Institute for Language, Logic and Information

SOME CHAPTERS ON INTERPRETABILITY LOGIC

Maarten de Rijke

ITLI Prepublication Series
X-90-02

University of Amsterdam

The ITLI Prepublication Series

1986
86-01 The Institute of Language, Logic and Information
86-02 Peter van Emde Boas — A Semantical Model for Integration and Modularization of Rules
86-03 Johan van Benthem — Categorial Grammar and Lambda Calculus
86-04 Reinhard Muskens — A Relational Formulation of the Theory of Types
86-05 Kenneth A. Bowen, Dick de Jongh — Some Complete Logics for Branched Time, Part I Well-founded Time, Forward looking Operators
86-06 Johan van Benthem — Logical Syntax

1987
87-01 Jeroen Groenendijk, Martin Stokhof — Type shifting Rules and the Semantics of Interrogatives
87-02 Renate Bartsch — Frame Representations and Discourse Representations
87-03 Jan Willem Klop, Roel de Vrijer — Unique Normal Forms for Lambda Calculus with Surjective Pairing
87-04 Johan van Benthem — Polyadic quantifiers
87-05 Víctor Sánchez Valencia — Traditional Logicians and de Morgan's Example
87-06 Eleonore Oversteegen — Temporal Adverbials in the Two Track Theory of Time
87-07 Johan van Benthem — Categorial Grammar and Type Theory
87-08 Renate Bartsch — The Construction of Properties under Perspectives
87-09 Herman Hendriks — Type Change in Semantics: The Scope of Quantification and Coordination

1988
Logic, Semantics and Philosophy of Language:
LP-88-01 Michiel van Lambalgen — Algorithmic Information Theory
LP-88-02 Yde Venema — Expressiveness and Completeness of an Interval Tense Logic
LP-88-03 — Year Report 1987
LP-88-04 Reinhard Muskens — Going partial in Montague Grammar
LP-88-05 Johan van Benthem — Logical Constants across Varying Types
LP-88-06 Johan van Benthem — Semantic Parallels in Natural Language and Computation
LP-88-07 Renate Bartsch — Tenses, Aspects, and their Scopes in Discourse
LP-88-08 Jeroen Groenendijk, Martin Stokhof — Context and Information in Dynamic Semantics
LP-88-09 Theo M.V. Janssen — A mathematical model for the CAT framework of Eurotra
LP-88-10 Anneke Kleppe — A Blissymbolics Translation Program

Mathematical Logic and Foundations:
ML-88-01 Jaap van Oosten — Lifschitz' Realizabiility
ML-88-02 M.D.G. Swaen — The Arithmetical Fragment of Martin Löf's Type Theories with weak Σ-elimination
ML-88-03 Dick de Jongh, Frank Veltman — Provability Logics for Relative Interpretability
ML-88-04 A.S. Troelstra — On the Early History of Intuitionistic Logic
ML-88-05 A.S. Troelstra — Remarks on Intuitionism and the Philosophy of Mathematics

Computation and Complexity Theory:
CT-88-01 Ming Li, Paul M.B.Vitanyi — Two Decades of Applied Kolmogorov Complexity
CT-88-02 Michiel H.M. Smid — General Lower Bounds for the Partitioning of Range Trees
CT-88-03 Michiel H.M. Smid, Mark H. Overmars, Leen Torenvliet, Peter van Emde Boas — Maintaining Multiple Representations of Dynamic Data Structures
CT-88-04 Dick de Jongh, Lex Hendriks, Gerard R. Renardel de Lavalette — Computations in Fragments of Intuitionistic Propositional Logic
CT-88-05 Peter van Emde Boas — Machine Models and Simulations (revised version)
CT-88-06 Michiel H.M. Smid — A Data Structure for the Union-find Problem having good Single-Operation Complexity
CT-88-07 Johan van Benthem — Time, Logic and Computation
CT-88-08 Michiel H.M. Smid, Mark H. Overmars, Leen Torenvliet, Peter van Emde Boas — Multiple Representations of Dynamic Data Structures
CT-88-09 Theo M.V. Janssen — Towards a Universal Parsing Algorithm for Functional Grammar
CT-88-10 Edith Spaan, Leen Torenvliet, Peter van Emde Boas — Nondeterminism, Fairness and a Fundamental Analogy
CT-88-11 Sieger van Denneheuvel, Peter van Emde Boas — Towards implementing RL

Other prepublications:
X-88-01 Marc Jumelet — On Solovay's Completeness Theorem

1989
Logic, Semantics and Philosophy of Language:
LP-89-01 Johan van Benthem — The Fine-Structure of Categorial Semantics
LP-89-02 Jeroen Groenendijk, Martin Stokhof — Dynamic Predicate Logic, towards a compositional, non-representational semantics of discourse
LP-89-03 Yde Venema — Two-dimensional Modal Logics for Relation Algebras and Temporal Logic of Intervals
LP-89-04 Johan van Benthem — Language in Action
LP-89-05 Johan van Benthem — Modal Logic as a Theory of Information
LP-89-06 Andreja Prijatelj — Intensional Lambek Calculi: Theory and Application
LP-89-07 Heinrich Wansing — The Adequacy Problem for Sequential Propositional Logic
LP-89-08 Víctor Sánchez Valencia — Peirce's Propositional Logic: From Algebra to Graphs
LP-89-09 Zhisheng Huang — Dependency of Belief in Distributed Systems

Mathematical Logic and Foundations:
ML-89-01 Dick de Jongh, Albert Visser — Explicit Fixed Points for Interpretability Logic
ML-89-02 Roel de Vrijer — Extending the Lambda Calculus with Surjective Pairing is conservative
ML-89-03 Dick de Jongh, Franco Montagna — Rosser Orderings and Free Variables
ML-89-04 Dick de Jongh, Marc Jumelet, Franco Montagna — On the Proof of Solovay's Theorem
ML-89-05 Rineke Verbrugge — Σ-completeness and Bounded Arithmetic
ML-89-06 Michiel van Lambalgen — The Axiomatization of Randomness
ML-89-07 Dirk Roorda — Elementary Inductive Definitions in HA: from Strictly Positive towards Monotone
ML-89-08 Dirk Roorda — Investigations into Classical Linear Logic
ML-89-09 Alessandra Carbone — Provable Fixed points in $I\Delta_0+\Omega_1$

Computation and Complexity Theory:
CT-89-01 Michiel H.M. Smid — Dynamic Deferred Data Structures
CT-89-02 Peter van Emde Boas — Machine Models and Simulations
CT-89-03 Ming Li, Herman Neuféglise, Leen Torenvliet, Peter van Emde Boas — On Space Efficient Simulations
CT-89-04 Harry Buhrman, Leen Torenvliet — A Comparison of Reductions on Nondeterministic Space
CT-89-05 Pieter H. Hartel, Michiel H.M. Smid, Leen Torenvliet, Willem G. Vree — A Parallel Functional Implementation of Range Queries
CT-89-06 H.W. Lenstra, Jr. — Finding Isomorphisms between Finite Fields
CT-89-07 Ming Li, Paul M.B. Vitanyi — A Theory of Learning Simple Concepts under Simple Distributions and Average Case Complexity for the Universal Distribution (Prel. Version)
CT-89-08 Harry Buhrman, Steven Homer, Leen Torenvliet — Honest Reductions, Completeness and Nondeterminstic Complexity Classes
CT-89-09 Harry Buhrman, Edith Spaan, Leen Torenvliet — On Adaptive Resource Bounded Computations
CT-89-10 Sieger van Denneheuvel — The Rule Language RL/1
CT-89-11 Zhisheng Huang, Sieger van Denneheuvel, Peter van Emde Boas — Towards Functional Classification of Recursive Query Processing

Other Prepublications:
X-89-01 Marianne Kalsbeek — An Orey Sentence for Predicative Arithmetic
X-89-02 G. Wagemakers — New Foundations: a Survey of Quine's Set Theory
X-89-03 A.S. Troelstra — Index of the Heyting Nachlass
X-89-04 Jeroen Groenendijk, Martin Stokhof — Dynamic Montague Grammar, a first sketch
X-89-05 Maarten de Rijke — The Modal Theory of Inequality
X-89-06 Peter van Emde Boas — Een Relationele Semantiek voor Conceptueel Modelleren: Het RL-project

1990 SEE INSIDE BACK COVER

Instituut voor Taal, Logica en Informatie
Institute for Language, Logic and Information

Faculteit der Wiskunde en Informatica
(Department of Mathematics and Computer Science)
Plantage Muidergracht 24
1018TV Amsterdam

Faculteit der Wijsbegeerte
(Department of Philosophy)
Nieuwe Doelenstraat 15
1012CP Amsterdam

SOME CHAPTERS ON INTERPRETABILITY LOGIC

Maarten de Rijke
Department of Mathematics and Computer Science
University of Amsterdam

Received February 1990

Master's Thesis
Supervisor D.H.J. de Jongh

Preface

Interpretability has been studied extensively; see e.g. Visser [1989] for numerous references. It has been used to establish results on decidability and undecidability of theories and on relative consistency, and to compare the strength of theories.

Given some fixed axiomatic theory T, interpretability over T may be considered in two ways. It may be studied as a unary predicate on axiomatic theories, say $\mathrm{UnInterp}_T(A)$, which stands for 'T interprets $T + A$'; or as a binary relation between axiomatic theories, say $\mathrm{BinInterp}_T(A, B)$, which stands for '$T + A$ interprets $T + B$'. Accordingly, the modal analysis of interpretability in the spirit of Solovay's analysis of provability may be undertaken using either a unary or a binary modal operator. Smoryński was the first one to treat the unary interpretability predicate as a modal operator ('$\mathbf{I}$'), and Švejdar was subsequently the first one to introduce a binary operator ('$\rhd$') to be interpreted as the binary interpretability relation.

Now, interpretability as a binary relation between theories seems to be the basic notion, since unary interpretability is reducible to it; in the above notation: $\mathrm{UnInterp}_T(A) \leftrightarrow \mathrm{BinInterp}_T(\top, A)$, or: $\mathbf{I}A \leftrightarrow \top \rhd A$. This thesis contains five loosely connected chapters on the modal logic of interpretability — all but one treat interpretability as a binary modal operator. These five chapters are preceded by the introductory Chapter 1, in which we briefly survey our notation, and list some axioms and definitions.

In their important 1990 paper Dick de Jongh and Frank Veltman prove the modal completeness of several interpretability logics. In Chapter 2 we present new proofs for the modal completeness of some of the systems de Jongh and Veltman deal with. In stead of using de Jongh and Veltman's method — which is based on the machinery of finite maximal consistent sets contained in some large adequate set — we use infinite maximal consistent sets and small adequate sets.

The main result in Chapter 3 builds on Albert Visser's proof of the arithmetic completeness of *ILP* with respect to finitely axiomatized sequential theories that extend $\mathrm{I}\Delta_0 + \mathrm{SupExp}$. By proving some additional propositions we will be able to turn Visser's proof into a proof of the arithmetic completeness of ILP^{ω}.

As we pointed out before, the main emphasis in this thesis will be on binary interpretability logic. However, for several of the binary systems introduced in Visser [1988] we will precisely determine their unary counterparts, i.e., their subsystems in which the formula A in $A \rhd B$ is in fact $\top$. This will be done in Chapter 4.

In Chapter 5 we introduce the notion of an internal definition; this is a schema of the form $f(A, B) \leftrightarrow A \rhd B$, where $f(A, B)$ is a formula in the modal language that contains '$\Box$' and '$\Diamond$' only. Using such schemas we will study the hierarchy of L-conservative extensions of *IL*.

Compared to most of the other chapters the sixth and final one has a somewhat light-hearted flavor. In it we make some general remarks on the modal theory of the language $\mathcal{L}(\Box, \rhd)$ of interpretability logic. We also characterize the first-order formulas that are equivalent to formulas in a natural extension of $\mathcal{L}(\Box, \rhd)$.

Acknowledgments

Special thanks go to Dick de Jongh. First of all because he taught me much (if not all I know about the subject) about provability logic, the metamathematics of arithmetic and interpretability logic. Secondly, I want to thank him for his supervision of my work on this thesis. Many of the investigations reported upon in the sequel were originally inspired by some questions and suggestions of his. He also gave useful comments on earlier versions of this thesis.

I am grateful to Johan van Benthem for a number of suggestions concerning Section 6.2.

