\vee}$, defined by either Kleene's or McCarthy's truth-table for $\vee$, but we have no information which of the two. Again this can be captured by using a static interpretation of the following "truth-table":

$\tilde{\vee}$	$\mathbf{f}$	$\mathbf{e}$	$\mathbf{t}$
$\mathbf{f}$	$\{\mathbf{f}\}$	$\{\mathbf{e}\}$	$\{\mathbf{t}\}$
$\mathbf{e}$	$\{\mathbf{e}\}$	$\{\mathbf{e}\}$	$\{\mathbf{e}, \mathbf{t}\}$
$\mathbf{t}$	$\{\mathbf{t}\}$	$\{\mathbf{t}\}$	$\{\mathbf{t}\}$

According to this static interpretation, the function $f_\vee : \{\mathbf{t}, \mathbf{f}, \mathbf{e}\}^2 \to \{\mathbf{t}, \mathbf{f}, \mathbf{e}\}$ used by the computer satisfies either $f_\vee(\mathbf{t}, \mathbf{e}) = \mathbf{t}$ (in case the computation is parallel) or $f_\vee(\mathbf{t}, \mathbf{e}) = \mathbf{e}$ (in case it is sequential). However, it is not known which of these two conditions is satisfied.

Incompleteness and inconsistency:
This example is taken from [Avron et. al., 2006; Avron et. al., 2008]. Suppose we have a framework for information collecting and processing, which consists of a set S of information sources and a processor P. The sources provide information about formulas over $\{\neg, \vee\}$, and we assume that

for each such formula ψ a source $s \in S$ can say that ψ is true (i.e., assigned the truth-value 1), ψ is false (i.e., assigned the truth-value 0), or that it has no knowledge about ψ. In turn, the processor collects information from the sources, combines it according to some strategy and defines the resulting combined valuation of formulas. Thus for every formula ψ the processor can encounter one of the four possible situations: (a) it has information that ψ is true, but no information that ψ is false, (b) it has information that ψ is false, but no information that ψ is true, (c) it has both information that ψ is true and information that it is false, and (d) it has no information on ψ at all. In view of this, it was suggested by Belnap in [Belnap, 1977] (following works and ideas of Dunn, e.g. [Dunn, 1976]) to account for incomplete and contradictory information by using the following four logical truth values:

$$\mathbf{t} = \{1\}, \mathbf{f} = \{0\}, \top = \{0,1\}, \bot = \emptyset$$

Here 1 and 0 represent "true" and "false" respectively, and so $\top$ represents inconsistent information, while $\bot$ represents absence of information.

The above scenario has many ramifications, corresponding to various assumptions regarding the kind of information provided by the sources and the strategy used by the processor to combine it. We assume that the processor respects at least the deterministic consequences (in both ways) of each of the classical truth tables. This assumption means that the values assigned by the processor to complex formulas and those it assigns to their immediate subformulas are interrelated according to the following principles derived from the classical truth-tables of $\neg$ and $\vee$:

1. The processor ascribes 1 to $\neg\varphi$ iff it ascribes 0 to φ.
2. The processor ascribes 0 to $\neg\varphi$ iff it ascribes 1 to φ.
3. If the processor ascribes 1 to either φ or ψ, then it ascribes 1 to $\varphi \vee \psi$.
4. The processor ascribes 0 to $\varphi \vee \psi$ iff it ascribes 0 to both φ and ψ.

Here the statement "the processor ascribes 0 to ψ" means that 0 is included in the subset of $\{0,1\}$ which is assigned by the processor to ψ (recall that the truth-values used by the processor correspond to subsets of $\{0,1\}$). It is crucial to note that the converse of (3) does *not* hold, since some source might inform the processor that $\varphi \vee \psi$ is true, without providing information about the truth/falsehood of either φ or ψ. Under the above assumptions, there can be a number of possible scenarios concerning the type of formulas evaluated by the sources. The case when the sources provide information only about atomic formulas has been considered in [Belnap, 1977]. This case is deterministic, and leads to the famous Dunn-Belnap four-valued logic. Now consider the case when the sources provide information about arbitrary formulas (also complex ones), but not necessarily all of them. In this case the assumptions above are reflected in the following *non-deterministic* truth-tables:

$\widetilde{\vee}$	$\mathbf{f}$	$\bot$	$\top$	$\mathbf{t}$
$\mathbf{f}$	$\{\mathbf{f}, \top\}$	$\{\mathbf{t}, \bot\}$	$\{\top\}$	$\{\mathbf{t}\}$
$\bot$	$\{\mathbf{t}, \bot\}$	$\{\mathbf{t}, \bot\}$	$\{\mathbf{t}\}$	$\{\mathbf{t}\}$
$\top$	$\{\top\}$	$\{\mathbf{t}\}$	$\{\top\}$	$\{\mathbf{t}\}$
$\mathbf{t}$	$\{\mathbf{t}\}$	$\{\mathbf{t}\}$	$\{\mathbf{t}\}$	$\{\mathbf{t}\}$

	$\widetilde{\neg}$
$\mathbf{f}$	$\{\mathbf{f}\}$
$\bot$	$\{\bot\}$
$\top$	$\{\top\}$
$\mathbf{t}$	$\{\mathbf{f}\}$

Note that the table for negation reflects the principles 1 and 2, while the table for disjunction reflects the principles 3 and 4. To see this, let us examine one of the most peculiar cases: the entry $\mathbf{f}\widetilde{\vee}\mathbf{f} = \{\mathbf{f}, \top\}$. Suppose that ψ and φ are both assigned the truth-value $\mathbf{f} = \{0\}$. Then by principle 4 above, the truth-value of $\psi \vee \varphi$ (which is a subset of $\{0, 1\}$) must include 0. If in addition one of the sources assigned 1 to $\psi \vee \varphi$, then the processor ascribes 1 to $\psi \vee \varphi$ too, and so the truth-value it assigned to $\psi \vee \varphi$ is in this case $\top$. Otherwise it is $\mathbf{f}$. This justifies the two options in the truth-table. The rest of the entries can be explained in a similar way.

1.3 Things To Come

The rest of this survey is divided into two parts. Part I describes the propositional framework of Nmatrices. We begin with some preliminaries and a review of many-valued matrices in Section 2. The basic definitions of the framework of Nmatrices are presented in Section 3. In Section 4 we introduce *canonical signed calculi*, a natural family of proof systems manipulating sets of signed formulae (Gentzen-type systems can be thought of as a specific instance of such calculi). The relation between Nmatrices and canonical calculi is then explored in two complementary directions. In Section 4.1 we provide a general proof theory for Nmatrices using canonical calculi. In Section 4.2 modular non-deterministic semantics is provided for every canonical calculus (satisfying a simple syntactic condition). We then proceed to describe further applications of Nmatrices. In Section 5 we extend the modular approach to two non-canonical families of Gentzen-type calculi: those that are obtained from the positive fragments of classical logic and intuitionistic logic by adding various natural Gentzen-type rules for negation. In Section 6 Nmatrices are used for yet another family of non-classical logics: *paraconsistent logics*, designed for reasoning in the presence of contradictions. In Part II we handle the extension of the framework of Nmatrices to the first-order level and beyond. In Section 7 we briefly review the two standard approaches to interpreting unary quantifiers in many-valued logics. In Section 8 we extend the propositional framework of Nmatrices to languages with such quantifiers and discuss the problems that this move reveals (and were not evident on the propositional level). Section 9 is devoted to the particular case of the usual first-order quantifiers. An application of this case is presented in Section 10, where we extend the results from Section 6, and provide semantics for a large family of first-order paraconsistent logics. Section 11 further generalizes the framework of Nmatrices to *multi-ary quantifiers* and extends the relation between Nmatrices and canonical signed calculi to languages with such quantifiers.

Due to lack of space, we omit in what follows most of the proofs, providing instead pointers to the relevant papers. Those of the proofs we do include are intended to give the reader a better insight into the nature of Nmatrices, and a flavour of the (mostly new) methods that can be employed in handling and applying them.

PART I: THE PROPOSITIONAL CASE

2 PRELIMINARIES

In what follows, $\mathcal{L}$ is a propositional language and $Frm_{\mathcal{L}}$ is its set of wffs. The metavariables ψ, φ range over $\mathcal{L}$-formulas, and Γ, Δ over sets of $\mathcal{L}$-formulas. For an $\mathcal{L}$-formula ψ, we denote by $\mathsf{Atoms}(\psi)$ the set of atomic formulas in ψ. We denote by $SF(\Gamma)$ the set of all subformulas of Γ.

2.1 Logics, Consequence Relations and Abstract Rules

DEFINITION 1.

1. A *Scott consequence relation* (*scr* for short) for a language $\mathcal{L}$ is a binary relation $\vdash$ between sets of formulas of $\mathcal{L}$ that satisfies the following three conditions:

strong reflexivity: if $\Gamma \cap \Delta \neq \emptyset$ then $\Gamma \vdash \Delta$.

monotonicity: if $\Gamma \vdash \Delta$ and $\Gamma \subseteq \Gamma'$, $\Delta \subseteq \Delta'$ then $\Gamma' \vdash \Delta'$.
Transitivity (*cut*): if $\Gamma \vdash \psi, \Delta$ and $\Gamma', \psi \vdash \Delta'$ then $\Gamma, \Gamma' \vdash \Delta, \Delta'$.

2. A Tarskian consequence relation (tcr) $\vdash^1$ for a language $\mathcal{L}$ is a binary relation between sets of L-formulas and L-formulas, that satisfies the following conditions:

   strong reflexivity: if $\psi \in \Gamma$ then $\Gamma \vdash^1 \psi$.
   monotonicity: if $\Gamma \vdash^1 \psi$ and $\Gamma \subseteq \Gamma'$, then $\Gamma' \vdash^1 \psi$.
   Transitivity (*cut*): if $\Gamma \vdash^1 \psi$ and $\Gamma', \psi \vdash^1 \varphi$ then $\Gamma, \Gamma' \vdash^1 \varphi$.

3. A tcr $\vdash$ for $\mathcal{L}$ is *structural* if for every uniform $\mathcal{L}$-substitution σ and every Γ and ψ, if $\Gamma \vdash \psi$ then $\sigma(\Gamma) \vdash \sigma(\psi)$. $\vdash$ is *finitary* if whenever $\Gamma \vdash \psi$, there exists some finite $\Gamma' \subseteq \Gamma$, such that $\Gamma' \vdash \psi$. $\vdash$ is *consistent* (or *non-trivial*) if there exist some non-empty Γ and some ψ s.t. $\Gamma \not\vdash \psi$. $\vdash$ is *uniform* if $\Gamma \vdash \psi$ whenever $\Gamma, \Delta \vdash \psi$, $\mathsf{Atoms}(\Gamma \cup \{\psi\}) \cap \mathsf{Atoms}(\Delta) = \emptyset$, and Δ is consistent (i.e. there exists φ such that $\Delta \not\vdash \varphi$). Similar properties can be defined for an scr.

4. *A Tarskian propositional logic* (*propositional logic*) is a pair $\langle \mathcal{L}, \vdash \rangle$, where $\mathcal{L}$ is a propositional language, and $\vdash$ is a structural and consistent tcr (scr) for $\mathcal{L}$. The logic $\langle \mathcal{L}, \vdash \rangle$ is finitary if $\vdash$ is finitary.

For the rest of this section, we focus on scrs. However, the properties below can be formulated in the context of tcrs as well.

There are several ways of defining consequence relations for a language $\mathcal{L}$. The two most common ones are the proof-theoretical and the model-theoretical approaches. In the former, the definition of a consequence relation is based on some notion of a *proof* in some formal calculus. In the latter approach, the definition is based on a notion of a *semantics* for $\mathcal{L}$. The general notion of an abstract semantics is rather opaque. One usually starts by defining a notion of a *valuation* as a certain type of partial functions from $Frm_{\mathcal{L}}$ to some set. Then ones defines what it means for a valuation to *satisfy* a formula (or to be a *model* of a formula). A semantics is then some set S of valuations, and the consequence relation induced by S is defined as follows: $\Gamma \vdash_{\mathsf{S}} \Delta$ if every *total* valuation in S which satisfies all the formulas in Γ, satisfies some formula in Δ as well (note that this always defines an scr). We say that a semantics S is *analytic*[2] if every partial valuation in S,

[2]The term 'effective' was used in [Avron, 2007a; Avron and Zamansky, 2007c; Avron and Zamansky, 2007a] instead of 'analytic'.

whose domain is closed under subformulas, can be extended to a full (i.e. total) valuation in S. This implies that the exact identity of the language $\mathcal{L}$ is not important, since analycity allows us to focus on some subset of its connectives. (See Remark 12 below for another important consequence of analycity.) We shall shortly see that both ordinary many-valued semantics and non-deterministic semantics based on propositional Nmatrices are always analytic. However this is not necessarily the case in general [3].

DEFINITION 2.

1. A *pure (abstract) rule* in a propositional language $\mathcal{L}$ is any ordered pair $\langle \Gamma, \Delta \rangle$, where Γ and Δ are finite sets of formulas in $\mathcal{L}$ (We shall usually denote such a rule by $\Gamma \Rightarrow \Delta$ rather than by $\langle \Gamma, \Delta \rangle$).

2. Let $\mathbf{L} = \langle \mathcal{L}, \vdash_1 \rangle$ be a propositional logic, and let S be a set of rules in a propositional language $\mathcal{L}'$. The *extension* $\mathbf{L}[S]$ of $\langle \mathcal{L}, \vdash_1 \rangle$ by S is[4] the logic $\langle \mathcal{L}^*, \vdash^* \rangle$, where $\mathcal{L}^* = \mathcal{L} \cup \mathcal{L}'$, and $\vdash^*$ is the least *structural* scr $\vdash$ such that $\Gamma \vdash \Delta$ whenever $\Gamma \vdash_1 \Delta$ or $\langle \Gamma, \Delta \rangle \in S$.

REMARK 3. It is easy to see that $\vdash^*$ is the closure under cuts and weakenings of the set of all pairs $\langle \sigma(\Gamma), \sigma(\Delta) \rangle$, where σ is a uniform substitution in $\mathcal{L}^*$, and either $\Gamma \vdash_1 \Delta$ or $\langle \Gamma, \Delta \rangle \in S$. This in turn implies that an extension of a finitary logic by a set of pure rules is again finitary.

CONVENTION 4. To emphasize the fact that the presence of a rule in a system means the presence of all its instances, we shall usually describe a rule using the metavariables φ, ψ, θ rather than the atomic formulas $p_1, p_2, ...$. Thus although formally $(\supset\Rightarrow)$ is the rule $p_1, p_1 \supset p_2 \Rightarrow p_2$, we shall write it as $\varphi, \varphi \supset \psi \Rightarrow \psi$.

REMARK 5. Suppose that the formula θ occurs in a pure rule of a logic $\mathcal{L}$, and we decide to select θ as the "principal formula" of that rule. Assume e.g. that the rule is of the form $\varphi_1, \ldots, \varphi_n \Rightarrow \psi_1, \ldots, \psi_k, \theta$ (the consideration in the other case is similar). Suppose further that $\Gamma_i \vdash \Delta_i, \varphi_i$ for $i = 1, \ldots, n$ and $\psi_j, \Gamma_j \vdash \Delta_j$ for $j = 1, \ldots, k$. Then $\Gamma_1, \ldots, \Gamma_n \vdash \Delta_1, \ldots, \Delta_k, \theta$ (by n+k

[3] For instance, in the bivaluations semantics and the possible translations semantics described in [Carnielli, 1998; Carnielli and Marcos, 2002; Carnielli et. al., 2007] no general theorem of analycity is available. Hence analycity should be proved from scratch for every useful instance of these types of semantics.

[4] Obviously, the extension of $\langle \mathcal{L}, \vdash_1 \rangle$ by S is well-defined (i.e. a logic) only if $\vdash^*$ is consistent. In all the cases we consider below this will easily be guaranteed by the semantics we provide (and so we shall not even mention it).

cuts). It follows that $\mathcal{L}$ is closed in this case under the Gentzen-type rule:

$$\frac{\Gamma_i \Rightarrow \Delta_i, \varphi_i \ \ (i = 1, \ldots, n) \qquad \psi_j, \Gamma_j \Rightarrow \Delta_j \ \ (j = 1, \ldots, k)}{\Gamma_1, \ldots, \Gamma_n \Rightarrow \Delta_1, \ldots, \Delta_k, \theta}$$

Conversely, if $\mathcal{L}$ is closed under this Gentzen-type rule then by applying it to the reflexivity axioms $\varphi_i \vdash \varphi_i$ $(i = 1, \ldots, n)$ and $\psi_j \vdash \psi_j$ $(j = 1, \ldots, k)$ we get $\varphi_1, \ldots, \varphi_n \vdash \psi_1, \ldots, \psi_k, \theta$. It follows that every pure rule in the sense of Definition 2 is equivalent to some *multiplicative* (in the terminology of [Girard, 1987]) or *pure* (in the terminology of [Avron, 1991]) Gentzen-type rule. Moreover: it is easy to see that most standard rules used in Gentzen-type systems are equivalent to finite sets of pure rules in the sense of Definition 2. For example: the usual $(\supset\Rightarrow)$ rule of classical logic is equivalent by what we have just shown to the pure rule $\varphi, \varphi \supset \psi \Rightarrow \psi$. The classical $(\Rightarrow\supset)$, in turn, can be split into the following two rules:

$$\frac{\Gamma, \varphi \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \varphi \supset \psi} \qquad \frac{\Gamma \Rightarrow \Delta, \psi}{\Gamma \Rightarrow \Delta, \varphi \supset \psi}$$

Hence $(\Rightarrow\supset)$ is equivalent to the set $\{\psi \Rightarrow \varphi \supset \psi, \quad \Rightarrow \varphi, \varphi \supset \psi\}$. [5]

2.2 Many-valued Matrices

The most standard general method for defining propositional logics is by using many-valued (deterministic) matrices ([Rosser and Turquette, 1952; Bolc and Borowik, 1992; Malinowski, 1993; Gottwald, 2001; Hähnle, 2001; Urquhart, 2001]):

DEFINITION 6.

1. A matrix for $\mathcal{L}$ is a tuple $\mathcal{P} = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$, where:

   - $\mathcal{V}$ is a non-empty set of truth values.
   - $\mathcal{D}$ (designated truth values) is a non-empty proper subset of $\mathcal{V}$.
   - For every n-ary connective $\diamond$ of $\mathcal{L}$, $\mathcal{O}$ includes a corresponding function $\tilde{\diamond} : \mathcal{V}^n \to \mathcal{V}$.

   We say that $\mathcal{P}$ is (in)finite if so is $\mathcal{V}$.

2. A partial valuation in $\mathcal{P}$ is a function v to $\mathcal{V}$ from some subset of $Frm_{\mathcal{L}}$ which is closed under subformulas, such that for each n-ary connective $\diamond$ of $\mathcal{L}$, the following holds for all $\psi_1, \ldots, \psi_n \in Frm_{\mathcal{L}}$:

$$v(\diamond(\psi_1, ..., \psi_n)) = \tilde{\diamond}(v(\psi_1), ..., v(\psi_n))$$

[5] Recall that formally we should have written here $\{p_2 \Rightarrow p_1 \supset p_2, \quad \Rightarrow p_1, p_1 \supset p_2\}$.

A partial valuation in $\mathcal{P}$ is a (full) *valuation* if its domain is $Frm_{\mathcal{L}}$. A partial valuation v in $\mathcal{P}$ satisfies a formula ψ ($v \models \psi$) if $v(\psi) \in \mathcal{D}$.

3. Let $\mathcal{P}$ be a matrix. We say that $\Gamma \vdash_{\mathcal{P}} \Delta$ if whenever a valuation in $\mathcal{P}$ satisfies all the formulas of Γ, it satisfies also at least one of the formulas of Δ. We say that $\Gamma \vdash^1_{\mathcal{P}} \psi$ if $\Gamma \vdash_{\mathcal{P}} \{\psi\}$. For a family of matrices F, we say that $\Gamma \vdash_F \Delta$ if $\Gamma \vdash_{\mathcal{P}} \Delta$ for every $\mathcal{P}$ in F.

4. A logic **L** is sound for a matrix $\mathcal{P}$ if $\vdash_{\mathbf{L}} \subseteq \vdash_{\mathcal{P}}$. **L** is complete for a matrix $\mathcal{P}$ if $\vdash_{\mathcal{P}} \subseteq \vdash_{\mathbf{L}}$. $\mathcal{P}$ is a characteristic matrix for a logic **L** if $\vdash_{\mathbf{L}} = \vdash_{\mathcal{P}}$. F is a characteristic set of matrices for **L** if $\vdash_{\mathbf{L}} = \vdash_F$.

The following well-known theorem can easily be proved:

THEOREM 7. *For every matrix $\mathcal{P}$ for $\mathcal{L}$, $\vdash_{\mathcal{P}}$ is a uniform propositional logic, and $\vdash^1_{\mathcal{P}}$ is a uniform Tarskian propositional logic.*

The converse of this theorem also holds (see [Urquhart, 2001]):

THEOREM 8. *Every (Tarskian) uniform structural logic has a characteristic matrix.*

REMARK 9. Although every Tarskian uniform structural logic has a characteristic matrix, it is often the case that this matrix is infinite, and is hard to find and use. We will shortly see that finite characteristic Nmatrices exist for many logics which have only infinite characteristic matrices (see Theorem 24).

THEOREM 10. **(Compactness)** ([Shoesmith, 1971]) *If $\mathcal{P}$ is a finite matrix then $\vdash_{\mathcal{P}}$ and $\vdash^1_{\mathcal{P}}$ are finitary.*

The next important result is again very easy to prove:

PROPOSITION 11. **(Analycity)** *Any partial valuation in a matrix $\mathcal{P}$ for $\mathcal{L}$, which is defined on a set of $\mathcal{L}$-formulas closed under subformulas, can be extended to a full valuation in $\mathcal{P}$.*

REMARK 12. At this point the importance of analycity should again be stressed. Because of this property $\vdash_{\mathcal{S}}$ is decidable whenever $\mathcal{S}$ is a finite matrix. Moreover, analycity guarantees semi-decidability of non-theoremhood even if a matrix $\mathcal{P}$ is infinite, provided that $\mathcal{P}$ is effective (i.e, the set of truth-values is countable, the interpretation functions of the connectives are computable, and the set of designated truth-values is decidable). Note that this implies decidability in case $\vdash_{\mathcal{S}}$ also has a corresponding sound and complete proof system.

REMARK 13. One of the main shortcomings of matrix-based semantics is its lack of modularity with respect to proof systems. To use this type of semantics, the rules and axioms of a system which are related to a given connective should be considered as a whole, and there is no method for separately determining the semantic effects of each rule alone. Take for example the standard Gentzen-type rules for negation:

$$\frac{\Gamma \Rightarrow \Delta, \psi}{\Gamma, \neg\psi \Rightarrow \Delta}\ (\neg \Rightarrow) \qquad \frac{\Gamma, \psi \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \neg\psi}\ (\Rightarrow \neg)$$

The corresponding truth-table is the classical one:

	$\neg$
t	**f**
f	**t**

However, if one of the negation rules is discarded, the resulting system has no finite characteristic matrix (this is a special case of Theorem 24 below). It follows that in the framework of (ordinary) matrices the semantic effects of each of the above two rules of negation cannot be analyzed separately. We will shortly see that in contrast, the semantics of non-deterministic matrices does allow a high degree of modularity: In many cases the effect of each syntactic rule or axiom alone can easily be determined, and the semantics of a proof system can then be constructed by straightforwardly combining the semantics of its various rules and axioms.

3 INTRODUCING NMATRICES

Nmatrices were introduced in [Avron and Lev, 2001; Avron and Lev, 2005; Avron and Konikowska, 2005]. The definitions below are taken from there.

DEFINITION 14. A *non-deterministic matrix (Nmatrix)* for $\mathcal{L}$ is a tuple $\mathcal{M} = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$, where:

- $\mathcal{V}$ is a non-empty set of truth values.
- $\mathcal{D}$ (designated truth values) is a non-empty proper subset of $\mathcal{V}$.
- For every n-ary connective $\diamond$ of $\mathcal{L}$, $\mathcal{O}$ includes a corresponding function $\tilde{\diamond} : \mathcal{V}^n \to 2^{\mathcal{V}} \setminus \{\emptyset\}$.

DEFINITION 15. Let $\mathcal{M} = (\mathcal{V}, \mathcal{D}, \mathcal{O})$ be an Nmatrix for $\mathcal{L}$.

1. A partial *dynamic valuation* in $\mathcal{M}$ (or an $\mathcal{M}$-legal partial dynamic valuation) is a function v from some subset of $Frm_{\mathcal{L}}$ to $\mathcal{V}$, which is closed under subformulas, such that for each n-ary connective $\diamond$ of $\mathcal{L}$, the following holds for all $\psi_1, \ldots, \psi_n \in Frm_{\mathcal{L}}$:

$$\text{(SLC)} \qquad v(\diamond(\psi_1, \ldots, \psi_n)) \in \widetilde{\diamond}(v(\psi_1), \ldots, v(\psi_n))$$

   A partial valuation in $\mathcal{M}$ is called a *valuation* if its domain is $Frm_{\mathcal{L}}$.

2. A (partial) *static valuation* in $\mathcal{M}$ (or an $\mathcal{M}$-legal (partial) static valuation) is a (partial) dynamic valuation which satisfies also the following compositionality (or functionality) principle (CMP): for each $\diamond$ of $\mathcal{L}$ and for every $\psi_1, \ldots, \psi_n, \varphi_1, \ldots, \varphi_n \in Frm_{\mathcal{L}}$,

$$v(\diamond(\psi_1, \ldots, \psi_n)) = v(\diamond(\varphi_1, \ldots, \varphi_n)) \text{ if } v(\psi_i) = v(\varphi_i) \ (i = 1 \ldots n)$$

REMARK 16. Ordinary (deterministic) matrices correspond to the case when each $\widetilde{\diamond}$ is a function taking singleton values only (then it can be treated as a function $\widetilde{\diamond} : \mathcal{V}^n \to \mathcal{V}$). In this case there is no difference between static and dynamic valuations, and we have full determinism.

REMARK 17. Like in usual multi-valued semantics, the principle here is that each formula has a definite logical value. This is why we exclude $\emptyset$ from being a value of $\widetilde{\diamond}$. However, the absence of any logical value for a formula can still be simulated in our formalism by introducing a special logical value $\perp$ representing exactly this case (which is a well-known procedure in the framework of partial logics ([Blamey, 1986])).

To understand the difference between ordinary matrices and Nmatrices, recall that in the deterministic case (see Defn. 6), the truth-value assigned by a valuation v to a complex formula is defined as follows: $v(\diamond(\psi_1, ..., \psi_n)) = \tilde{\diamond}(v(\psi_1), ..., v(\psi_n))$. Thus the truth-value assigned to $\diamond(\psi_1, ..., \psi_n)$ is uniquely determined by the truth-values of its subformulas: $v(\psi_1), ..., v(\psi_n)$. This, however, is not the case in dynamic valuations in Nmatrices: in general the truth-values assigned to $\psi_1, ..., \psi_n$ do not uniquely determine the truth-value assigned to $\diamond(\psi_1, ..., \psi_n)$ because v makes a non-deterministic choice out of the set of options $\tilde{\diamond}(v(\psi_1), ..., v(\psi_n))$. Therefore the non-deterministic semantics is non-truth-functional, as opposed to the deterministic one.

DEFINITION 18.

1. A (partial) valuation v in $\mathcal{M}$ *satisfies* a formula ψ ($v \models \psi$) if ($v(\psi)$ is defined and) $v(\psi) \in \mathcal{D}$. It is a *model of* Γ ($v \models \Gamma$) if it satisfies every formula in Γ.

2. We say that ψ is dynamically (statically) valid in $\mathcal{M}$, in symbols $\models^d_{\mathcal{M}} \psi$ ($\models^s_{\mathcal{M}} \psi$), if $v \models \psi$ for each dynamic (static) valuation v in $\mathcal{M}$.

3. A logic $\mathbf{L}$ is *dynamically (statically) weakly sound for an Nmatrix* $\mathcal{M}$ if $\vdash_{\mathbf{L}} \psi$ implies $\models^d_{\mathcal{M}} \psi$ ($\models^s_{\mathcal{M}} \psi$). A logic $\mathbf{L}$ is *dynamically (statically) weakly complete for* $\mathcal{M}$ if $\models^d_{\mathcal{M}} \psi$ ($\models^s_{\mathcal{M}} \psi$) implies $\vdash_{\mathbf{L}} \psi$. $\mathcal{M}$ is a *dynamically (statically) weakly characteristic* for $\mathbf{L}$ if $\mathbf{L}$ is dynamically (statically) both weakly sound and weakly complete for $\mathcal{M}$.

4. $\vdash^d_{\mathcal{M}}$ ($\vdash^s_{\mathcal{M}}$), the *dynamic (static) consequence relation induced by* $\mathcal{M}$, is defined as follows: $\Gamma \vdash^d_{\mathcal{M}} \Delta$ ($\Gamma \vdash^s_{\mathcal{M}} \Delta$), if every dynamic (static) model v in $\mathcal{M}$ of Γ satisfies some $\psi \in \Delta$.

5. A logic $\mathbf{L} = \langle \vdash_{\mathbf{L}}, \mathcal{L} \rangle$ is dynamically (statically) sound for an Nmatrix $\mathcal{M}$ for $\mathcal{L}$ if $\vdash_{\mathbf{L}} \subseteq \vdash^d_{\mathcal{M}}$ ($\vdash_{\mathbf{L}} \subseteq \vdash^s_{\mathcal{M}}$). $\mathbf{L}$ is dynamically (statically) complete for $\mathcal{M}$ if $\vdash^d_{\mathcal{M}} \subseteq \vdash_{\mathbf{L}}$ ($\vdash^s_{\mathcal{M}} \subseteq \vdash_{\mathbf{L}}$). $\mathcal{M}$ is dynamically (statically) characteristic for $\mathbf{L}$ if $\vdash^d_{\mathcal{M}} = \vdash_{\mathbf{L}}$ ($\vdash^s_{\mathcal{M}} = \vdash_{\mathbf{L}}$).

REMARK 19. Obviously, the static consequence relation includes the dynamic one, i.e. $\vdash^s_{\mathcal{M}} \supseteq \vdash^d_{\mathcal{M}}$. Also, for ordinary matrices $\vdash^s_{\mathcal{M}} = \vdash^d_{\mathcal{M}}$.

CONVENTION 20. We shall denote $\mathcal{F} = \mathcal{V} \setminus \mathcal{D}$, and shall usually identify singletons of truth-values with the truth-values themselves.

EXAMPLE 21. Assume that $\mathcal{L}$ has binary connectives $\vee$, $\wedge$, and $\supset$ interpreted classically, and a unary connective $\neg$, for which the law of contradiction obtains, but not necessarily the law of excluded middle. This leads to the Nmatrix $\mathcal{M}^2 = (\mathcal{V}, \mathcal{D}, \mathcal{O})$ for $\mathcal{L}$, where Let $\mathcal{V} = \{\mathbf{f}, \mathbf{t}\}, \mathcal{D} = \{\mathbf{t}\}$, and $\mathcal{O}$ is given by:

		$\widetilde{\vee}$	$\widetilde{\wedge}$	$\widetilde{\supset}$
$\mathbf{t}$	$\mathbf{t}$	$\mathbf{t}$	$\mathbf{t}$	$\mathbf{t}$
$\mathbf{t}$	$\mathbf{f}$	$\mathbf{t}$	$\mathbf{f}$	$\mathbf{f}$
$\mathbf{f}$	$\mathbf{t}$	$\mathbf{t}$	$\mathbf{f}$	$\mathbf{t}$
$\mathbf{f}$	$\mathbf{f}$	$\mathbf{f}$	$\mathbf{f}$	$\mathbf{t}$

	$\widetilde{\neg}$
$\mathbf{t}$	$\mathbf{f}$
$\mathbf{f}$	$\{\mathbf{t}, \mathbf{f}\}$

Note that classical negation can be defined in $\mathcal{M}^2$ by: $\sim\psi = \psi \supset \neg\psi$ (this is a semantic counterpart of the observation made in [Béziau, 1999]).

EXAMPLE 22. Consider the following two 3-valued Nmatrices $\mathcal{M}^3_L, \mathcal{M}^3_S$. In both we have $\mathcal{V} = \{\mathbf{f}, \top, \mathbf{t}\}, \mathcal{D} = \{\top, \mathbf{t}\}$. Also the interpretations of disjunction, conjunction and implication are the same in both of them, and correspond to those in positive classical logic:

$$a \widetilde{\vee} b = \begin{cases} \mathcal{D} & \text{if either } a \in \mathcal{D} \text{ or } b \in \mathcal{D} \\ \mathcal{F} & \text{if } a, b \in \mathcal{F} \end{cases}$$

$$a\tilde{\wedge}b = \begin{cases} \mathcal{D} & \text{if } a, b \in \mathcal{D} \\ \mathcal{F} & \text{if either } a \in \mathcal{F} \text{ or } b \in \mathcal{F} \end{cases}$$

$$a\tilde{\supset}b = \begin{cases} \mathcal{D} & \text{if either } a \in \mathcal{F} \text{ or } b \in \mathcal{D} \\ \mathcal{F} & \text{if } a \in \mathcal{D} \text{ and } b \in \mathcal{F} \end{cases}$$

However, negation is interpreted differently: more liberally in $\mathcal{M}^3_L$, and more strictly in $\mathcal{M}^3_S$:

$\mathcal{M}^3_L$:

	$\tilde{\neg}$
t	**f**
$\top$	$\mathcal{V}$
f	**t**

$\mathcal{M}^3_S$:

	$\tilde{\neg}$
t	**f**
$\top$	$\mathcal{D}$
f	**t**

EXAMPLE 23. After considering 2-valued Nmatrices and 3-valued Nmatrices, our last example is the 4-valued Nmatrix $\mathcal{M}_4 = (\mathcal{V}, \mathcal{D}, \mathcal{O})$, where $\mathcal{V} = \{\mathbf{f}, \bot, \top, \mathbf{t}\}$, $\mathcal{D} = \{\top, \mathbf{t}\}$, $\wedge, \vee, \supset$ are defined by the general rules given in Example 22 (applied, however, to the sets $\mathcal{D}$ and $\mathcal{F} = \mathcal{V} \setminus \mathcal{D}$ appearing in the current example), while $\neg$ is the negation of the bilattice $\mathcal{FOUR}$ ([Belnap, 1977; Ginsberg, 1988; Fitting, 1994; Arieli and Avron, 1996]):

	$\tilde{\neg}$
t	**f**
$\top$	$\top$
$\bot$	$\bot$
f	**t**

At this point it is natural to ask whether finite Nmatrices can be used for characterizing logics that cannot be characterized by finite ordinary matrices. The next theorem provides a positive answer to this question:

THEOREM 24. *Let $\mathcal{M}$ be a two-valued Nmatrix which has at least one proper non-deterministic operation. Then there is no finite family of finite ordinary matrices F, such that $\vdash^d_{\mathcal{M}} = \vdash_F$. If in addition $\mathcal{M}$ includes the classical implication, then there is no finite family of ordinary matrices F, such that $\vdash^d_{\mathcal{M}} \psi$ iff $\vdash_F \psi$.*

Proof: a straightforward modification of the proof of Theorem 3.4 in [Avron and Lev, 2005].

As the next easy theorem shows, things are different in the case of the static semantics:

THEOREM 25. *For every (finite) Nmatrix $\mathcal{M}$, there is a (finite) family of ordinary matrices, such that* $\vdash^s_{\mathcal{M}} = \vdash_F$.

Thus only the expressive power of the dynamic semantics based on Nmatrices is stronger than that of ordinary matrices. For this reason (after providing general proof theory for both kinds of semantics in the next subsection) our main focus will be on this semantics and what it induces. Accordingly, *we shall usually write simply* $\vdash_{\mathcal{M}}$ *instead of* $\vdash^d_{\mathcal{M}}$.

The following theorem from [Avron and Lev, 2005] is a generalization of Theorem 10 to the case of Nmatrices:

THEOREM 26. **(Compactness)** $\vdash_{\mathcal{M}}$ *is finitary for any finite Nmatrix* $\mathcal{M}$.

Later we shall prove a stronger version of this theorem (see Theorem 53). The proof of the next important result is as easy for Nmatrices as it is for ordinary matrices:

PROPOSITION 27. **(Analycity)** *Let* $\mathcal{M} = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$ *be an Nmatrix for* $\mathcal{L}$, *and let* v' *be a partial valuation in* $\mathcal{M}$. *Then* v' *can be extended to a (full) valuation in* $\mathcal{M}$.

It is easy to show that like in the case of ordinary matrices (see Remark 12), Proposition 27 implies the following Theorem:

THEOREM 28. *Non-theoremhood of a logic which has an effective characteristic Nmatrix* $\mathcal{M}$ *is semi-decidable. If* $\mathcal{M}$ *is finite, or* L *also has a sound and complete formal proof system, then* L *is decidable.*

The following is an easy analogue for Nmatrices of Theorem 7:

PROPOSITION 29. *For any Nmatrix* $\mathcal{M}$, $\vdash_{\mathcal{M}}$ *is uniform.*

We end this subsection by introducing the notion of a *refinement*:

DEFINITION 30. Let $\mathcal{M}_1 = \langle \mathcal{V}_1, \mathcal{D}_1, \mathcal{O}_1 \rangle$ and $\mathcal{M}_2 = \langle \mathcal{V}_2, \mathcal{D}_2, \mathcal{O}_2 \rangle$ be Nmatrices for a language $\mathcal{L}$.

1. A reduction of $\mathcal{M}_1$ to $\mathcal{M}_2$ is a function $F : \mathcal{V}_1 \to \mathcal{V}_2$ such that:
   (a) For every $x \in \mathcal{V}_1$, $x \in \mathcal{D}_1$ iff $F(x) \in \mathcal{D}_2$.
   (b) $F(y) \in \widetilde{\diamond}_{\mathcal{M}_2}(F(x_1), \ldots, F(x_n))$ for every n-ary connective $\diamond$ of $\mathcal{L}$ and every $x_1, \ldots, x_n, y \in \mathcal{V}_1$ such that $y \in \widetilde{\diamond}_{\mathcal{M}_1}(x_1, \ldots, x_n)$.
2. $\mathcal{M}_1$ is a *refinement* of $\mathcal{M}_2$ if there exists a reduction of $\mathcal{M}_1$ to $\mathcal{M}_2$.

THEOREM 31. *If $\mathcal{M}_1$ is a refinement of $\mathcal{M}_2$ then* $\vdash_{\mathcal{M}_2} \subseteq \vdash_{\mathcal{M}_1}$.

REMARK 32. An important case in which $\mathcal{M}_1 = \langle \mathcal{V}_1, \mathcal{D}_1, \mathcal{O}_1 \rangle$ is a refinement of $\mathcal{M}_2 = \langle \mathcal{V}_2, \mathcal{D}_2, \mathcal{O}_2 \rangle$ is when $\mathcal{V}_1 \subseteq \mathcal{V}_2$, $\mathcal{D}_1 = \mathcal{D}_2 \cap \mathcal{V}_1$, and $\widetilde{\diamond}_{\mathcal{M}_1}(\vec{x}) \subseteq \widetilde{\diamond}_{\mathcal{M}_2}(\vec{x})$ for every n-ary connective $\diamond$ of $\mathcal{L}$ and every $\vec{x} \in \mathcal{V}_1^n$. It is easy to see that the identity function on $\mathcal{V}_1$ is in this case a reduction of $\mathcal{M}_1$ to $\mathcal{M}_2$. A refinement of this sort will be called *simple*.

4 CANONICAL DEDUCTION SYSTEMS AND NMATRICES

The idea of "canonical" systems implicitly underlies a long tradition in the philosophy of logic, established by G. Gentzen in his classical paper [Gentzen, 1969]. According to this tradition, the meaning of a connective is determined by the introduction and the elimination rules which are associated with it (see, e.g., [Zucker, 1978a; Zucker, 1978b]). The supporters of this thesis usually have in mind Natural Deduction systems of an ideal type. In this type of "canonical systems" each connective $\diamond$ has its own introduction and elimination rules, in each of which $\diamond$ is mentioned exactly once, and no other connective is involved. The rules should also be pure in the sense of [Avron, 1991]. Unfortunately, already the handling of negation requires rules which are not canonical in this sense. This problem was solved by Gentzen himself by moving to what is now known as (multiple-conclusion) Gentzen-type calculi, which instead of introduction and elimination rules use left and right introduction rules. The intuitive notion of a "canonical rule" can be adapted to such systems in a straightforward way, and it is well-known that the usual classical connectives can indeed be fully characterized in this framework by such rules. Moreover, the cut-elimination theorem obtains in all the known Gentzen-type calculi for propositional classical logic (or some fragment of it) which employ only rules of this type. These facts were generalized in [Avron and Lev, 2005], where the notion of a canonical propositional Gentzen-type system has been introduced. This notion was further generalized in [Avron and Konikowska, 2005; Avron and Zamansky, 2009] to *canonical signed calculi.* These calculi and their intimate connections with finite Nmatrices are the subject of the present section.

Signed calculi consist of rules operating on finite sets of *signed formulas*, and axioms being sets of such formulas. The deduction formalism we use here for them is similar to the Rasiowa-Sikorski (R-S) systems ([Rasiowa and Sikorski, 1963; Konikowska, 2002]), known also as dual tableaux ([Baaz et. al., 1993; Hähnle, 1999]).

Henceforth (until the end of Section 4) $\mathcal{V}$ *denotes some finite set of signs.*

DEFINITION 33. A *signed formula* for $(\mathcal{L}, \mathcal{V})$ is an expression of the form $s : \psi$, where $s \in \mathcal{V}$ and $\psi \in Frm_{\mathcal{L}}$. A signed formula $s : \psi$ is atomic if ψ is an atomic formula. A *sequent* for $(\mathcal{L}, \mathcal{V})$ is a finite set of signed formulas for $(\mathcal{L}, \mathcal{V})$. A *clause* is a sequent consisting of atomic signed formulas.

REMARK 34. The usual (two-sided) sequent notation $\Gamma \Rightarrow \Delta$ can be interpreted as $\{f : \Gamma\} \cup \{t : \Delta\}$, i.e. a sequent in the sense of Defn. 33 over the two signs $\{t, f\}$.

DEFINITION 35. Let v be a function from the set of formulas of $\mathcal{L}$ to $\mathcal{V}$.

1. v satisfies a signed formula $\gamma = (l : \psi)$, denoted by $v \models (l : \psi)$, if $v(\psi) = l$.

2. v satisfies a set of signed formulas Υ, denoted by $v \models \Upsilon$, if there is some $\gamma \in \Upsilon$, such that $v \models \gamma$.

CONVENTION 36. Formulas will be denoted by φ, ψ, signed formulas - by $\alpha, \beta, \gamma, \delta$, sets of signed formulas - by Υ, Λ, sequents - by Ω, Σ, Π, sets of sets of signed formulas - by Φ, Ψ and sets of sequents - by Θ, Ξ. We write $S : \psi$ instead of $\{s : \psi \mid s \in S\}$, and $S : \Delta$ instead of $\{s : \psi \mid s \in S, \psi \in \Delta\}$.

DEFINITION 37. A *signed canonical (propositional) rule of arity n* for $(\mathcal{L}, \mathcal{V})$ is an expression of the form $[\Theta/S : \diamond(p_1, \ldots, p_n)]$, where S is a non-empty subset of $\mathcal{V}$, $\diamond$ is an n-ary connective of $\mathcal{L}$ and $\Theta = \{\Sigma_1, ..., \Sigma_m\}$, where $m \geq 0$ and for every $1 \leq j \leq m$, Σ_j are clauses (see Definition 33) consisting of signed formulas of the form $a : p_k$, where $a \in \mathcal{V}$ and $1 \leq k \leq n$. An *application* of a rule $[\{\Sigma_1, ..., \Sigma_m\}/S : \diamond(p_1, \ldots, p_n)]$ is any inference of the form:

$$\frac{\Omega \cup \Sigma_1^* \quad \ldots \quad \Omega \cup \Sigma_m^*}{\Omega \cup S : \diamond(\psi_1, ..., \psi_n)}$$

where $\psi_1, ..., \psi_n$ are $\mathcal{L}$-formulas, Ω is a sequent, and for all $1 \leq i \leq m$: Σ_i^* is obtained from Σ_i by replacing p_j by ψ_j for every $1 \leq j \leq n$.

EXAMPLE 38.

1. The standard Gentzen-style introduction rules for the classical conjunction are usually defined as follows:

$$\frac{\Gamma, \psi, \varphi \Rightarrow \Delta}{\Gamma, \psi \wedge \varphi \Rightarrow \Delta} \qquad \frac{\Gamma \Rightarrow \Delta, \psi \quad \Gamma \Rightarrow \Delta, \varphi}{\Gamma \Rightarrow \Delta, \psi \wedge \varphi}$$

Using the notation in Remark 34, we can write $\{f : \Gamma\} \cup \{t : \Delta\}$ (that is, ψ occurs with a sign 't' if $\psi \in \Gamma$ and with a sign 'f' if $\psi \in \Delta$), thus the canonical representation of the rules above is as follows:

$$[\{\{f : p_1, f : p_2\}\}/\{f\} : p_1 \wedge p_2] \qquad [\{\{t : p_1\}, \{t : p_2\}\}/\{t\} : p_1 \wedge p_2]$$

Applications of these rules have the forms:

$$\frac{\Omega \cup \{f : \psi_1, f : \psi_2\}}{\Omega \cup \{f : \psi_1 \wedge \psi_2\}} \qquad \frac{\Omega \cup \{t : \psi_1\} \quad \Omega \cup \{t : \psi_2\}}{\Omega \cup \{t : \psi_1 \wedge \psi_2\}}$$

2. Consider a calculus over $\mathcal{V} = \{a, b, c\}$ with the following introduction rules for a ternary connective $\circ$:

$$[\{\{a : p_1, c : p_2\}, \{a : p_3, b : p_2\}\}/\{a, c\} : \circ(p_1, p_2, p_3)]$$

$$[\{\{c : p_2\}, \{a : p_3, b : p_3\}, \{c : p_1\}\}/\{b, c\} : \circ(p_1, p_2, p_3)]$$

Their applications are of the forms:

$$\frac{\Omega \cup \{a : \psi_1, c : \psi_2\} \quad \Omega \cup \{a : \psi_3, b : \psi_2\}}{\Omega \cup \{a : \circ(\psi_1, \psi_2, \psi_3), c : \circ(\psi_1, \psi_2, \psi_3)\}}$$

$$\frac{\Omega \cup \{c : \psi_2\} \quad \Omega \cup \{a : \psi_3, b : \psi_3\} \quad \Omega \cup \{c : \psi_1\}}{\Omega \cup \{b : \circ(\psi_1, \psi_2, \psi_3), c : \circ(\psi_1, \psi_2, \psi_3)\}}$$

DEFINITION 39. Let $\mathcal{V}$ be a finite set of signs.

1. A logical axiom for $\mathcal{V}$ is a sequent of the form: $\{l : \psi \mid l \in \mathcal{V}\}$.

2. The cut and weakening rules for $\mathcal{V}$ are defined as follows:

$$\frac{\Omega \cup \{l : \psi \mid l \in L_1\} \quad \Omega \cup \{l : \psi \mid l \in L_2\}}{\Omega \cup \{l : \psi \mid l \in L_1 \cap L_2\}}\ CUT$$

$$\frac{\Omega}{\Omega, l : \psi}\ WEAK$$

where $L_1, L_2 \subseteq \mathcal{V}$ and $l \in \mathcal{V}$.

The following proposition follows from the completeness of many-valued resolution (see [Baaz et. al., 1995]):

PROPOSITION 40. *Let* Θ *be a set of clauses and* Ω *- a clause. Then* Ω *follows from* Θ *(in the sense that every* v *which satisfies* Θ *also satisfies* Ω*) iff there is some* $\Omega' \subseteq \Omega$*, such that* Ω' *is derivable from* Θ *by cuts.*

COROLLARY 41. *Let* Θ *be a set of clauses. The empty sequent is derivable from* Θ *by cuts iff* Θ *is not satisfiable.*

Now we are ready to define "canonical signed calculi" in precise terms:

DEFINITION 42. A signed calculus over a language $\mathcal{L}$ and a finite set of signs $\mathcal{V}$ is *canonical* if it consists of:

1. All logical axioms for $\mathcal{V}$.
2. The rules of cut and weakening from Defn. 39.
3. Any number of signed canonical inference rules.

Not all canonical calculi are useful. Of interest are only those of them which "define" the semantic meaning of the logical connectives they introduce. It turned out that this property can be captured syntactically by a simple syntactic criterion called *coherence*, introduced in [Avron and Lev, 2005] for canonical Gentzen-type systems, and extended in [Avron and Zamansky, 2009] to signed calculi.

DEFINITION 43. A canonical calculus G is *coherent* if $\Theta_1 \cup ... \cup \Theta_m$ is unsatisfiable whenever $\{[\Theta_1/S_1 : \psi], ..., [\Theta_m/S_m : \psi]\}$ is a set of rules of G such that $S_1 \cap ... \cap S_m = \emptyset$ (here $\psi = \diamond(p_1, \ldots, p_n)$ for some n-ary $\diamond$).

Obviously, coherence is a decidable property of canonical calculi. Note also that by Corollary 41, a canonical calculus G is *coherent* if whenever $\{[\Theta_1/S_1 : \psi], ..., [\Theta_m/S_m : \psi]\}$ is a set of rules of G, and $S_1 \cap ... \cap S_m = \emptyset$, we have that $\Theta_1 \cup ... \cup \Theta_m$ is inconsistent (i.e. the empty sequent can be derived from it using cuts).

EXAMPLE 44.

1. Consider the canonical calculus G_1 over $\mathcal{L} = \{\wedge\}$ and $\mathcal{V} = \{t, f\}$, the canonical rules of which are the two rules for $\wedge$ from Example 38. We can derive the empty sequent from $\{\{t : p_1\}, \{t : p_2\}, \{f : p_1, f : p_2\}\}$ as follows:

$$\dfrac{\dfrac{\{t : p_1\} \quad \{f : p_1, f : p_2\}}{\{f : p_2\}}\ CUT \qquad \{t : p_2\}}{\emptyset}\ CUT$$

Thus G_1 is coherent.

2. Consider the canonical calculus G_2 over $\mathcal{V} = \{a, b, c\}$ with the following introduction rules for the ternary connective $\circ$:

$$[\{\{a : p_1\}, \{b : p_2\}\}/\{a, b\} : \circ(p_1, p_2, p_3)]$$

$$[\{\{a : p_2, c : p_3\}\}/\{c\} : \circ(p_1, p_2, p_3)]$$

Clearly, the set $\{\{a : p_1\}, \{b : p_2\}, \{a : p_2, c : p_3\}\}$ is satisfiable, thus G_2 is not coherent.

REMARK 45. [Ciabattoni and Terui, 2006a] investigates a general class of single-conclusion two-sided (sequent) calculi called *simple calculi.* These calculi may include any set of structural rules, and so the two-sided canonical calculi are a particular instance of simple calculi which include all of the standard structural rules. The reductivity condition of [Ciabattoni and Terui, 2006a] can be shown to be equivalent to our coherence criterion in the context of two-sided canonical systems.

Next we define some notions of cut-elimination[6] in canonical calculi:

DEFINITION 46. Let G be a canonical signed calculus and let Θ be some set of sequents.

1. A cut is called a Θ-*cut* if the cut formula occurs in Θ. We say that a proof is Θ-cut-free if the only cuts in it are Θ-cuts.

2. A cut is called Θ-*analytic* if the cut formula is a subformula of some formula occurring in Θ. A proof is called Θ-analytic[7] if all cuts in it are Θ-analytic.

3. A canonical calculus G admits *(standard) cut-elimination* if whenever $\vdash_G \Omega$, Ω has a cut-free proof in G. G admits *strong cut-elimination*[8] if whenever $\Theta \vdash_G \Omega$, Ω has in G a Θ-cut-free proof from Θ.

4. G admits *strong analytic cut-elimination* if whenever $\Theta \vdash_G \Omega$, Ω has in G a $\Theta \cup \{\Omega\}$-analytic proof from Θ. G admits *analytic cut-elimination* if whenever $\vdash_G \Omega$, Ω has in G a $\{\Omega\}$-analytic proof.

[6] We note that by 'cut-elimination' we mean here just the *existence* of proofs without (certain forms of) cuts, rather than an algorithm to transform a given proof to a cut-free one (for the assumptions-free case the term "cut-admissibility" is sometimes used, but this notion is too weak for our purposes).

[7] This is a generalization of the notion of analytic cut (see e.g. [Baaz et. al., 2001]).

[8] The notion of strong cut-elimination from [Avron, 1993] was studied in the context of canonical Gentzen-type systems in [Avron and Zamansky, 2007b].

EXAMPLE 47. Consider the following calculus G' for a language with a binary connective $\circ$ and $\mathcal{V} = \{a, b, c\}$. The rules of G' are as follows:

$$R_1 = \{\{a : p_1\}\}/\{a, b\} : p_1 \circ p_2\} \qquad R_2 = \{\{a : p_1\}\}/\{b, c\} : p_1 \circ p_2\}$$

In the following proof in G', the cut in the final step is analytic:

$$\dfrac{\dfrac{a : p_1, b : p_1, c : p_1}{b : p_1, c : p_1, b : (p_1 \circ p_2), c : (p_1 \circ p_2)} \qquad \dfrac{a : p_1, b : p_1, c : p_1}{b : p_1, c : p_1, a : (p_1 \circ p_2), b : (p_1 \circ p_2)}}{b : p_1, c : p_1, b : (p_1 \circ p_2)}$$

4.1 Canonical Calculi for Nmatrices

There are numerous works on proof theory for logics based on finite *ordinary* matrices, mainly using many-placed sequent calculi or tableaux systems with truth values as signs (cf. [Baaz et. al., 1993; Borowik, 1986; Carnielli, 1991; Rousseau, 1967; Takahashi, 1967; Hähnle, 1999; Baaz et. al., 2000]). In this section we present analogous canonical signed calculi for logics based on finite Nmatrices (developed in [Avron and Konikowska, 2005]).

DEFINITION 48. Let $\mathcal{M} = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$ be an Nmatrix for $\mathcal{L}$.

1. $\Phi \vdash^d_{\mathcal{M}} \Upsilon$ ($\Phi \vdash^s_{\mathcal{M}} \Upsilon$) if $v \models \Upsilon$ for every $\mathcal{M}$-legal dynamic (static) valuation v which satisfies all the sets in Φ.

2. Let G be a deduction system based on sequents. G is *dynamically (statically) strongly sound for* $\mathcal{M}$ if $\Theta \vdash_G \Omega$ implies that $\Theta \vdash^d_{\mathcal{M}} \Omega$ ($\Theta \vdash^s_{\mathcal{M}} \Omega$). G is *dynamically (statically) strongly complete for* $\mathcal{M}$ if $\Theta \vdash^d_{\mathcal{M}} \Omega$ ($\Theta \vdash^s_{\mathcal{M}} \Omega$) implies $\Theta \vdash_G \Omega$. $\mathcal{M}$ is a *dynamically (statically) strongly characteristic Nmatrix for* G if G is dynamically (statically) strongly sound and strongly complete for $\mathcal{M}$. (The notions of soundness, completeness and a characteristic Nmatrix are defined similarly by setting $\Theta = \emptyset$.)

It should be noted that the set of designated values $\mathcal{D}$ in an Nmatrix $\mathcal{M} = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$ has not been used in the semantic definitions above. This is because the consequence relations defined above are between sets of sequents and sequents. Recall, however, that the set $\mathcal{D}$ *is* used in Definition 18, where the consequence relations between *sets of formulas* are defined. The following easy observations are the key for using proof systems based on sets of signed formulas for characterizing logics induced by Nmatrices:

PROPOSITION 49. *Let $\mathcal{M} = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$ be an Nmatrix for $\mathcal{L}$. Then:*

$$\Gamma \vdash^d_{\mathcal{M}} \Delta \textit{ iff } \{\mathcal{D} : \psi \mid \psi \in \Gamma\} \cup \{\mathcal{F} : \psi \mid \psi \in \Delta\} \vdash^d_{\mathcal{M}} \emptyset \textit{ iff } \vdash^d_{\mathcal{M}} \mathcal{F} : \Gamma \cup \mathcal{D} : \Delta$$

$$\Gamma \vdash^s_{\mathcal{M}} \Delta \textit{ iff } \{\mathcal{D} : \psi \mid \psi \in \Gamma\} \cup \{\mathcal{F} : \psi \mid \psi \in \Delta\} \vdash^s_{\mathcal{M}} \emptyset \textit{ iff } \vdash^s_{\mathcal{M}} \mathcal{F} : \Gamma \cup \mathcal{D} : \Delta$$

DEFINITION 50. The proof system $SF^d_{\mathcal{M}}$ for the dynamic semantics of a finite-valued Nmatrix $\mathcal{M} = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$ is the canonical signed calculus for $(\mathcal{L}, \mathcal{V})$ which for every n-ary $\diamond$, and every $a_1, \ldots, a_m, b_1, \ldots, b_k \in \mathcal{V}$ such that $\widetilde{\diamond}(a_1, \ldots, a_m) = \{b_1, \ldots, b_k\}$, includes the rule:

$$[\{a_1 : p_1\}, \ldots, \{a_m : p_m\} / \{b_1, \ldots, b_k\} : \diamond(p_1, \ldots, p_n)]$$

or in a more conventional formulation:

$$(\diamond\text{-D}) \qquad \frac{\Omega \cup \{a_1 : \varphi_1\} \quad \ldots \quad \Omega \cup \{a_m : \varphi_m\}}{\Omega \cup \{b_1 : \diamond(\varphi_1, \ldots, \varphi_m), \ldots, b_k : \diamond(\varphi_1, \ldots, \varphi_m)\}}$$

The following theorem is a generalization of a result first shown in [Avron and Konikowska, 2005]. Its proof requires just a straightforward extension of the argument given there:

THEOREM 51. *$SF^d_{\mathcal{M}}$ is dynamically strongly characteristic for $\mathcal{M}$.*

The following corollary follows from Prop. 49 and the above theorem:

COROLLARY 52. *$\Gamma \vdash^d_{\mathcal{M}} \Delta$ iff $\{\mathcal{D} : \psi \mid \psi \in \Gamma\} \cup \{\mathcal{F} : \psi \mid \psi \in \Delta\} \vdash_{SF^d_{\mathcal{M}}} \emptyset$. Moreover, if Γ and Δ are finite then $\Gamma \vdash^d_{\mathcal{M}} \Delta$ iff $\vdash_{SF^d_{\mathcal{M}}} \mathcal{F} : \Gamma \cup \mathcal{D} : \Delta$.*

Now we are ready to prove the general compactness theorem mentioned immediately after Theorem 26:

THEOREM 53. **(Compactness)**

1. *Let Θ be a set of sequents and Ω a sequent. If $\Theta \vdash^d_{\mathcal{M}} \Omega$, then there is some finite $\Theta' \subseteq \Theta$, such that $\Theta' \vdash^d_{\mathcal{M}} \Omega$.*

2. *Let Γ, Δ be two sets of $\mathcal{L}$-formulas. If $\Gamma \vdash^d_{\mathcal{M}} \Delta$, then there are some finite $\Gamma' \subseteq \Gamma$ and $\Delta' \subseteq \Delta$, such that $\Gamma' \vdash^d_{\mathcal{M}} \Delta'$.*

Proof. For the first part, assume that $\Theta \vdash^d_{\mathcal{M}} \Omega$. Then $\Theta \vdash_{SF^d_{\mathcal{M}}} \Omega$ by Theorem 51, and so there is some finite $\Theta' \subseteq \Theta$, such that $\Theta' \vdash_{SF^d_{\mathcal{M}}} \Omega$. Hence (again by Theorem 51) $\Theta' \vdash^d_{\mathcal{M}} \Omega$. For the second part, suppose that $\Gamma \vdash^d_{\mathcal{M}} \Delta$. Then by Proposition 49, $\{\mathcal{D} : \psi \mid \psi \in \Gamma\} \cup \{\mathcal{F} : \psi \mid \psi \in \Delta\} \vdash^d_{\mathcal{M}} \emptyset$. By the first part, there are some finite $\Gamma' \subseteq \Gamma$ and $\Delta' \subseteq \Delta$, such that $\{\mathcal{D} : \psi \mid \psi \in \Gamma'\} \cup \{\mathcal{F} : \psi \mid \psi \in \Delta'\} \vdash^d_{\mathcal{M}} \emptyset$. By Proposition 49, $\Gamma' \vdash^d_{\mathcal{M}} \Delta'$. ■

DEFINITION 54. The proof system $SF^s_{\mathcal{M}}$ for the static semantics of $\mathcal{M}$ is obtained from the system $SF^d_{\mathcal{M}}$ by adding, for any m-ary connective $\diamond$ of $\mathcal{L}$ and any $a_1, \dots, a_m, b \in \mathcal{V}$ such that $b \in \widetilde{\diamond}(a_1, \dots, a_m)$, the rule ($\diamond$-S):

$$\frac{\{\Omega \cup \{a_j : \varphi_j\}\}_{1 \le j \le m} \quad \{\Omega \cup \{a_j : \psi_j\}\}_{1 \le j \le m} \quad \Omega \cup \{b : \diamond(\psi_1, \dots, \psi_m)\}}{\Omega \cup \{b : \diamond(\varphi_1, \dots, \varphi_m)\}}$$

Obviously, these $(2m+1)$-premise inference rules are not very convenient. More importantly: they are not analytic[9]. However, they can be simplified at the price of extending the language with a constant $\underline{a}$ for every $a \in \mathcal{V}$. In that case we can also resign from repeating the inference rules from the dynamic semantics, adding instead equivalent axioms for the constants:

DEFINITION 55. The proof system $SF^{sc}_{\mathcal{M}}$ for the static semantics of the language featuring constants consists of:

- **Axioms**: Each set of signed formulas containing either:
    1. $\{a : \varphi \mid a \in \mathcal{V}\}$, where φ is any formula in $\mathcal{W}$; or
    2. $\{a : \underline{a}\}$, for any $a \in \mathcal{V}$; or
    3. $\{b_1 : \diamond(\underline{a_1}, \dots, \underline{a_m}), \dots, b_k : \diamond(\underline{a_1}, \dots, \underline{a_m})\}$ for any m-ary connective $\diamond$ of $\mathcal{L}$ and any $a_1, \dots, a_m,\ b_1, \dots, b_k \in \mathcal{V}$ such that $\widetilde{\diamond}(a_1, \dots, a_m) = \{b_1, \dots, b_k\}$.
- **Inference rules:** For any $a_1, \dots, a_m, b \in \mathcal{V}$ and any m-ary connective $\diamond$ such that $b \in \widetilde{\diamond}(a_1, \dots, a_m)$, the rule ($\diamond$-SC):

$$\frac{\Omega \cup \{a_1 : \varphi_1\} \quad \dots \quad \Omega \cup \{a_m : \varphi_m\} \quad \Omega \cup \{b : \diamond(\underline{a_1}, \dots, \underline{a_m})\}}{\Omega \cup \{b : \diamond(\varphi_1, \dots, \varphi_m)\}}$$

[9] By an analytic rule we mean a rule which has some kind of a subformula property (see, e.g. [Baaz et. al., 2000]). This should not be confused with analycity of semantics (see Remark 12).