At a very late stage of the writing of this thesis my two-months old hard disk broke down; when, after three weeks, it was finally repaired, it only worked for about three days and broke down again. I am grateful to Dick de Jongh, Connie Rawejai and Paul de Rijke for helping me out.

Amstelveen,
January 8, 1990

Contents

1. Axioms and Models

The first section contains the necessary preliminaries. The systems *IL*, *ILP* and *ILM* are introduced as well as their semantics. In Section 2 we discuss some theorems and derived rules of *IL*.

1 Introduction

The languages we consider all extend the language $\mathcal{L}$ of propositional logic, i.e. they contain an infinite supply of proposition letters p, q, r, ... as well as the usual connectives. We use $\mathcal{L}(\odot_1, \ldots, \odot_n)$ to denote the modal language with operators $\odot_1, \ldots, \odot_n$. Upper case letters will be used to denote modal formulas. Mostly, we will be concerned with logics in the languages $\mathcal{L}(\Box, \rhd)$ and $\mathcal{L}(\Box, \mathbf{I})$. Here '$\Box$' is the usual provability operator, while the intended interpretation of a formula $A \rhd B$ (in an arithmetical theory T) is '$T+B$ is relatively interpretable in $T+A$'. The unary operators '$\Box^+$' and '$\mathbf{I}$' are defined by '$\Box^+ A := A \wedge \Box A$' and '$\mathbf{I}A := \top \rhd A$', respectively.

We now introduce the basic interpretability logics; they are all devised by Albert Visser, and discussed extensively in Visser [1988] and [1990].

1.1 Definition. (i) The provability logic L denotes Propositional Logic plus

$(R1)$ $\vdash A, \vdash A \to B \Longrightarrow \vdash B$,
$(R2)$ $\vdash A \Longrightarrow \vdash \Box A$,
$(L1)$ $\Box(A \to B) \to (\Box A \to \Box B)$,
$(L2)$ $\Box A \to \Box\Box A$,
$(L3)$ $\Box(\Box A \to A) \to \Box A$.

(ii) The interpretability logic *IL* is obtained from L by adding the axioms

$(J1)$ $\Box(A \to B) \to (A \rhd B)$,
$(J2)$ $(A \rhd B) \wedge (B \rhd C) \to A \rhd C$,
$(J3)$ $(A \rhd C) \wedge (B \rhd C) \to A \vee B \rhd C$,
$(J4)$ $A \rhd B \to (\Diamond A \to \Diamond B)$,
$(J5)$ $\Diamond A \rhd A$.

(iii) Several axioms have special names:

(F) $(A \rhd \Diamond A) \to \Box\neg A$,
(W) $A \rhd B \to A \rhd (B \wedge \Box\neg A)$,
(P) $A \rhd B \to \Box(A \rhd B)$,
(M) $A \rhd B \to (A \wedge \Box C) \rhd (B \wedge \Box C)$,
(M_0) $A \rhd B \to (\Diamond A \wedge \Box C) \rhd (B \wedge \Box C)$.

If X is the name of some axiom, then *ILX* denotes the system $IL + X$.

Next we define the so-called Veltman models for interpretability logic:

1.2 Definition. (i) An *L-frame* is a pair $\langle W, R\rangle$, where W is a non-empty set and R is a transitive, conversely well-founded relation on W.

(ii) If R is a binary relation on some set W, and $w \in W$, then $wR = \{w' \in W \mid wRw'\}$. Moreover, $\breve{R} := \{\langle y, x\rangle \mid xRy\}$ and $\underline{R} := \{\langle x, y\rangle \mid xRy \text{ or } x = y\}$.

(iii) An *IL-frame* is an L-frame $\langle W, R\rangle$ with an additional relation S_w, for each $w \in W$, which satisfies:

(1) S_w is a relation on wR,
(2) S_w is reflexive and transitive,
(3) if w', $w'' \in wR$ and $w'Rw''$, then $w'S_w w''$.

We will often write S for $\{S_w \mid w \in W\}$. To save words we assume from now on that $\mathcal{F}$ denotes the frame $\langle W, R, S\rangle$. Moreover, primes, sub- and superscripts of $\mathcal{F}$ are supposed to distribute over W, R, S, e.g., $\mathcal{F}^1 = \langle W^1, R^1, S^1\rangle$.

1.3 Definition. An *IL-model* is given by an *IL*-frame $\mathcal{F}$ together with a forcing relation $\Vdash$ satisfying:

$$u \Vdash \Box A \iff \forall v\,(uRv \Longrightarrow v \Vdash A),$$
$$u \Vdash A \rhd B \iff \forall v\,(uRv \text{ and } v \Vdash A \Longrightarrow \exists w\,(vS_u w \text{ and } w \Vdash B)).$$

We write $\mathcal{F} \models A$ if $w \Vdash A$ for every $\Vdash$ on $\mathcal{F}$ and $w \in W$. To save words we assume from now on that $\mathcal{M}$ denotes the model $\langle W, R, S, \Vdash\rangle$. Moreover, primes, sub- and superscripts of $\mathcal{M}$ are supposed to distribute over W, R, S and $\Vdash$.

1.4 Definition. (i) An *ILP-model* is an *IL*-model that satisfies the extra condition: if $wRw'RuS_w v$ then $uS_{w'}v$.

(ii) An *ILM-model* is an *IL*-model satisfying the extra condition: if $uS_w vRz$ then uRz.

In their 1990 paper de Jongh and Veltman prove the following (modal) completeness results for *IL*, *ILP* and *ILM*:

1.5 Theorem. (i) *IL is complete with respect to finite IL-models.*
(ii) *ILP is complete with respect to finite ILP-models.*
(iii) *ILM is complete with respect to finite ILM-models.*

2 Some Theorems and Derived Rules of *IL*

We prove some theorems in *IL* and use these to give an alternative, $\Box$-free axiomatization of *IL*. Then we use the modal completeness results mentioned in Section 1 to show the *IL*-validity of some rules.

2.1 Proposition. (i) $IL \vdash \Box A \leftrightarrow \neg A \rhd \bot$;
(ii) $IL \vdash \Diamond A \leftrightarrow \neg(A \rhd \bot)$;
(iii) *Let* $J5' \equiv A \rhd A \wedge \Box\neg A$. *Then* $J5$ *and* $J5'$ *are equivalent over* $IL \setminus J5$.

Proof. Clearly, (ii) is immediate from (i). To prove (i), note

$$\begin{aligned} IL \vdash \Box A &\rightarrow \Box(\top \rightarrow A) \\ &\rightarrow \Box(\neg A \rightarrow \bot) \\ &\rightarrow \neg A \rhd \bot, \text{ by } J1. \end{aligned}$$

Conversely,

$$\begin{aligned} IL \vdash \neg A \rhd \bot &\to (\Diamond\neg A \to \Diamond\bot), \text{ by } J4 \\ &\to (\Box\top \to \Box A) \\ &\to \Box A. \end{aligned}$$

(iii) Observe

$$L \vdash A \to (A \wedge \Box\neg A) \vee \Diamond(A \wedge \Box\neg A)$$
$$L \vdash \Box(A \to (A \wedge \Box\neg A) \vee \Diamond(A \wedge \Box\neg A))$$
$$IL \setminus J5 \vdash A \rhd (A \wedge \Box\neg A) \vee \Diamond(A \wedge \Box\neg A), \text{ by } J1$$

Now

$$IL \setminus J5 \vdash A \wedge \Box\neg A \rhd A \wedge \Box\neg A, \text{ and}$$
$$IL \setminus J5 + J5 \vdash \Diamond(A \wedge \Box\neg A) \rhd A \wedge \Box\neg A$$

so $J2$ yields

$$IL \setminus J5 + J5 \vdash A \rhd A \wedge \Box\neg A.$$

Conversely, assuming $J5'$, we have

$$IL \setminus J5 \vdash A \vee \Diamond A \rhd (A \vee \Diamond A) \wedge \Box\neg(A \vee \Diamond A). \qquad (\star)$$

Furthermore,

$$L \vdash \Box\neg(A \vee \Diamond A) \to \Box\neg A.$$

So

$$\begin{aligned} L \vdash (A \vee \Diamond A) \wedge \Box\neg(A \vee \Diamond A) &\to (A \vee \Diamond A) \wedge \Box\neg A \\ &\to A. \end{aligned}$$

By $R2$ and $J1$ this implies $IL \setminus J5 \vdash (A \vee \Diamond A) \wedge \Box\neg(A \vee \Diamond A) \rhd A$. Together with $(\star)$ and $J2$ this yields $IL \setminus J5 + J5' \vdash A \vee \Diamond A \rhd A$. Finally, $L \vdash \Diamond A \to A \vee \Diamond A$ implies $IL \setminus J5 \vdash \Diamond A \rhd A \vee \Diamond A$, so by $J2$ we have $IL \setminus J5 + J5' \vdash \Diamond A \rhd A$. ∎

Using the equivalence $\Box A \leftrightarrow \neg A \rhd \bot$ we provide an alternative, $\Box$-free axiomatization of IL.

2.2 Definition. IL^* denotes Propositional Logic plus

($R1^*$) $\vdash A, \vdash A \to B \Rightarrow \vdash A$;
($R2^*$) $\vdash A \to B \Rightarrow \vdash A \rhd B$;
($J1^*$) $(A \wedge B) \rhd \bot \to A \rhd \neg B$;
($J2^*$) $(A \rhd B) \wedge (B \rhd C) \to A \rhd C$;
($J3^*$) $(A \rhd C) \wedge (B \rhd C) \to (A \vee B) \rhd C$;
($J5^*$) $A \rhd A \wedge (A \rhd B)$.

In addition, IL^* contains the definition $\Box A := \neg A \rhd \bot$.

2.3 Proposition. *If $IL^* \vdash A$ then $IL \vdash A$.*

Proof. It is sufficient to show that IL is closed under the IL^*-rules and that IL proves all IL^*-axioms. Now, closure under $R1^*$ is trivial, and if $IL \vdash A \to B$, then $IL \vdash A \rhd B$ by $R2$ and $J1$, so IL is closed under $R2^*$. It is clear that $J2^*$, $J3^*$ are IL-derivable, and given the fact that $IL \vdash \Box A \leftrightarrow \neg A \rhd \bot$, it is clear that $J1^*$ is. So it remains to be proved that $IL \vdash J5^*$. We have

$$\begin{aligned} IL \vdash A \wedge \Box\neg A &\to A \wedge \Box(A \to B) \\ &\to A \wedge (A \rhd B), \text{ by } J1, \end{aligned}$$

so

$$IL \vdash A \wedge \Box\neg A \rhd A \wedge (A \rhd B), \text{ by } R2 \text{ and } J1.$$

Together with Proposition 1.3.(iii) and $J5$ this implies that $IL \vdash J5^*$. ■

2.4 Proposition. *If $IL \vdash A$ then $IL^* \vdash A$.*

Proof. We show that IL^* proves all IL-axioms, and that it is closed under all IL-rules. Closure under $R1$ is trivial. As to $R2$, note that $IL^* \vdash A$ implies $IL^* \vdash \neg A \to \bot$ implies $IL^* \vdash \neg A \rhd \bot$ — but this means $IL^* \vdash \Box A$.

To show that IL^* proves $L1$, we have to show that $IL^* \vdash ((A \wedge \neg B) \rhd \bot) \wedge (\neg A \rhd \bot) \to \neg B \rhd \bot$. Now,

$$IL^* \vdash ((A \wedge \neg B) \rhd \bot) \wedge (\neg A \rhd \bot) \to ((A \wedge \neg B) \vee \neg A) \rhd \bot, \text{ by } J3^* \qquad (\star)$$

Furthermore, $IL^* \vdash \neg B \to (A \wedge \neg B) \vee \neg A$, so $IL^* \vdash \neg B \rhd (A \wedge \neg B) \vee \neg A$, by $R2^*$. Thus, by $(\star)$ and $J2^*$ it follows that $IL^* \vdash (A \wedge \neg B) \rhd \bot \wedge (\neg A \rhd \bot) \to \neg B \rhd \bot$, as required.

As is well-known, to show that IL^* proves $L2$ it suffices to show that IL^* proves $L3$. Now

$$\begin{aligned} IL^* \vdash (\neg A \wedge (\neg A \rhd \bot)) \rhd \bot \to \neg A \rhd (\neg A \wedge (\neg A \rhd \bot)) \wedge (\neg A \wedge (\neg A \rhd \bot)) \rhd \bot, \text{ by } J5^* \\ \to \neg A \rhd \bot, \text{ by } J2^*, \end{aligned}$$

which means that $IL^* \vdash L3$.