REMARK 56. Examining the generic deduction systems given above, we can easily observe that the inference rules of the static semantics really differ from those of the dynamic semantics only in case of truly non-deterministic values of the connectives. Indeed, if the value of the connective is a singleton, i.e. $\widetilde{\diamond}(\underline{a_1}, \ldots, \underline{a_m}) = \{b\}$, the rule ($\diamond$-S) is just a weaker version of ($\diamond$-D), and so need not be included in $SF^s_{\mathcal{M}}$. As for $SF^{sc}_{\mathcal{M}}$, the last premise of rule ($\diamond$-SC) is derivable in the system by virtue of the singleton set $\{b : \diamond(\underline{a_1}, \ldots, \underline{a_m})\}$ being an axiom — hence it can be skipped. As the other premises of the "static" and "dynamic" rules coincide, and so do the conclusions in such a "singleton" case, the rules can be considered identical. in this case the "static" Axiom 3 corresponding to such a singleton value of the connective can be deleted too, since it is derivable from rule ($\diamond$-D) and the basic axioms for the constants ("static" Axiom 2).

REMARK 57. It can easily be proved that the weakening rule is admissible in $SF^{sc}_{\mathcal{M}}$. This is the reason why it is not necessary to officially include it among the rules of this system.

The following generalizes a theorem from [Avron and Konikowska, 2005]:

THEOREM 58. *$SF^{sc}_{\mathcal{M}}$ is statically strongly characteristic for $\mathcal{M}$.*

COROLLARY 59. *$\Gamma \vdash^s_{\mathcal{M}} \Delta$ iff $\{\mathcal{D} : \psi \mid \psi \in \Gamma\} \cup \{\mathcal{F} : \psi \mid \psi \in \Delta\} \vdash_{SF^s_{\mathcal{M}}} \emptyset$. Moreover, if Γ and Δ are finite, then $\Gamma \vdash^s_{\mathcal{M}} \Delta$ iff $\vdash_{SF^{sc}_{\mathcal{M}}} \mathcal{F} : \Gamma \cup \mathcal{D} : \Delta$.*

4.2 Nmatrices for Canonical Calculi

In this subsection we provide, in a modular way, finite non-deterministic semantics for signed canonical calculi. Moreover, we show that there is an exact correspondence between the coherence of a canonical calculus G, the existence of a strongly characteristic Nmatrix for G, and analytic cut-elimination (Definition 46) in G. Then we focus on stronger notions of cut-elimination and show that coherence is not a sufficient condition for them. Therefore we define a stronger criterion of *density* which is a necessary and sufficient condition for strong cut-elimination in canonical calculi. Finally, we focus on the special case of Gentzen-type (two-signed) canonical calculi and show how the correspondence theorem can be used to provide a solution to the well-known "Tonk" problem of Prior ([Prior, 1960]).

Modular Semantics for Signed Canonical Calculi [10]

[10]This subsection is based on [Avron and Zamansky, 2009], where all proofs can be found.

We start by defining semantics for the simplest canonical calculus: the one without any canonical rules.

DEFINITION 60. $G_0^{(\mathcal{L},\mathcal{V})}$ is the canonical calculus over a language $\mathcal{L}$ and a set of signs $\mathcal{V}$, whose set of canonical rules is empty.

In the rest of this section we assume that our language $\mathcal{L}$, the set of signs $\mathcal{V}$, and the set of designated signs $\mathcal{D}$, are fixed. Accordingly, we shall write G_0 instead of $G_0^{(\mathcal{L},\mathcal{V})}$. It is easy to see that G_0 is (trivially) coherent. We now define a strongly characteristic Nmatrix for G_0. It has the maximal degree of non-determinism in interpreting the connectives of $\mathcal{L}$.

DEFINITION 61. $\mathcal{M}_0 = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$ is the Nmatrix in which $\tilde{\diamond}(a_1, ..., a_n) = \mathcal{V}$ for every n-ary connective $\diamond$ of $\mathcal{L}$ and $a_1, ..., a_n \in \mathcal{V}$.

THEOREM 62. *$\mathcal{M}_0$ is (dynamically) strongly characteristic for G_0.*

Now to the modular effects of canonical rules. The idea is that each rule which is added to G_0 imposes a certain semantic condition on refinements of $\mathcal{M}_0$, while coherence guarantees that these semantic conditions are not contradictory. This can be formalized as follows:

DEFINITION 63. For $\langle a_1, ..., a_n \rangle \in \mathcal{V}^n$, the set of clauses $\mathsf{C}_{\langle a_1,...,a_n \rangle}$ is defined as follows:

$$\mathsf{C}_{\langle a_1,...,a_n \rangle} = \{\{a_1 : p_1\}, \{a_2 : p_2\}, ..., \{a_n : p_n\}\}$$

DEFINITION 64. Let R be a canonical rule of the form $[\Theta/S : \diamond]$. $\mathsf{C}(\mathsf{R})$, the refining condition induced by R, is defined as follows:

$\mathsf{C}(\mathsf{R})$: For $a_1, ..., a_n \in \mathcal{V}$, if $\mathsf{C}_{\langle a_1,...,a_n \rangle} \cup \Theta$ is consistent, then $\tilde{\diamond}(a_1, ..., a_n) \subseteq S$.

Intuitively, a rule $[\Theta/S : \diamond]$ leads to the deletion from $\tilde{\diamond}(a_1, ..., a_n)$ of all the truth-values which are not in S. If some rules $[\Theta_1/S_1 : \diamond], ..., [\Theta_m/S_2 : \diamond]$ "overlap", their overall effect leads to $S_1 \cap ... \cap S_m$ (the coherence of a calculus guarantees that $S_1 \cap ... \cap S_m$ is not empty in such a case).

DEFINITION 65. Let G be a canonical calculus for $(\mathcal{L}, \mathcal{V})$.

1. Define an application of a rule $[\Theta/S : \diamond]$ of G for some n-ary connective $\diamond$ on $a_1, ..., a_n \in \mathcal{V}$ as follows:

$$[\Theta/S : \diamond](a_1, ..., a_n) = \begin{cases} S & \text{if } \Theta \cup \mathsf{C}_{\langle a_1,...,a_n \rangle} \text{ is consistent} \\ \mathcal{V} & \textit{otherwise} \end{cases}$$

2. $\mathcal{M}_G = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$ is any Nmatrix, such that for every n-ary connective $\diamond$ for $\mathcal{L}$ and every $a_1, ..., a_n \in \mathcal{V}$:

$$\tilde{\diamond}_{\mathcal{M}_G}(a_1, ..., a_n) = \bigcap \{[\Theta/S : \diamond](a_1, ..., a_n) \mid [\Theta/S : \diamond] \in G\}$$

PROPOSITION 66. *If G is coherent, then $\mathcal{M}_G$ is well-defined.*

REMARK 67. It is easy to see that for a coherent calculus G, $\mathcal{M}_G$ is the weakest refinement of $\mathcal{M}_0$, in which all the conditions induced by the rules of G are satisfied. Thus if G' is a coherent calculus obtained from G by adding a new canonical rule, $\mathcal{M}'_G$ can be straightforwardly obtained from $\mathcal{M}_G$ by some deletions of options as dictated by the condition which corresponds to the new rule.

THEOREM 68. *For every coherent canonical calculus G, $\mathcal{M}_G$ is a (dynamically) strongly characteristic Nmatrix for G.*

REMARK 69. The last theorem provides the converse of Theorem 51.

The next theorem is the most important result of this subsection. It establishes a quadruple correspondence between coherence of canonical calculi, non-deterministic matrices and analytic cut-elimination.

THEOREM 70. *Let G be a canonical calculus. The following statements concerning G are equivalent.*

1. *G is coherent.*
2. *G has a strongly characteristic Nmatrix.*
3. *G admits strong analytic cut-elimination.*
4. *G admits analytic cut-elimination.*

What about full (strong) cut-elimination? The next example shows that coherence is *not* a sufficient condition for it. Therefore a stronger condition is provided in the definition that follows that example.

EXAMPLE 71. Consider the calculus G' from Example 47. G' is obviously coherent. A proof of the sequent $\{b : p_1, c : p_1, b : (p_1 \circ p_2)\}$ is given in that example. However, this sequent clearly has no cut-free proof in G'.

DEFINITION 72. A canonical calculus G is *dense* if for every $a_1, ..., a_n \in \mathcal{V}$ and every two rules of G of the forms $[\Theta_1/S_1 : \diamond]$ and $[\Theta_2/S_2 : \diamond]$, such that $\Theta_1 \cup \Theta_2 \cup \mathsf{C}_{\langle a_1,...,a_n\rangle}$ is consistent, there is some rule $[\Theta/S : \diamond]$ in G, such that $\Theta \cup \mathsf{C}_{\langle a_1,...,a_n\rangle}$ is consistent and $S \subseteq S_1 \cap S_2$.

LEMMA 73. *Every dense canonical calculus is coherent.*

THEOREM 74. *Let G be a canonical calculus. Then the following statements concerning G are equivalent:*

1. *G is dense.*
2. *G admits cut-elimination.*
3. *G admits strong cut-elimination.*

Canonical Gentzen-type Calculi and Tonk

A very important class of canonical signed calculi is the class of canonical *ordinary* Gentzen-type calculi ([Avron and Lev, 2001; Avron and Lev, 2005]), i.e. calculi employing ordinary sequents of the form $\Gamma \Rightarrow \Delta$. As noted above, such calculi can be thought of as a special case of canonical signed calculi in which the set $\mathcal{V}$ of signs is $\{t, f\}$, and $\mathcal{D}$ is $\{t\}$. For this particular class, the criteria of coherence and density can be simplified, because the next proposition can easily be proved:

PROPOSITION 75. *A canonical ordinary Gentzen-type calculus G is* coherent *iff for every two canonical rules of G of the form $\Theta_1/\{t\} : \diamond$ and $\Theta_2/\{f\} : \diamond$, the set of clauses $\Theta_1 \cup \Theta_2$ is classically inconsistent (and so the empty sequent can be derived from it using cuts). Moreover, such a calculus is dense iff it is coherent.*

The following characterization theorem[11] easily follows from Theorems 70, 74, and Proposition 75:

THEOREM 76. *Let G be a canonical calculus with the set of signs $\mathcal{V} = \{t, f\}$. Then the following statements concerning G are equivalent:*

1. *G is coherent.*

[11]With the exception of the last item (concerning *strong* cut-elimination), this theorem was originally proved in [Avron and Lev, 2005].

2. *G is non-trivial.*

3. *G has a characteristic two-valued Nmatrix.*

4. *G admits cut-elimination.*

5. *G admits strong cut-elimination.*

Theorem 76 was used in [Avron and Lev, 2001; Avron and Lev, 2005] to provide a complete solution for the old "Tonk" problem of Prior in the multiple conclusion framework (the single conclusion case is handled in [Avron, 2010]). In [Prior, 1960] Prior strongly challenged the above mentioned Gentzen's thesis that the semantic meaning of a connective is determined by its introduction and elimination rules. He did that by introducing his famous binary "connective" Tonk (denoted below by T), which has two rules of the "ideal" type. The introduction rule allows to infer $\varphi\mathsf{T}\psi$ from φ. The elimination rule allows to infer ψ from $\varphi\mathsf{T}\psi$. In the presence of Tonk, every formula can be derived from any other formula, making trivial the "logic" that is "defined" by any system which includes this "connective". Prior's paper made it clear that not every combination of "ideal" introduction and elimination rules can be used for defining the semantic meaning of a connective, and some constraints should be imposed on the set of rules. Such a constraint was indeed suggested by Belnap in [Belnap, 1962]: the rules for a connective $\diamond$ should be *conservative*, in the sense that if $T \vdash \psi$ is derivable using them, and $\diamond$ does not occur in $T \cup \{\psi\}$, then $T \vdash \psi$ can also be derived without using the rules for $\diamond$. However, Belnap did not provide any effective necessary and sufficient criterion for checking whether a given set of rules is conservative in the above sense. Moreover: he formulated the condition of conservativity only with respect to the basic deduction framework, in which no connectives are assumed. Accordingly, nothing in what he wrote excludes the possibility of a system G having two connectives, each of them "defined" by a set of rules which is conservative over the basic system, while G itself is not conservative over it. To prevent this situation one should demand a much stronger conservativity condition than Belnap's, and it might not even be clear how it should be formulated. Later attempts of solutions of the Tonk problem insisted on closer connections between the introduction and the elimination rules for a given connective than those implicit in Belnap's condition of conservativity. Usually it is demanded that the introduction and elimination rules should precisely "match" (see, e.g., [Sundholm, 2002; Hodges, 2001]) in the sense that the elimination rules could be derived from the introduction rules by some syntactic procedure. From Theorem 76 it follows that this condition is too strong. What should be required from the set of rules is only *coherence*, which is an absolute (and minimal) condition for non-triviality. Tonk's rules indeed do not meet this condition: in the framework of canonical Gentzen-type systems its rules are translated into the following pair of rules: $\{\{f : p_1\}\}/\{f\} : \mathsf{T}$ and $\{\{t : p_2\}\}/\{t\} : \mathsf{T}$. This pair is not coherent, since the set $\{\{f : p_1\}, \{t : p_2\}\}$ is classically consistent. It is no wonder therefore that the resulting calculus is inconsistent. On the other hand every coherent set of canonical rules does indeed define a *unique* non-deterministic connective over $\{t, f\}$. This proves Gentzen's thesis at least in the multiple-conclusion canonical case. For further discussion and generalizations, we refer the reader to [Avron, 2010].

5 USING NMATRICES FOR NON-CANONICAL SYSTEMS

In the previous section we have applied finite Nmatrices in a modular way to characterize canonical calculi. The goal of this section is to show that the modular approach can be further extended and fruitfully applied (at least in many important cases) also to *non-canonical* Gentzen-type calculi. As our example we take the most common type of non-canonical rules that can be found in the literature: those which involve a combination of negation with other connectives. We investigate the semantic effects of rules of this type in the context of two major families of non-canonical Gentzen-type calculi: those that are obtained from the positive fragments of classical logic and intuitionistic logic by adding various natural Gentzen-type rules for negation. Not surprisingly, while Nmatrices suffice for providing adequate semantics for the first family, for the second one we need a combination of Nmatrices with intuitionistic Kripke frames. We demonstrate the power of this semantic tool by using it for solving the following important problem: given a system from the second family, determine whether or not it is a conservative extension of the positive fragment of intuitionistic logic.

The material of this section is based on [Avron, 2007b; Avron, 2005a].

5.1 Extensions of Classical Logic

In this section $\mathcal{L}$ denotes the propositional language $\{\wedge, \vee, \supset, \neg\}$, while $\mathcal{L}_{\mathbf{ff}}$ is the language obtained from $\mathcal{L}$ by adding the constant $\mathbf{ff}$. LK^+ denotes positive classical logic taken over $\mathcal{L}$, and LK denotes positive classical logic taken over $\mathcal{L}_{\mathbf{ff}}$. $\mathbf{G}[LK^+]$, the standard Gentzen-type (canonical) for LK^+, is given in Figure 2. $\mathbf{G}[LK]$, the Gentzen-type system for LK, is obtained from $\mathbf{G}[LK^+]$ by adding the sequent $\mathbf{ff} \Rightarrow$ as an additional axiom.

Axioms:

$$A \Rightarrow A$$

Structural Rules:

Cut, Weakening

Logical Rules:

$$\frac{\Gamma \Rightarrow \Delta, \psi \quad \varphi, \Gamma \Rightarrow \Delta}{\Gamma, \psi \supset \varphi \Rightarrow \Delta}\ (\supset\Rightarrow) \qquad \frac{\Gamma, \psi \Rightarrow \varphi, \Delta}{\Gamma \Rightarrow \psi \supset \varphi, \Delta}\ (\Rightarrow\supset)$$

$$\frac{\Gamma, \psi, \varphi \Rightarrow \Delta}{\Gamma, \psi \wedge \varphi \Rightarrow \Delta}\ (\wedge \Rightarrow) \qquad \frac{\Gamma \Rightarrow \psi, \Delta \quad \Gamma \Rightarrow \varphi, \Delta}{\Gamma \Rightarrow \psi \wedge \varphi, \Delta}\ (\Rightarrow \wedge)$$

$$\frac{\Gamma, \psi \Rightarrow \Delta \quad \Gamma, \varphi \Rightarrow \Delta}{\Gamma, \psi \vee \varphi \Rightarrow \Delta}\ (\vee \Rightarrow) \qquad \frac{\Gamma \Rightarrow \psi, \varphi, \Delta}{\Gamma \Rightarrow \psi \vee \varphi, \Delta}\ (\Rightarrow \vee)$$

Figure 2. The system LK

The table in Figure 3 lists the most common and natural logical rules for formulas involving negation and its combinations with other connectives (along with corresponding Hilbert-style axioms and Gentzen-style rules). Note that only the first two rules in this table are canonical.

DEFINITION 77. Denote by NIR the set of rules in Figure 3. For a logic $\mathbf{L}$ and $S \subseteq NIR$, let $\mathbf{L}[S]$ be the extension of $\mathbf{L}$ by S.

CONVENTION 78. For a rule R, denote by $\mathbf{H_R}$ its corresponding Hilbert-style axiom, and by $\mathbf{G_R}$ its corresponding Gentzen-style rule.

REMARK 79. It is easy to see that for every $S \subseteq NIR$ and $\mathbf{L} \in \{LK^+, LK\}$, a sound and complete Hilbert-style axiomatization for $\mathbf{L}[S]$ can be obtained by adding to some axiomatization of $\mathbf{L}$ the set of axioms $\{\mathbf{H_R} | \ \mathbf{R} \in S\}$, and similarly for Gentzen-type axiomatizations. We denote the resulting systems by $\mathbf{HL}[S]$ and $\mathbf{GL}[S]$, respectively.

Now we provide semantics for the systems introduced in Defn. 77. The basic idea is to let the value assigned to a sentence φ provide information not only about the truth/falsity of φ, but also about the truth/falsity of its negation. This leads to the use of elements from $\{0, 1\}^2$ as our truth-values, where the intended intuitive meaning of $v(\varphi) = \langle x, y\rangle$ is the following:

Rule	Abstract form	Hilbert-style axiom	Gentzen-style rule
$(\neg \Rightarrow)$	$\neg\varphi, \varphi \vdash$	$\neg\varphi \supset (\varphi \supset \psi)$	$\dfrac{\Gamma \Rightarrow \Delta, \varphi}{\Gamma, \neg\varphi \Rightarrow \Delta}$
$(\Rightarrow \neg)$	$\vdash \neg\varphi, \varphi$	$\neg\varphi \vee \varphi$	$\dfrac{\Gamma, \varphi \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \neg\varphi}$
$(\neg\neg \Rightarrow)$	$\neg\neg\varphi \vdash \varphi$	$\neg\neg\varphi \supset \varphi$	$\dfrac{\Gamma, \varphi \Rightarrow \Delta}{\Gamma, \neg\neg\varphi \Rightarrow \Delta}$
$(\Rightarrow \neg\neg)$	$\varphi \vdash \neg\neg\varphi$	$\varphi \supset \neg\neg\varphi$	$\dfrac{\Gamma \Rightarrow \Delta, \varphi}{\Gamma \Rightarrow \Delta, \neg\neg\varphi}$
$(\neg \supset\Rightarrow)_1$	$\neg(\varphi \supset \psi) \vdash \varphi$	$\neg(\varphi \supset \psi) \supset \varphi$	$\dfrac{\Gamma, \varphi \Rightarrow \Delta}{\Gamma, \neg(\varphi \supset \psi) \Rightarrow \Delta}$
$(\neg \supset\Rightarrow)_2$	$\neg(\varphi \supset \psi) \vdash \neg\psi$	$\neg(\varphi \supset \psi) \supset \neg\psi$	$\dfrac{\Gamma, \neg\psi \Rightarrow \Delta}{\Gamma, \neg(\varphi \supset \psi) \Rightarrow \Delta}$
$(\Rightarrow \neg \supset)$	$\varphi, \neg\psi \vdash \neg(\varphi \supset \psi)$	$\varphi \supset (\neg\psi \supset \neg(\varphi \supset \psi))$	$\dfrac{\Gamma \Rightarrow \Delta, \varphi \quad \Gamma \Rightarrow \neg\psi}{\Gamma \Rightarrow \Delta, \neg(\varphi \supset \psi)}$
$(\neg\vee \Rightarrow)_1$	$\neg(\varphi \vee \psi) \vdash \neg\varphi$	$\neg(\varphi \vee \psi) \supset \neg\varphi$	$\dfrac{\Gamma, \neg\varphi \Rightarrow \Delta}{\Gamma, \neg(\varphi \vee \psi) \Rightarrow \Delta}$
$(\neg\vee \Rightarrow)_2$	$\neg(\varphi \vee \psi) \vdash \neg\psi$	$\neg(\varphi \vee \psi) \supset \neg\psi$	$\dfrac{\Gamma, \neg\psi \Rightarrow \Delta}{\Gamma, \neg(\varphi \vee \psi) \Rightarrow \Delta}$
$(\Rightarrow \neg\vee)$	$\neg\varphi, \neg\psi \vdash \neg(\psi \vee \varphi)$	$(\neg\varphi \wedge \neg\psi) \supset \neg(\varphi \vee \psi)$	$\dfrac{\Gamma \Rightarrow \Delta, \neg\psi \quad \Gamma \Rightarrow \Delta, \neg\varphi}{\Gamma \Rightarrow \Delta, \neg(\varphi \vee \psi)}$
$(\neg\wedge \Rightarrow)$	$\neg(\varphi \wedge \psi) \vdash \neg\varphi, \neg\psi$	$\neg(\varphi \wedge \psi) \supset (\neg\varphi \vee \neg\psi)$	$\dfrac{\Gamma, \neg\psi \Rightarrow \Delta \quad \Gamma, \neg\varphi \Rightarrow \Delta}{\Gamma, \neg(\varphi \wedge \psi) \Rightarrow \Delta}$
$(\Rightarrow \neg\wedge)_1$	$\neg\varphi \vdash \neg(\varphi \wedge \psi)$	$\neg\varphi \supset \neg(\varphi \wedge \psi)$	$\dfrac{\Gamma \Rightarrow \Delta, \neg\varphi}{\Gamma \Rightarrow \Delta, \neg(\varphi \wedge \psi)}$
$(\Rightarrow \neg\wedge)_2$	$\neg\psi \vdash \neg(\varphi \wedge \psi)$	$\neg\psi \supset \neg(\varphi \wedge \psi)$	$\dfrac{\Gamma \Rightarrow \Delta, \neg\psi}{\Gamma \Rightarrow \Delta, \neg(\varphi \wedge \psi)}$

Figure 3. The set of rules NIR

- $x = 1$ iff φ is "true" (i.e. $v(\varphi) \in \mathcal{D}$).
- $y = 1$ iff $\neg\varphi$ is "true" (i.e. $v(\neg\varphi) \in \mathcal{D}$).

This interpretation of the truth-values dictates the following constraint on any valuation v (where $P_1(\langle x_1, x_2\rangle) = x_1$, and $P_2(\langle x_1, x_2\rangle) = x_2$):

$$P_1(v(\neg\varphi)) = P_2(v(\varphi))$$

In terms of Nmatrices this constraint translates into the condition:

$$\text{(NEG)} \qquad \widetilde{\neg} a \subseteq \{y \mid P_1(y) = P_2(a)\}$$

We start our semantic investigation of NIR with the weakest Nmatrix which satisfies Condition (NEG) and has the standard interpretation of $\supset, \wedge$, and $\vee$ (since the standard rules for these connectives are in LK^+).

DEFINITION 80. Let $\mathcal{M}_4^B = \langle \mathcal{V}_4, \mathcal{D}_4, \mathcal{O}_4\rangle$ be the following Nmatrix for $\mathcal{L}$:

- $\mathcal{V}_4 = \{t, \top, \bot, f\}$[12] where:

$$\begin{array}{rcl} t & = & \langle 1, 0\rangle \\ \top & = & \langle 1, 1\rangle \\ \bot & = & \langle 0, 0\rangle \\ f & = & \langle 0, 1\rangle \end{array}$$

- $\mathcal{D}_4 = \{a \in \mathcal{V}_4 \mid P_1(a) = 1\} = \{t, \top\}$
- Let $\mathcal{V} = \mathcal{V}_4$, $\mathcal{D} = \mathcal{D}_4$, $\mathcal{F} = \mathcal{V}_4 - \mathcal{D}$. The operations in $\mathcal{O}_4$ are:

$$\widetilde{\neg} a = \begin{cases} \mathcal{D} & \text{if } P_2(a) = 1 \quad (\text{i.e. } a \in \{f, \top\}) \\ \mathcal{F} & \text{if } P_2(a) = 0 \quad (\text{i.e. } a \in \{t, \bot\}) \end{cases}$$

$$a \widetilde{\supset} b = \begin{cases} \mathcal{D} & \text{if } a \in \mathcal{F} \text{ or } b \in \mathcal{D} \\ \mathcal{F} & \text{otherwise} \end{cases}$$

$$a \widetilde{\vee} b = \begin{cases} \mathcal{D} & \text{if } a \in \mathcal{D} \text{ or } b \in \mathcal{D} \\ \mathcal{F} & \text{otherwise} \end{cases}$$

$$a \widetilde{\wedge} b = \begin{cases} \mathcal{D} & \text{if } a \in \mathcal{D} \text{ and } b \in \mathcal{D} \\ \mathcal{F} & \text{otherwise} \end{cases}$$

[12] The intuition behind these four truth-values is like in Dunn-Belnap's logic, see the end of the Introduction.

$\mathcal{M}_4^{B\mathbf{ff}}$ (for $\mathcal{L}_{\mathbf{ff}}$) is obtained from $\mathcal{M}_4^B$ by adding the condition: $\widetilde{\mathbf{ff}} \in \mathcal{F}$.

THEOREM 81. *$\mathcal{M}_4^B$ ($\mathcal{M}_4^{B\mathbf{ff}}$) is a (dynamically) characteristic Nmatrix for LK^+ (LK).*

Now each rule of NIR induces a semantic condition, and $\mathbf{L}[S]$ ($\mathbf{L} \in \{LK^+, LK\}$) is characterized by the simple refinement (Remark 32) of $\mathcal{M}_4^B/\mathcal{M}_4^{B\mathbf{ff}}$, induced by the conditions that correspond to the rules in S.

DEFINITION 82. *The refining conditions induced by the rules in NIR:*

C($\neg \Rightarrow$) : If $P_1(a) = 1$ then $P_2(a) = 0$

C($\Rightarrow \neg$) : If $P_1(a) = 0$ then $P_2(a) = 1$

C($\Rightarrow \neg\neg$) : If $P_1(a) = 1$ then $\widetilde{\neg} a \subseteq \{x \mid P_2(x) = 1\}$

C($\neg\neg \Rightarrow$) : If $P_1(a) = 0$ then $\widetilde{\neg} a \subseteq \{x \mid P_2(x) = 0\}$

C($\neg \supset\Rightarrow$)$_1$: If $P_1(a) = 0$ then $a \widetilde{\supset} b \subseteq \{x \mid P_2(x) = 0\}$

C($\neg \supset\Rightarrow$)$_2$: If $P_2(b) = 0$ then $a \widetilde{\supset} b \subseteq \{x \mid P_2(x) = 0\}$

C($\Rightarrow \neg \supset$) : If $P_1(a) = 1$ and $P_2(b) = 1$ then $a \widetilde{\supset} b \subseteq \{x \mid P_2(x) = 1\}$

C($\neg\vee \Rightarrow$)$_1$: If $P_2(a) = 0$ then $a \widetilde{\vee} b \subseteq \{x \mid P_2(x) = 0\}$

C($\neg\vee \Rightarrow$)$_2$: If $P_2(b) = 0$ then $a \widetilde{\vee} b \subseteq \{x \mid P_2(x) = 0\}$

C($\Rightarrow \neg\vee$) : If $P_2(a) = 1$ and $P_2(b) = 1$ then $a \widetilde{\vee} b \subseteq \{x \mid P_2(x) = 1\}$

C($\Rightarrow \neg\wedge$)$_1$: If $P_2(a) = 1$ then $a \widetilde{\wedge} b \subseteq \{x \mid P_2(x) = 1\}$

C($\Rightarrow \neg\wedge$)$_2$: If $P_2(b) = 1$ then $a \widetilde{\wedge} b \subseteq \{x \mid P_2(x) = 1\}$

C($\neg\wedge \Rightarrow$) : If $P_2(a) = 0$ and $P_2(b) = 0$ then $a \widetilde{\vee} b \subseteq \{x \mid P_2(x) = 0\}$

As an example how these conditions have been derived, take ($\neg \supset\Rightarrow$)$_2$. This rule is valid if $\neg(a \supset b)$ is in $\mathcal{F}$ whenever $\neg b$ is in $\mathcal{F}$ (where $x \supset y$ denotes some element in $x \widetilde{\supset} y$, and $\neg x$ denotes some element in $\widetilde{\neg} x$). This is equivalent to: if $P_2(b) = 0$ then $P_2(a \supset b) = 0$, which is exactly C($\neg \supset\Rightarrow$)$_2$.

REMARK 83. With the obvious extensions of P_1 and P_2, The above formulation of the conditions in $C(NIR)$ can be applied whenever the truth-values are finite sequences of 0's and 1's, the designated elements are those for which the first component is 1, and condition (NEG) is satisfied. However, these conditions can be simplified in case exactly $\{t, f, \top, \bot\}$ are used. Thus the conditions involving $\neg$ and $\supset$ can be reformulated as follows:

C$(\neg \Rightarrow)$: Use only t, f and $\bot$

C$(\Rightarrow \neg)$: Use only t, f and $\top$

C$(\Rightarrow \neg\neg)$: $\tilde{\neg} t = \{f\}$, $\tilde{\neg}\top = \{\top\}$

C$(\neg\neg \Rightarrow)$: $\tilde{\neg} f = \{t\}$, $\tilde{\neg} \bot = \{\bot\}$

C$(\neg \supset\Rightarrow)_1$: If $a \in \mathcal{F}$ then $a \tilde{\supset} b \subseteq \{t, \bot\}$

C$(\neg \supset\Rightarrow)_2$: If $b \in \{t, \bot\}$ then $a \tilde{\supset} b \subseteq \{t, \bot\}$

C$(\Rightarrow \neg \supset)$: If $a \in \mathcal{D}$ and $b \in \{\top, f\}$ then $a \tilde{\supset} b \subseteq \{\top, f\}$

Moreover, if we consider only simple refinements of $\mathcal{M}_4^B$, then the three last conditions can be further transformed into more specific ones:

C$(\neg \supset\Rightarrow)_1$: If $a \in \mathcal{F}$ then $a \tilde{\supset} b = \{t\}$

C$(\neg \supset\Rightarrow)_2$: If $b = t$ then $a \tilde{\supset} b = \{t\}$
If $b = \bot$ and $a \in \mathcal{F}$ then $a \tilde{\supset} b = \{t\}$
If $b = \bot$ and $a \in \mathcal{D}$ then $a \tilde{\supset} b = \{\bot\}$

C$(\Rightarrow \neg \supset)$: If $a \in \mathcal{D}$ and $b \in \{\top, f\}$ then $a \tilde{\supset} b = \{b\}$

DEFINITION 84.

1. For $S \subseteq NIR$, let $C(S) = \{\mathrm{C}r \mid r \in S\}$
2. For $S \subseteq NIR$, let $\mathcal{M}_S$ ($\mathcal{M}_S^{\mathbf{ff}}$) be the weakest simple refinement of $\mathcal{M}_4^B$ ($\mathcal{M}_4^{B\mathbf{ff}}$) in which the conditions in $C(S)$ are all satisfied. In other words: $\mathcal{M}_S = \langle \mathcal{V}_S, \mathcal{D}_S, \mathcal{O}_S \rangle$, where $\mathcal{V}_S$ is the set of values from $\mathcal{V}_4$ which are not rejected by any condition in S, $\mathcal{D}_S = \mathcal{D}_4 \cap \mathcal{V}_S$, and for any connective $\diamond$ and $\vec{x} \in \mathcal{V}_S^n$ (where n is the arity of $\diamond$), the interpretation in $\mathcal{O}$ of $\diamond$ assigns to $\vec{x}$ the set of all the values in $\tilde{\diamond}(\vec{x})$ which are not forbidden by any condition in $C(S)$ (it is easy to check that for $S \subseteq NIR$ this set is never empty. The same is true for $\mathcal{D}_S$).

EXAMPLE 85.

1. Let $C_{min} = LK^+[\{(\Rightarrow \neg), (\neg\neg \Rightarrow)\}]$. Then $\mathcal{M}_{C_{min}}$ is the three-valued Nmatrix[13] $\langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$, where:
   - $\mathcal{V} = \{t, \top, f\}$ (the rule $(\Rightarrow \neg)$ causes the deletion of $\bot$)

[13] C_{min} is studied in [Carnielli and Marcos, 1999]. The 3-valued Nmatrix for this logic described here was first introduced in [Avron and Lev, 2005].

- $\mathcal{D} = \{t, \top\}$
- The operations in $\mathcal{O}$ are:

$$\tilde{\neg} a = \begin{cases} \mathcal{D} & \text{if } a = \top \\ \{f\} & \text{if } a = t \\ \{t\} & \text{if } a = f \end{cases}$$

$$a \tilde{\supset} b = \begin{cases} \mathcal{D} & \text{if } a = f \text{ or } b \in \mathcal{D} \\ \{f\} & \text{otherwise} \end{cases}$$

$$a \tilde{\vee} b = \begin{cases} \mathcal{D} & \text{if } a \in \mathcal{D} \text{ or } b \in \mathcal{D} \\ \{f\} & \text{otherwise} \end{cases}$$

$$a \tilde{\wedge} b = \begin{cases} \mathcal{D} & \text{if } a, b \in \mathcal{D} \\ \{f\} & \text{otherwise} \end{cases}$$

2. Let $\mathcal{FOUR} = NIR - \{(\neg \Rightarrow), (\Rightarrow \neg)\}$. Then $\mathcal{M}_{\mathcal{FOUR}}$ is a 4-valued *deterministic* Nmatrix (i.e. an ordinary *matrix*). The operations in this matrix are defined as follows (where $a \leq_t \bot, \top \leq_t t$): [14]

$$\tilde{\neg} t = f \quad \tilde{\neg} f = t \quad \tilde{\neg} \top = \top \quad \tilde{\neg} \bot = \bot$$

$$a \tilde{\supset} b = \begin{cases} t & \text{if } a \notin \mathcal{D} \\ b & \text{otherwise} \end{cases}$$

$$a \tilde{\wedge} b = inf_{\leq_t}\{a, b\} \qquad a \tilde{\vee} b = sup_{\leq_t}\{a, b\}$$

Now the modular character of the semantics of Nmatrices allows us to formulate and prove together soundness and completeness theorems for 2^{13} systems (most of which define different logics):

THEOREM 86. *For $S \subseteq NIR$, $\mathcal{M}_S$ ($\mathcal{M}_S^{\mathbf{ff}}$) is a (dynamically) characteristic Nmatrix for $LK^+[S]$ ($LK[S]$).*

Proof. We give an outline of the completeness part. So let $\mathbf{L} \in \{LK^+, LK\}$, and assume that $\mathcal{T} \not\vdash_{\mathbf{L}[S]} \psi_0$. We construct a model of $\mathcal{T}$ in $\mathcal{M}_S$ which is not a model of ψ_0. For this extend $\mathcal{T}$ to a maximal set $\mathcal{T}^*$ of formulas such that $\mathcal{T}^* \not\vdash_{\mathbf{L}[S]} \psi_0$. Then $\varphi \notin \mathcal{T}^*$ iff $\mathcal{T}^*, \varphi \vdash_{\mathbf{L}[S]} \psi_0$. Define now a valuation v by $v(\varphi) = \langle x(\varphi), y(\varphi) \rangle$, where:

$$x(\varphi) = \begin{cases} 1 & \varphi \in \mathcal{T}^* \\ 0 & \varphi \notin \mathcal{T}^* \end{cases} \qquad y(\varphi) = \begin{cases} 1 & \neg\varphi \in \mathcal{T}^* \\ 0 & \neg\varphi \notin \mathcal{T}^* \end{cases}$$

[14] Without $\tilde{\supset}$, $\mathcal{M}_{\mathcal{FOUR}}$ is the famous 4-valued matrix of Dunn and Belnap ([Dunn, 1976; Belnap, 1977]). The connective $\tilde{\supset}$ of $\mathcal{O}_4$ was introduced in [Arieli and Avron, 1996]. The soundness and completeness of the logic $\mathcal{FOUR}$ for $\mathcal{M}_{\mathcal{FOUR}}$ was also first stated and proved there.

It is not difficult to see that v is a legal valuation in $\mathcal{M}_4^B$. To show that it is also a legal valuation in $\mathcal{M}_S$, we need to check that it respects the conditions in $C(S)$ (as formulated in Remark 83). We do some of the cases, leaving the rest for the reader:

C($\neg \Rightarrow$) : Assume $(\neg \Rightarrow) \in S$. Then there can be no sentence φ such that $\{\varphi, \neg\varphi\} \subseteq \mathcal{T}^*$. Hence $v(\varphi) \neq \top$ for all φ.

C($\Rightarrow \neg$) : Assume $(\Rightarrow \neg) \in S$, but $v(\varphi) = \perp$ for some φ. Then $\varphi \notin \mathcal{T}^*$ and $\neg\varphi \notin \mathcal{T}^*$. It follows that $\mathcal{T}^*, \varphi \vdash_{\mathbf{L}[S]} \psi_0$ and $\mathcal{T}^*, \neg\varphi \vdash_{\mathbf{L}[S]} \psi_0$. Hence $\mathcal{T}^* \vdash_{\mathbf{L}[S]} \varphi \supset \psi_0$, and $\mathcal{T}^* \vdash_{\mathbf{L}[S]} \neg\varphi \supset \psi_0$. This contradicts the fact that $\mathcal{T}^* \not\vdash_{\mathbf{L}[S]} \psi_0$, since $\varphi \supset \psi_0, \neg\varphi \supset \psi_0 \vdash_{\mathbf{L}[S]} \psi_0$ in case $(\neg \Rightarrow) \in S$.

C($\Rightarrow \neg\neg$) : Assume $(\Rightarrow \neg\neg) \in S$.

- Suppose $v(\varphi) = t$. Then $\varphi \in \mathcal{T}^*$ and $\neg\varphi \notin \mathcal{T}^*$. By $(\Rightarrow \neg\neg)$, also $\neg\neg\varphi \in \mathcal{T}^*$. Hence $v(\neg\varphi) = f$ by definition of v.
- Suppose $v(\varphi) = \top$. Then $\varphi \in \mathcal{T}^*$ and $\neg\varphi \in \mathcal{T}^*$. By $(\Rightarrow \neg\neg)$, also $\neg\neg\varphi \in \mathcal{T}^*$. Hence $v(\neg\varphi) = \top$ by definition of v.

C($\neg \supset\Rightarrow$)$_1$: Assume $(\neg \supset\Rightarrow)_1 \in S$. Suppose that $v(\varphi) \notin \mathcal{D}$. Then $\varphi \notin \mathcal{T}^*$, and so also $\neg(\varphi \supset \psi) \notin \mathcal{T}^*$. It follows that $v(\varphi \supset \psi) \in \{t, \perp\}$.

Obviously, v is a model of $\mathcal{T}$ in $\mathcal{M}_S$ which is not a model of ψ_0. ■

The following corollary is implied by the above theorem and the analycity of Nmatrices:

COROLLARY 87. *$LK^+[S]$ and $LK[S]$ are decidable for every $S \subseteq NIR$.*

5.2 Extensions of Intuitionistic Logic

Let LJ denote propositional intuitionistic logic (over $\{\wedge, \vee, \supset, \mathbf{ff}\}$), and let LJ^+ be its positive fragment (i.e. its $\{\wedge, \vee, \supset\}$-fragment). Next we investigate extensions of LJ^+ and LJ by a negation connective $\neg$.[15] Now, it is well known that it is impossible to conservatively add to LJ^+ or LJ a connective $\neg$ which is both explosive (i.e.: $\neg A, A \vdash B$ for all A, B) and satisfies the law

[15]Positive intuitionistic logic might be a better starting point for investigating negations than positive classical logic (especially *constructive* negations), because its valid sentences are all intuitively correct. LK^+, in contrast, includes counterintuitive tautologies like $(A \wedge B \supset C) \supset (A \supset C) \vee (B \supset C)$ or $A \vee (A \supset B)$. Moreover: the classical natural deduction rules for the positive connectives ($\wedge, \vee$ and $\supset$) define LJ^+, not LK^+. It is only with the aid of the classical rules for (the classical) negation that one can prove the counterintuitive positive tautologies mentioned above.

of excluded middle LEM. With such an addition we get classical logic. The intuitionists indeed reject LEM, retaining the explosive nature of negation (which is usually defined by $\sim\varphi =^{Def} \varphi \supset \mathbf{ff}$). In this subsection we show that this is not the only possible choice. The main problem we shall solve is: Which of the logics $LJ^+[S]$ ($S \subseteq NIR$) is conservative over LJ^+ (and similarly for LJ)? We believe that each such logic is entitled to be called "a logic with a constructive negation".

REMARK 88. $\mathbf{G}[LJ^+]$ ($\mathbf{G}[LJ]$), a multiple-conclusioned Gentzen-type system for LJ^+ (LJ), is obtained from $\mathbf{G}[LK^+]$ ($\mathbf{G}[LK]$) by replacing the ($\Rightarrow\supset$) rule with the following impure (single-conclusion) rule:

$$\frac{\Gamma, A \Rightarrow B}{\Gamma \Rightarrow A \supset B}\ (\Rightarrow\supset)$$

It is again easy to see that for every $S \subseteq NIR$ and $\mathbf{L} \in \{LJ^+, LJ\}$, a Gentzen-type system $\mathbf{GL}[S]$ which is sound and complete for $\mathbf{L}[S]$ can be obtained by adding to $\mathbf{GL}$ the Gentzen-type versions of the rules in S. In what follows we identify $\mathbf{L}[S]$ and $\mathbf{GL}[S]$.

Like in the classical case, we start by generalizing the standard, two-valued semantics of LJ^+ (or LJ). Recall that this semantics is usually provided by the class of all Kripke frames of the form $\mathcal{W} = \langle W, \leq, v\rangle$ [16], where $\langle W, \leq\rangle$ is a nonempty partially ordered set (of "worlds"), and v is a function from $W \times Frm_{\mathcal{L}}$ to $\mathcal{V}$ that satisfies the following conditions:

1. If $y \geq x$ and $v(x,\varphi) = t$ then $v(y,\varphi) = t$.[17]
2.
   - $v(x, \varphi \wedge \psi) = t$ iff $v(x,\varphi) = t$ and $v(x,\psi) = t$
   - $v(x, \varphi \vee \psi) = t$ iff $v(x,\varphi) = t$ or $v(x,\psi) = t$
   - $v(x, \mathbf{ff}) = f$ (if $\mathbf{ff}$ is in the language).
3. $v(x, \mathbf{ff}) = f$ (if $\mathbf{ff}$ is in the language).
4. $v(x, \varphi \supset \psi) = t$ iff $v(y,\psi) = t$ for every $y \geq x$ such that $v(y,\varphi) = t$

[16] In the literature by a "frame" one usually means just the pair $\langle W, \leq\rangle$. Here we have found it convenient to use this technical term differently, so that the valuation v is an integral part of it.

[17] For the language of LJ it suffices to demand this condition for atomic formulas only; then one can prove that every formula has this property. This is not the case for the nondeterministic generalizations with $\neg$ that we present below.

Obviously, if $\mathcal{W} = \langle W, \leq, v\rangle$ is a frame, then for every $x \in W$ the function $\lambda\varphi.v(x,\varphi)$ behaves like an ordinary classical valuation with respect to all the connectives except $\supset$. The treatment of $\supset$ is indeed what distinguishes between classical logic and intuitionistic logic. This observation leads to the following nondeterministic generalization of intuitionistic Kripke frames:

DEFINITION 89. Let $\supset$ be one of the connectives of a propositional language $\mathcal{L}$, and let $\mathcal{M} = \langle \mathcal{V}, \mathcal{D}, \mathcal{O}\rangle$ be an Nmatrix for $\mathcal{L} - \{\supset\}$. An $\mathcal{M}$-*frame* for $\mathcal{L}$ is a triple $\mathcal{W} = \langle W, \leq, v\rangle$ such that:

1. $\langle W, \leq\rangle$ is a nonempty partially ordered set
2. $v : W \times Frm_{\mathcal{L}} \to \mathcal{V}$ satisfies the following conditions:
   - *Persistence:* if $y \geq x$ and $v(x,\varphi) \in \mathcal{D}$ then $v(y,\varphi) \in \mathcal{D}$
   - For every $x \in W$, $\lambda\varphi.v(x,\varphi)$ is a legal $\mathcal{M}$-valuation.
   - $v(x, \varphi \supset \psi) \in \mathcal{D}$ iff $v(y,\psi) \in \mathcal{D}$ for every $y \geq x$ such that $v(y,\varphi) \in \mathcal{D}$

We say that a formula φ is *true* in a world $x \in W$ of a frame $\mathcal{W}$ if $v(x,\varphi) \in \mathcal{D}$. A sequent $\Gamma \Rightarrow \Delta$ is *valid* in $\mathcal{W}$ if for every $x \in W$ there is either $\varphi \in \Gamma$ such that φ is not true in x, or $\psi \in \Delta$ such that ψ is true in x.

Obviously, if $\mathcal{M}_1$ is a refinement of $\mathcal{M}_2$, then any $\mathcal{M}_1$-frame is also an $\mathcal{M}_2$-frame, and every sequent valid in $\mathcal{M}_2$ is also valid in $\mathcal{M}_1$.

DEFINITION 90.

1. Let $\mathcal{M} = \langle \mathcal{V}, \mathcal{D}, \mathcal{O}\rangle$ be an Nmatrix for a language which includes the language of LJ^+. We say that $\mathcal{M}$ is *suitable* for LJ^+ if the following conditions are satisfied (where again $\mathcal{V} - \mathcal{D}$ is denoted by $\mathcal{F}$):
   - If $a \in \mathcal{D}$ and $b \in \mathcal{D}$ then $a \wedge b \subseteq \mathcal{D}$
   - If $a \notin \mathcal{D}$ then $a \wedge b \subseteq \mathcal{F}$
   - If $b \notin \mathcal{D}$ then $a \wedge b \subseteq \mathcal{F}$

   - If $a \in \mathcal{D}$ then $a \vee b \subseteq \mathcal{D}$
   - If $b \in \mathcal{D}$ then $a \vee b \subseteq \mathcal{D}$
   - If $a \notin \mathcal{D}$ and $b \notin \mathcal{D}$ then $a \vee b \subseteq \mathcal{F}$

   - If $b \in \mathcal{D}$ then $a \supset b \subseteq \mathcal{D}$
   - If $a \in \mathcal{D}$ and $b \notin \mathcal{D}$ then $a \supset b \subseteq \mathcal{F}$

2. Let $\mathcal{M} = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$ be an Nmatrix for a language which includes the language of LJ. We say that $\mathcal{M}$ is *suitable* for LJ if it is suitable for LJ^+, and the following condition is satisfied:
   - $\mathbf{ff} \subseteq \mathcal{F}$

THEOREM 91. *Assume $\mathcal{W}$ is an $\mathcal{M}$-frame, where $\mathcal{M}$ is suitable for LJ^+ (LJ). Then any sequent provable in LJ^+ (LJ) is valid in $\mathcal{W}$.*

Below we concentrate on the systems $LJ^+(S)$ for $S \subseteq NIR$ (obtaining similar results for $LJ(S)$ causes no further difficulties).

DEFINITION 92. Let $\mathcal{M}_4^{IB}$ be the following Nmatrix $\langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$ for $\mathcal{L}$:

- $\mathcal{V} = \{t, \top, f, \bot\}$
- $\mathcal{D} = \{t, \top\}$
- $a \supset b = \begin{cases} \mathcal{D} & b \in \mathcal{D} \\ \mathcal{F} & b \notin \mathcal{D}, a \in \mathcal{D} \\ \mathcal{V} & a, b \in \mathcal{F} \end{cases}$

  $a \vee b = \begin{cases} \mathcal{D} & a \in \mathcal{D} \text{ or } b \in \mathcal{D} \\ \mathcal{F} & \text{otherwise} \end{cases}$

  $a \wedge b = \begin{cases} \mathcal{D} & a, b \in \mathcal{D} \\ \mathcal{F} & \text{otherwise} \end{cases}$

  $\neg t = \neg \bot = \mathcal{F} \quad \neg f = \neg \top = \mathcal{D}$

REMARK 93. The only difference between $\mathcal{M}_4^{IB}$ and $\mathcal{M}_4^{B}$ (recall Defn. 80) is that in $\mathcal{M}_4^{IB}$ we have $a \supset b = \mathcal{V}$ in case $a, b \in \mathcal{F} = \mathcal{V} - \mathcal{D}$, while in $\mathcal{M}_4^{B}$ $a \supset b = \mathcal{D}$ in this case.

PROPOSITION 94. *Let $\mathcal{M}$ be a refinement of $\mathcal{M}_4^{IB}$. Then LJ^+ is sound for every $\mathcal{M}$-frame.*

Now we turn to the effects of the various negation rules in the context of our semantics for LJ^+ and its extensions. The conditions we associated with these conditions in the previous subsection lead this time to refinements of $\mathcal{M}_4^{IB}$ on which the corresponding frames are based.

DEFINITION 95. For $S \subseteq NIR$, let $\mathcal{M}_S^I$ be the weakest refinement of $\mathcal{M}_4^{IB}$ in which the conditions in $C(S)$ are satisfied.[18]

[18]It is advisable here to read again the first part of Remark 83.

THEOREM 96. *If $S \subseteq NIR$ then $LJ^+(S)$ is sound and strongly complete for $\mathcal{M}^I_S$-frames: $\Gamma \vdash_{LJ^+(S)} \psi$ iff for every $\mathcal{M}^I_S$-frame $\mathcal{W} = \langle W, \leq, v \rangle$, and every $x \in W$, if $v(x, \varphi) \in \mathcal{D}$ for every $\varphi \in \Gamma$ then also $v(x, \psi) \in \mathcal{D}$.*

EXAMPLE 97. The system C_ω of da Costa ([da Costa, 1974]) is identical to $LJ^+(\{(\Rightarrow \neg), (\neg\neg \Rightarrow)\})$. Theorem 96 provides illuminating semantics for it which is much simpler than the Kripke-type semantics given in [Baaz, 1986] and the bivaluations semantics of [Loparić, 1986] (and can be used for a decision procedure — see Corrolary 104 below). Here is a compact description of this semantics: A frame for C_ω is a triple $\langle W, \leq, v \rangle$ such that $\langle W, \leq \rangle$ is a nonempty partially ordered set, and $v : W \times \mathcal{F} \to \{t, f, \top\}$ is a valuation which satisfies the following conditions:

- If $x \leq y$ then $v(x, \varphi) \leq_k v(y, \varphi)$
- $v(x, \varphi \wedge \psi) = f$ iff $v(x, \varphi) = f$ or $v(x, \psi) = f$
- $v(x, \varphi \vee \psi) = f$ iff $v(x, \varphi) = f$ and $v(x, \psi) = f$
- $v(x, \varphi \supset \psi) = f$ iff for some $y \geq x$, $v(y, \varphi) \neq f$ while $v(y, \psi) = f$
- $v(x, \neg\varphi) = f$ iff $v(x, \varphi) = t$
- If $v(x, \varphi) = f$ then $v(x, \neg\varphi) = t$

A frame is a model of a formula φ if $v(x, \varphi) \neq f$ for every $x \in W$.

Theorem 96 does not have much value in itself. Indeed, it does not guarantee that $LJ^+(S)$ is conservative over LJ^+, and neither does it provide a decision procedure for $LJ^+(S)$. The reason for this is that the current semantic framework (of Nmatrices combined with intuitionistic frames) is not always analytic (recall Remark 12). Next we provide a definition of this notion which is suitable for the present context. For this we need first the following important observation.

PROPOSITION 98. *Let $\mathcal{M}$ be a refinement of $\mathcal{M}^{IB}_4$. Then the persistence condition for $\mathcal{M}$ is equivalent to the following* monotonicity condition*:*

- If $x \leq y$ then $v(x, \varphi) \leq_k v(y, \varphi)$, where the partial order $\leq_k$ on $\mathcal{V}_4$ is defined by: $\bot \leq_k t, f \leq_k \top$. [19]

[19] $\leq_k$ had a crucial role already in [Belnap, 1977]. The structure obtained by equipping $\mathcal{V}_4$ with both $\leq_t$ and $\leq_k$ is nowadays known as the basic (distributive) *bilattice* (see [Ginsberg, 1988; Fitting, 1994; Arieli and Avron, 1996]).

Analycity for the semantics of frames can now be defined as follows:

DEFINITION 99. Let $\mathcal{M} = \mathcal{M}_S^I$ for some $S \subseteq NIR$.

1. An $\mathcal{M}$-*semiframe* is a triple $\mathcal{W} = \langle W, \leq, v' \rangle$ such that:
   (a) $\langle W, \leq \rangle$ is a nonempty partially ordered set.
   (b) $v' : W \times \mathcal{F}' \to \mathcal{V}$ is a partial valuation such that:
      - $\mathcal{F}'$ is a subset of $Frm_{\mathcal{L}}$ which is closed under subformulas.
      - v' satisfies the monotonicity condition: if $y \geq x$ and $\varphi \in \mathcal{F}'$, then $v'(x, \varphi) \leq_k v'(y, \varphi)$.
      - v' respects $\mathcal{M}$: If $\diamond(\psi_1, \ldots, \psi_n) \in \mathcal{F}'$, then $v'(x, \diamond(\psi_1, \ldots, \psi_n))$ is in $\tilde{\diamond}(v'(x, \psi_1), \ldots, v'(x, \psi_n))$.
      - If $\varphi \supset \psi \in \mathcal{F}'$ then $v'(x, \varphi \supset \psi) \in \mathcal{D}$ iff $v'(y, \psi) \in \mathcal{D}$ for every $y \geq x$ such that $v'(y, \varphi) \in \mathcal{D}$.
2. $\mathcal{M}_S^I$ is *analytic* if for any $\mathcal{M}_S^I$- semiframe $\langle W, \leq, v' \rangle$ there exists an $\mathcal{M}_S^I$-frame $\langle W, \leq, v \rangle$ such that v extends v'.

The next theorem provides the conditions under which $\mathcal{M}_S^I$ is analytic:

THEOREM 100. *$\mathcal{M}_S^I$ ($S \subseteq NIR$) is analytic iff either* $\{(\Rightarrow \neg), (\neg \Rightarrow)\} \subseteq S$ *or* $\{(\Rightarrow \neg), (\neg \supset\Rightarrow)_1\} \not\subseteq S$.

REMARK 101. In [Avron, 2005a] it is shown that Theorem 100 would have failed had the definition of a semiframe included the persistence condition rather than the monotonicity condition.

REMARK 102. The problem with the combination $\{(\Rightarrow \neg), (\neg \supset\Rightarrow)_1\}$ is that the condition imposed by $(\neg \supset\Rightarrow)_1$ is not consistent with the condition of monotonicity in case $\perp$ is not available.

As an immediate application, Theorem 100 can be used to determine for which $S \subseteq NIR$ the system $LJ^+(S)$ is conservative over LJ^+. This is done in the next theorem. The proof of this theorem nicely demonstrates how our semantic framework can be used, as well as the crucial role of the analycity property. Therefore we include here this proof.