$J1$ is immediate from $J1^*$; $J2$ and $J3$ are $J2^*$ and $J3^*$. To show that $IL^* \vdash J4$, it is sufficient to show that $IL^* \vdash A \rhd B \to (B \rhd \bot \to A \rhd \bot)$ — and this is just a special case of $J2^*$. Finally, to show that $IL^* \vdash J5$, we only have to show, by Proposition 2.1.(iii), that $IL^* \vdash A \rhd A \wedge (A \rhd \bot)$ — which is just a special case of $J5^*$. ■

Next we derive closure of IL under various rules of inference. We need the following notions:

2.5 Definition. (i) A frame $\mathcal{F}_1$ is a *subframe* of a frame F_2 if (1) $W_1 \subseteq W_2$, (2) $R_1 = R_2 \cap (W_1 \times W_1)$ and (3) for all $w \in W_1$, $S_{1w} = S_{2w} \cap (W_1 \times W_1)$. $\mathcal{F}$ is a *generated subframe* of $\mathcal{F}_2$ if $\mathcal{F}_1$ is a subframe of $\mathcal{F}_2$ such that, for all $w \in W_1$, $v \in W_2$, if wR_2v then $v \in W_1$, and such that for all $w, v \in W_1$, $u \in W_2$, if $vS_{2w}u$ then $u \in W_1$.

(ii) A model $\mathcal{M}_1$ is a *submodel* of $\mathcal{M}_2$ if $\mathcal{F}_1$ is a subframe of $\mathcal{F}_2$, and for each proposition letter p, $\{ w \in W_1 \mid w \Vdash p \} = \{ w \in W_2 \mid w \Vdash p \} \cap W_1$. If, in addition, $\mathcal{F}_1$ is a generated subframe of $\mathcal{F}_2$, then $\mathcal{M}_1$ is a *generated submodel* of $\mathcal{M}_2$.

2.6 Proposition. *If $\mathcal{M}_1$ is a generated submodel of $\mathcal{M}_2$, then, for all $w \in W_1$ and all formulas A, $w \Vdash_1 A$ iff $w \Vdash_2 A$.*

Proof. Induction on A. ■

2.7 Proposition. (i) $IL \vdash A$ *iff* $IL \vdash \Box A$;
(ii) $IL \vdash A \rhd B$ *iff* $IL \vdash A \to (B \vee \Diamond B)$;
(iii) $IL \vdash A \rhd B$ *iff* $IL \vdash \Diamond A \to \Diamond B$.

Proof. (i) One direction is an application of rule $R2$, while the other direction is an easy application of the preceding 'Generation Theorem'.

(ii) One direction is easy:

$$\begin{aligned} IL \vdash A \to B \vee \Diamond B &\Rightarrow IL \vdash \Box(A \to B \vee \Diamond B), \text{ by } R2, \\ &\Rightarrow IL \vdash A \rhd B, \text{ by } J1, J3 \text{ and } J5. \end{aligned}$$

The converse direction may proved semantically, using a trick known as 'Smoryński's trick'. Assume that $IL \nvdash A \to B \vee \Diamond B$. By the completeness theorem for IL, there is an IL-model $\mathcal{M}_1 \not\models A \to B \vee \Diamond B$ (Figure 1):

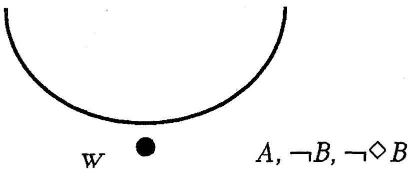


Figure 1.

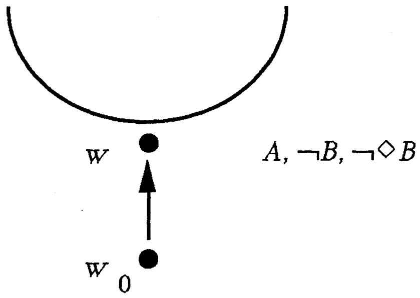


Figure 2.

We may safely assume that $\mathcal{M}_1$ is generated by w. Define $\mathcal{M}_2$ by appending a new root w_0 to $\mathcal{M}_1$ as follows (see Figure 2):

- $W_2 = \{ w_0 \} \cup W_1$;
- $R_2 = R_1 \cup \{ \langle w_0, v \rangle \mid v \in W_1 \}$. So for all $v \in W_1$, $vR_1 = vR_2$;
- for all old points v, S_v remains the same, while S_{w_0} is the reflexive closure of $R_2 \cup \{ \langle w, w \rangle \}$ on w_0R;
- for all old points v, $v \Vdash_2 p$ iff $v \Vdash_1 p$, and $w_0 \Vdash p$ for all p (or for some p, or ...).

Then $\mathcal{M}_2$ is an *IL*-model, and for all C and $v \in W_1$, $v \Vdash_2 C$ iff $v \Vdash_1 C$. Moreover, let $wS_{2\,w_0}v$ then $v = w$ or wRv; in both cases we find $v \Vdash_1 \neg B$, so $v \Vdash_2 \neg B$. Hence, we have w_0R_2w and $w \Vdash_2 A$, while for no $v \in wS_{1\,w_0}$, $v \Vdash_2 B$. Therefore, $w_0 \Vdash_2 \neg(A \rhd B)$ and $IL \nvdash A \rhd B$.
(iii) The direction from left to right is immediate by axiom $J4$ and rule $R2$. The proof of the converse uses Smoryński's trick, and is similar to the second half of (ii) above. ∎

If, in the above proofs, the 'input-model' in the application of Smoryński's trick is an *ILP* (or *ILM*-) model, then so is the 'output-model'. Therefore, if *ILS* denotes either *IL*, *ILP* or *ILM*, then:

2.8 Proposition. (i) $ILS \vdash A$ *iff* $ILS \vdash \Box A$;
(ii) $ILS \vdash A \rhd B$ *iff* $ILS \vdash A \to (B \vee \Diamond B)$;
(iii) $ILS \vdash A \rhd B$ *iff* $ILS \vdash \Diamond A \to \Diamond B$.

2. A New Approach to Modal Completeness Proofs

In this chapter we give new modal completeness proofs for several well-known systems. These new proofs use *infinite* maximal consistent sets in stead of the finite ones used in de Jongh and Veltman [1990]. Our approach has the advantage that it can do without the large adequate sets employed in that paper.

Section 1 contains the necessary preliminaries. Section 2 contains the modal completeness proofs of *IL* and *ILP*. We have not been able to prove completeness for *ILM* using our method — the problems besetting our attempt are set out in Section 3. The reader whose main interest lies with modal logic may move directly to Chapter 4 after having read this chapter; there, he or she will find further applications of the method described below.

1 Introduction

We start with some definitions. Let *ILS* denote either *IL* or *ILP*.

1.1 Definition. Let Γ, Δ be two maximal *ILS*-consistent sets. Then $\Gamma \prec \Delta$ (Δ is a *successor* of Γ), if

(i) $A, \Box A \in \Delta$ for each $\Box A \in \Gamma$;
(ii) $\Box A \in \Delta$ for some $\Box A \notin \Gamma$.

We write $\Gamma \preceq \Delta$ for $\Gamma = \Delta$ or $\Gamma \prec \Delta$.

1.2 Definition. Let Γ, Δ be two maximal *ILS*-consistent sets. Then Δ is a *C-critical* successor of Γ if

(i) $\Gamma \prec \Delta$;
(ii) $\neg A, \Box\neg A \in \Delta$ for each A such that $A \rhd C \in \Gamma$.

We repeat two results from de Jongh and Veltman [1990]. Originally they were proved for *finite* maximal *ILS*-consistent sets contained in some finite large adequate set. By a simple compactness argument, however, their proofs apply equally well to the infinite sets considered here, so we will state the two results without proof, and in the form in which we will need them.

1.3 Proposition. *Suppose Γ is a maximal ILS-consistent set. If $\neg(B \rhd C) \in \Gamma$, then there exists a C-critical successor Δ of Γ, such that $B, \Box B \in \Delta$.*

1.4 Proposition. *Suppose Γ is a maximal ILS-consistent set with $B \rhd C \in \Gamma$. If there is an E-critical successor Δ of Γ with $B \in \Delta$, then there exists an E-critical successor Δ' of Γ with $C, \Box\neg C \in \Delta'$.*

1.5 Proposition. $IL \vdash \neg(B \rhd C) \to \Diamond B.$

One final definition is needed before we can prove the completeness of *IL*:

1.6 Definition. A set of formulas Φ is *adequate* if

(i) if $B \in \Phi$, and C is a subformula of B, then $C \in \Phi$,
(ii) if $B \in \Phi$, and B is no negation, then $\neg B \in \Phi$.

Let Φ be a adequate set. Then we say that a formula $\Diamond B$ is *almost in* Φ, if, for some C, $B \rhd C \in \Phi$ or $C \rhd B \in \Phi$, or if, for some C, $B \equiv \neg C$ and $\Box C \in \Phi$.

Assuming that negations $\neg A$ are written as $A \to \bot$, any non-empty adequate set contains $\bot$.

2 The Modal Completeness of *IL* and *ILP*

Although we prove the modal completeness of *IL* and *ILP* separately, the definition of the structure $\langle W_\Gamma, R\rangle$ is common to both proofs. We therefore introduce it independently of the completeness proofs.

Given some infinite maximal *ILS*-consistent set Γ, and a finite adequate set Φ, $\langle W_\Gamma, R\rangle$ consists of pairs $\langle \Delta, \tau\rangle$, where the maximal conistent sets Δ are needed to handle the truth definition for formulas in $\Gamma \cap \Phi$. We use the sequences of formulas τ to carefully index the pairs we put into W_Γ. In this way we make sure that $\langle W_\Gamma, R\rangle$ will be a finite tree.

Let us get to work now. For the time being, let Γ be an infinite *ILS*-maximal consistent set, and let Φ be a finite adequate set. We use $\bar{w}, \bar{v}, \ldots$ to denote pairs $\langle \Delta, \tau\rangle$. And if $\bar{w} = \langle \Delta, \tau\rangle$, then $(\bar{w})_0 = \Delta$ and $(\bar{w})_1 = \tau$. We write $\sigma \subseteq \tau$ for σ is an initial segment of τ, and $\sigma \subset \tau$ if σ is a proper initial segment of τ.

2.1 Definition. Define W_Γ to be a minimal set of pairs $\langle \Delta, \tau\rangle$ such that

(i) $\langle \Gamma, \langle\langle\rangle\rangle\rangle \in W_\Gamma$;
(ii) if $\langle \Delta, \tau\rangle \in W_\Gamma$, $\Diamond B \in \Delta$ is almost in Φ and $C \in \Phi$, and if there is a C-critical successor Δ' of Δ with $B, \Box\neg B \in \Delta'$, then $\langle \Delta', \tau^\frown\langle\langle B, C\rangle\rangle\rangle \in W_\Gamma$ for *one* such Δ'.

Define R on W_Γ by putting $\bar{w} R \bar{v}$ iff $(\bar{w})_1 \subset (\bar{v})_1$.

In a series of Propositions we now establish the main facts about $\langle W_\Gamma, R\rangle$. First of all, our 'indexing mechanism' ensures that W_Γ is finite:

2.2 Proposition. *W_Γ is finite.*

Proof. Since $|\Phi| = m < \omega$ is finite it follows that $|\{\Diamond B \in \Gamma \mid \Diamond B \text{ is almost in } \Phi\}| = n < \omega$, for some n. So Γ gives rise to adding at most $n \cdot m$ new elements to W_Γ. Now each of these new elements will contain one less formula of the form $\Diamond B$, where $\Diamond B$ is almost in Φ. So each new element will give rise to adding at most $(n-1) \cdot m$ elements. Continuing in this way we see that $|W_\Gamma| \leq 1 + \prod_{i=0}^{n-1}((n-i) \cdot m) < \omega$. ∎

2.3 Proposition. *If $\langle \Delta, \tau\rangle \in W_\Gamma$ and E occurs as the second component in some pair in τ, then $\neg E \in \Delta$.*

Proof. Show by induction on the construction of W_Γ that if $\langle \Delta, \tau\rangle$ and E are as stated, then $\neg E, \Box\neg E \in \Delta$. ∎

2.4 Proposition. *If $\bar{w}$, $\bar{v} \in W_\Gamma$ and $(\bar{w})_1 \subseteq (\bar{v})_1$, then $(\bar{w})_0 \preceq (\bar{v})_0$.*

Proof. The proof is by induction on $n = \max\big(\mathrm{lh}((\bar{w})_1), \mathrm{lh}((\bar{v})_1)\big)$. The case $n = 1$ is trivial. So let $(\bar{w})_1 \subseteq (\bar{v})_1$, $\mathrm{lh}((\bar{v})_1) = n + 1$ and $(\bar{v})_1 = \big\langle \langle B_1, C_1\rangle, \ldots, \langle B_n, C_n\rangle, \langle B_{n+1}, C_{n+1}\rangle \big\rangle$.