THEOREM 103. *Let $S \subseteq NIR$. If neither* $\{(\Rightarrow \neg), (\neg \Rightarrow)\} \subseteq S$ *nor* $\{(\Rightarrow \neg), (\neg \supset\Rightarrow)_1\} \subseteq S$, *then $LJ^+(S)$ is a conservative extension of LJ^+. Otherwise $LJ^+(S) = LK^+(S)$.*

Proof. It is easy to see that the two conditions are necessary. Let $SN = NIR - \{(\Rightarrow \neg)\}$, $SP = NIR - \{(\neg \Rightarrow), (\neg \supset\Rightarrow)_1\}$. To show that the two conditions together are also sufficient, it suffices to show that both $LJ^+(SN)$ and $LJ^+(SP)$ are conservative over LJ^+. So let ψ be a sentence in the language of LJ^+ which is not provable in LJ^+. We show that ψ is provable in neither $LJ^+(SN)$ nor $LJ^+(SP)$. Since $\not\vdash_{LJ^+} \psi$, there is an *ordinary* two-valued Kripke frame $\langle W, \leq, u\rangle$ (where $u : W \times Frm_{\mathcal{L}} \to \{t, f\}$) in which ψ is not valid (i.e. $u(x_0, \psi) = f$ for some $x_0 \in W$). Now we define the corresponding semiframes for $LJ^+(SN)$ and $LJ^+(SP)$. Let $\mathcal{F}'$ be the set of formulas in the language of LJ^+.

$LJ^+(SN)$**:** Define v'_N on $W \times \mathcal{F}'$ by:

$$v'_N(x, \varphi) = \begin{cases} t & \text{if } u(x, \varphi) = t \\ \perp & \text{if } u(x, \varphi) = f \end{cases}$$

It is straightforward to check that $\langle W, \leq, v'_N\rangle$ is an $\mathcal{M}_{IP}[SN]$-semiframe (note that any condition concerning $\neg$ is vacuously satisfied, since there is no sentence of the form $\neg\varphi$ in $\mathcal{F}'$).

$LJ^+(SP)$**:** Define v'_P on $W \times \mathcal{F}'$ by:

$$v'_P(x, \varphi) = \begin{cases} \top & \text{if } u(x, \varphi) = t \\ f & \text{if } u(x, \varphi) = f \end{cases}$$

Again, it is easy to check that $\langle W, \leq, v'_P\rangle$ is an $\mathcal{M}_{IP}[SP]$-semiframe.

By Theorem 100, $\langle W, \leq, v'_N\rangle$ and $\langle W, \leq, v'_P\rangle$ can respectively be extended to an $\mathcal{M}_{IP}[SN]$-frame $\langle W, \leq, v_N\rangle$ and an $\mathcal{M}_{IP}[SP]$-frame $\langle W, \leq, v_P\rangle$. Since $v_N(x_0, \psi) = v'_N(x_0, \psi) = \perp$, ψ is not valid in $\langle W, \leq, v'_N\rangle$, and so it is not provable in $LJ^+(SN)$. Similarly, $v_P(x_0, \psi) = v'_P(x_0, \psi) = f$. Hence ψ is not valid in $\langle W, \leq, v'_P\rangle$, and so is not provable in $LJ^+(SP)$. ■

Theorems 100, 103 and Corollary 87 immediately entail:

COROLLARY 104. *$LJ^+(S)$ is decidable for every $S \subseteq NIR$.*

It follows from Theorem 103 that $LJ^+(SN)$ and $LJ^+(SP)$ are the two maximal logics in the family $\{LJ^+(S) \mid S \subseteq NIR\}$ which are conservative extensions of constructive positive logic. Now the first is the well-known system **N** of Nelson ([Almukdad and Nelson, 1984]) and von Kutschera ([von Kutschera, 1969]). The other, in contrast, is new. However, it is a very attractive system for constructive negation. First: it is paraconsistent (i.e.: a single contradiction does not imply everything in it). Second: LEM is valid in it. In fact, $LJ^+(SP)$ is obtained from **N** by replacing two of its axioms by LEM. Now, while LEM is very intuitive, the two axioms it replaces are not. Indeed, one of them, $\neg\varphi \supset (\varphi \supset \psi)$, intuitively means that if φ is false then it implies everything. The second, $\neg(\varphi \supset \psi) \supset \varphi$, intuitively means that if there is something that φ does not imply, then φ should be true (i.e.: it cannot be false). Obviously, these two principles are similar — and counterintuitive. It is no wonder that from a constructive point of view, each of them is inconsistent with LEM, and is rejected in $LJ^+(SP)$. It is also worth noting that despite the paraconsistent nature of $LJ^+(SP)$, the basic intuitive law of contradiction $\neg(\varphi \wedge \neg\varphi)$ *is* valid in it.

Next we turn to another application of Theorems 96 and 100: eliminations of cuts. It was shown in [Avron, 2007b] that in general the cut-elimination theorem does not hold for the Gentzen-type systems presented in this subsection. Moreover: examples have been given there of a subset S of NIR and a sequent which is provable in $LJ^+(S)$, but any proof of it there should contain a non-analytic cut. This is perhaps not surprising, since our logical rules themselves do not have the strict subformula property: some of them involve negations of subformulas of their conclusion, even though those negations are not subformulas themselves. Therefore, it is reasonable to expect the same from cuts. This leads to the following theorem from [Avron, 2005a]:

THEOREM 105. *Assume that $S \subseteq NIR$, and $\{(\Rightarrow \neg), (\neg \supset \Rightarrow)_1\} \not\subseteq S$. Then for every sequent s in the language of LJ^+ there is either a* finite *$\mathcal{M}^I_S$-frame in which s is not valid, or a proof in $LJ^+(S)$ in which every cut is either on a subformula of s or on a negation of such a subformula.*

6 NMATRICES FOR LOGICS OF FORMAL INCONSISTENCY

In this section we apply the framework of Nmatrices to provide modular semantics for yet another family of non-classical logics: da Costa's paraconsistent logics. A paraconsistent logic is a logic which allows non-trivial inconsistent theories. One of the oldest and best known approaches to the problem of designing useful paraconsistent logics is da Costa's approach. This approach is based on two main ideas. The first is to limit the applicability of the classical (and intuitionistic) rule $\neg\varphi, \varphi \vdash \psi$ to the case where φ is "consistent". The second is to express this assumption of consistency of φ within the language. The easiest way to implement these ideas is to include in the language a special connective $\circ$, with the intended meaning of $\circ\varphi$ being "φ is consistent". Then one can explicitly add the assumption of the consistency of φ to the problematic (from a paraconsistent point of view)

Name of rule	Abstract form	Hilbert-style axiom
$(\mathbf{b})$	$\circ\varphi, \neg\varphi, \varphi \vdash$	$(\circ\varphi \wedge \neg\varphi \wedge \varphi) \supset \psi$
$(\mathbf{k1})$	$\vdash \circ\varphi, \varphi$	$\circ\varphi \vee \varphi$
$(\mathbf{k2})$	$\vdash \circ\varphi, \neg\varphi$	$\circ\varphi \vee \neg\varphi$
$(\mathbf{i1})$	$\neg\circ\varphi \vdash \varphi$	$\neg\circ\varphi \supset \varphi$
$(\mathbf{i2})$	$\neg\circ\varphi \vdash \neg\varphi$	$\neg\circ\varphi \supset \neg\varphi$
$(\mathbf{a}_\neg)$	$\circ\varphi \vdash \circ\neg\varphi$	$\circ\varphi \supset \circ\neg\varphi$
$(\mathbf{a}_\diamond)$	$\circ\varphi, \circ\psi \vdash \circ(\varphi \diamond \psi)$	$\circ\varphi \supset (\circ\psi \supset \circ(\varphi \diamond \psi))$
$(\mathbf{o}^{\mathbf{1}}_\diamond)$	$\circ\varphi \vdash \circ(\varphi \diamond \psi)$	$\circ\varphi \supset \circ(\varphi \diamond \psi)$
$(\mathbf{o}^{\mathbf{2}}_\diamond)$	$\circ\psi \vdash \circ(\varphi \diamond \psi)$	$\circ\psi \supset \circ(\varphi \diamond \psi)$
$(\mathbf{l})$	$\neg(\varphi \wedge \neg\varphi) \vdash \circ\varphi$	$\neg(\varphi \wedge \neg\varphi) \supset \circ\varphi$

Figure 4. Schemata involving $\circ$

rule, getting the rule called $(\mathbf{b})$ below. Other rules concerning $\neg$ and $\circ$ can then be added, leading to a large family of logics known as "Logics of Formal Inconsistency" (LFIs - see [da Costa, 1974; Carnielli and Marcos, 2002; Carnielli et. al., 2007]). In this chapter we investigate those that are obtained using the rules in NIR, as well as the main rules involving the consistency operator that have been studied in the literature on LFIs. The latter rules are listed in Figure 4 (in which $\diamond \in \{\wedge, \vee, \supset\}$). The material of this section is based on [Avron, 2007a]. Throughout it, we fix the language $\mathcal{L}_C = \{\neg, \circ, \supset, \wedge, \vee\}$. Again our basic system will be LK^+ (the positive fragment of classical logic).

6.1 LFIs with Finite Characteristic Nmatrices

DEFINITION 106.

1. Let FCR be the set of all the rules in the table above *except the last one* $(\mathbf{l})$. We shall write $(\mathbf{i})$ instead of the combination of $(\mathbf{i1})$ and $(\mathbf{i2})$, $(\mathbf{a})$ instead of $\{(\mathbf{a}_\diamond) \mid \diamond \in \{\wedge, \vee, \supset\}\}$ and similarly for $(\mathbf{o})$.

2. Let $LFIR = NIR \cup FCR$. We denote by $HLFIR$ the set of Hilbert-style axioms corresponding to the rules in $LFIR$.

3. For $S \subseteq LFIR$ let $LK^+[S]$ be the extension of LK^+ by S.

The basic idea in providing semantics for $LK^+[S]$ (where $S \subseteq LFIR$) is this time to let the value assigned to a sentence φ provide information not only about the truth/falsity of φ and $\neg\varphi$, but also about the truth/falsity of $\circ\varphi$. This leads to the use of elements from $\{0,1\}^3$ as our truth-values, where the intended intuitive meaning of $v(\varphi) = \langle x, y, z\rangle$ is now:

- $x = 1$ iff φ is "true" (i.e. $v(\varphi) \in \mathcal{D}$).
- $y = 1$ iff $\neg\varphi$ is "true" (i.e. $v(\neg\varphi) \in \mathcal{D}$).
- $z = 1$ iff $\circ\varphi$ is "true" (i.e. $v(\circ\varphi) \in \mathcal{D}$).

In addition to (NEG), which remains unchanged, this interpretation dictates also the following condition:

$$\text{(CON)} \qquad \tilde{\circ}a \subseteq \{y \mid P_1(y) = P_3(a)\}$$

Accordingly, this time we start our semantic investigation of $LFIR$ with the weakest Nmatrix which satisfies both (NEG) and (CON). Then we show that every logic which is defined by some subset of $LFIR$ is characterized by some (easily computable) simple refinement of that Nmatrix.

DEFINITION 107. The Nmatrix $\mathcal{M}_8^B = \langle \mathcal{V}_8, \mathcal{D}_8, \mathcal{O}_8\rangle$ is defined as follows:

- $\mathcal{V}_8 = \{0,1\}^3$
- $\mathcal{D}_8 = \{a \in \mathcal{V}_8 \mid P_1(a) = 1\}$
- Let $\mathcal{V} = \mathcal{V}_8$, $\mathcal{D} = \mathcal{D}_8$, $\mathcal{F} = \mathcal{V}_8 - \mathcal{D}$. The operations in $\mathcal{O}_8$ are:

$$\tilde{\neg}a = \begin{cases} \mathcal{D} & \text{if } P_2(a) = 1 \\ \mathcal{F} & \text{if } P_2(a) = 0 \end{cases}$$

$$\tilde{\circ}a = \begin{cases} \mathcal{D} & \text{if } P_3(a) = 1 \\ \mathcal{F} & \text{if } P_3(a) = 0 \end{cases}$$

$$a \tilde{\vee} b = \begin{cases} \mathcal{D} & \text{if either } a \in \mathcal{D} \text{ or } b \in \mathcal{D}, \\ \mathcal{F} & \text{if } a, b \in \mathcal{F} \end{cases}$$

$$a \tilde{\supset} b = \begin{cases} \mathcal{D} & \text{if either } a \in \mathcal{F} \text{ or } b \in \mathcal{D} \\ \mathcal{F} & \text{if } a \in \mathcal{D} \text{ and } b \in \mathcal{F} \end{cases}$$

$$a \tilde{\wedge} b = \begin{cases} \mathcal{F} & \text{if either } a \in \mathcal{F} \text{ or } b \in \mathcal{F} \\ \mathcal{D} & \text{otherwise} \end{cases}$$

THEOREM 108. *([Avron, 2007a]) $\mathcal{M}_8^B$ is a (dynamically) characteristic Nmatrix for LK^+.*

DEFINITION 109.

1. The general refining conditions induced by the conditions in NIR are identical to those given in Definition 82.

2. The general refining conditions induced by the conditions in FCR are:

   C(**b**): If $P_1(a) = 1$ and $P_2(a) = 1$ then $P_3(a) = 0$

   C(**k1**): If $P_1(a) = 0$ then $P_3(a) = 1$

   C(**k2**): If $P_2(a) = 0$ then $P_3(a) = 1$

   C(**i1**): If $P_1(a) = 0$ then $\widetilde{\circ} a \subseteq \{x \mid P_2(x) = 0\}$

   C(**i2**): If $P_2(a) = 0$ then $\widetilde{\circ} a \subseteq \{x \mid P_2(x) = 0\}$

   C($\mathbf{a}_\neg$): If $P_3(a) = 1$ then $\widetilde{\neg} a \subseteq \{x \mid P_3(x) = 1\}$

   C($\mathbf{a}_\diamond$): If $P_3(a) = 1$ and $P_3(b) = 1$ then $a \widetilde{\diamond} b \subseteq \{x \mid P_3(x) = 1\}$

   C($\mathbf{o}^1_\diamond$): If $P_3(a) = 1$ then $a \widetilde{\diamond} b \subseteq \{x \mid P_3(x) = 1\}$

   C($\mathbf{o}^2_\diamond$): If $P_3(b) = 1$ then $a \widetilde{\diamond} b \subseteq \{x \mid P_3(x) = 1\}$

3. For $S \subseteq LFIR$, let $C(S) = \{\mathrm{C}r \mid r \in S\}$, and let $\mathcal{M}_S$ be the weakest simple refinement of $\mathcal{M}^B_8$ in which the conditions in $C(S)$ are all satisfied (again it is not difficult to check that this is well-defined for every $S \subseteq LFIR$).

THEOREM 110. *$\mathcal{M}_S$ ($S \subseteq LFIR$) is a characteristic Nmatrix for $LK^+[S]$.*

COROLLARY 111. *$LK^+[S]$ is decidable for every $S \subseteq LFIR$.*

EXAMPLE 112.

1. Let $\mathbf{B} = LK^+[\{(\Rightarrow \neg), (\mathbf{b})\}]$. This logic is the basic logic of formal inconsistency from [Carnielli and Marcos, 2002; Carnielli et. al., 2007] (where it is called mbC). By Theorem 110, the following Nmatrix $\mathcal{M}^B_5 = \langle \mathcal{V}_5, \mathcal{D}_5, \mathcal{O}_5 \rangle$ is characteristic for it:

   - $\mathcal{V}_5 = \{t, t_I, I, f_I, f\}$ where:

   $$\begin{array}{lcl} t & = & \langle 1,0,1 \rangle \\ t_I & = & \langle 1,0,0 \rangle \\ I & = & \langle 1,1,0 \rangle \\ f & = & \langle 0,1,1 \rangle \\ f_I & = & \langle 0,1,0 \rangle \end{array}$$

   - $\mathcal{D}_5 = \{t, I, t_I\}$ $(= \{\langle x, y, z \rangle \in \mathcal{V}_5 \mid x = 1\})$.

- Let $\mathcal{D} = \mathcal{D}_5$, $\mathcal{F} = \mathcal{V}_5 - \mathcal{D}$. The operations in $\mathcal{O}_5$ are defined by:

$$\tilde{\neg} a = \begin{cases} \mathcal{D} & \text{if } a \in \{I, f, f_I\} \\ \mathcal{F} & \text{if } a \in \{t, t_I\} \end{cases}$$

$$\tilde{\circ} a = \begin{cases} \mathcal{D} & \text{if } a \in \{t, f\} \\ \mathcal{F} & \text{if } a \in \{I, t_I, f_I\} \end{cases}$$

  The rest of the operations are defined like in Definition 107.

2. Let $S = \{(\Rightarrow \neg), (\mathbf{b}), (\Rightarrow \neg \supset), (\mathbf{i1}), (\mathbf{a}_\neg)\}$. Then $\mathcal{M}_S = \langle \mathcal{V}_S, \mathcal{D}_S, \mathcal{O}_S \rangle$, where:

   - $\mathcal{V}_S = \{t, t_I, I, f\}$
   - $\mathcal{D}_S = \{t, I, t_I\}$
   - $a \tilde{\supset} b = \begin{cases} \mathcal{D}_S & \text{if either } a = f \text{ or } b \in \{t, t_I\} \\ \{I\} & \text{if } a \in \mathcal{D}_S \text{ and } b = I \\ \{f\} & \text{if } a \in \mathcal{D}_S \text{ and } b = f \end{cases}$
   - $\tilde{\neg} t = \tilde{\neg} t_I = \{f\} \quad \tilde{\neg} I = \mathcal{D}_S \quad \tilde{\neg} f = \{t\}$
   - $\tilde{\circ} t = \mathcal{D}_S \quad \tilde{\circ} t_I = \tilde{\circ} I = \{f\} \quad \tilde{\circ} f = \{t, t_I\}$

3. Let $\mathbf{Cia} = \{(\Rightarrow \neg), (\mathbf{b}), (\neg\neg \Rightarrow), (\mathbf{i}), (\mathbf{a})\}$. $\mathcal{M}_{Cia} = \langle \mathcal{V}_{Cia}, \mathcal{D}_{Cia}, \mathcal{O}_{Cia} \rangle$, where:

   - $\mathcal{V}_{Cia} = \{t, I, f\}$
   - $\mathcal{D}_{Cia} = \{t, I\}$
   - $a \tilde{\supset} b = \begin{cases} \{f\} & \text{if } a \in \{t, I\} \text{ and } b = f \\ \{t\} & \text{if either } a = f,\ b \in \{f, t\} \text{ or } a = t,\ b = t \\ \{t, I\} & \text{otherwise} \end{cases}$
   - $a \tilde{\vee} b = \begin{cases} \{f\} & \text{if } a = f \text{ and } b = f \\ \{t\} & \text{if either } a = t,\ b \in \{f, t\} \text{ or } b = t,\ a \in \{f, t\} \\ \{t, I\} & \text{otherwise} \end{cases}$
   - $a \tilde{\wedge} b = \begin{cases} \{f\} & \text{if } a = f \text{ or } b = f \\ \{t\} & \text{if } a = t \text{ and } b = t \\ \{t, I\} & \text{otherwise} \end{cases}$
   - $\tilde{\neg} t = \{f\} \quad \tilde{\neg} I = \{I\} \quad \tilde{\neg} f = \{t\}$
   - $\tilde{\circ} t = \tilde{\circ} f = \{t\} \quad \tilde{\circ} I = \{f\}$

6.2 LFIs with Infinite Characteristic Nmatrices

The family of LFIs for which we provided semantics in the previous subsection does not include the well-known da Costa's original logic C_1 from ([da Costa, 1974]). Now C_1 is just the $\circ$-free fragment of **Cila**, the logic which is obtained by adding the rule **(l)** from Figure 4 to the system **Cia** from Example 112. This rule is problematic, because of the following theorem:

THEOREM 113. *No system between* **Bl** *and* $\mathbf{Bl}[(\Rightarrow \neg\neg), (\neg\neg \Rightarrow), (\mathbf{i}), (\mathbf{o})]$ *has a finite characteristic Nmatrix (and so none of them has a finite characteristic ordinary matrix).* [20]

It follows that the method used in the previous subsection cannot work for logics like **Cila**. As a reasonable useful substitute, in this subsection we present *infinite* (but still effective) characteristic Nmatrices for a family of such systems (which includes **Cila**). Then we show that these Nmatrices can still be used to provide decision procedures for the logics they characterize.

As usual, we start with the basic LFI which includes (**l**), and find first a characteristic Nmatrix for it.

DEFINITION 114. The system **Bl** is obtained from the basic system **B** (from Example 112) by adding (**l**) as an axiom.

Now the validity of (**l**) in an Nmatrix means that whenever $\circ\varphi$ is "false", so is $\neg(\varphi \wedge \neg\varphi)$. Accordingly, Nmatrices appropriate for **Bl** should be able to distinguish between conjunctions of an "inconsistent" formula with its negation from other types of conjunctions. Therefore such Nmatrices should enforce an intimate connection between the truth-value of an "inconsistent" formula and the truth-value of its negation. This in turn requires a supply of infinitely many truth-values, corresponding to the potentially infinite number of "inconsistent" formulas. But from where will we take these truth-values, and how should we define the operations on them? A key observation in our path to solve these problems is that **(k1)** and **(k2)** are theorems of **Bl**. Hence **Bl** extends $\mathbf{B}[\{(\mathbf{b}), (\Rightarrow \neg), (\mathbf{k1}), (\mathbf{k2})\}]$. Accordingly, the Nmatrices which we will use for characterizing **Bl** and its extensions will be refinements (see Definition 30) of $\mathcal{M}_{\{(\mathbf{b}),(\Rightarrow\neg),(\mathbf{k1}),(\mathbf{k2})\}}$. The latter is an Nmatrix with three truth-values: those that were denoted above by t, f, and I. Now one of the most productive method of refining a given *Nmatrix* $\mathcal{M}$ (which is

[20] It easily follows from this theorem that C_1 has no finite characteristic Nmatrix. Now it has been known before that C_1 and some other LFIs have no characteristic *ordinary* matrices (see e.g. [Carnielli and Marcos, 2002; Carnielli et. al., 2007]). However, the result of Theorem 113 is much stronger.

not available in the framework of ordinary matrices!) is to first *duplicate* its elements: we can construct an Nmatrix $\mathcal{M}'$ which is completely equivalent to $\mathcal{M}$ by replacing each element a by a nonempty set of "copies", and then defining the operations in $\mathcal{M}'$ to be "the same" as in the original $\mathcal{M}$, but without distinguishing between two copies of the same element of $\mathcal{M}$. In other words: if $b', a'_1, \ldots, a'_n$ are copies in $\mathcal{M}'$ of $b, a_1 \ldots, a_n$ (respectively), then $b' \in \tilde{\diamond}(a'_1 \ldots, a'_n)$ in $\mathcal{M}'$ iff $b \in \tilde{\diamond}(a_1 \ldots, a_n)$ in $\mathcal{M}$. [21] What we shall do in order to construct an Nmatrix for **Bl** is first to duplicate the elements of $\mathcal{M}_{\{(\mathbf{b}),(\Rightarrow\neg),(\mathbf{k1}),(\mathbf{k2})\}}$ (actually only t and I) infinitely many times. Then we shall refine the resulting Nmatrix in the way hinted above so that axiom $(\mathbf{l})$ becomes valid.

DEFINITION 115. Let $\mathcal{T} = \{t_i^j \mid i \geq 0, j \geq 0\}$, $\mathcal{I} = \{I_i^j \mid i \geq 0, j \geq 0\}$, $\mathcal{F} = \{f\}$. The Nmatrix $\mathcal{M}_{\mathbf{Bl}} = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$ is defined as follows:

1. $\mathcal{V} = \mathcal{T} \cup \mathcal{I} \cup \mathcal{F}$ and $\mathcal{D} = \mathcal{T} \cup \mathcal{I}$.

2. $\mathcal{O}$ is defined by:

$$a \tilde{\vee} b = \begin{cases} \mathcal{D} & \text{if either } a \in \mathcal{D} \text{ or } b \in \mathcal{D} \\ \mathcal{F} & \text{if } a, b \in \mathcal{F} \end{cases}$$

$$a \tilde{\supset} b = \begin{cases} \mathcal{D} & \text{if either } a \in \mathcal{F} \text{ or } b \in \mathcal{D} \\ \mathcal{F} & \text{if } a \in \mathcal{D} \text{ and } b \in \mathcal{F} \end{cases}$$

$$a \tilde{\wedge} b = \begin{cases} \mathcal{F} & \text{if either } a \in \mathcal{F} \text{ or } b \in \mathcal{F} \\ \mathcal{T} & \text{if } a = I_i^j \text{ and } b \in \{I_i^{j+1}, t_i^{j+1}\} \\ \mathcal{D} & \text{otherwise} \end{cases}$$

$$\tilde{\neg} a = \begin{cases} \mathcal{F} & \text{if } a \in \mathcal{T} \\ \mathcal{D} & \text{if } a \in \mathcal{F} \\ \{I_i^{j+1}, t_i^{j+1}\} & \text{if } a = I_i^j \end{cases}$$

$$\tilde{\circ} a = \begin{cases} \mathcal{D} & \text{if } a \in \mathcal{F} \cup \mathcal{T} \\ \mathcal{F} & \text{if } a \in \mathcal{I} \end{cases}$$

THEOREM 116. *$\mathcal{M}_{\mathbf{Bl}}$ is a characteristic Nmatrix for* **Bl**.

Now we turn to the extensions of **Bl** with axioms.

DEFINITION 117.

[21] Actually, we have already implicitly used this method above several times. Thus from the point of view of the positive classical connectives, $\mathcal{M}_8^B$ is just a duplication of the classical two-valued matrix: all elements of $\mathcal{D}$ are copies of "true", all elements of $\mathcal{F}$ are copies of "false".

1. Let $LFIR^l = LFIR - \{(\Rightarrow \neg), (\neg \Rightarrow), (\mathbf{b}), (\mathbf{k_1}), (\mathbf{k_2})\}$.
2. For $S \subseteq LFIR^l$, the system $\mathbf{Bl}[S]$ is obtained from $\mathbf{Bl}$ by adding the schemata in S.

Like in the previous subsection, each of the schemata in $LFIR^l$ corresponds to some easily computed semantic condition, this time on simple refinements of the basic Nmatrix $\mathcal{M}_{\mathbf{Bl}}$. These conditions are in fact identical to the conditions that correspond to these axioms in refinements of $\mathcal{M}_{\{(\mathbf{b}),(\Rightarrow\neg),(\mathbf{k1}),(\mathbf{k2})\}}$, but with t replaced by $\mathcal{T}$, and I replaced by $\mathcal{I}$.

DEFINITION 118. For $S \subseteq LFIR^l$, $\mathcal{M}_{\mathbf{Bl}[S]}$ is the weakest simple refinement of $\mathcal{M}_{\mathbf{Bl}}$ which satisfies the following conditions:

1. If $(\neg\neg \Rightarrow) \in S$ then $a \in \mathcal{F} \Rightarrow \tilde{\neg}(a) \subseteq \mathcal{T}$
2. If $(\Rightarrow \neg\neg) \in S$ then $a \in \mathcal{I} \Rightarrow \tilde{\neg}(a) \subseteq \mathcal{I}$
3. If $(\mathbf{i1}) \in S$ then $a \in \mathcal{T} \Rightarrow \tilde{\circ}(a) \subseteq \mathcal{T}$
4. If $(\mathbf{i1}) \in S$ then $a \in \mathcal{F} \Rightarrow \tilde{\circ}(a) \subseteq \mathcal{T}$
5. If $(\mathbf{a}_\neg) \in S$ then $a \in \mathcal{I} \Rightarrow \tilde{\circ}a \subseteq \mathcal{I}$
6. If $(\mathbf{a}_\diamond) \in S$ then $a \in \mathcal{F} \cup \mathcal{I}, b \in \mathcal{F} \cup \mathcal{I} \Rightarrow a\tilde{\diamond}b \subseteq \mathcal{F} \cup \mathcal{I}$
7. If $(\mathbf{o}^{\mathbf{1}}_\diamond) \in S$ then $a \in \mathcal{F} \cup \mathcal{T} \Rightarrow a\tilde{\diamond}b \subseteq \mathcal{F} \cup \mathcal{T}$
8. If $(\mathbf{o}^{\mathbf{2}}_\diamond) \in S$ then $b \in \mathcal{F} \cup \mathcal{T} \Rightarrow a\tilde{\diamond}b \subseteq \mathcal{F} \cup \mathcal{T}$
9. If $(\neg \supset\Rightarrow)_1 \in S$ then $a \in \mathcal{F} \Rightarrow (a\tilde{\supset}b) \subseteq \mathcal{T}$
10. If $(\neg \supset\Rightarrow)_2 \in S$ then $b \in \mathcal{T} \Rightarrow (a\tilde{\supset}b) \subseteq \mathcal{T}$
11. If $(\Rightarrow \neg \supset) \in S$ then $a \in \mathcal{D}, b \in \mathcal{F} \cup \mathcal{I} \Rightarrow a\tilde{\supset}b \subseteq \mathcal{F} \cup \mathcal{I}$

THEOREM 119. *For $S \subseteq LFIR$, $\mathcal{M}_{\mathbf{Bl}[S]}$ is a (dynamically) characteristic Nmatrix for* $\mathbf{Bl}[S]$.

COROLLARY 120. *For every $S \subseteq LFIR^l$, the logic* $\mathbf{Bl}[S]$ *is decidable.*

Proof. The proof of Theorem 119 in [Avron, 2007a] implies that to check whether a given formula φ is provable in $\mathbf{L}$, it suffices to check all legal partial valuations v in $\mathcal{M}_L$ which assign to subformulas of φ values in

$$\{f\} \cup \{t_i^j \mid 0 \leq i \leq n(\varphi), 0 \leq j \leq k(\varphi)\} \cup \{I_i^j \mid 0 \leq i \leq n(\varphi), 0 \leq j \leq k(\varphi)\}$$

where $n(\varphi)$ is the number of subformulas of φ which do not begin with $\neg$, and $k(\varphi)$ is the maximal number of consecutive negation symbols occurring within φ. This is a finite process.

COROLLARY 121. *da Costa's system C_1 is decidable,* [22] *and it has a characteristic Nmatrix $\mathcal{M}_{C_1}$, in which the sets of truth-values and designated truth-values are like in $\mathcal{M}_{\mathbf{Bl}}$, and the interpretations of the connectives are defined as follows:*

$$a\widetilde{\supset}b = \begin{cases} \mathcal{F} & a \in \mathcal{D},\ b \in \mathcal{F} \\ \mathcal{T} & a \in \mathcal{F},\ b \notin \mathcal{I} \\ \mathcal{T} & b \in \mathcal{T},\ a \notin \mathcal{I} \\ \mathcal{D} & otherwise \end{cases} \qquad a\widetilde{\wedge}b = \begin{cases} \mathcal{F} & a \in \mathcal{F} \text{ or } b \in \mathcal{F} \\ \mathcal{T} & a \in \mathcal{T},\ b \in \mathcal{T} \\ \mathcal{T} & a = I_i^j,\ b \in \{I_i^{j+1}, t_i^{j+1}\} \\ \mathcal{D} & otherwise \end{cases}$$

$$\widetilde{\neg}a = \begin{cases} \mathcal{F} & a \in \mathcal{T} \\ \mathcal{T} & a \in \mathcal{F} \\ \{I_i^{j+1}, t_i^{j+1}\} & a = I_i^j \end{cases} \qquad a\widetilde{\vee}b = \begin{cases} \mathcal{F} & a \in \mathcal{F},\ b \in \mathcal{F} \\ \mathcal{T} & a \in \mathcal{T},\ b \notin \mathcal{I} \\ \mathcal{T} & b \in \mathcal{T},\ a \notin \mathcal{I} \\ \mathcal{D} & otherwise \end{cases}$$

PART II: THE FIRST-ORDER CASE AND BEYOND

In the first part we have described the semantic framework of Nmatrices on the propositional level and presented a number of applications of this framework. However, no semantic framework can be considered really useful unless it can be naturally extended at least to the first-order level. Accordingly, this part is devoted to extending the framework of Nmatrices to languages with quantifiers.

The simplest and most well-known quantifiers are of course the first-order quantifiers $\forall$ and $\exists$ (and we shall devote Section 9 to them). However, we start by exploring a more general notion of quantifiers. By a (unary) quantifier we mean a logical constant which (may) bind a variable when applied to a formula. In other words, if $\mathcal{Q}$ is a quantifier, x is a variable and ψ is a formula, then $\mathcal{Q}x\psi$ is a formula in which all occurrences of x are bound by $\mathcal{Q}$. It should be noted that this notion can be further generalized to *multi-ary quantifiers*, which are logical constants that can be applied to more than one formula. If $\mathcal{Q}$ is an n-ary quantifier, x is a variable and $\psi_1, ..., \psi_n$ are formulas, then $\mathcal{Q}x(\psi_1, ..., \psi_n)$ is a formula in which all occurrences of x are

[22] The decidability of C_1, as well as of most of the systems presented here is not new (see, e.g. [Carnielli and Marcos, 2002; Carnielli et. al., 2007]).

bound by $\mathcal{Q}$. In this context the ordinary quantifiers can be thought of as unary quantifiers, while the bounded universal and existential quantifiers $\overline{\forall}$ and $\overline{\exists}$ used in syllogistic reasoning are examples of binary quantifiers[23].

7 MANY-VALUED MATRICES WITH QUANTIFERS

We start with ordinary (unary) quantifiers and their treatment in the framework of standard many-valued matrices. In what follows, L is a language, which includes a set of propositional connectives, a set of quantifiers, a countable set of variables, and a signature, consisting of a non-empty set of predicate symbols, a set of function symbols, and a set of constants. Frm_L is the set of (standardly defined) wffs of L, and Frm_L^{cl} is its set of closed wffs. Trm_L is the set of terms of L, and Trm_L^{cl} is its set of closed terms. In ordinary (deterministic) many-valued matrices (unary) quantifiers are standardly interpreted using the notion of *distributions*. This notion is due to Mostowski ([Mostowski, 1961]; the term 'distribution' was later coined in [Carnielli, 1987].

DEFINITION 122. Given a set of truth values $\mathcal{V}$, a *distribution* of a quantifier $\mathcal{Q}$ is a function $\lambda_{\mathcal{Q}} : (2^{\mathcal{V}} \setminus \{\emptyset\}) \to \mathcal{V}$.

The following is a standard definition (see, e.g. [Urquhart, 2001]) of a deterministic matrix with distribution quantifiers:

DEFINITION 123. A matrix for L is a tuple $\mathcal{P} = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$, where:

- $\mathcal{V}$ is a non-empty set of truth-values,
- $\mathcal{D}$ is a non-empty proper set of $\mathcal{V}$,
- $\mathcal{O}$ includes a function $\tilde{\diamond} : \mathcal{V}^n \to \mathcal{V}$ for every n-ary connective of L, and a function $\tilde{\mathcal{Q}} : 2^{\mathcal{V}} \setminus \{\emptyset\} \to \mathcal{V}$ for every quantifier of L.

EXAMPLE 124. Consider the matrix $\mathcal{P} = \langle \{\mathbf{t}, \mathbf{f}\}, \{\mathbf{t}\}, \mathcal{O} \rangle$ for a first-order language L, where $\mathcal{O}$ contains the following (standard) interpretations of $\forall$ and $\exists$:

$\mathbf{H}$	$\tilde{\forall}(\mathbf{H})$	$\tilde{\exists}(\mathbf{H})$
$\{\mathbf{t}\}$	$\mathbf{t}$	$\mathbf{t}$
$\{\mathbf{t}, \mathbf{f}\}$	$\mathbf{f}$	$\mathbf{t}$
$\{\mathbf{f}\}$	$\mathbf{f}$	$\mathbf{f}$

[23]The respective meanings of $\overline{\forall}x(\psi_1, \psi_2)$ and $\overline{\exists}x(\psi_1, \psi_2)$ are $\forall x(\psi_1 \to \psi_2)$ and $\exists x(\psi_1 \wedge \psi_2)$.

The notion of a structure is defined standardly:

DEFINITION 125. Let $\mathcal{P} = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$ be a matrix for L. An L-structure S for $\mathcal{P}$ is a pair $\langle D, I \rangle$ where D is a (non-empty) domain and I is an interpretation of constants, predicate symbols and function symbols of L, which satisfies:

- For every constant c of L: $I(c) \in D$.
- For every n-ary predicate symbol p of L: $I(p) \in D^n \to \mathcal{V}$.
- For every n-ary function symbol f of L: $I(f) \in D^n \to D$.

There are two main approaches to interpreting quantified formulas: the objectual (referential) approach, which uses *assignments*, and the substitutional approach ([Leblanc, 2001]), which is based on *substitutions*. Below we shortly review these two approaches. In the better known objectual approach (used in most standard textbooks on classical first-order logic, like [Mendelson, 1964; Enderton, 1972; van Dalen, 1980]), a variable is thought of as ranging over a set of objects from the domain, and assignments map variables to elements of the domain. In the context of many-valued deterministic matrices this is usually formalized as follows (see e.g. [Urquhart, 2001; Hähnle, 1999]).

DEFINITION 126. Given an L-structure $S = \langle D, I \rangle$, an *assignment* G in S is any function mapping the variables of L to D. For any $a \in D$ we denote by $G[x := a]$ the assignment which is similar to G, except that it assigns a to x. G is extended to L-terms as follows: $G(c) = I(c)$ for every constant c of L and $G(f(\mathbf{t}_1, ..., \mathbf{t}_n)) = I(f)(G(\mathbf{t}_1), ..., G(\mathbf{t}_n))$ for every n-ary function symbol f of L and $\mathbf{t}_1, ..., \mathbf{t}_n \in Trm_L$.

DEFINITION 127. Let S be an L-structure for a matrix $\mathcal{P}$ and let G be an assignment in S. The valuation $v_{S,G} : Frm_L \to \mathcal{V}$ is defined as follows:

- $v_{S,G}(p(\mathbf{t}_1, ..., \mathbf{t}_n)) = I(p)(G(\mathbf{t}_1), ..., G(\mathbf{t}_n))$.
- $v_{S,G}(\diamond(\psi_1, ..., \psi_n)) = \tilde{\diamond}(v_{S,G}(\psi_1), ..., v_{S,G}(\psi_n))$.
- $v_{S,G}(\mathcal{Q}x\psi) = \tilde{\mathcal{Q}}(\{v_{S,G[x:=a]}(\psi) \mid a \in D\})$.

In the alternative substitutional approach to quantification (used e.g. for first-order classical logic in [Shoenfield, 1967]) a variable is thought of as ranging over syntactical (closed) terms rather than over elements of the domain. Accordingly, the key notion in this approach is that of a *substitution instance* (rather than an assignment):

DEFINITION 128. For any formula ψ, a *substitution L-instance of* ψ is a formula $\psi\{\mathbf{t}_1/x_1, ..., \mathbf{t}_n/x_n\}$, where for all $1 \leq i \leq n$, $\mathbf{t}_i$ is an L-term free for x_i in ψ. A *substitution L-instance of* Γ is a set $\{\psi\{\mathbf{t}_1/x_1, ..., \mathbf{t}_n/x_n\} \mid \psi \in \Gamma\}$ for some $\mathbf{t}_1, ..., \mathbf{t}_n \in Trm_L$ which are free for $x_1, ..., x_n$ (respectively) in all the formulas of Γ.

The main idea of the substitutional approach is that a formula $\forall\psi x$ ($\exists x\psi$) is true if and only if each (at least one) of the closed substitution instances of ψ is true. To apply this approach, we need to assume that every element of the domain has a closed term referring to it. This condition can be satisfied by extending the language with individual constants:

DEFINITION 129. For an L-structure $S = \langle D, I\rangle$ for $\mathcal{P}$, $L(D)$ is the language obtained from L by adding to it the set of *individual constants* $\{\overline{a} \mid a \in D\}$. The $L(D)$-structure which is induced by S is $\langle D, I'\rangle$, where I' is the unique extension of I to $L(D)$ such that $I'(\overline{a}) = a$.

Henceforth we shall identify an L-structure S with the $L(D)$-structure which is induced by S.

Here is the substitutional counterpart of the notion of a valuation given in Definition 127:

DEFINITION 130. Let $S = \langle D, I\rangle$ be an L-structure for a matrix $\mathcal{P} = \langle \mathcal{V}, \mathcal{D}, \mathcal{O}\rangle$. The *valuation* $v_S : Frm^{\mathsf{cl}}_{L(D)} \to \mathcal{V}$ is defined as follows:

- $v_S(p(\mathbf{t}_1, ..., \mathbf{t}_n)) = I(p)(I(\mathbf{t}_1), ..., I(\mathbf{t}_n))$
- $v_S(\diamond(\psi_1, ..., \psi_n)) = \tilde{\diamond}(v(\psi_1), ..., v(\psi_n))$
- $v_S(\mathcal{Q}x\psi) = \tilde{\mathcal{Q}}(\{v_S(\psi\{\overline{a}/x\}) \mid a \in D\})$

For reasons that will become clear in the sequel, in what follows we shall use the substitutional approach to define the consequence relations we are interested in, and not the objectual one.

DEFINITION 131. Let $S = \langle D, I\rangle$ be an L-structure for a matrix $\mathcal{P} = \langle \mathcal{V}, \mathcal{D}, \mathcal{O}\rangle$.

- The valuation v_S *satisfies* a sentence ψ (denoted by $v_S \models \psi$), if $v_S(\psi) \in \mathcal{D}$. v_S is a *model* of $\Gamma \subseteq Frm^{\mathsf{cl}}_{L(D)}$ (denoted by $v_S \models \Gamma$), if $v_S(\psi) \in \mathcal{D}$ for every $\psi \in \Gamma$.

- v_S satisfies a formula $\varphi \in Frm_L$, denoted by $v_S \models \varphi$, if for every closed $L(D)$-instance φ' of φ, ($v_S(\varphi')$ is defined and) $v_S(\varphi') \in \mathcal{D}$. v_S satisfies a set of formulas $\Gamma \subseteq Frm_L$, denoted by $v_S \models \Gamma$, if for every closed $L(D)$-instance Γ' of Γ, $v_S \models \Gamma'$.

In analogy to the propositional case (recall Definition 1), a (Taskian) logic $\mathbf{L}$ is a pair $\langle L, \vdash \rangle$, where L is a language and $\vdash$ is a structural and consistent scr (tcr) for L.[24] However, unlike in the propositional case, when variables and quantifiers are involved, there is more than one natural way of defining consequence relations induced by a given matrix. Two such relations which are usually associated with first-order logic are the truth and the validity consequence relations ([Avron, 1991]). Using the substitutional approach they can be generalized to the context of many-valued matrices as follows:

DEFINITION 132.

- For sets of L-formulas Γ, Δ, we say that $\Gamma \vdash^t_{\mathcal{P}} \Delta$ if for every L-structure S and every closed $L(D)$-instance $\Gamma' \cup \Delta'$ of $\Gamma \cup \Delta$: $v_S \models \Gamma'$ implies that $v_S \models \psi$ for some $\psi \in \Delta'$.

- We say that $\Gamma \vdash^v_{\mathcal{P}} \Delta$ if for every L-structure S: $v_S \models \Gamma$ implies that $v_S \models \psi$ for some $\psi \in \Delta$.

To demonstrate the difference between the validity and the truth consequence relations, consider a matrix $\mathcal{P}$ for a first-order language L with the standard interpretations of the quantifiers $\forall$ and $\exists$ from Example 124. Then $p(x) \vdash^v_{\mathcal{P}} \forall x p(x)$, but $p(x) \not\vdash^t_{\mathcal{P}} \forall x p(x)$. On the other hand, the classical deduction theorem holds for $\vdash^t_{\mathcal{P}}$, but not for $\vdash^v_{\mathcal{P}}$. However, the two consequence relations are identical from the point of view of theoremhood (i.e., $\vdash^t_{\mathcal{P}} \psi$ iff $\vdash^v_{\mathcal{P}} \psi$). This is a special case of the second part of the following well-known proposition:

PROPOSITION 133. *Let $\mathcal{P}$ be a matrix for L.*

1. $\Gamma \vdash^t_{\mathcal{P}} \psi$ *implies* $\Gamma \vdash^v_{\mathcal{P}} \psi$.

2. *If* $\Gamma \subseteq Frm^{\mathsf{cl}}_L$ *(i.e,* Γ *contsists of* sentences*), then* $\Gamma \vdash^t_{\mathcal{P}} \psi$ *iff* $\Gamma \vdash^v_{\mathcal{P}} \psi$.

[24] The extension of the notion of structurality to languages with quantifiers is not an immediate matter. We omit the technical details.

8 NMATRICES WITH QUANTIFIERS

The extension of Nmatrices to languages with quantifiers is a natural generalization of Definition 123:

DEFINITION 134. An Nmatrix for L is a tuple $\mathcal{M} = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$, where:

- $\mathcal{V}$ is a non-empty set of truth-values,
- $\mathcal{D}$ is a non-empty proper set of $\mathcal{V}$,
- $\mathcal{O}$ includes a function $\tilde{\diamond} : \mathcal{V}^n \to 2^{\mathcal{V}} \setminus \{\emptyset\}$ for every n-ary connective of L, and a function $\tilde{\mathcal{Q}} : 2^{\mathcal{V}} \setminus \{\emptyset\} \to 2^{\mathcal{V}} \setminus \{\emptyset\}$ for every quantifier of L.

EXAMPLE 135. Consider the Nmatrix $\mathcal{M} = \langle \{\mathbf{t}, \mathbf{f}\}, \{\mathbf{t}\}, \mathcal{O} \rangle$ for a first-order language L, where $\mathcal{O}$ contains the following (non-standard) interpretations of $\forall$ and $\exists$:

$\mathbf{H}$	$\tilde{\forall}(\mathbf{H})$	$\tilde{\exists}(\mathbf{H})$
$\{\mathbf{t}\}$	$\{\mathbf{t}, \mathbf{f}\}$	$\{\mathbf{t}\}$
$\{\mathbf{t}, \mathbf{f}\}$	$\{\mathbf{f}\}$	$\{\mathbf{t}, \mathbf{f}\}$
$\{\mathbf{f}\}$	$\{\mathbf{f}\}$	$\{\mathbf{f}\}$

L-structures for Nmatrices are defined like in Definition 125. However, it seems difficult to apply the objectual approach to quantification in the context of Nmatrices. The reason for this is that unlike the deterministic case (recall Definition 127), an L-structure S and an assignment G *do not uniquely determine* the valuation $v_{S,G}$ in a Nmatrix $\mathcal{M}$. Thus the expression $v_{S,G[x:=a]}$ (used in Definition 127) is not well-defined. The substitutional approach, in contrast, *is* suitable for the non-deterministic context.

DEFINITION 136. Let $S = \langle D, I \rangle$ be an L-structure.

1. A set of sentences $W \subseteq Frm^{\mathsf{cl}}_{L(D)}$ is *closed under subsentences with respect to* S if (i) for every n-ary connective $\diamond$ of L: $\psi_1, ..., \psi_n \in W$ whenever $\diamond(\psi_1, ..., \psi_n) \in W$, and (ii) for every quantifier $\mathcal{Q}$ of L and every $a \in D$: if $\mathcal{Q}x\psi \in W$, then $\psi\{\overline{a}/x\} \in W$.

2. Let $W \subseteq Frm^{\mathsf{cl}}_{L(D)}$ be some set of sentences closed under subsentences with respect to S. We say that a partial S-valuation $v : W \to \mathcal{V}$ is *semi-legal in* $\mathcal{M}$ if it satisfies the following conditions:

    - $v(p(\mathbf{t}_1, ..., \mathbf{t}_n)) = I(p)(I(\mathbf{t}_1), ..., I(\mathbf{t}_n))$
    - $v(\diamond(\psi_1, ..., \psi_n)) \in \tilde{\diamond}_{\mathcal{M}}(v(\psi_1), ..., v(\psi_n))$

- $v(\mathcal{Q}x\psi) \in \tilde{\mathcal{Q}}(\{v(\psi\{\overline{a}/x\}) \mid a \in D\})$

A partial S-valuation v in $\mathcal{M}$ is a (full) S-valuation if its domain is $Frm^{\mathsf{cl}}_{L(D)}$.

It is easy to see that the above notion of a valuation is now well-defined. This is due to the fact that the truth-value $v(\mathcal{Q}x\psi)$ depends on the truth-values assigned by v *itself* to the subsentences of $\mathcal{Q}x\psi$ (unlike in our previous attempt using objectual quantification, where $v_{S,G[x:=a]}$ was used in the definition of $v_{S,G}$).

REMARK 137. It is important to stress the difference between our use of notation in the above definition and the one used in Definition 130. Given a (deterministic) matrix $\mathcal{P}$ and an L-structure S, the valuation v_S is uniquely determined by S and $\mathcal{P}$. However, this is not the case for non-deterministic valuations in an Nmatrix $\mathcal{M}$ (although S does determine the truth-values of the atomic sentences), and so we write "an S-valuation v" (compare to "the valuation v_S").

DEFINITION 138. Let $S = \langle D, I\rangle$ be an L-structure for an Nmatrix $\mathcal{M} = \langle \mathcal{V}, \mathcal{D}, \mathcal{O}\rangle$. Let $W \subseteq Frm^{\mathsf{cl}}_{L(D)}$ be some set of sentences closed under subsentences with respect to S, and let $v : W \to \mathcal{V}$ be a partial S-valuation.

- v *satisfies* a sentence $\psi \in W$ (denoted by $v \models \psi$), if $v(\psi) \in \mathcal{D}$. v is a *model* of $\Gamma \subseteq W$ (denoted by $v \models \Gamma$), if $v(\psi) \in \mathcal{D}$ for every $\psi \in \Gamma$.

- v satisfies a formula $\varphi \in Frm_L$ (denoted by $v \models \varphi$), if for every closed $L(D)$-instance φ' of φ, ($v(\varphi')$ is defined and) $v(\varphi') \in \mathcal{D}$. v is a *model* of $\Gamma \subseteq Frm_L$ (denoted by $v \models \Gamma$), if for every closed $L(D)$-instance Γ' of Γ, $v \models \Gamma'$.

The following analycity property is analogous to that given in Proposition 27 for the propositional case:

PROPOSITION 139. *Let $\mathcal{M}$ be an Nmatrix for L and S an L-structure for $\mathcal{M}$. Any partial S-valuation v, which is semi-legal in $\mathcal{M}$ can be extended to a full S-valuation, which is semi-legal in $\mathcal{M}$.*

At this point we note two important problems concerning the above naive semantics, which do not arise on the propositional level. The first problem is related to the principle of α-equivalence, capturing the idea that the names of bound variables are immaterial. It is of course quite reasonable to expect that in any useful semantics two α-equivalent sentences are always assigned the same truth-value. However, this is not necessarily the case for valuations in Nmatrices as defined above. As an example, consider a language L_a with the unary connective $\neg$ and the quantifier $\forall$. Let $\mathcal{M}_a = \langle\{\mathbf{t}, \mathbf{f}\}, \{\mathbf{t}\}, \mathcal{O}\rangle$ be the Nmatrix for L_a with the standard (deterministic) interpretation of $\forall$ and the non-deterministic interpretation of $\neg$ given in Example 21. Let $S_a = \langle\{a\}, I_a\rangle$ be the simple L_a-structure, such that $I_a(c_a) = a$ and $I_a(p)(a) = \mathbf{t}$. Clearly, there is a $\mathcal{M}_a$-semi-legal S_a-valuation v, such that $v(\neg\forall xp(x)) = \mathbf{t}$ and $v(\neg\forall yp(y)) = \mathbf{f}$. Hence two α-equivalent formulas are not necessarily assigned the same truth-value by a $\mathcal{M}_a$-semi-legal S_a-valuation![25] The second problem is related to the nature of identity and becomes really crucial if equality is added to the language. Suppose we have two terms, denoting the same object. It is again reasonable to expect that we should be able to use these terms interchangeably, or substitute one term for another in any context. Returning to our example, suppose we add another constant d_a to the language L_a and extend the structure S_a to interpret it: $I(d_a) = a$. Thus the constants d_a and c_a refer to the same element a, but there is a $\mathcal{M}_a$-legal valuation v, such that $v(\neg p(c_a)) = \mathbf{t}$ and $v(\neg p(d_a)) = \mathbf{f}$.

These problems are directly related to introducing a new level of freedom by the non-deterministic choice of truth-values for quantified formulas. In view of these issues, further limitations need to be imposed on this choice. This can be done by introducing the following congruence relation, capturing these principles.

DEFINITION 140. Let $S = \langle D, I\rangle$ be an L-structure for an Nmatrix $\mathcal{M}$. The relation $\sim^S$ between terms of $L(D)$ is defined as follows:

- $x \sim^S x$ for every variable x of L.
- If $\mathbf{t}, \mathbf{t}' \in Trm^{\mathsf{cl}}_{L(D)}$ and $I[\mathbf{t}] = I[\mathbf{t}']$, then $\mathbf{t} \sim^S \mathbf{t}'$.
- If $\mathbf{t}_1 \sim^S \mathbf{t}'_1, ..., \mathbf{t}_n \sim^S \mathbf{t}'_n$, then $f(\mathbf{t}_1, ..., \mathbf{t}_n) \sim^S f(\mathbf{t}'_1, ..., \mathbf{t}'_n)$.

The relation $\sim^S$ between formulas of $L(D)$ is defined as follows:

- If $\mathbf{t}_1 \sim^S \mathbf{t}'_1, \mathbf{t}_2 \sim^S \mathbf{t}'_2, ..., \mathbf{t}_n \sim^S \mathbf{t}'_n$, then $p(\mathbf{t}_1, ..., \mathbf{t}_n) \sim^S p(\mathbf{t}'_1, ..., \mathbf{t}'_n)$.
- If $\psi_i \sim^S \varphi_i$ for all $1 \leq i \leq n$, then $\diamond(\psi_1, ..., \psi_n) \sim^S \diamond(\varphi_1, ..., \varphi_n)$ for every n-ary connective $\diamond$ of $\mathcal{L}$.
- If $\psi\{z/x\} \sim^S \varphi\{z/y\}$, where x, y are distinct variables and z is a new variable, then $\mathcal{Q}x\psi \sim^S \mathcal{Q}y\varphi$ for every quantifier $\mathcal{Q}$ of L.

[25]Of course, two different occurrences of the same formula are still assigned the same truth-value, since a valuation is a mapping from *formulas* to truth-values.

The following lemma is easy to prove:

LEMMA 141. *Let S be an L-structure, and let $\boldsymbol{t}_1, \boldsymbol{t}_2$ be closed terms of $L(D)$ such that $\boldsymbol{t}_1 \sim^S \boldsymbol{t}_2$. Let ψ_1, ψ_2 be $L(D)$-formulas such that $\psi_1 \sim^S \psi_2$. Then $\psi_1\{\boldsymbol{t}/x\} \sim^S \psi_2\{\boldsymbol{t}_2/x\}$.*

Using the above congruence relation, we can now modify Definition 136 as follows:

DEFINITION 142. Let S be an L-structure and $\mathcal{M}$ an Nmatrix for L. Let $W \subseteq Frm^{\mathsf{cl}}_{L(D)}$ be some set of sentences closed under subsentences with respect to S. A partial S-valuation $v : W \to \mathcal{V}$ is *$\sim^S$-legal in $\mathcal{M}$* if it is semi-legal in $\mathcal{M}$ and for every $\psi, \varphi \in W$: $\psi \sim^S \varphi$ implies $v(\psi) = v(\varphi)$.

Now we come to the definition of consequence relations induced by Nmatrices, analogous to Definition 132:

DEFINITION 143.

- For sets of L-formulas Γ, Δ, we say that $\Gamma \vdash^t_{\mathcal{M}} \Delta$ if for every L-structure S, every S-valuation v which is $\sim^S$-legal in $\mathcal{M}$, and every closed $L(D)$-instance $\Gamma' \cup \Delta'$ of $\Gamma \cup \Delta$: $v \models \Gamma'$ implies $v \models \psi$ for some $\psi \in \Delta'$.
- We say that $\Gamma \vdash^v_{\mathcal{M}} \Delta$ if for every L-structure S and S-valuation v which is $\sim^S$-legal in $\mathcal{M}$: $v \models \Gamma$ implies $v \models \psi$ for some $\psi \in \Delta$.

The following extension of Proposition 133 to the context of Nmatrices can be easily proved:

PROPOSITION 144. *Let $\mathcal{M}$ be an Nmatrix for L.*

1. *$\Gamma \vdash^t_{\mathcal{M}} \psi$ implies $\Gamma \vdash^v_{\mathcal{M}} \psi$.*
2. *If $\Gamma \subseteq Frm^{\mathsf{cl}}_L$ (i.e, Γ contains only closed formulas), then $\Gamma \vdash^t_{\mathcal{M}} \psi$ iff $\Gamma \vdash^v_{\mathcal{M}} \psi$.*

As for analycity, the following analogue of Proposition 139 can be proved (the presence of the $\sim^S$-relation makes its proof less trivial):

PROPOSITION 145. *Let $\mathcal{M}$ be an Nmatrix for L and S an L-structure. Then any partial S-valuation which is $\sim^S$-legal in $\mathcal{M}$ can be extended to a full S-valuation which is $\sim^S$-legal in $\mathcal{M}$.*

We end this section by generalizing the notions of reduction and refinement from Definition 30 to languages with quantifiers:

DEFINITION 146. Let $\mathcal{M}_1 = \langle \mathcal{V}_1, \mathcal{D}_1, \mathcal{O}_1 \rangle$ and $\mathcal{M}_2 = \langle \mathcal{V}_2, \mathcal{D}_2, \mathcal{O}_2 \rangle$ be two Nmatrices for L.

1. A reduction of $\mathcal{M}_1$ to $\mathcal{M}_2$ is a function $F : \mathcal{V}_1 \to \mathcal{V}_2$, such that:
   - For every $x \in \mathcal{V}_1$, $x \in \mathcal{D}_1$ iff $F(x) \in \mathcal{D}_2$.
   - $F(y) \in \tilde{\diamond}_{\mathcal{M}_2}(F(x_1), ..., F(x_n))$ for every n-ary connective $\diamond$ of L and every $x_1, ..., x_n, y \in \mathcal{V}_1$, such that $y \in \tilde{\diamond}_{\mathcal{M}_1}(x_1, ..., x_n)$.
   - $F(y) \in \tilde{\mathcal{Q}}_{\mathcal{M}_2}(\{F(z) \mid z \in H\})$ for every quantifier $\mathcal{Q}$ of L, every $y \in \mathcal{V}_1$ and $H \in 2^{\mathcal{V}_1} \setminus \{\emptyset\}$, such that $y \in \tilde{\mathcal{Q}}_{\mathcal{M}_1}(H)$.
2. $\mathcal{M}_1$ is a *refinement* of $\mathcal{M}_2$ if there exists a reduction of $\mathcal{M}_1$ to $\mathcal{M}_2$.

THEOREM 147. *Let $\mathcal{M}_1$ be a refinement of $\mathcal{M}_2$. Then $\vdash^t_{\mathcal{M}_2} \subseteq \vdash^t_{\mathcal{M}_1}$ and $\vdash^v_{\mathcal{M}_2} \subseteq \vdash^v_{\mathcal{M}_1}$.*

REMARK 148. Again an important case in which $\mathcal{M}_1 = \langle \mathcal{V}_1, \mathcal{D}_1, \mathcal{O}_1 \rangle$ is a refinement of $\mathcal{M}_2 = \langle \mathcal{V}_2, \mathcal{D}_2, \mathcal{O}_2 \rangle$ is when $\mathcal{V}_1 \subseteq \mathcal{V}_2$, $\mathcal{D}_1 = \mathcal{D}_2 \cap \mathcal{V}_1$, $\widetilde{\diamond}_{\mathcal{M}_1}(\vec{x}) \subseteq \widetilde{\diamond}_{\mathcal{M}_2}(\vec{x})$ for every n-ary connective $\diamond$ of L and every $\vec{x} \in \mathcal{V}_1^n$, and $\tilde{\mathcal{Q}}_{\mathcal{M}_1}(H) \subseteq \tilde{\mathcal{Q}}_{\mathcal{M}_2}(H)$ for every quantifier $\mathcal{Q}$ of L and every $H \in 2^{\mathcal{V}_1} \setminus \{\emptyset\}$. It is easy to see that the identity function on $\mathcal{V}_1$ is in this case a reduction of $\mathcal{M}_1$ to $\mathcal{M}_2$. We will refer to this kind of refinement as *simple*.