First assume that $(\bar{w})_1 = (\bar{v})_1$. By construction W_Γ contains $\bar{w}'$, $\bar{v}'$ such that $(\bar{w})_1 = (\bar{w}')_1 {}^\frown \langle B_{n+1}, C_{n+1}\rangle$, $(\bar{w})_0 \succ (\bar{w}')_0$ and $(\bar{v})_1 = (\bar{v}')_1 {}^\frown \langle B_{n+1}, C_{n+1}\rangle$, $(\bar{v})_0 \succ (\bar{v}')_0$. By the induction hypothesis we have $(\bar{v}')_0 \preceq (\bar{w}')_0$ and $(\bar{w}')_0 \preceq (\bar{v}')_0$. It follows that $(\bar{v}')_0 = (\bar{w}')_0$, and hence $\bar{v}' = \bar{w}'$. Now for each pair $\langle B, C\rangle$ at most one C-critical $\bar{u}$ with $B, \Box\neg B \in (\bar{u})_0 \succ (\bar{v}')_0$ is added to W_Γ. Therefore $\bar{w} = \bar{v}$ and $(\bar{w})_0 \preceq (\bar{v})_0$ as required.

Next assume that $(\bar{w})_1 \neq (\bar{v})_1$. Let $\bar{v}'$ be an immediate predecessor of $\bar{v}$. Then $(\bar{w})_1 \subseteq (\bar{v}')_1$, so by the induction hypothesis $(\bar{w})_0 \preceq (\bar{v}')_0 \prec (\bar{v})_0$ as required. ■

2.5 Proposition. *$\langle W_\Gamma, R\rangle$ is a tree.*

Proof. Since transitivity and asymmetry are straightforward, we only have to prove that for each $\bar{w} \in W_\Gamma$ the set of its R-predecessors is finite and linear. Finiteness follows from Proposition 2.2. To prove linearity, assume that $\bar{u}$, $\bar{v}$ are two R-predecessors of $\bar{w}$. Then $(\bar{u})_1 \subset (\bar{w})_1$ and $(\bar{v})_1 \subset (\bar{w})_1$, so $(\bar{u})_1 \subseteq (\bar{v})_1$ or $(\bar{v})_1 \subseteq (\bar{u})_1$ $(\star)$. If $\bar{u} \neq \bar{v}$, then $(\bar{u})_0 \not\preceq (\bar{v})_0$ or $(\bar{v})_0 \not\preceq (\bar{u})_0$ or $(\bar{u})_1 \not\subseteq (\bar{v})_1$ or $(\bar{v})_1 \not\subseteq (\bar{u})_1$. By the preceding Proposition it follows that (i) $(\bar{u})_1 \not\subseteq (\bar{v})_1$ or (ii) $(\bar{v})_1 \not\subseteq (\bar{u})_1$. If (i) holds, then by $(\star)$ we have $(\bar{v})_1 \subset (\bar{u})_1$, hence $\bar{v}R\bar{u}$. Similarly, if (ii) holds then $\bar{u}R\bar{v}$, as required. ■

2.6 Theorem. *$IL \vdash A$ iff for all finite IL-models $\mathcal{M}$, $\mathcal{M} \models A$.*

Proof. Soundness is immediate. To prove completeness, assume that $IL \nvdash A$. Let Φ be a *finite* adequate set containing $\neg A$, and let Γ be a maximal IL-consistent set with $\neg A \in \Gamma$. Construct $\langle W_\Gamma, R\rangle$ as above — using infinite maximal IL-consistent sets. Then R has all the properties required. Define S on W by putting $\bar{v}S_{\bar{w}}\bar{u}$ iff

$$(\bar{v})_1 = (\bar{w})_1{}^\frown\langle\langle B, C\rangle\rangle{}^\frown\tau \text{ and } (\bar{u})_1 = (\bar{w})_1{}^\frown\langle\langle B', C\rangle\rangle{}^\frown\sigma, \text{ for some } B,\ B',\ C,\ \tau \text{ and } \sigma.$$

Then S has all the properties required. We complete the proof by putting $\bar{w} \Vdash p$ iff $p \in (\bar{w})_0$, and proving that for all $F \in \Phi$, $\bar{w} \in W$ we have $\bar{w} \Vdash F$ iff $F \in (\bar{w})_0$. The proof is by induction on F. We first consider the case $F \equiv B \rhd C$.

First assume that $B \rhd C \notin (\bar{w})_0$. Then we have to show that $\exists \bar{v}\,(\bar{w}R\bar{v} \wedge B \in (\bar{v})_0 \wedge \forall \bar{u}\,(\bar{v}S_{\bar{w}}\bar{u} \to \neg C \in (\bar{u})_0))$. Now $B \rhd C \notin (\bar{w})_0$ implies $\neg(B \rhd C) \in (\bar{w})_0$, and so $\Diamond B \in (\bar{w})_0$, by Proposition 1.5. Moreover, $\Diamond B$ is almost in Φ. By Proposition 1.3 there is a C-critical successor Δ' of $(\bar{w})_0$ with $B, \Box\neg B \in \Delta'$. We may safely assume that $\bar{v} = \big\langle \Delta', (\bar{w})_1{}^\frown\langle\langle B, C\rangle\rangle\big\rangle \in W_\Gamma$. Then $\bar{w}R\bar{v}$ and $B \in (\bar{v})_0$. Moreover, if $\bar{v}S_{\bar{w}}\bar{u}$, for some $\bar{u} \in W_\Gamma$, then C occurs as the second component in some pair in $(\bar{u})_1$. So, by Proposition 2.3, $\neg C \in (u)_0$ as required.

If, conversely, $B \rhd C \in (\bar{w})_0$, then we have to prove that $\forall \bar{v}\,(\bar{w}R\bar{v} \wedge B \in (\bar{v})_0 \to \exists \bar{u}\,(\bar{v}S_{\bar{w}}\bar{u} \wedge C \in (\bar{u})_0))$. So let $\bar{v} \in W_\Gamma$ be such that $\bar{w}R\bar{v}$. By construction $\bar{v}$ is E-critical for some $E \in \Phi$. According to Proposition 1.4 there is an E-critical successor Δ' of $(\bar{w})_0$ that contains $C, \Box\neg C$. Since $B \rhd C \in (\bar{w})_0$ and $B \in (\bar{v})_0 \succ (\bar{w})_0$, it follows that $\Diamond B \in (\bar{w})_0$, and therefore $\Diamond C \in (\bar{w})_0$ by axiom $J4$. Since $\Diamond C$ is almost in Φ, we may assume that $\bar{u} = \big\langle \Delta', (\bar{w})_1{}^\frown\langle\langle C, E\rangle\rangle\big\rangle \in W_\Gamma$. Clearly, we have $\bar{v}S_{\bar{w}}\bar{u}$ and $C \in (\bar{u})_0$ as required.

Now, the case $F \equiv \Box B$ follows from the previous case, since IL proves $\Box B \leftrightarrow (\neg B \rhd \bot)$, and if $\Box B \in \Phi$, then $\neg\Box B, \bot \in \Phi$ and $\Diamond\neg B$ is almost in Φ. ■

The same apparatus can be used almost without modification to prove the modal completeness of *ILP* using small adequate sets.

2.7 Theorem. *$ILP \vdash A$ iff for all finite ILP-models $\mathcal{M}$, $\mathcal{M} \models A$.*

Proof. Soundness is immediate. To prove completeness, assume that $ILP \nvdash A$. Let Φ be a *finite* adequate set containing $\neg A$, and let Γ be a maximal ILP-consistent set with $\neg A \in \Gamma$. Construct

$\langle W_\Gamma, R\rangle$ as above — this time using infinite maximal *ILP*-consistent sets. Then R has all the properties required. Moreover, every $\bar{w} \in W_\Gamma$ differing from $\langle \Gamma, \langle\langle\rangle\rangle\rangle$ has exactly one immediate predecessor. S is defined on W_Γ by putting $\bar{v} S_{\bar{w}} \bar{u}$ iff

$$(\bar{v})_1 = (\bar{w})_1{}^\frown\tau^\frown\langle\langle B, C\rangle\rangle \text{ and } (\bar{u})_1 = (\bar{w})_1{}^\frown\tau^\frown\langle\langle B', C\rangle\rangle^\frown\sigma, \text{ for some } B,\ B',\ C,\ \tau \text{ and } \sigma.$$

One easily verifies that S has all the required properties. We complete the proof by putting $\bar{w} \Vdash p$ iff $p \in (\bar{w})_0$, and by proving that for all $F \in \Phi$, $\bar{w} \in W_\Gamma$, we have $\bar{w} \Vdash F$ iff $F \in (\bar{w})_0$. Once again the proof is by induction on F. And once again, the case $F \equiv \Box B$ follows from the case $F \equiv B \rhd C$, so we may restrict ourselves to the latter case.

The case that $B \rhd C \notin (\bar{w})_0$ is entirely analogous to the corresponding case in the completeness proof for *IL*.

Finally, assume that $B \rhd C \in (\bar{w})_0$. Then we have to show that $\forall \bar{v}\,(\bar{w} R \bar{v} \wedge B \in (\bar{v})_0 \rightarrow \exists \bar{u}\,(\bar{v} S_{\bar{w}} \bar{u} \wedge C \in (\bar{u})_0))$. So assume that $\bar{w} R \bar{v}$ and $B \in (\bar{v})_0$. Since $\langle W_\Gamma, R\rangle$ is a tree, we can find a unique immediate predecessor $\bar{w}'$ of $\bar{v}$. Then $B \rhd C \in (\bar{w}')_0$ and since $B \in (\bar{v})_0 \succ (\bar{w}')_0$, we also have $\Diamond B \in (\bar{w}')_0$ and therefore $\Diamond C \in (\bar{w}')_)$, by axiom *J*4. By construction $(\bar{v})_0$ is an E-critical successor of $(\bar{w}')_0$ for some $E \in \Phi$, and $(\bar{v})_1 = (\bar{w}')_1{}^\frown\langle\langle D, E\rangle\rangle$ for some $D \in \Phi$. Now Proposition 1.4 yields an E-critical successor Δ' of $(\bar{w}')_0$ with $C, \Box\neg C \in \Delta'$. Since $\Diamond C$ is almost in Φ, we may assume that $\bar{u} = \langle \Delta', (\bar{w}')_1{}^\frown\langle\langle C, E\rangle\rangle\rangle \in W_\Gamma$. Clearly, $\bar{v} S_{\bar{w}} \bar{u}$ and $C \in (\bar{u})_0$ as required. ■

3 On the Modal Completeness of *ILM*

The construction used above to prove *IL* and *ILP* complete, cannot be used without modification to prove the modal completeness of *ILM*. There are several reasons for this.

In general, *ILM*-models are not trees. But, as we pointed out before, the 'indexing mechanism' used in Section 2 makes $\langle W_\Gamma, R\rangle$ into a tree. This problem may probably be overcome by choosing some other indexing mechanism. But there is a more serious difficulty.

Let Γ be some maximal *ILM*-consistent set, and suppose that $(A \rhd B) \wedge (B \rhd A) \in \Gamma$. By axiom M, $A \wedge \Box(\neg A \wedge \neg B) \rhd B \wedge \Box(\neg A \wedge \neg B) \in \Gamma$. Now assume that we have an R-successor Δ_1 with $A \wedge \Box(\neg A \wedge \neg B) \in \Delta_1$. Then there must be some S_Γ-successor Δ_2 of Δ_1 with $B \wedge \Box(\neg A \wedge \neg B) \in \Delta_2$. Clearly, Δ_2 cannot be an R-successor of Δ_1. Similarly, Δ_2 asks for an S_Γ-successor Δ_3 with $A \wedge \Box(\neg A \wedge \neg B) \in \Delta_3$, which cannot be an R-successor of Δ_2, etc. At present I see no way to avoid having to introduce infinite S-chains in situations like these.