9 THE FIRST-ORDER CASE

Next we focus on the first-order quantifiers $\forall$ and $\exists$ with their natural interpretations. Throughout this section we assume that $\forall$ and $\exists$ are in L. In Example 124 we have seen the standard interpretation of these quantifiers in the two-valued case. This can be generalized to an arbitrary number of truth-values as follows:

DEFINITION 149. Let $\mathcal{M} = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$ be an Nmatrix for L. We say that a quantifier $\mathcal{Q}$ is *universally interpreted in* $\mathcal{M}$ if for all $H \in 2^{\mathcal{V}} \setminus \{\emptyset\}$:

$$\widetilde{\mathcal{Q}}(H) \subseteq \begin{cases} \mathcal{D} & \text{if } H \subseteq \mathcal{D} \\ \mathcal{F} & \text{otherwise} \end{cases}$$

A quantifier $\mathcal{Q}$ is *existentially interpreted in* $\mathcal{M}$ if for all $H \in 2^{\mathcal{V}} \setminus \{\emptyset\}$:

$$\widetilde{\mathcal{Q}}(H) \subseteq \begin{cases} \mathcal{D} & \text{if } H \cap \mathcal{D} \neq \emptyset \\ \mathcal{F} & \text{otherwise} \end{cases}$$

At this point we note a problem, the nature of which is very similar to the problems of the α-equivalence and identity principles which we handled in the previous section. Namely, in the context of universally and existentially interpreted quantifiers, one would expect the equivalence of two formulas, where one is obtained from the other by deletion or addition of *void* quantifiers (by a void quantifier we mean the case then a variable is bound vacuously). For instance, we expect $\neg\forall xp(c)$ and $\neg p(c)$ to be equivalent. This, however, is not always the case, again due to the degree of freedom introduced by the non-deterministic choice in our semantic framework. For an example, consider again the Nmatrix $\mathcal{M}_a = \langle\{\mathbf{t},\mathbf{f}\},\{\mathbf{t}\},\mathcal{O}\rangle$ discussed in the previous section, where $\neg$ is interpreted like in Example 21, and $\forall$ and $\exists$ have the universal and the existential interpretations in $\mathcal{M}_a$ (respectively). Then there exists an L-structure S and an S-valuation v legal in $\mathcal{M}_a$, such that $v(\neg\forall xp(c)) = \mathbf{t}$, but $v(\neg p(c)) = \mathbf{f}$.
The solution is similar to the one in the previous section: we extend the congruence relation $\sim^S$ to capture the principle of void quantification:

DEFINITION 150. Let L be a language which includes the quantifiers $\forall$ and $\exists$ and let $S = \langle D, I\rangle$ be an L-structure. $\sim^S_{\forall\exists}$ is the minimal congruence relation between $L(D)$-formulas, which satisfies: (i) $\sim^S \subseteq \sim^S_{\forall\exists}$, and (ii) If $\psi \sim^S_{\forall\exists} \psi'$ and x does not occur free in ψ, then $\mathcal{Q}x\psi \sim^S_{\forall\exists} \psi'$ for $\mathcal{Q} \in \{\forall,\exists\}$.

The following extension of Lemma 141 is again easy to prove:

LEMMA 151. *Let S be an L-structure, and let $\boldsymbol{t}_1, \boldsymbol{t}_2$ be closed terms of $L(D)$ such that $\boldsymbol{t}_1 \sim^S \boldsymbol{t}_2$. Let ψ_1, ψ_2 be $L(D)$-formulas such that $\psi_1 \sim^S_{\forall\exists} \psi_2$. Then $\psi_1\{\boldsymbol{t}/x\} \sim^S_{\forall\exists} \psi_2\{\boldsymbol{t}_2/x\}$.*

DEFINITION 152. Let S be an L-structure and $\mathcal{M}$ an Nmatrix for L. Let $W \subseteq Frm^{\mathsf{cl}}_{L(D)}$ be some set of sentences closed under subsentences with respect to S. A partial S-valuation $v : W \to \mathcal{V}$ is $\sim^S_{\forall\exists}$*-legal in* $\mathcal{M}$ if it is semi-legal in $\mathcal{M}$ and for every $\psi, \varphi \in W$: $\psi \sim^S_{\forall\exists} \varphi$ implies $v(\psi) = v(\varphi)$.

Using the above definition, we can now modify the notions of truth- and validity-based consequence relations from Definition 143:

DEFINITION 153. The consequence relations $\vdash^t_{\mathcal{M},\forall\exists}$ and $\vdash^v_{\mathcal{M},\forall\exists}$ are defined like $\vdash^t_{\mathcal{M}}$ and $\vdash^v_{\mathcal{M}}$ (respectively), but using $\sim^S_{\forall\exists}$ rather than $\sim^S$.

PROPOSITION 154. *Let $\mathcal{M}$ be an Nmatrix for L.*

1. $\Gamma \vdash^t_{\mathcal{M},\forall\exists} \psi$ *implies* $\Gamma \vdash^v_{\mathcal{M},\forall\exists} \psi$.
2. *If* $\Gamma \subseteq Frm^{\mathsf{cl}}_L$ *(i.e, Γ contains only closed formulas), then* $\Gamma \vdash^t_{\mathcal{M},\forall\exists} \psi$ *iff* $\Gamma \vdash^v_{\mathcal{M},\forall\exists} \psi$.

It should be noted that analycity for $\sim^S_{\forall\exists}$ is *not* always guaranteed. Consider, for instance, an Nmatrix $\mathcal{M}_v = \langle\{\mathbf{t},\mathbf{f}\},\{\mathbf{t}\},\mathcal{O}\rangle$ for some first-order language L, with the following interpretation of $\forall$: $\tilde{\forall}[\{H\}] = \{\mathbf{t}\}$ for every $H \subseteq P^+(\{\mathbf{t},\mathbf{f}\})$. Let $S = \langle\{a\}, I\rangle$ be an L-structure, such that $I(c) = a$ and $I(p) = \emptyset$. Let $W = \{p(c)\}$. Then no partial valuation on W can be extended to a full $\mathcal{M}$-legal valuation v which respects $\sim^S_{\forall\exists}$. Next we characterize those Nmatrices in which this problem does not occur.

DEFINITION 155. Let L include propositional connectives and (at most) the quantifiers $\forall$ and $\exists$. An Nmatrix $\mathcal{M}$ for L is $\{\forall,\exists\}$-*analytic* if every L-structure S has the property that every partial S-valuation which is $\sim^S_{\forall\exists}$-legal in $\mathcal{M}$ can be extended to a full S-valuation which is $\sim^S_{\forall\exists}$-legal in $\mathcal{M}$.

THEOREM 156. *Let $\mathcal{M} = \langle\mathcal{V},\mathcal{D},\mathcal{O}\rangle$ be an Nmatrix for a language which in addition to propositional connectives includes (at most) the quantifiers $\forall$ and $\exists$. $\mathcal{M}$ is $\{\forall,\exists\}$-analytic iff for every $a \in \mathcal{V}$: $a \in \tilde{\mathcal{Q}}[\{a\}]$ for $\mathcal{Q} \in \{\forall,\exists\}$.*

Next we turn to the problem of extending a propositional formal system having a nondeterministic semantics to the first-order level. We take as an example $\mathbf{H}LK^+$, the Hilbert-type system which corresponds to the basic Nmatrix $\mathcal{M}^B_4$ from Definition 80 (see Remark 79).

DEFINITION 157. $\mathbf{QHL_0}$ is obtained by adding to $\mathbf{H}LK^+$ the following standard axioms and inference rules for $\forall$ and $\exists$:

$$\forall x\psi \supset \psi\{\mathbf{t}/x\} \qquad \psi\{\mathbf{t}/x\} \supset \exists x\psi$$

$$\frac{(\varphi \supset \theta)}{(\varphi \supset \forall x\theta)} \qquad \frac{(\theta \supset \varphi)}{(\exists x\theta \supset \varphi)}$$

where $\mathbf{t}$ is any term free for x in ψ, and x does not occur free in φ.

Unfortunately, $\mathbf{QHL_0}$ is not very useful. Due to the absence of axioms for negation, neither the α-equivalence principle, nor the void quantification principle, are derivable in it. For instance, $\nvdash_{\mathbf{QHL_0}} \neg\forall xp(x) \leftrightarrow \neg\forall yp(y)$, and $\nvdash_{\mathbf{QHL_0}} (\neg\forall xp(c)) \leftrightarrow \neg p(c)$. To handle this, we follow da Costa's approach from [da Costa, 1974]:

DEFINITION 158. $\sim^{dc}$ is the minimal congruence relation between formulas, which satisfies for $\mathcal{Q} \in \{\forall,\exists\}$:

- If $\psi\{z/x\} \sim^{dc} \psi'\{z/y\}$, where z is fresh, then $\mathcal{Q}x\psi \sim^{dc} \mathcal{Q}y\psi'$.
- If $\psi \sim^{dc} \psi'$ and x does not occur free in ψ, then $\mathcal{Q}x\psi \sim^{dc} \psi'$.

DEFINITION 159. Let $\mathbf{QHL}$ be the system obtained from $\mathbf{QHL_0}$ by adding the axiom **(DC)** $\psi \supset \psi'$ whenever $\psi \sim^{dc} \psi'$.

DEFINITION 160. Let the Nmatrix $\mathcal{QM}_4^B$ be the extension of the Nmatrix $\mathcal{M}_4^B$ (Definition 80) with the following interpretations of $\forall$ and $\exists$:

$$\widetilde{\forall}(H) = \begin{cases} \mathcal{D} & \text{if } H \subseteq \mathcal{D} \\ \mathcal{F} & \text{otherwise} \end{cases}$$

$$\widetilde{\exists}(H) = \begin{cases} \mathcal{D} & \text{if } H \cap \mathcal{D} \neq \emptyset \\ \mathcal{F} & \text{otherwise} \end{cases}$$

PROPOSITION 161. $\Gamma \vdash_{\mathbf{QHL}} \psi$ *iff* $\Gamma \vdash^v_{\mathcal{QM}_4^B, \forall\exists} \psi$.

The proof is very similar to the proof of Theorem 164 below.

10 AN APPLICATION: NMATRICES FOR FIRST-ORDER LOGICS OF FORMAL INCONSISTENCY

In this section we further apply the framework of Nmatrices with first-order quantifiers to provide semantics for first-order LFIs (the propositional fragments of which were already handled in section 6). For simplicity of presentation, we formulate these logics in terms of Hilbert-style systems, rather than in terms of abstract consequence relations. The results of this section are mainly taken from [Avron and Zamansky, 2007c; Zamansky and Avron, 2007]. Throughout it, we let $L_C = \{\vee, \wedge, \supset, \neg, \circ, \forall, \exists\}$.

Our starting point will be the basic paraconsistent system **QHB**, obtained from **QHL** (Definition 159) by the addition of the following schemata:

$$(\Rightarrow \neg)\ \varphi \vee \neg\varphi \qquad (\mathbf{b})\ (\circ\varphi \wedge \neg\varphi \wedge \varphi) \supset \psi$$

CONVENTION 162. **QHB** is the obvious first-order extension of the Hilbert-style axiomatization of the logic **B** from Example 112. Accordingly, in this section we shall refer to **QHB** simply as **B**.

We obtain a large family of first-order LFIs by extending **B** with various combinations of axioms from $HLFIR$ (Definition 106), to which we add the following quantifier-related versions of the axioms (see, e.g. [Carnielli et. al., 2000]) (**a**) and (**o**) which were considered in section 6: [26]

$$(\mathbf{a}_{\mathcal{Q}})\ \forall x \circ\varphi \supset (\circ(\mathcal{Q}x\varphi)) \qquad (\mathbf{o}_{\mathcal{Q}})\ \exists x \circ\varphi \supset (\circ(\mathcal{Q}x\varphi)) \qquad (\mathcal{Q} \in \{\forall, \exists\})$$

[26]See [Zamansky and Avron, 2006b; Avron and Zamansky, 2007c; Zamansky and Avron, 2007] for other quantifier-related axioms treated in the context of Nmatrices.

DEFINITION 163. Let $\mathcal{QR} = HLFIR \cup \{(\mathbf{a}_\forall), (\mathbf{a}_\exists), (\mathbf{o}_\forall), (\mathbf{o}_\exists)\}$. For a set $S \subseteq \mathcal{QR}$, $\mathbf{B}[S]$ is the system obtained by adding the axioms in S to $\mathbf{B}$.

Our Nmatrix for $\mathbf{B}$ is a straightforward extentension of the Nmatrix $\mathcal{M}_5^B$ from Example 112:

THEOREM 164. *Let $\mathcal{QM}_5^B$ be the extension of $\mathcal{M}_5^B$ with the following interpretations of quantifiers:*

$$\widetilde{\forall}(H) = \begin{cases} \mathcal{D} & \text{if } H \subseteq \mathcal{D} \\ \mathcal{F} & \text{otherwise} \end{cases} \qquad \widetilde{\exists}(H) = \begin{cases} \mathcal{D} & \text{if } H \cap \mathcal{D} \neq \emptyset \\ \mathcal{F} & \text{otherwise} \end{cases}$$

Then $\Gamma \vdash^v_{\mathcal{QM}_5^B, \forall\exists} \psi_0$ iff $\Gamma \vdash_{\mathbf{B}} \psi_0$.

Proof. The proof of soundness is not hard and is left to the reader. For completeness, we first note that by definition of the interpretation of $\forall$ in $\mathcal{QM}_5^B$, $\forall x\varphi \vdash_{\mathcal{QM}_5^B} \varphi$ and $\varphi \vdash_{\mathcal{QM}_5^B} \forall x\varphi$ for every formula φ and every variable x. Obviously the same relations hold between φ and $\forall x\varphi$ also in $\mathbf{B}$. It follows that we may assume that all formulas in $\Gamma \cup \{\psi_0\}$ are sentences. It is also easy to see that we may restrict ourselves to sentences in σ_r, the signature consisting of all the constants, function, and predicate symbols occurring in $\Gamma \cup \{\psi_0\}$. Now suppose that $\Gamma \not\vdash_{\mathbf{B}} \psi_0$. We will construct an σ_r-structure S and a $\mathcal{QM}_5^B$-legal S-valuation v, such that $v \models \Gamma$, but $v \not\models \psi_0$. Let L' be the language obtained from σ_r by adding a countably infinite set of new constants. It is a standard matter to show (using a usual Henkin-type construction) that Γ can be extended to a maximal set Γ^* of sentences in L', such that: (i) $\Gamma^* \not\vdash_{\mathbf{B}} \psi_0$, (ii) $\Gamma \subseteq \Gamma^*$, (iii) For every L'-sentence $\exists x\psi \in \Gamma^*$ there is a constant $\mathbf{c}$ of L', such that $\psi\{\mathbf{c}/x\} \in \Gamma^*$, and (iv) For every L'-sentence $\forall x\psi \notin \Gamma^*$, there is a constant $\mathbf{c}$ of L', such that $\psi\{\mathbf{c}/x\} \notin \Gamma^*$. (The last property follows from property (iii), the deduction theorem for $\mathbf{B}$, and the fact that for any $x \notin Fv(\varphi)$, $(\forall x\psi \supset \varphi) \supset \exists x(\psi \supset \varphi)$ is provable in $\mathbf{B}$.) It is now easy to show that Γ^* has the following properties: (1) If $\psi \notin \Gamma^*$, then $\psi \supset \psi_0 \in \Gamma^*$, (2) $\psi \vee \varphi \in \Gamma^*$ iff either $\varphi \in \Gamma^*$ or $\psi \in \Gamma^*$, (3) $\psi \wedge \varphi \in \Gamma^*$ iff both $\varphi \in \Gamma^*$ and $\psi \in \Gamma^*$, (4) $\varphi \supset \psi \in \Gamma^*$ iff either $\varphi \notin \Gamma^*$ or $\psi \in \Gamma^*$, (5) Either $\psi \in \Gamma^*$ or $\neg\psi \in \Gamma^*$, (6) If ψ and $\neg\psi$ are both in Γ^*, then $\circ\psi \notin \Gamma^*$, (7) If $\psi \in \Gamma^*$, then for every L'-sentence ψ' such that $\psi' \sim^{dc} \psi$: $\psi' \in \Gamma^*$, (8) If $\forall x\theta \in \Gamma^*$, then for every closed L'-term $\mathbf{t}$: $\theta\{\mathbf{t}/x\} \in \Gamma^*$. If $\forall x\theta \notin \Gamma^*$, then there is some closed term $\mathbf{t}_\theta$ of L', such that $\theta\{\mathbf{t}_\theta/x\} \notin \Gamma^*$, (9) If $\exists x\theta \in \Gamma^*$, then there is some closed term $\mathbf{t}_\theta$ of L, such that $\theta\{\mathbf{t}_\theta/x\} \in \Gamma^*$. If $\exists x\theta \notin \Gamma^*$, then for every closed term $\mathbf{t}$ of L':

$\theta\{\mathbf{t}/x\} \notin \Gamma^*$.
The L'-structure $S = \langle D, I\rangle$ is defined as follows:

- D is the set of all the closed terms of L'.
- For every constant c of L': $I(c) = c$.
- For every $\mathbf{t}_1, ..., \mathbf{t}_n \in D$: $I(f)(\mathbf{t}_1, ..., \mathbf{t}_n) = f(\mathbf{t}_1, ..., \mathbf{t}_n)$.
- For every $\mathbf{t}_1, ..., \mathbf{t}_n \in D$: $I(p)(\mathbf{t}_1, ..., \mathbf{t}_n) = \langle x, y, z\rangle$, where $x, y, z \in \{0, 1\}$ and (i) $x = 1$ iff $p(\mathbf{t}_1, ..., \mathbf{t}_n) \in \Gamma^*$, (ii) $y = 1$ iff $\neg p(\mathbf{t}_1, ..., \mathbf{t}_n) \in \Gamma^*$, (iii) $z = 1$ iff $\circ p(\mathbf{t}_1, ..., \mathbf{t}_n) \in \Gamma^*$.

Given an $L'(D)$-sentence ψ, let the sentence $\widetilde{\psi}$ be obtained by replacing all individual constants $\overline{\mathbf{t}}$ occurring in ψ by the respective (closed) term $\mathbf{t}$. Then the following lemma is easy to prove:

LEMMA 165. *For any $\psi, \varphi \in Frm^{\mathsf{cl}}_{L'(D)}$: if $\psi \sim^S_{\forall\exists} \varphi$, then $\widetilde{\psi} \sim^{dc} \widetilde{\varphi}$.*

The refuting S-valuation $v : Frm^{\mathsf{cl}}_{L'(D)} \to \mathcal{V}$ is defined as follows:

$$v(\psi) = \langle x_\psi, y_\psi, z_\psi\rangle$$

where $x_\psi, y_\psi, z_\psi \in \{0, 1\}$ and: (i) $x_\psi = 1$ iff $\widetilde{\psi} \in \Gamma^*$, (ii) $y_\psi = 1$ iff $\widetilde{\neg\psi} \in \Gamma^*$, (iii) $z_\psi = 1$ iff $\widetilde{\circ\psi} \in \Gamma^*$.

Let ψ, ψ' be two $L'(D)$-sentences, such that $\psi \sim^S_{\forall\exists} \psi'$. Then by lemma 165, $\widetilde{\psi} \sim^{dc} \widetilde{\psi'}$, and by property 7 of Γ^*, $\widetilde{\psi} \in \Gamma^*$ iff $\widetilde{\psi'} \in \Gamma^*$. Similarly, since $\neg\psi \sim^S_{\forall\exists} \neg\psi'$ and $\circ\psi \sim^S_{\forall\exists} \circ\psi'$, $\neg\widetilde{\psi} = \widetilde{\neg\psi} \sim^{dc} \widetilde{\neg\psi'} = \neg\widetilde{\psi'}$ and $\widetilde{\circ\psi} \sim^{dc} \widetilde{\circ\psi'}$. Thus $\widetilde{\neg\psi} \in \Gamma^*$ iff $\widetilde{\neg\psi'} \in \Gamma^*$ and $\widetilde{\circ\psi} \in \Gamma^*$ iff $\widetilde{\circ\psi'} \in \Gamma^*$. Hence $v(\psi) = v(\psi')$ and so v respects the $\sim^S_{\forall\exists}$ relation.
It remains to check that v respects the interpretations of the connectives and quantifiers in $\mathcal{QM}_5$. This is guaranteed by the properties of Γ^*. We prove this for the case of $\forall$:

- Let $\forall x\psi$ be an $L'(D)$-sentence, such that $\{v(\psi\{\overline{a}/x\}) \mid a \in D\} \subseteq \mathcal{D}$. Suppose by contradiction that $v(\forall x\psi) \notin \mathcal{D}$. Then $\widetilde{\forall x\psi} = \forall x\widetilde{\psi} \notin \Gamma^*$. By property 8 of Γ^*, there exists some closed L'-term $\mathbf{t}$, such that $\widetilde{\psi}\{\mathbf{t}/x\} \notin \Gamma^*$. Then $v(\widetilde{\psi}\{\mathbf{t}/x\}) \notin \mathcal{D}$. Since $\psi \sim^S_{\forall\exists} \widetilde{\psi}$, by lemma 151 also $\psi\{\mathbf{t}/x\} \sim^S_{\forall\exists} \widetilde{\psi}\{\mathbf{t}/x\}$. We have already shown that v respects the $\sim^S_{\forall\exists}$ relation, and so $v(\psi\{\mathbf{t}/x\}) \notin \mathcal{D}$. By lemma 151 again, $\psi\{\mathbf{t}/x\} \sim^S_{\forall\exists} \psi\{\overline{\mathbf{t}}/x\}$, and so $v(\psi\{\overline{\mathbf{t}}/x\}) \notin \mathcal{D}$, in contradiction to our assumption.

- Let $\forall x\psi$ be an $L'(D)$-sentence, such that $\{v(\psi\{\overline{a}/x\}) \mid a \in D\} \cap \mathcal{F} \neq \emptyset$. Suppose by contradiction that $v(\forall x\psi) \notin \mathcal{F}$. Then $\widetilde{\forall x\psi} \in \Gamma^*$. By property 8 of Γ^*, for every closed L'-term $\mathbf{t}$: $\widetilde{\psi}\{\mathbf{t}/x\} \in \Gamma^*$. Then $v(\widetilde{\psi}\{\mathbf{t}/x\}) \in \mathcal{D}$. Similarly to the previous case, we get that $v(\psi\{\overline{a}/x\}) \in \mathcal{D}$ for every $a \in D$, in contradiction to our assumption.

Now for every L'-sentence ψ: $v(\psi) \in \mathcal{D}$ iff $\psi \in \Gamma^*$. So $v \models \Gamma$ (recall that $\Gamma \subseteq \Gamma^*$), but $v \not\models \psi_0$. ■

Like in the propositional case, the systems obtained by adding some set of axioms from $\mathcal{QR}$ to $\mathbf{B}$ can be characterized by the simple refinement of $\mathcal{QM}_5^B$ induced by the conditions corresponding to the axioms from $\mathcal{QR}$:

DEFINITION 166.

1. Let $\mathsf{Con} = \{\langle x, y, 1\rangle \mid x, y \in \{0,1\}\}$.
   - For $r \in HLFIR$, $C(r)$ is defined like in Definition 82 (for NIR) or Definition 109 (for FCR).
   - C($\mathbf{a}_{\mathcal{Q}}$): If $H \subseteq \mathsf{Con}$, then $\tilde{\mathcal{Q}}(H) \subseteq \mathsf{Con}$
   - C($\mathbf{o}_{\mathcal{Q}}$): If $H \cap \mathsf{Con} \neq \emptyset$, then $\tilde{\mathcal{Q}}(H) \subseteq \mathsf{Con}$
2. For $S \subseteq \mathcal{QR}$, $C(S) = \{Cr \mid r \in S\}$, and $\mathcal{QM}_S$ is the weakest simple refinement of $\mathcal{QM}_5^B$ in which the conditions in $C(S)$ are all satisfied.

EXAMPLE 167. Let $S_i = \{(\mathbf{i})\}$, $S_a = S_i \cup \{(\mathbf{a})\}$ and $S_o = S_i \cup \{(\mathbf{o})\}$. The interpretations of $\forall$ and $\exists$ are defined in $\mathcal{QM}_{S_i}$, $\mathcal{QM}_{S_a}$ and $\mathcal{QM}_{S_o}$ (respectively) as follows:[27]

$\mathcal{QM}_{S_i}$:

H	$\widetilde{\forall}[H]$	$\widetilde{\exists}[H]$
$\{t\}$	$\{t,I\}$	$\{t,I\}$
$\{f\}$	$\{f\}$	$\{f\}$
$\{I\}$	$\{t,I\}$	$\{t,I\}$
$\{t,f\}$	$\{f\}$	$\{t,I\}$
$\{t,I\}$	$\{t,I\}$	$\{t,I\}$
$\{f,I\}$	$\{f\}$	$\{t,I\}$
$\{t,f,I\}$	$\{f\}$	$\{t,I\}$

[27]Recall that by C($\mathbf{i_1}$) and C($\mathbf{i_2}$) the truth-values t_I and f_I are deleted and we are left with only three truth-values: t, f and I.

$\mathcal{QM}_{S_o}$:

H	$\widetilde{\forall}[H]$	$\widetilde{\exists}[H]$
$\{t\}$	$\{t\}$	$\{t\}$
$\{f\}$	$\{f\}$	$\{f\}$
$\{I\}$	$\{t, I\}$	$\{t, I\}$
$\{t, f\}$	$\{f\}$	$\{t\}$
$\{t, I\}$	$\{t\}$	$\{t\}$
$\{f, I\}$	$\{f\}$	$\{t\}$
$\{t, f, I\}$	$\{f\}$	$\{t\}$

$\mathcal{QM}_{S_a}$:

H	$\widetilde{\forall}[H]$	$\widetilde{\exists}[H]$
$\{t\}$	$\{t\}$	$\{t\}$
$\{f\}$	$\{f\}$	$\{f\}$
$\{I\}$	$\{t, I\}$	$\{t, I\}$
$\{t, f\}$	$\{f\}$	$\{t\}$
$\{t, I\}$	$\{t, I\}$	$\{t, I\}$
$\{f, I\}$	$\{f\}$	$\{t, I\}$
$\{t, f, I\}$	$\{f\}$	$\{t, I\}$

THEOREM 168. *For $S \subseteq \mathcal{QR}$, $\Gamma \vdash^v_{\mathcal{QM}_S,\forall\exists} \psi$ iff $\Gamma \vdash_{\mathbf{B}[S]} \psi$.*

And what about systems which include the problematic axiom (**l**) (see Figure 4 and Section 6.2)? It suffices to say that they can be handled in a way which is very similar to the systems discussed so far in this section. The only difference is that their semantics is based on the Nmatrix $\mathcal{M}_{\mathbf{Bl}}$ from Definition 115 rather than on $\mathcal{M}_5^B$.

EXAMPLE 169. da Costa's well-known first-order logic C_1^* is the $\circ$-free fragment of $\mathbf{B}[\{(\mathbf{i}), (\mathbf{c}), (\mathbf{a})\}]$ (note that the axioms $(\mathbf{a}_\forall)$ and $(\mathbf{a}_\exists)$ are also included). Let $\mathcal{M}_{C_1^*}$ be the Nmatrix which extends $\mathcal{M}_{C_1}$ from Corollary 121 with the following interpretations of quantifiers:

$$\widetilde{\forall}(H) = \begin{cases} \mathcal{T} & \text{if } H \subseteq \mathcal{T} \\ \mathcal{D} & \text{if } H \subseteq \mathcal{D} \text{ and } H \cap \mathcal{I} \neq \emptyset \\ \mathcal{F} & \text{otherwise} \end{cases}$$

$$\widetilde{\exists}(H) = \begin{cases} \mathcal{T} & \text{if } H \subseteq \mathcal{T} \cup \mathcal{F} \text{ and } H \cap \mathcal{T} \neq \emptyset \\ \mathcal{D} & \text{if } H \cap \mathcal{I} \neq \emptyset \\ \mathcal{F} & \text{otherwise} \end{cases}$$

Then $\Gamma \vdash^v_{\mathcal{QM}_{C_1^*},\forall\exists} \psi$ iff $\Gamma \vdash_{C_1^*} \psi$.

11 CANONICAL DEDUCTION SYSTEMS AND NMATRICES WITH MORE GENERAL QUANTIFIERS

The main goal of this section is to extend the notion of coherent canonical calculi from the propositional case to the level of multi-ary quantifiers. After that we briefly summarize the main related results, omitting the (quite complicated) technical details, which can be found in [Avron and Zamansky, 2010].

Henceforth L is a language with multi-ary quantifiers[28]. In order to work with signed formulas, we need to extend the semantic notions from Definitions 35 and 48 to languages with multi-ary quantifiers. This is done by replacing "$\mathcal{M}$-legal valuation v" by "a structure S and an S-valuation v which is $\sim^S$-legal in $\mathcal{M}$", and $\vdash^d_{\mathcal{M}}$ by $\vdash^t_{\mathcal{M}}$. We then have the following counterpart of Proposition 49:

PROPOSITION 170. *Let $\mathcal{M} = \langle \mathcal{V}, \mathcal{D}, \mathcal{O} \rangle$ be an Nmatrix for L. Then:*

$$\Gamma \vdash^t_{\mathcal{M}} \Delta \textit{ iff } \{\mathcal{D} : \psi \mid \psi \in \Gamma\} \cup \{\mathcal{F} : \psi \mid \psi \in \Delta\} \vdash^t_{\mathcal{M}} \emptyset \textit{ iff } \vdash^t_{\mathcal{M}} \mathcal{F} : \Gamma \cup \mathcal{D} : \Delta$$

In order to represent canonical rules with multi-ary quantifiers, we shall use a simplified language which abstracts over the internal structure of L-formulas. For a single canonical rule introducing some n-ary quantifier, this representation language includes the unary predicate symbols $p_1, ..., p_n$ and some finite sets of variables and constants: a constant signifies the case of a term variable, while a variable signifies an eigenvariable.