(It may be worthwhile to solve these problems. Consider the system $ILWM_0$. Visser [1989] conjectures that it embodies precisely the principles valid in all Σ_1^b-axiomatized theories with designated natural numbers satisfying $I\Delta_0 + \Omega_1$. Now, we cannot use the method based on the machinery of finite maximal consistent sets contained in a large adequate set to prove $ILWM_0$ modally complete. For, if we want to apply the axiom M_0 we need to have $\Diamond A \wedge \Box C \rhd B \wedge \Box C$ available for any $A \rhd B$ and $\Box C$ in the adequate set; we leave it to the reader to check that this may turn the adequate set into an infinite one, that contains infinitely many non-equivalent formulas. So we have to use another method to prove $ILWM_0$ complete w.r.t. some class of finite models. It may be that some variation on the method employed in Section 2 will work; but to find this out, we first have to solve the problems mentioned in this section.)

3. The Arithmetic Completeness of ILP^{ω}

In his Visser [1990] Albert Visser proves that *ILP* completely axiomatizes the interpretability logic of finitely axiomatized sequential theories extending $I\Delta_0 + \text{SupExp}$. To introduce the topic of this chapter, we briefly mention some steps in the proof of Visser's completeness theorem.

The completeness part is proved by contraposition. That is, if $ILP \nvdash A$ then, by the modal completeness of *ILP*, there is a so-called Friedman model (of *ILP*) in which A fails. Using this model (and a Solovay-like function) an interpretation $(\cdot)^*$ of $\mathcal{L}(\Box, \rhd)$ in the language of arithmetic is then defined. After that, and relative to some fixed finitely axiomatized sequential theory U that extends $I\Delta_0 + \text{SupExp}$, it is shown that A^* is not derivable. One of the key results needed in showing this, is a result by Friedman that gives a characterization of interpretability in terms of consistency:

> Let T and S be finitely axiomatized sequential theories, then T interprets S iff $I\Delta_0 + \text{Exp}$ proves that the consistency of T (with respect to cut free proofs) implies the consistency of S (with respect to cut free proofs).

A proof of this result may be found in Visser [1990], Section 7.4; before Visser, Smoryński gave a proof in his Smoryński [1985b].

Now, let U be some finitely axiomatized sequential theory extending $I\Delta_0 + \text{SupExp}$. Given the above characterization, one naturally defines $(\cdot)^*$ to be a map which assigns to $A \rhd B$ a formalization of the right hand side of this characterization, where $T = U + A$ and $S = U + B$. (Of course $(\cdot)^*$ is defined on other formulas in the usual way.)

The main aim of this chapter is to show that ILP^{ω} axiomatizes all $A \in \mathcal{L}(\Box, \rhd)$ such that for all interpretations $(\cdot)^*$ of the kind described above, A^* is true (i.e., true on the standard model).

This chapter is organized as follows. The first section contains a characterization of derivability in IL^{ω}, ILP^{ω} and ILM^{ω} in terms of derivability in *IL*, *ILP* and *ILM* respectively. Except for some things that take up too much space, we repeat must of the preliminary assumptions, definitions and results from Visser [1990]; this is done in Section 2. Then, in Section 3, we prove our completeness result.

1 ω-Versions of Several Systems

It is well-known that L can be axiomatized without the rule $R2$. An inspection of the simple proof of this fact — as given in e.g. Smoryński [1985a] — shows that *IL* (and *ILP*, *ILM*, ...) has an $R2$-free formulation as well.

In the sequel we consider the system IL^{ω} which is the extension of *IL* (in its $R2$-free formulation) by the axiom schema of *Reflection*: $\Box A \to A$. The systems ILP^{ω} and ILM^{ω} are defined in a similar way. Our main goal is to characterize derivability in IL^{ω} in terms of derivability in *IL* — as a corollary to this result we obtain such characterizations for ILP^{ω} and ILM^{ω}. In working towards this goal we follow Smoryński [1985a] Ch. 2 Sect. 4 rather closely.

Let Sub(A) denote the set of subformulas of A.

1.1 Definition. Let $A \in \mathcal{L}(\Box, \rhd)$. An *IL*-model $\mathcal{M}$ with root w_0 is called *A-sound* if (i) for every B such that $\Box B \in \text{Sub}(A)$ we have $w_0 \Vdash \Box B \rightarrow B$, and (ii) for every C such that $C \rhd D \in \text{Sub}(A)$, for some D, we have $w_0 \Vdash C \rightarrow \Diamond C$.

We now adapt Smoryński's derived models to the present context.

1.2 Definition. Let $\mathcal{M}$ be an *IL*-model with root w_0. The *derived model* $\mathcal{M}'$ is defined by

$W' = W \cup \{ w_{-1} \}$ (w_{-1} new);
R' is the transitive closure of $R \cup \{ \langle w_{-1}, w_0 \rangle \}$;
S': for $v \in W$, $S'_v = S_v$, and $S_{w_{-1}}$ is the transitive closure of $S_{w_0} \cup \{ \langle w_0, w_0 \rangle \} \cup \{ \langle w_0, v \rangle \mid w_0 R v \}$;
$\Vdash'$: if $v \in W$, then $v \Vdash' p$ iff $v \Vdash p$, and $w_{-1} \Vdash p$ iff $w_0 \Vdash p$.

A sequence of successive derivations is denoted $\mathcal{M}^{(1)}$, $\mathcal{M}^{(2)}, \ldots$ with respective roots w_{-1}, w_{-2}, ...

Notice the following:

(i) if $v \in W$, $E \in \mathcal{L}(\Box, \rhd)$, then $v \Vdash E$ iff $v \Vdash^{(n)} E$;
(ii) if $v, u \in W$, then vRu iff $vR^{(n)}u$;
(iii) if $v \in W$, then $S_v = S_v^{(n)}$.

(i), (ii) and (iii) together allow us to omit (nearly) all superscripts. Some other properties are

(iv) $S_{w_0} \subseteq S_{w_{-1}} \subseteq \ldots \subseteq S_{w_{-n}} \subseteq \ldots$;
(v) $S_{w_{-(n+1)}} \restriction (R[w_{-n}] \times R[w_{-n}]) = S_{w_{-n}}$;
(vi) if $w_{-n} R v S_{w_{-(n+1)}} u$ then $u \neq w_{-n}$, since $w_{-n} R v$ implies $w_{-n} \neq v$ and since the only $S_{w_{-(n+1)}}$-arrow that arrives at w_{-n} also starts at w_{-n};
(vii) if $w_{-n} R v S_{w_{-(n+1)}} z$, then $w_{-n} R z$.

1.3 Proposition. *Let $\mathcal{M}$ be an A-sound model, and $\mathcal{M}^{(n)}$ the n-th derived model. Then for all $E \in \text{Sub}(A)$, $w_0 \Vdash E$ iff $w_{-n} \Vdash E$.*

Proof. This is by induction on n. Of course it is enough to prove the case $n = 1$. This is by induction on E. We only consider the case $E \equiv C \rhd D$. Suppose $w_0 \Vdash C \rhd D$, and $w_{-1} R v$, $v \Vdash C$. Then we must find a z with $v S_{w_{-1}} z$ and $z \Vdash D$. Now $w_{-1} R v$ implies $w_0 = v$ or $w_0 R v$. If $w_0 = v$, then $w_0 \Vdash C$, so $w_0 \Vdash \Diamond C$, by A-soundness. So

$$\begin{aligned} & u \Vdash C, \text{ for some } u \breve{R} w_0 \\ & \quad \Rightarrow z \Vdash D, \text{ for some } z \text{ with } u S_{w_0} z, \text{ since } w_0 \Vdash C \rhd D \\ & \quad \Rightarrow z \Vdash D, \text{ for some } z \text{ with } u S_{w_{-1}} z, \text{ since } S_{w_0} \subseteq S_{w_{-1}}. \end{aligned}$$

Now $w_{-1} R v R u$ implies $v S_{w_{-1}} u$, and by the transitivity of $S_{w_{-1}}$ this implies $v S_{w_{-1}} z$ as required. If $w_0 R v$, we can proceed in a similar way.

Next suppose that $w_{-1} \Vdash C \rhd D$, and $w_0 R v, v \Vdash C$. Then we must find a z with $v S_{w_0} z$ and $z \Vdash D$. By definition $w_0 R v$ implies $w_{-1} R v$, so we find a z with $v S_{w_{-1}} z$ and $z \Vdash D$. By remark (vii) above $w_0 R v S_{w_{-1}} z$ implies $w_0 R z$. But then we have $v S_{w_0} z$, because $S_{w_{-1}} \restriction (R[w_0] \times R[w_0]) = S_{w_0}$. ∎

1.4 Proposition. *If $\mathcal{M}$ is A-sound, then so is $\mathcal{M}^{(n)}$.*

Proof. This is by induction on n. It suffices to prove the case $n = 1$. Assume that $\Box B \in \mathrm{Sub}(A)$, then we have to show that $w_{-1} \Vdash \Box B \to B$. So suppose that $w_{-1} \Vdash \Box B$, then $w_0 \Vdash B$. By Proposition 1.3 it follows that $w_{-1} \Vdash B$.

Next, assume $C \rhd D \in \mathrm{Sub}(A)$ and $w_{-1} \Vdash C$. We have to show that $w_{-1} \Vdash \Diamond C$. By Proposition 1.3 we have $w_0 \Vdash C$, and therefore $w_{-1} \Vdash \Diamond C$. ∎

We say that a formula A is *true* on a model $\mathcal{M}$ with root w_0, if $w_0 \Vdash A$. A is *valid* on a model if it is forced by each point in that model.

1.5 Proposition. *Let $A \in \mathcal{L}(\Box, \rhd)$. Then*

(i) *$IL^\omega \vdash A$ iff A is true on all A-sound IL-models;*

(ii) *$IL \vdash A$ iff A is valid on all A-sound IL-models.*

Proof. (i) If A is true on all A-sound IL-models then

$$IL + \Big(\bigwedge_{\Box B \in \mathrm{Sub}(A)} (\Box B \to B) \wedge \bigwedge_{C \rhd D \in \mathrm{Sub}(A)} (C \to \Diamond C) \Big) \vdash A$$

so

$$IL + \Big(\bigwedge_{\Box B \in \mathrm{Sub}(A)} (\Box B \to B) \wedge \bigwedge_{C \rhd D \in \mathrm{Sub}(A)} (\Box\neg C \to \neg C) \Big) \vdash A,$$

whence $IL^\omega \vdash A$.

To prove the converse, suppose that A is false on an A-sound model $\mathcal{M}$ with root w_0. We show that $IL^\omega \nvdash A$, by showing that for any finite set X, $IL + \bigwedge_{\Box B \in X}(\Box B \to B) \nvdash A$. By Proposition 1.4 each model in the sequence $\mathcal{M}^{(0)}, \mathcal{M}^{(1)}, \ldots$ is an A-sound model on which A is false. It suffices to show, for any given finite set X, there to be some n_0 such that $w_{n_0} \Vdash \bigwedge_{\Box B \in X}(\Box B \to B)$. But this is simple: we have either $w_{-n} \Vdash \Box B$ for all n, whence $w_{-n} \Vdash B$ and $w_{-n} \Vdash \Box B \to B$, for all n, *or* $w_{n_0} \nVdash \Box B$ for some n_0, whence $w_{-n} \nVdash \Box B$ and $w_{-n} \Vdash \Box B \to B$ for all $n \geq n_0$.

(ii) If $IL \vdash A$ then A is valid on all IL-models and, hence, on all A-sound IL-models.

To prove the converse, assume $IL \nvdash A$. Then we find an IL-model $\mathcal{M}$ with root w_0 such that $w_0 \nVdash A$. If $\mathcal{M}$ is A-sound, then we are done. So assume that $\mathcal{M}$ is not A-sound. Then $w_0 \nVdash \Box B \to B$, for some $\Box B \in \mathrm{Sub}(A)$, or $w_0 \nVdash C \to \Diamond C$ for some $C \rhd D \in \mathrm{Sub}(A)$. In the first case $w_0 \nVdash B$, hence $w_{-n} \nVdash \Box B$ and $w_{-n} \Vdash \Box B \to B$, for all $n > 0$. In the second case $w_0 \Vdash C$, hence $w_{-n} \Vdash \Diamond C$ and $w_{-n} \Vdash C \to \Diamond C$, for all $n > 0$.

It follows that if $\mathcal{M}^{(1)}$ is not A-sound, then this is witnessed by formulas $\Box B_1, \ldots, \Box B_m$ and $C_1 \rhd D_1, \ldots, C_k \rhd D_k$ different from the ones that caused $\mathcal{M}$ to be *A-unsound*. Reasoning as above, we find that $w_{-n} \Vdash \neg\Box B_1 \wedge \ldots \wedge \neg\Box B_m \wedge \Diamond C_1 \wedge \ldots \Diamond C_k$, for all $n > 1$. Since $\mathrm{Sub}(A)$ is finite, we find an n_0 such that $\mathcal{M}^{(n_0)}$ is A-sound. Since we clearly still have $w_0 \nVdash A$ in that model, we have proved (ii). ∎

We can now prove the main result:

1.6 Proposition. *Let $A \in \mathcal{L}(\Box, \rhd)$. Then the following are equivalent:*

(i) $IL^\omega \vdash A$

(ii) $IL \vdash \Big(\bigwedge_{\Box B \in \mathrm{Sub}(A)}(\Box B \to B) \wedge \bigwedge_{C \rhd D \in \mathrm{Sub}(A)}(C \to \Diamond C)\Big) \to A.$

Proof. By Proposition 1.5 we have $IL^\omega \vdash A$ iff A is true on all A-sound IL-models, i.e. on all models of $IL + (\bigwedge_{\Box B \in \mathrm{Sub}(A)}(\Box B \to B) \wedge \bigwedge_{C \rhd D \in \mathrm{Sub}(A)}(C \to \Diamond C))$. So $IL^\omega \vdash A$ iff $IL \vdash (\bigwedge_{\Box B \in \mathrm{Sub}(A)}(\Box B \to B) \wedge \bigwedge_{C \rhd D \in \mathrm{Sub}(A)}(C \to \Diamond C)) \to A$. ∎

Proposition 1.5 can also be used to obtain the following reduction of IL to IL^ω:

1.7 Proposition. *Let $A \in \mathcal{L}(\Box, \rhd)$. Then $IL \vdash A$ iff $IL^\omega \vdash \Box A$.*

Proof. $IL \vdash A$ implies $IL \vdash \Box A$ implies $IL^\omega \vdash \Box A$. And, conversely, by Proposition 1.5.(i) $IL^\omega \vdash \Box A$ implies that $\Box A$ is true and hence valid on all $\Box A$-sound models. Part (ii) of Proposition 1.5 yields $IL \vdash \Box A$, so $IL \vdash A$. ■

An inspection of the proof of Proposition 1.6 shows that IL may be replaced by ILP, and also by ILM:

1.8 Proposition. *Let $A \in \mathcal{L}(\Box, \rhd)$.*
(1) *The following are equivalent:*
  (i) $ILP^\omega \vdash A$
  (ii) $ILP \vdash \left(\bigwedge_{\Box B \in \mathrm{Sub}(A)}(\Box B \to B) \wedge \bigwedge_{C \rhd D \in \mathrm{Sub}(A)}(C \to \Diamond C)\right) \to A.$
(2) *The following are equivalent:*
  (i) $ILM^\omega \vdash A$
  (ii) $ILM \vdash \left(\bigwedge_{\Box B \in \mathrm{Sub}(A)}(\Box B \to B) \wedge \bigwedge_{C \rhd D \in \mathrm{Sub}(A)}(C \to \Diamond C)\right) \to A.$

2 Arithmetic Completeness: Preliminaries

In this section we briefly review the various notions needed and assumptions made to prove ILP^ω arithmetically complete. We will not provide full details; these may be looked up in Pudlák [1985], Visser [1988], [1990] and Wilkie and Paris [1987].

Officially we will be working in the language $\mathcal{L}^*$ — which is the relational version of the language of arithmetic — in which successor, addition and multiplication are (2-, 3- and 3-place) relation symbols. So, the only terms of $\mathcal{L}^*$ are the variables and $\underline{0}$. We will, however, *pretend* that we are working with function symbols.

An $\mathcal{L}^*$-formula is called Δ_0 if all its quantifiers are bounded. $I\Delta_0$ is PA with induction restricted to Δ_0-formulas. Now in $I\Delta_0$ we can prove almost all the basic properties of natural numbers. However, $I\Delta_0$ does not prove the existence of fast growing functions, like exponentiation. It is well-known that it is possible to define '$y = 2^x$' in $I\Delta_0$, at least as a partial function (cf. Pudlák [1985]). Moreover, for reasonable functions f like exponentiation, $I\Delta_0$ + 'f is total' implies $I\Delta_0(f)$ — where the bounding terms also involve f.

Define $|x| := \mu y.\,(2^{Sy} > Sx)$, and $\omega_1(x) := 2^{(|x|^2)}$. Ω_1 denotes the axiom $\forall x \exists y\,(y = \omega_1(x))$; Exp denotes the axiom $\forall x \exists y\,(y = 2^x)$; supexp is the 'stack of twos' function: $\mathrm{supexp}(0) = 1$, $\mathrm{supexp}(n+1) = 2^{\mathrm{supexp}(n)}$; SupExp denotes the axiom $\forall x \exists y\,(y = \mathrm{supexp}(x))$.

We make the following assumptions on the theories T we will be considering. We assume that T is given by an R_1^+-formula $\alpha_T(x)$ having just x free plus the relevant information on what the set of natural numbers of T is; α_T gives the set of codes of non-logical axioms of the theory (cf. Wilkie and Paris [1987]). We assume that the numbers of T satisfy $I\Delta_0 + \Omega_1$. Furthermore, the theories we will consider will be assumed to be finitely axiomatized and sequential:

2.1 Definition. A theory T is *finitely axiomatized* if its axiom set is given by a disjunction of formulas of the form $x = \underline{n}$, where $\underline{n}$ codes a formula.

The notion of sequentiality is due to Pudlák (in a slightly different form):

2.2 Definition. Let T be a theory such that the numbers of T satisfy $I\Delta_0 + \Omega_1$. T is called *sequential* if in it one can form sequences of any of its objects, i.e., if there is a relation $(s)_x = a$ such that T proves

(i) $\forall s, x, a, b\,((s)_x = a \wedge (s)_x = b \rightarrow a = b)$;
(ii) $\forall s \exists x \forall y\,(\exists b\,((s)_y = b \leftrightarrow y < x)$;
(iii) $\exists s \forall x, a\,(\neg(s)_x = a)$;
(iv) $\forall s, a, x\,(\forall y < x\, \exists b\,((s)_y = b) \rightarrow$
$\rightarrow \exists s' \forall b \forall y \leq x\,((s')_y = b \leftrightarrow ((y < x \wedge (s)_y = b) \vee (y = x \wedge a = b)))$.

Wilkie and Paris [1987] show that $I\Delta_0 + \Omega_1$ is a completely adequate theory for arithmetizing syntax. E.g., if T is a theory satisfying the assumptions made above, we can formalize in $I\Delta_0 + \Omega_1$ (as an R_1^+-formula) $\mathrm{Proof}_T(x, y)$, which represents the relation 'y is a proof of the formula x from T'. We can further define $\mathrm{Prov}_T(x) := \exists y\, \mathrm{Proof}_T(x, y)$, and $\Box_T(x) := \mathrm{Prov}_T(x)$, $\Diamond_T \equiv \neg\Box_T\neg$.

To apply Friedman's characterization of interpretability for finitely axiomatized sequential theories as mentioned in the introduction to this chapter, we need a notion of cut free proof. We follow Visser [1990] (which follows Wilkie and Paris [1987]) in choosing tableaux provability. We let $\mathrm{TabInconPr}(T, x)$ denote the formalization of 'x is a tableau proof from T of a contradiction'. We write $\mathrm{TabProof}_T(x, y)$ for $\mathrm{TabInconPr}(U, x)$, where U is T plus the negation of the formula coded by y. Furthermore, $\Delta_T(y) := \exists x\, \mathrm{TabProof}_T(x, y)$; $\nabla_T \equiv \neg \Delta_T \neg$, and $\mathrm{TabCon}(T) := \forall x\, \neg \mathrm{TabProof}_T(x, \ulcorner \bot \urcorner)$.

It is well-known that T is inconsistent in the usual sense just if there is a tableau proof from T of a contradiction. The advantage of such tableau proofs is that they only contain subformulas of the sentences in T. The disadvantage however is that they are in general 'iteratedly exponentially longer' than conventional proofs. So in general $I\Delta_0 + \mathrm{Exp}$ will not prove $\Box_T\varphi \rightarrow \Delta_T\varphi$. On the other hand $I\Delta_0 + \mathrm{SupExp}$ does prove this implication, i.e., $I\Delta_0 + \mathrm{SupExp}$ proves cut elimination.

Using the notation introduced in the previous paragraphs, we rephrase Friedman's characterization of interpretability: let U and V be finitely axiomatized sequential theories. Then

$$I\Delta_0 + \mathrm{Exp} \vdash U \rhd V \leftrightarrow \Delta_{Exp}(\mathrm{TabCon}(U) \rightarrow \mathrm{TabCon}(V)).$$

This result inspired Albert Visser to define an alternative semantics for *ILP*:

2.3 Definition. A *Friedman structure* is a tuple $\mathcal{F} = \langle W, b, P, Q\rangle$, where

(i) $W \neq \emptyset$;
(ii) $b \in W$ and for all $x \in W$, bQx;
(iii) $Q \subseteq W^2$ is transitive, irreflexive and upwards well-founded;
(iv) $P \subseteq Q$;
(v) for all x, y, z, if $xQyPz$ then xPz.

Moreover, we let $R := Q \circ P$, i.e., xRy iff $\exists z\, xQzPy$.

2.4 Definition. A *Friedman model* is a pair $\langle \mathcal{F}, \Vdash\rangle$ where $\mathcal{F}$ is a Friedman structure, and $\Vdash$ satisfies the usual clauses, with R as the accessibility relation for '$\Box$', and

$$x \Vdash A \rhd B \quad \Longleftrightarrow \quad \forall u\,(xQu \Rightarrow (\exists y\,(uPy \wedge y \Vdash A) \Rightarrow \exists z\,(uPz \wedge z \Vdash B))).$$

2.5 Theorem. *ILP is complete with respect to finite Friedman models.*

Proof. Visser [1990], Theorem 8.1. ∎

In fact, in his completeness proof Albert Visser shows that we can take Q to be a tree, and that we may assume P to be given in 'Carlson-style'. That is, we may assume that there is a set $X \subseteq W$ such that the root b is an element of X, and such that xPy iff xQy and $y \in X$. Moreover, the elements $x \in X$ have the additional property that xPy implies xRy.

3 Arithmetic Completeness: The Main Result

For the remainder of this chapter, let U be a Δ_2-sound finitely axiomatized sequential extension of $I\Delta_0 + \mathrm{SupExp}$.

3.1 Definition. An *interpretation* $(\cdot)^*$ of $\mathcal{L}(\Box, \rhd)$ in the language of arithmetic assigns to every $A \in \mathcal{L}(\Box, \rhd)$ a sentence A^* of the language of arithmetic, such that for all A, B, p

(i) if p is a proposition letter, then p^* is a sentence of the language of arithmetic;
(ii) $\top^*$ is '$0 = 0$'; $\bot^*$ is '$0 = 1$';
(iii) $(A \circ B)^*$ is $A^* \circ B^*$, for $\circ \in \{\wedge, \vee, \rightarrow\}$;
(iv) $(\neg A)^*$ is $\neg(A^*)$;
(v) $(\Box A)^*$ is $\Box_U(A^*)$;
(vi) $(A \rhd B)^*$ is $\triangle_{Exp}(\nabla_U(A^*) \rightarrow \nabla_U(B^*))$.

3.2 Theorem. *Let $A \in \mathcal{L}(\Box, \rhd)$. Then the following are equivalent:*

(i) $ILP \vdash \Big(\bigwedge_{\Box B \in \mathrm{Sub}(A)}(\Box B \rightarrow B) \wedge \bigwedge_{C \rhd D \in \mathrm{Sub}(A)}(C \rightarrow \Diamond C)\Big) \rightarrow A$
(ii) $ILP^{\omega} \vdash A$
(iii) *A^* is true for all interpretations $(\cdot)^*$.*

The equivalence of (i) and (ii) is Proposition 1.8.(1). The implication (ii) $\Rightarrow$ (iii) is clear. So we need only prove (iii) $\Rightarrow$ (i) to complete the proof of the Theorem. This is by contraposition.