DEFINITION 171. For $n \geq 1$ and a set of constants Con, $QL_n(Con)$ is the first-order language with n unary predicate symbols $p_1, ..., p_n$ and the set of constants Con ($QL_n(Con)$ contains no quantifiers or logical connectives).

CONVENTION 172. In case the set Con is clear from context, we will write QL_n instead of $QL_n(Con)$.

DEFINITION 173. A *signed canonical quantifier rule of arity* n over a finite set of signs $\mathcal{V}$ is an expression of the form $[\Theta / \ S : \mathcal{Q}]$, where $\mathcal{Q}$ is an n-ary quantifier, $S \subseteq \mathcal{V}$, and $\Theta = \{\Sigma_1, ..., \Sigma_m\}$, where for all $1 \leq j \leq m$, Σ_j is a clause over QL_n (i.e. it consists of signed formulas of the form $s : p_i(x)$ or $s : p_i(c)$, where $s \in \mathcal{V}$ and $1 \leq i \leq n$).

EXAMPLE 174. Using the notation in Remark 34, applications of the standard Gentzen-type introduction rules for $\forall$ have the following forms:

$$\frac{\Omega, t : \psi\{z/w\}}{\Omega, t : \forall w \psi} \qquad \frac{\Omega, f : \psi\{\mathbf{t}/w\}}{\Omega, f : \forall w \psi}$$

[28] For simplicity of presentation, we assume that the language L does not include any propositional connectives. The latter can anyway be thought of as multi-ary quantifiers which bind no variables.

where z and $\mathbf{t}$ are free for w in ψ and z does not occur free in $\Omega \cup \{\forall w \psi\}$. The canonical representation of these rules will be:

$$[\{\{t : p_1(x)\}\}/\ \{t\} : \forall] \qquad [\{\{f : p_1(c_1)\}\}/\{f\} : \forall]$$

This shows that for instantiating a canonical rule we need a context and some notion of a mapping from the terms and formulas of QL_n to the terms and formulas of L, which handles with care the choice of terms and variables of L, so that they satisfy the appropriate conditions.

DEFINITION 175. For a canonical rule $R = [\Theta/\ S : \mathcal{Q}]$ and a sequent Ω over L, an $\langle R, \Omega, z\rangle$-*mapping* is any function χ from the predicate symbols, terms and formulas of QL_n to formulas and terms of L, satisfying the following conditions:

- For every $1 \leq i \leq n$, $\chi(p_i)$ is an L-formula.
- $\chi(y)$ is a variable of L.
- $\chi(x) \neq \chi(y)$ for every two variables $x \neq y$ of QL_n.
- $\chi(c)$ is an L-term, such that $\chi(x)$ does not occur in $\chi(c)$ for any variable x occurring in Θ.
- For every $1 \leq i \leq n$, if $p_i(\mathbf{t})$ occurs in Θ, $\chi(\mathbf{t})$ is a term free for z in $\chi(p_i)$, and if $\mathbf{t}$ is a variable, then $\chi(\mathbf{t})$ does not occur free in $\Omega \cup \{\mathcal{Q}z(\chi(p_1), ..., \chi(p_n))\}$.
- $\chi(p_i(\mathbf{t})) = \chi(p_i)\{\chi(\mathbf{t})/z\}$.

χ is extended to sequents as follows: $\chi(\Sigma) = \{a : \chi(\psi) \mid a : \psi \in \Sigma\}$.

DEFINITION 176. Let $\mathcal{Q}$ be an n-ary quantifier. An *application of a canonical quantifier rule* $R = [\{\Sigma_1, ..., \Sigma_m\}/\ S : \mathcal{Q}]$ is any inference step of the form:

$$\frac{\Omega \cup \chi(\Sigma_1) \quad \dots \quad \Omega \cup \chi(\Sigma_m)}{\Omega \cup S : \mathcal{Q}x(\chi(p_1), \dots, \chi(p_n))}$$

where Ω is a sequent and χ is some $\langle R, \Omega, x\rangle$-mapping.

EXAMPLE 177. The introduction rules for the bounded universal binary quantifier $\overline{\forall}$ over $\mathcal{V} = \langle t, \top, f, \bot\rangle$ can be formulated as follows (taking t and $\top$ as the designated truth-values, this is a natural generalization of its classical interpretation):

$$[\ \{\ \{f : p_1(x), \bot : p_1(x), t : p_2(x), \top : p_2(x)\}\ \}\ /\ \{t, \top\} : \overline{\forall}\ ]$$

$$[\ \{\ \{t : p_1(c_1), \top : p_1(c_1)\},\ \{f : p_2(c_1), \bot : p_2(c_1)\}\ \}\ /\ \{f, \bot\} : \overline{\forall}\ ]$$

Their applications have the forms:

$$\frac{\Omega \cup \{f : \psi_1\{y/z\}, \bot : \psi_1\{y/z\}, t : \psi_2\{y/z\}, \top : \psi_2\{y/z\}\}}{\Omega \cup \{t : \overline{\forall} z(\psi_1, \psi_2), \top : \overline{\forall} z(\psi_1, \psi_2)\}}$$

$$\frac{\Omega \cup \{t : \psi_1\{\mathbf{t}/z\}, \top : \psi_1\{\mathbf{t}/z\}\} \quad \Omega \cup \{f : \psi_2\{\mathbf{t}/z\}, \bot : \psi_2\{\mathbf{t}/z\}\}}{\Omega \cup \{f : \overline{\forall} z(\psi_1, \psi_2), \bot : \overline{\forall} z(\psi_1, \psi_2)\}}$$

On the level of quantifiers two new elements are added to canonical calculi: the axiom of α-equivalence and the rule of substitution.

DEFINITION 178. Let $\mathcal{V} = \{l_1, ..., l_n\}$ be a finite set of signs.

1. A *logical axiom*[29] for $\mathcal{V}$ is any sequent $\{l_1 : \psi_1, l_2 : \psi_2 ..., l_n : \psi_n\}$, where $\psi_1 \equiv_\alpha \psi_2 ... \equiv_\alpha \psi_n$.

2. The substitution rule for $\mathcal{V}$ is defined as follows:

$$\frac{\Omega}{\Omega'}\ \mathbf{Sub}$$

where Ω' is obtained from Ω by legal substitutions of terms for free variables.

The following proposition follows from the completeness of many-valued resolution ([Baaz et. al., 1995]):

PROPOSITION 179. *Let Θ be a set of clauses. The empty sequent can be derived from Θ using cuts and substitutions iff Θ is not satisfiable.*

DEFINITION 180. We say that a signed calculus over $\mathcal{V}$ is *canonical* if it consists of: (i) All logical axioms for $\mathcal{V}$, (ii) The rules of cut, weakening and substitution, and (iii) A finite number of signed canonical quantifier rules.

Next we extend the propositional criterion of coherence to canonical calculi with multi-ary quantifiers.

DEFINITION 181. For sets of clauses $\Theta_1, ..., \Theta_m$, $\mathsf{Rnm}(\Theta_1 \cup ... \cup \Theta_m)$ is a set $\Theta'_1 \cup ... \cup \Theta'_m$, such that for all $1 \leq i \leq n$, Θ'_i is obtained from Θ_i by

[29]This is an extension of the α-axiom from [Zamansky and Avron, 2006c]

renaming the constants and variables which occur in Θ_i, and no constant or variable occur in both Θ'_i and Θ'_j in case $i \neq j$.

DEFINITION 182. A canonical calculus G is *coherent* if $\mathsf{Rnm}(\Theta_1 \cup ... \cup \Theta_m)$ is unsatisfiable whenever $[\Theta_1/\ S_1 : \mathcal{Q}], ..., [\Theta_m/\ S_m : \mathcal{Q}]$ is a set of rules of G, such that $S_1 \cap ... \cap S_m = \emptyset$.

Note that by Proposition 179, the above definition of coherence can be translated into a purely syntactic one.

PROPOSITION 183. *The coherence of a canonical calculus is decidable.*

EXAMPLE 184. Consider a canonical calculus over $\mathcal{V} = \{t, \top, f\}$ with the following rules for a unary quantifier $\mathcal{Q}$:

$$\begin{aligned} R_1 &= [\ \{\ \{t,\top\} : p_1(x)\ \}\ /\ \{t,\top\} : \mathcal{Q}\ ] \\ R_2 &= [\ \{\ \{\top, f\} : p_1(y)\ \}\ /\ \{\top, f\} : \mathcal{Q}\ ] \\ R_3 &= [\ \{\ \{t, f\} : p_1(c_1)\ \}\ /\ \{t, f\} : \mathcal{Q}\ ] \end{aligned}$$

Since $\{t,\top\} \cap \{\top, f\} \cap \{t, f\} = \emptyset$, we need to check whether the empty sequent is derivable (using cuts and substitutions) from the set of premises of these rules:

$$\dfrac{\dfrac{\{t,f\} : p_1(c_1) \quad \dfrac{\{t,\top\} : p_1(x)}{\{t,\top\} : p_1(c_1)}\,\textbf{Sub}}{\{t\} : p_1(c_1)}\,\textbf{Cut} \qquad \dfrac{\{\top,f\} : p_1(y)}{\{\top,f\} : p_1(c_1)}\,\textbf{Sub}}{\emptyset}\,\textbf{Cut}$$

Thus this calculus is coherent (note that each pair of premises is consistent, but the three of them together are not).

Below we briefly review the main results related to the connection between canonical calculi and finite Nmatrices. What follows is an extension of the results for the propositional case from Sections 4.1 and 4.2.

The notions of standard, analytic and strong cut-elimination from Definition 46 can be naturally extended to calculi with multi-ary quantifiers.

THEOREM 185. *A coherent canonical calculus which admits strong cut-elimination can be constructed for every finite Nmatrix.*

As a corollary, we have the following extension of Theorem 53:

COROLLARY 186. **(Compactness)** *Let Θ be a set of sequents and Ω a sequent.*

1. *If $\Theta \vdash^t_{\mathcal{M}} \Omega$, then there is some finite $\Theta' \subseteq \Theta$, such that $\Theta' \vdash^t_{\mathcal{M}} \Omega$.*

2. *Let Γ, Δ be two sets of L-formulas. If $\Gamma \vdash^t_{\mathcal{M}} \Delta$, then there are some finite $\Gamma' \subseteq \Gamma$ and $\Delta' \subseteq \Delta$, such that $\Gamma' \vdash^t_{\mathcal{M}} \Delta'$.*

In the converse direction, every coherent calculus has a corresponding finite Nmatrix. Moreover, there is a direct correspondence[30] between analytic cut-elimination, coherence and finite Nmatrices:

THEOREM 187. *Let G be a canonical calculus for a language L and a finite set of signs $\mathcal{V}$. The following statements concerning G are equivalent:*

1. *G is coherent.*
2. *G has a strongly characteristic finite Nmatrix.*
3. *G admits strong analytic cut-elimination.*

BIBLIOGRAPHY

[Almukdad and Nelson, 1984] Almukdad A. and D. Nelson, Constructible falsity and inexact predicates, *Journal of Symbolic Logic*, 49:231–333, 1984.

[Arieli and Avron, 1996] Arieli O. and A. Avron, Reasoning with Logical Bilattices, *Journal of Logic, Language and Information*, 5:25–63, 1996.

[Artemov et.al. (eds.), 2005] S. Artemov, H. Barringer, A.S. d'Avila Garcez, L.C. Lamb and J. Woods, *We Will Show Them! Essays in Honour of Dov Gabbay*, vol. 1, College Publications, 2005.

[Avron, 1991] Avron, A., Simple consequence relations, *Information and Computation*, 92:105–139, 1991.

[Avron, 1993] Avron, A., Gentzen-Type Systems, Resolution and Tableaux, *Journal of Automated Reasoning*, 10:265–281, 1993.

[Avron and Lev, 2001] Avron, A. and I. Lev, Canonical Propositional Gentzen-type Systems, *Proceedings of the 1st International Joint Conference on Automated Reasoning (IJCAR 2001)*, R. Gore, A. Leitsch, T. Nipkow (eds.), Springer Verlag, LNAI 2083, 529-544, Springer Verlag, 2001.

[Avron and Lev, 2005] Avron, A. and I. Lev, Non-deterministic Multi-valued Structures, *Journal of Logic and Computation*, 15:241–261, 2005.

[Avron, 2005a] Avron, A., A Nondeterministic View on Nonclassical Negations, *Studia Logica*, 80:159–194, 2005.

[Avron, 2005b] Avron, A., Non-deterministic Matrices and Modular Semantics of Rules, *Logica Universalis*, J.-Y. Beziau ed., Birkhűser Verlag, 149–167, 2005.

[Avron, 2005c] Avron, A. Logical Non-determinism as a Tool for Logical Modularity: An Introduction, [Artemov et.al. (eds.), 2005], 105–124, 2005.

[Avron and Konikowska, 2005] Avron, A. and B. Konikowska, Proof Systems for Logics Based on Non-deterministic Multiple-valued Structures, *Logic Journal of the IGPL*, 13:365-387, 2005.

[30] Cut-elimination for a general family of sequent calculi with generalized quantifiers (of which the canonical calculi are specific instances) is investigated in [Ciabattoni and Terui, 2006b] (extending [Ciabattoni and Terui, 2006a], see Remark 45). Their reductivity condition can again be shown to be equivalent to coherence.

[Avron and Zamansky, 2005] Avron, A. and A. Zamansky, Quantification in Non-deterministic Multi-valued Structures, *Proceedings of the 35th IEEE International Symposium on Multiple-Valued Logics*, 296–301, 2005.

[Avron et. al., 2006] Avron, A., J. Ben-Naim and B. Konikowska, Cut-free Ordinary Sequent Calculi for Logics Having Generalized Finite-Valued Semantics, *Logica Universalis* 1:41–69, 2006.

[Avron, 2007a] Avron, A., Non-deterministic semantics for logics with a consistency operator, *Journal of Approximate Reasoning*, 45:271–287, 2007.

[Avron, 2007b] Avron, A., Non-deterministic Semantics for Families of Paraconsistent Logics, *Handbook of Paraconsistency*, J.-Y. Beziau, W. Carnielli and D. Gabbay (eds.), 285-320, Studies in Logic 9, College Publications, 2007.

[Avron and Zamansky, 2007a] Avron, A. and A. Zamansky, Non-deterministic Multi-valued Matrices for First-order Logics of Formal Inconsistency, *Proceedings of the 37th IEEE International Symposium on Multiple-Valued Logics*, 2007.

[Avron and Zamansky, 2007b] Avron, A. and A. Zamansky, Generalized Non-deterministic matrices and (n,k)-ary quantifiers, *Proceedings of the Symposium on Logical Foundations of Computer Science*, S. Artemov and A. Nerode (eds.), LNCS 4514, 26–40, Springer, 2007.

[Avron and Zamansky, 2007c] Avron, A. and A. Zamansky, Many-valued non-deterministic semantics for first-order Logics of Formal (In)consistency, *Algebraic and Proof-theoretic Aspects of Non-classical Logics*, S. Aguzzoli, A. Ciabattoni, B. Gerla, C. Manara, and V. Marra, (eds.), 1–24, LNAI 4460, Springer, 2007.

[Avron et. al., 2008] Avron, A., J. Ben-Naim, and B. Konikowska, Processing Information from a Set of Sources, Forthcoming in *Towards Mathematical Philosophy*, Trends in Logic series, 2008.

[Avron and Zamansky, 2008] Avron, A. and A. Zamansky, Canonical calculi with (n,k)-ary quantifiers, *Journal of Logical Methods in Computer Science*, 10.2168/LMCS-4(3:2), 2008 (an extended version of [Zamansky and Avron, 2006c]).

[Avron and Zamansky, 2009] Zamansky, A. and A. Avron, Signed Calculi, Non-deterministic Matrices and Cut-elimination, *Proceedings of the Symposium on Logical Foundations of Computer Science* LNCS 5407, 31 – 46, Springer, 2009.

[Avron and Zamansky, 2010] Zamansky, A. and A. Avron, Signed Canonical Calculi with Quantifiers (an extended version of [Avron and Zamansky, 2009]), to appear.

[Avron, 2010] Avron, A., Tonk - a full mathematical solution, to appear in Ruth Manor's Festschrift.

[Baaz, 1986] Baaz M.,Kripke-Type Semantics for da Costa's paraconsistent Logic C_ω,*Notre Dame Journal of Formal Logic*, 27:523–527, 1986.

[Loparić, 1986] Loparić A.,A Semantical Study of some Propositional Calculi, *The Journal of Non-Classical Logics*, 3:73–95, 1986.

[Baaz et. al., 1993] Baaz M., C.G. Fermüller, G. Salzer and R.Zach, Dual systems of sequents and tableaux for many-valued logics, *Bull. EATCS*, 51:192–197, 1993.

[Baaz et. al., 1995] Baaz M. and C.G. Fermüller, Resolution-based theorem-proving for many-valued logics, *Journal of Symbolic Computation*, 19(4): 353–391, 1995.

[Baaz et. al., 1998] Baaz M., C.G. Fermüller, G. Salzer and R .Zach, Labeled Calculi and Finite-valued Logics, *Studia Logica*, 61:7–33, 1998.

[Baaz et. al., 2000] Baaz M., C. G. Fermüller, and G. Salzer, Automated Deduction for Many-valued Logics, *Handbook of Automated Reasoning*, A. Robinson and A. Voronkov (eds.), 1355–1400, Elsevier Science Publishers, 2000.

[Baaz et. al., 2001] Baaz M., U. Egly and A. Leitsch, Normal Form Transformations, *Handbook of Automated Reasoning*, 273–333, 2001.

[Batens et. al., 1999] Batens D., K. de Clerq and N. Kurtonina, Embedding and interpolation for some paralogics. The propositional case, *Reports on Mathematical Logic* 33:29–44, 1999.

[Belnap, 1962] Belnap N.D, Tonk, plonk and plink, *Analysis*, 22:130–134, 1962.

[Belnap, 1977] Belnap N.D, A useful four-valued logic, *Modern uses of multiple-valued logic*, G. Epstein and J. M. Dunn, (eds.) Reidel, Dordrecht, NL, 5–37, 1977.

[Béziau, 1999] Béziau J.-Y., Classical negation can be expressed by one of its halves, *Logic Journal of the Interest Group in Pure and Applied Logics*, 7:145–151, 1999.

[Blamey, 1986] Blamey S., Partial Logic, *Handbook of Philosophical Logic*, D. Gabbay, F. Guenthner (eds.), 3:1–70, Reidel , Dordrecht, 1986.

[Bolc and Borowik, 1992] Bolc L. and P.Borowik, *Many-Valued Logics 1. Theoretical Foundations*, Springer, Berlin, 1992.

[Borowik, 1986] Borowik P., Multi-valued n-sequential propositional logic, *Journal of Symbolic Logic*, 52:309–310, 1985.

[Carnielli, 1987] Carnielli W.A., Systematization of Finite Many-valued Logics through the method of Tableaux, *Journal of Symbolic Logic*, 52 (2):473–493, 1987.

[Carnielli, 1991] Carnielli W.A., On sequents and tableaux for many-valued logics, *Journal of non-classical logics*, 8(1):59–76, 1991.

[Carnielli, 1998] Carnielli W.A., Possible-translations semantics for paraconsistent logics, *Frontiers in paraconsistent logic: Proceedings of the I World Congress on Paraconsistency*, 159–72, Ghent, 1998.

[Carnielli and Marcos, 1999] Carnielli W.A. and J. Marcos, Limits of Paraconsistent Calculi, Notre Dame Journal of Formal Logic, 40:375–390, 1999.

[Carnielli et. al., 2000] Carnielli W.A, J. Marcos and S. de Amo, Formal inconsistency and evolutionary databases, *Logic and logical philosophy*, 8:115–152, Ghent, 2000.

[Carnielli et. al., 2002] Carnielli W.A., M.E. Coniglio and I.M.L. D'Ottaviano, *Paraconsistency - the Logical Way to the Inconsistent, Lecture notes in pure and appliced mathematics*, vol. 228, Marcel Dekker, New York, 2002.

[Carnielli and Marcos, 2002] Carnielli W.A. and J. Marcos, A Taxonomy of C-systems, [Carnielli et. al., 2002], 1–94, 2002.

[Carnielli and Coniglio, 2005] Carnielli W.A. and M.E. Coniglio,Splitting Logics, [Artemov et.al. (eds.), 2005], 389–414, 2005.

[Carnielli et. al., 2007] Carnielli W.A., M.E. Coniglio and J. Marcos, Logics of Formal Inconsistency, *Handbook of Philosophical Logic*, 2nd Edition, D. Gabbay and F. Guenther (eds.), 14:15–107, Springer, 2007.

[Church, 1940] Church A., A formulation of the simple theory of types, *Journal of Symbolic Logic*, 5:56–68, 1940.

[Ciabattoni and Terui, 2006a] Ciabattoni A. and Terui K., Towards a semantic characterization of cut elimination, *Studia Logica*, 82(1):95–119, 2006.

[Ciabattoni and Terui, 2006b] Ciabattoni A. and Terui K., Modular cut-elimination: finding proofs or counter-examples, *Proceedings of the 13-th International Conference of Logic for Programming AI and Reasoning (LPAR06)*, LNAI 4246, 135–149, 2006.

[da Costa, 1974] da Costa N.C.A, On the theory of inconsistent formal systems, *Notre Dame Journal of Formal Logic*, 15:497–510, 1974.

[Dunn, 1976] Dunn J.M., Intuitive semantics for first-degree entailments and coupled trees, *Philosophical Studies*, 29:149–168, 1976.

[Enderton, 1972] Enderton H., *A Mathematical Introduction to Logic*, Academic Press, 1972.

[Fitting, 1994] Fitting M., Kleene's Three-valued Logics and Their Children, *Fundamenta Informaticae* 20:113-131, 1994.

[Gentzen, 1969] Gentzen, G., Investigations into Logical Deduction, *The collected works of Gerhard Gentzen*, M.E. Szabo, ed., 68–131, North Holland, Amsterdam , 1969.

[Ginsberg, 1988] Ginsberg M.L, Multivalued logics: A Uniform Approach to Reasoning in AI, *Computer Intelligence*, 4:256-316, 1988.

[Girard, 1987] J. Y. Girard, Linear logic, *Theoretical Computer Science*, 50:1–101, 1987.

[Gottwald, 2001] Gottwald S., A Treatise on Many-Valued Logics, *Studies in Logic and Computation*, vol. 9, Research Studies Press, Baldock, 2001.

[Hähnle, 1994] Hähnle, R., *Automated Deduction in Multiple-valued Logics*, Oxford University Press, 1994.

[Hähnle, 1998] Hähnle, R., Commodious Axiomatization of Quantifiers in Many-valued Logic, *Studia Logica*, 61:101–121, 1998.

[Hähnle, 1999] Hähnle, R. Tableaux for Multiple-valued Logics, *Handbook of Tableau Methods*, M. D'Agostino, D. Gabbay, R. Hähnle and J. Posegga, (eds.), 529–580, Kluwer Academic Publishers, Dordrecht, 1999.

[Hähnle, 2001] Hähnle, R., Advanced many-valued logics, *Handbook of Philosophical Logic*, D. Gabbay, F. Guenthner (eds.), 2nd Edition, Kluwer Academic Publishers, Dordrecht, 2:297–395, 2001.

[Henkin, 1961] Henkin, L., Some remarks on infinitely long formulas, *Infinistic Methods*, 167–183. Pergamon Press, Oxford, 1961.

[Hodges, 2001] Hodges, W., Elementary Predicate Logic, *Handbook of Philosophical Logic*, D. Gabbay, F. Guenthner (eds.), 2nd Edition, Kluwer Academic Publishers, Dordrecht, 1:1–129, 2001.

[Kalish and Montague, 1964] Kalish, D. and R. Montague, *Logic: Techniques of Formal Reasoning*, NY: Harcourt, Brace&World, 1964.

[Kleene, 1938] Kleene, S.C, On a notation for ordinal numbers, *Journal of Symbolic Logic*, 3:150–155, 1938.

[Konikowska, 2002] Konikowska B., Rasiowa-Sikorski Deduction Systems in Computer Science Applications, *Theoretical Computer Science* 286:323-266, 2002.

[Leblanc, 2001] Leblanc H., Alternatives to Standard First-order Semantics, *Handbook of Philosophical Logic*, 2:53–133, 2001.

[Malinowski, 1993] Malinowski G., *Many-Valued Logics*, Clarendon Press, Oxford, 1993.

[McCarthy, 1963] McCarthy, J., A basis for a mathematical theory of computation, *Computer Programming and Formal Systems*, P.Braffort and D.Hirshberg (eds.), 33–70, North Holland, Amsterdam , 1963.

[Mendelson, 1964] Mendelson E., Introduction to Mathematical Logic, Chapman and Hall, fourth edition, 1964.

[Mostowski, 1961] Mostowski, A., Axiomatizability of some Many-valued Predicate Calculi, *Fundamenta Mathematicae,* , 15:165–190, North Holland, Amsterdam , 1961.

[Prior, 1960] Prior, A.N, The runabout inference ticket, *Analysis*, 21:38–9, 1960.

[Rabaey et. al, 2003] Rabaey J.M., A. Chandrakasan and B. Nikolic , *Digital Integrated Cicruits: A Design Perspective*, Prentice-Hall, 2003.

[Rasiowa and Sikorski, 1963] Rasiowa H. and R. Sikorski, *The Mathematics of Metamathematics*, PWN (Polish Scientific Publishers), Warsaw 1963.

[Rosser and Turquette, 1952] Rosser J.B. and A.R. Turquette, *Many-Valued Logics*, North-Holland Publ. Comp., Amsterdam, 1952.

[Rousseau, 1967] Rousseau G., Sequents in many-valued logic 1, *Fundamenta Mathematicae*, LX:23–33, 1967.

[Salzer, 1996] Salzer, G., Optimal axiomatizations for multiple-valued operators and quantifiers based on semilattices, *Proceedings of 13-th CADE, Springer, M. McRobbie and J. Slaney (eds.)*, 1104:688–702, 1996.

[Schroeder-Heister, 1985] Schroeder-Heister, P. Natural deduction calculi with rules of higher levels, *Journal of Symbolic Logic*, 50:275–276, 1985.

[Shoenfield, 1967] Shoenfield, J.R., *Mathematical Logic*, *Association for Symbolic Logic*, 1967.

[Shoesmith, 1971] Shoesmith D.J. and Smiley T.J, Deducibility and many-valuedness, *Journal of Symbolic Logic*, 36:610–622, 1971.

[Sundholm, 2002] Sundholm, G., Proof Theory and Meaning, *Handbook of Philosophical Logic*, D. Gabbay, F. Guenthner (eds.), 2nd Edition, Kluwer Academic Publishers, Dordrecht, 9:165–198, 2002.

[Takahashi, 1967] Takahashi M., Many-valued logics of extended Gentzen style I, *Sci. Rep. Tokio Kyoiku Daigaku Sec. A*, 9(231):95–110, 1967.

[Troelstra and Schwichtenberg, 2000] Troelstra, A.S. and H. Schwichtenberg, Basic Proof Theory, *Cambridge University Press*, 126–130, 2000.

[Urquhart, 2001] Urquhart A., Many-valued Logic, *Handbook of Philosophical Logic*, D. Gabbay and F. Guenthner (eds.), 2nd Edition, Kluwer Academic Publishers, Dordrecht, 2:249–295, 2001.

[van Dalen, 1980] van Dalen D., *Logic and Structure*, Springer, 1980.

[von Kutschera, 1969] von Kutschera F., Ein verallgemeinerter widerlegungsbegriff fur Gentzenkalkule, *Archiv fur Mathematische Logik und Gurndlagenforschung*, 12:104–118, 1969.

[Wójcicki, 1988] Wójcicki R.,*Theory of Logical Calculi: basic theory of consequence operations*, Kluwer Academic Publishers, Dordrecht, 1988.

[Zamansky and Avron, 2006a] Zamansky, A. and A. Avron, Cut Elimination and Quantification in Canonical Systems, *Studia Logica*, 82(1):157–176, 2006.

[Zamansky and Avron, 2006b] Zamansky A. and A. Avron, Non-deterministic semantics for first-order paraconsistent logics, *Proceedings of the 10th International Conference on Principles of Knowledge Representation and Reasoning*, 431–440, 2006.

[Zamansky and Avron, 2006c] Zamansky, A. and A. Avron, Canonical Gentzen-type calculi with (n,k)-ary quantifiers , *Proceedings of the Third International Joint Conference in Automated Reasoning (IJCAR06)*, U. Furbach and N. Shankar (eds.), LNAI 4130, 251 – 265, Springer, 2006.

[Zamansky and Avron, 2007] Zamansky, A. and A. Avron, Effective non-deterministic semantics for first-order LFIs, *Journal of Multiple-valued Logic and Soft Computing* (an extended version of [Avron and Zamansky, 2007a]), 13:4-6, 335 - 352, 2007.

[Zucker, 1978a] Zucker, J.I., The adequacy problem for classical logic, *Journal of Philosophical Logic*, 7:517–535, 1978.

[Zucker, 1978b] Zucker, J.I. and R.S. Tragesser, The adequacy problem for inferential logic, *Journal of Philosophical Logic*, 7:501–516, 1978.

INDEX

D. Gabbay and F. Guenthner (eds.),
Handbook of Philosophical Logic, Volume 16, 305–310.
DOI 10.1007/978-94-007-0479-4,