If (i) does not hold, then, by the modal completeness of *ILP* with respect to finite Friedman models, there is such a model $\mathcal{M}_0 = \langle \{2, \ldots, n\}, 2, P_0, Q_0, \Vdash_0 \rangle$ with

$$2 \Vdash_0 \Big(\bigwedge_{\Box B \in \mathrm{Sub}(A)} (\Box B \rightarrow B) \wedge \bigwedge_{C \rhd D \in \mathrm{Sub}(A)} (C \rightarrow \Diamond C)\Big) \text{ and } 2 \nVdash_0 A.$$

The fact that $2 \Vdash_0 (\bigwedge_{\Box B \in \mathrm{Sub}(A)}(\Box B \rightarrow B) \wedge \bigwedge_{C \rhd D \in \mathrm{Sub}(A)}(C \rightarrow \Diamond C))$ means of course that $\mathcal{M}_0$ is A-sound.

As we pointed out in Section 2, we may assume that Q_0 is a tree, and that P_0 is given in 'Carlson-style' by a set X_0: xP_0y iff xQ_0y and $y \in X_0$, and $2 \in X_0$. $\mathcal{M}$ is the result of appending a *double* new root to $\mathcal{M}_0$. That is, $\mathcal{M} = \langle \{0, 1, 2 \ldots, n\}, 0, P, Q, \Vdash \rangle$, where

- xQy iff $\begin{cases} x = 0 \text{ and } y \neq 0, & \text{or} \\ x = 1 \text{ and } y > 1, & \text{or} \\ xQ_0y. \end{cases}$
- xPy iff xQy and $y \in X := X_0 \cup \{0\}$
- $x \Vdash p$ iff $\begin{cases} x > 1 \text{ and } x \Vdash_0 p, & \text{or} \\ x = 0 \text{ and } 2 \Vdash_0 p. \end{cases}$

3.3 Proposition. *Let $B \in \mathrm{Sub}(A)$. Then $2 \Vdash_0 B$ iff $2 \Vdash B$ iff $0 \Vdash B$.*

Proof. The first equivalence is trivial; the proof of the second one is similar to the proof of Proposition 1.3, and uses the A-soundness of $\mathcal{M}_0$. ∎

Next we define by the Recursion Theorem:

$$\begin{aligned} H(0) &= 0 \\ H(x+1) &= \begin{cases} y, & \text{if } H(x)Py \text{ and } \mathrm{TabProof}_U(x, L \neq y), \\ y, & \text{if } H(x)Qy \text{ and } \mathrm{TabProof}_{Exp}(x, L \neq y), \\ H(x), & \text{otherwise.} \end{cases} \\ L &:= \text{ the unique } x \text{ such that } \exists y \, \forall z > y \, (H(z) = x). \end{aligned}$$

3.4 Proposition. *The formula '$H(x) = u$' is $\Delta_0(2^x)$.*

3.5 Proposition. *Let $0 \leq x, y \leq n$. Then*

(i) $\mathrm{I}\Delta_0 + \mathrm{Exp} \vdash x < y \to H(x)\underline{Q}H(y)$;
(ii) $\mathrm{I}\Delta_0 + \mathrm{Exp} \vdash$ 'L exists';
(iii) $\mathrm{I}\Delta_0 + \mathrm{Exp} \vdash L = x \leftrightarrow \exists y \, (H(y) = x) \wedge \forall uv \, (H(u) = x \wedge v > u \to H(v) = x)$.

Proof. This is part of the proof of Theorem 8.2 in Visser [1990]. ∎

Define, for proposition letters $p \in \mathrm{Sub}(A)$,

$$p^* := \bigvee \{ L = \underline{i} \mid 0 \leq i \leq n \text{ and } i \Vdash p \},$$

and let p^* be arbitrary for $p \notin \mathrm{Sub}(A)$. Obviously, this completely determines $(\cdot)^*$.

3.6 Proposition. *Let $\psi \in \Pi^0_2$. Then* $\mathrm{I}\Delta_0 + \mathrm{Exp} \vdash \Delta_{Exp}\psi \to \Delta_U\psi$.

Proof. Visser [1990], Lemma 8.2. ∎

3.7 Proposition. $U \vdash L \in X$.

Proof. Reason in U: by Proposition 3.5.(ii) L exists. So assume $L = \underline{i} \notin X$. Then, by the definition of H, $i > 0$ and $\Delta_{Exp} L \neq \underline{i}$.

Now, step outside U for a minute and recall that U extends $\mathrm{I}\Delta_0 + \mathrm{SupExp}$. By Consequence 7.3.7 in Visser [1990], $\mathrm{I}\Delta_0 + \mathrm{SupExp}$ proves Π^0_2-reflection for $\mathrm{I}\Delta_0 + \mathrm{Exp}$. Therefore, back inside U we have $L \neq \underline{i}$ — a contradiction. ∎

3.8 Proposition. *Let $0 \leq j \leq n$. Then*

(i) $\mathrm{I}\Delta_0 + \mathrm{Exp} \vdash L = \underline{0} \wedge \underline{0}P\underline{j} \to \nabla_U L = \underline{j}$;
(ii) $\mathrm{I}\Delta_0 + \mathrm{Exp} \vdash L = \underline{0} \wedge \underline{0}Q\underline{j} \to \nabla_{Exp} L = \underline{j}$.

Proof. (i) Reason inside $\mathrm{I}\Delta_0 + \mathrm{Exp}$. Assume $L = \underline{0} \wedge \underline{0}P\underline{j} \wedge \Delta_U L \neq \underline{j}$. $L = \underline{0}$ implies $\forall z \geq 0 \, (H(z) = 0)$, by Proposition 3.5.(iii). So if $\mathrm{TabProof}_U(u, L \neq \underline{j})$ then $H(u) = \underline{0}P\underline{j}$, and hence $H(u+1) = \underline{i}$ — a contradiction.
(ii) Similarly. ∎

3.9 Proposition. *Let $0 < i \leq n$, and $B \in \mathrm{Sub}(A)$. Then*

(i) $i \Vdash B \Rightarrow \mathrm{I}\Delta_0 + \mathrm{Exp} \vdash L = \underline{i} \to B^*$;
(ii) $i \nVdash B \Rightarrow \mathrm{I}\Delta_0 + \mathrm{Exp} \vdash L = \underline{i} \to \neg B^*$.

Proof. This is part of the proof of Theorem 8.2 in Visser [1990]. Let us only note that in the course of his proof the author proves

Claim 1. if for all j with iPj, $j \Vdash C$, then $\mathrm{I}\Delta_0 + \mathrm{Exp} \vdash L = \underline{i} \to \Delta_U C^*$;
Claim 2. if for some j with iPj, $j \nVdash C$, then $\mathrm{I}\Delta_0 + \mathrm{Exp} \vdash L = \underline{i} \to \neg\Delta_U C^*$. ∎

3.10 Proposition. *Let $B \in \mathrm{Sub}(A)$. Then*

(i) $0 \Vdash B \Rightarrow \mathrm{I}\Delta_0 + \mathrm{Exp} \vdash L = \underline{0} \rightarrow B^*$;
(ii) $0 \nVdash B \Rightarrow \mathrm{I}\Delta_0 + \mathrm{Exp} \vdash L = \underline{0} \rightarrow \neg B^*$.

Proof. Needless to say, this is by induction on B. The cases B is a proposition letter, $B \equiv \neg C$, $B \equiv C \wedge D$ are trivial. The case $B \equiv \Box C$ follows from the case $B \equiv C \rhd D$. Before proving this case, we assume that the induction hypothesis holds for C and prove two claims:

Claim 1. if for all j with $0Pj$, $j \Vdash C$, then $\mathrm{I}\Delta_0 + \mathrm{Exp} \vdash L = \underline{0} \rightarrow \Delta_U C^*$;
Claim 2. if for some j with $0Pj$, $j \nVdash C$, then $\mathrm{I}\Delta_0 + \mathrm{Exp} \vdash L = \underline{0} \rightarrow \neg\Delta_U C^*$.

Proof of Claim 1. By the preceding Proposition we have

$$\bigwedge_{0Pj} j \Vdash C \Rightarrow \bigwedge_{1Pj} \mathrm{I}\Delta_0 + \mathrm{Exp} \vdash L = \underline{j} \rightarrow C^*.$$

Moreover,

$$\begin{aligned} \bigwedge_{0Pj} j \Vdash C &\Rightarrow 2 \Vdash C, \\ &\Rightarrow 0 \Vdash C, \text{ by Proposition 3.3,} \\ &\Rightarrow \mathrm{I}\Delta_0 + \mathrm{Exp} \vdash L = \underline{0} \rightarrow C^*, \text{ by the induction hypothesis.} \end{aligned}$$

So

$$\begin{aligned} \bigwedge_{0Pj} j \Vdash C &\Rightarrow \bigwedge_{0\underline{P}j} \mathrm{I}\Delta_0 + \mathrm{Exp} \vdash L = \underline{j} \rightarrow C^*, \\ &\Rightarrow \mathrm{I}\Delta_0 + \mathrm{Exp} \vdash \bigvee_{0\underline{P}j} L = \underline{j} \rightarrow C^*, \\ &\Rightarrow \mathrm{I}\Delta_0 + \mathrm{Exp} \vdash \Delta_U \bigvee_{0\underline{P}j} L = \underline{j} \rightarrow \Delta_U C^*. \end{aligned} \qquad (\star)$$

Now, by Proposition 3.7 we have $U \vdash L \in X$, so $\mathrm{I}\Delta_0 + \mathrm{Exp} \vdash \Delta_U \bigvee_{0\underline{P}j} L = \underline{j}$. By $(\star)$ it follows that

$$\begin{aligned} &\mathrm{I}\Delta_0 + \mathrm{Exp} \vdash \Delta_U C^*, \text{and} \\ &\mathrm{I}\Delta_0 + \mathrm{Exp} \vdash L = \underline{0} \rightarrow \Delta_U C^*, \end{aligned}$$

as desired. ■*Claim 1*

Proof of Claim 2. Suppose $0Pj$, $j \nVdash C$. Then

$$\begin{aligned} \exists j\,(0Pj \wedge j \nVdash C) &\Rightarrow \exists j\,(0Pj \wedge \mathrm{I}\Delta_0 + \mathrm{Exp} \vdash L = \underline{j} \rightarrow \neg C^*), \text{ by Proposition 3.9} \\ &\Rightarrow \exists j\,(0Pj \wedge \mathrm{I}\Delta_0 + \mathrm{Exp} \vdash \Delta_U(L = \underline{j} \rightarrow \neg C^*)) \\ &\Rightarrow \exists j\,(0Pj \wedge \mathrm{I}\Delta_0 + \mathrm{Exp} \vdash \nabla_U L = \underline{j} \rightarrow \neg\Delta_U C^*). \end{aligned}$$

Now apply Proposition 3.8.(i):

$$\mathrm{I}\Delta_0 + \mathrm{Exp} \vdash L = \underline{0} \rightarrow \neg\Delta_U C^*.$$ ■*Claim 2*

Let us get back on track, and continue with the proof of the Proposition. Assume $B \equiv C \rhd D$ and $0 \Vdash B$. Then we have by Claims 1 and 2 of the preceding Proposition that for every j with $0Qj$,

$$\mathrm{I}\Delta_0 + \mathrm{Exp} \vdash L = \underline{j} \rightarrow (\nabla_U C^* \rightarrow \nabla_U D^*). \qquad (\star\star)$$

By Claims 1 and 2 of this Proposition, we also have

$$\mathrm{I}\Delta_0 + \mathrm{Exp} \vdash L = \underline{0} \rightarrow (\nabla_U C^* \rightarrow \nabla_U D^*), \qquad (\star\star\star)$$

because $0Py$ implies $0Ry$.

Furthermore,

$$\begin{aligned} I\Delta_0 + \mathrm{Exp} \vdash \text{`}L \text{ exists'} &\Rightarrow I\Delta_0 + \mathrm{Exp} \vdash \Delta_{Exp} \bigvee_{0Qj} L = \underline{j} \\ &\Rightarrow I\Delta_0 + \mathrm{Exp} \vdash \Delta_{Exp}(\nabla_U C^* \to \nabla_U D^*), \text{ by } (\star\star) \text{ and } (\star\star\star), \\ &\Rightarrow I\Delta_0 + \mathrm{Exp} \vdash B^* \\ &\Rightarrow I\Delta_0 + \mathrm{Exp} \vdash L = \underline{0} \to B^*. \end{aligned}$$

Finally, we assume that $B \equiv C \rhd D$ and $0 \not\Vdash B$. Then there must be some j with $0Qj$, and for some k with jPk, $k \Vdash C$, while for all k' with jPk', $k' \not\Vdash D$. Then by the Claims of the preceding Proposition, $I\Delta_0 + \mathrm{Exp} \vdash L = \underline{j} \to \nabla_U C^*$ and $I\Delta_0 + \mathrm{Exp} \vdash L = \underline{j} \to \Delta_U \neg D^*$. So

$$\begin{aligned} \exists j\,(0Qj \wedge I\Delta_0 + \mathrm{Exp} &\vdash \Delta_{Exp}(L = \underline{j} \to \neg(\nabla_U C^* \to \nabla_U D^*)) \\ &\vdash \Delta_{Exp}(\nabla_U C^* \to \nabla_U D^*) \to \Delta_{Exp} L \neq \underline{j} \\ &\vdash \nabla_{Exp} L = \underline{j} \to \neg\Delta_{Exp}(\nabla_U C^* \to \nabla_U D^*)). \end{aligned}$$

Now $I\Delta_0 + \mathrm{Exp} \vdash L = \underline{0} \to \nabla_{Exp} L = \underline{j}$, for j with $0Qj$, by Proposition 3.8.(ii). Therefore

$$I\Delta_0 + \mathrm{Exp} \vdash L = \underline{0} \to \neg\Delta_{Exp}(\nabla_U C^* \to \nabla_U D^*).$$ ■

One final Proposition before we can complete the proof of Theorem 3.2:

3.11 Proposition. *In the standard model $L = 0$.*

Proof. We first show that $I\Delta_0 + \mathrm{Exp} \vdash L = \underline{i} \wedge \underline{i} > \underline{0} \to \Delta_U L \neq \underline{i}$. Reason in $I\Delta_0 + \mathrm{Exp}$: assume $L = \underline{i}$ and $\underline{i} > \underline{0}$. By the Least Number Principle for Δ_0-formulas it follows that $\exists z\,(H(z+1) = \underline{i} \wedge H(z) \neq \underline{i})$. By the definition of H we have $\Delta_{Exp} L \neq \underline{i}$ or $\Delta_U L \neq \underline{i}$. By Proposition 3.6 it follows that $\Delta_U L \neq \underline{i}$.

By Proposition 3.5.(ii) L exists. If $j > 0$ then our remarks above yield

$$\begin{aligned} L = j &\Rightarrow U \vdash L \neq \underline{i} \\ &\Rightarrow L \neq j, \end{aligned}$$

by the Δ_2-soundness of U. This is a contradiction, so $L = 0$. ■

Proof of Theorem 3.2. This is almost immediate now:

$$\begin{aligned} 2 \not\Vdash_0 A &\Rightarrow 0 \not\Vdash A, \text{ by Proposition 3.3}, \\ &\Rightarrow I\Delta_0 + \mathrm{Exp} \vdash L = \underline{0} \to \neg A^*, \text{ by Proposition 3.10.(ii)} \end{aligned}$$

Since $I\Delta_0 + \mathrm{Exp}$ proves only true theorems and $L = \underline{0}$ is true by Proposition 3.11, it follows that $\neg A^*$ is true i.e., A^* is false. ■

4. Unary Interpretability Logic

In Smoryński [1989] the author asks what the logics of unary relative interpretability are. I.e., if $\mathbf{I}A$ abbreviates $\top \rhd A$ then what are the logics in $\mathcal{L}(\Box, \mathbf{I})$? We device systems *ils* in $\mathcal{L}(\Box, \mathbf{I})$ such that for all $A \in \mathcal{L}(\Box, \mathbf{I})$, $ILS \vdash A$ iff $ils \vdash A$ — where $S = \top$, P, M, W.

First we present our axiomatizations and describe some of their properties. After that, we prove each of these systems *ils* complete with respect to *ILS*-models. From this, we immediately obtain conservation and arithmetic completeness results. Finally we show the existence of unique and explicitly definable fixed points for these systems.

1 Introduction

In this section we introduce the various systems and state some of their properties. We start with some definitions.

1.1 Definition. (i) The unary interpretability logic *il* is obtained from the provability logic L by adding the axioms

- $(I1)$ $\mathbf{I}\Box\bot$,
- $(I2)$ $\mathbf{I}A \wedge \Box(A \rightarrow B) \rightarrow \mathbf{I}B$,
- $(I3)$ $\mathbf{I}(A \vee \Diamond A) \rightarrow \mathbf{I}A$,
- $(I4)$ $\mathbf{I}A \wedge \Diamond\top \rightarrow \Diamond A$.

(ii) Several axioms have special names:

- (p) $\mathbf{I}A \rightarrow \Box\mathbf{I}A$,
- (m) $\mathbf{I}A \rightarrow \mathbf{I}(A \wedge \Box\bot)$.

ilp denotes the system $il + p$ and *ilm* denotes the system $il + m$. For other axiom schemes X we will simply refer to $ILX \cap \mathcal{L}(\Box, \mathbf{I})$ as *ilx*.

One easily shows that for all $A \in \mathcal{L}(\Box, \mathbf{I})$ we have, if $ils \vdash A$ then $ILS \vdash A$, by proving that all *ils*-axioms are derivable in *ILS*:

1.2 Proposition. *For all $A \in \mathcal{L}(\Box, \mathbf{I})$, if $ils \vdash A$ then $ILS \vdash A$.*

Proof. First we show that $IL \vdash I1, I2, I3, I4$.

$I1$ Notice that by $J1$ and $J5$ we have $IL \vdash \Box\bot \rhd \Box\bot$ and $IL \vdash \Diamond\Box\bot \rhd \Box\bot$, so by $J3$ we get

$$IL \vdash \Box\bot \vee \Diamond\Box\bot \rhd \Box\bot. \qquad (\star)$$

Furthermore

$$\begin{aligned} IL \vdash &\Box(\top \rightarrow (\top \wedge \Box\bot) \vee \Diamond(\top \wedge \Box\bot)), \\ \Rightarrow\ & IL \vdash \Box(\top \rightarrow \Box\bot \vee \Diamond\Box\bot), \\ \Rightarrow\ & IL \vdash \top \rhd \Box\bot \vee \Diamond\Box\bot, \quad \text{by } J1, \\ \Rightarrow\ & IL \vdash \top \rhd \Box\bot, \quad \text{by } J2 \text{ and } (\star), \\ \Rightarrow\ & IL \vdash \mathbf{I}\Box\bot. \end{aligned}$$

I2 By *J1* we have

$$\begin{aligned} IL \vdash & \Box(A \to B) \to A \rhd B, \\ & \Rightarrow IL \vdash \top \rhd A \wedge \Box(A \to B) \to \top \rhd B, \quad \text{by } J2 \\ & \Rightarrow IL \vdash \mathbf{I}A \wedge \Box(A \to B) \to \mathbf{I}B. \end{aligned}$$

I3 As before we find

$$\begin{aligned} IL \vdash & A \vee \Diamond A \rhd A, \\ & \Rightarrow IL \vdash \top \rhd A \vee \Diamond A \to \top \rhd A, \quad \text{by } J2, \\ & \Rightarrow IL \vdash \mathbf{I}(A \vee \Diamond A) \to \mathbf{I}A. \end{aligned}$$

I4 By *J4* we have $IL \vdash \mathbf{I}A \wedge \Diamond\top \to \Diamond A$.

Let us check that $ILP \vdash p$ and $ILM \vdash m$. Clearly, we have $ILP \vdash p$. To prove $ILM \vdash m$, recall first that $ILM \vdash W$. It follows that $ILM \vdash \top \rhd A \to \top \rhd (A \wedge \Box\bot)$, i.e. $ILM \vdash m$.

Now, suppose that $A \in \mathcal{L}(\Box, \mathbf{I})$ and $il \vdash A$; since we have $IL \vdash I1, I2, I3, I4$, it follows that $IL \vdash A$. Similarly, if $ilp \vdash A$ then $ILP \vdash A$, and if $ilm \vdash A$ then $ILM \vdash A$. ∎

Before proceeding to prove the converse of the Proposition, let us try and discover some of the theorems and derived rules of *il*.

1.3 Proposition. (i) *If* $il \vdash A$, *then* $il \vdash \mathbf{I}A$; *so in particular,* $il \vdash \mathbf{I}\top$;
(ii) $il \vdash \Box A \to \mathbf{I}A$
(iii) $il \vdash \mathbf{I}A \to \mathbf{I}(A \wedge \Box\neg A)$
(iv) $ilp \vdash m$.

Proof. (i) Note

$$\begin{aligned} il \vdash A & \Rightarrow il \vdash \Box\bot \to A \\ & \Rightarrow il \vdash \Box(\Box\bot \to A), \text{ by } R2, \\ & \Rightarrow il \vdash \mathbf{I}\Box\bot \wedge \Box(\Box\bot \to A), \text{ by } I1 \\ & \Rightarrow il \vdash \mathbf{I}A, \text{ by } I4. \end{aligned}$$

(ii) This is immediate from (i) and *I2*.
(iii) Observe

$$\begin{aligned} il \vdash \mathbf{I}A & \to \mathbf{I}A \wedge \Box(A \to (A \wedge \Box\neg A) \vee \Diamond(A \wedge \Box\neg A)), \text{ since } L \subseteq il, \\ & \to \mathbf{I}((A \wedge \Box\neg A) \vee \Diamond(A \wedge \Box\neg A)), \text{ by } I2, \\ & \to \mathbf{I}(A \wedge \Box\neg A), \text{ by } I3. \end{aligned}$$

(iv) We have

$$\begin{aligned} ilp \vdash \mathbf{I}A & \to \Box\mathbf{I}A \\ & \to \Box(\Diamond\top \to \Diamond A), \text{ by } I4, \\ & \to \Box(\Box\neg A \to \Box\bot) \\ & \to \Box(A \wedge \Box\neg A \to A \wedge \Box\bot) \\ & \to \mathbf{I}(A \wedge \Box\neg A) \wedge \Box(A \wedge \Box\neg A \to A \wedge \Box\bot), \text{ by (iii)} \\ & \to \mathbf{I}(A \wedge \Box\bot), \text{ by } I2. \end{aligned}$$

∎

It is clear that by part (iv) of the Proposition $ilm \subseteq ilp$. In the sequel we will show that $ilp \not\subseteq ilm$. It will also appear that *ilw* can be taken to be *ilm*.

The fact that *il* is closed under the rule $\vdash A \Rightarrow \vdash \mathbf{I}A$ might lead one to expect that the $\mathbf{I}$-operator is just an ordinary unary modal operator, say like $\Box$, and that its semantics may be based on a binary relation. This is not the case, however.

To show this, let us restrict ourselves for a moment to the $\Box$-free fragment of $\mathcal{L}(\Box, \mathbf{I})$, and assume that $\langle W, T, \Vdash \rangle$, with $T \subseteq W^2$, is an appropriate model for that fragment, i.e., $x \Vdash \mathbf{I}A$ iff $y \Vdash A$, for all y with xTy. Then one easily verifies that $\langle W, T, \Vdash \rangle \models \mathbf{I}(A \to B) \to (\mathbf{I}A \to \mathbf{I}B)$. But, if *il* is supposed to axiomatize the provable $\mathcal{L}(\Box, \mathbf{I})$-formulas in *IL*, then we must have $il \nvdash \mathbf{I}(A \to B) \to (\mathbf{I}A \to \mathbf{I}B)$, since $IL \nvdash \mathbf{I}(A \to B) \to (\mathbf{I}A \to \mathbf{I}B)$. So the semantics of the $\mathbf{I}$-operator cannot be based on a binary operator.

Assuming that *il* does indeed axiomatize $\mathcal{L}(\Box, \mathbf{I}) \cap il$, we find that Proposition 1.2.7.(ii) implies $il \vdash \mathbf{I}A$ iff $il \vdash A \vee \Diamond A$. Moreover, $\vdash \mathbf{I}A \Rightarrow \vdash A$ is not a derived rule of *il*: we have $il \vdash \mathbf{I}\Box\bot$, but $il \nvdash \Box\bot$ because $IL \nvdash \Box\bot$.

2 Modal Completeness: Preliminaries

To prove the converse of Proposition 1.2 we show that *ils* is complete with respect to finite *ILS*-models. For $s = \top$ and $s = p$ we use the method employed in Chapter 2 to prove the completeness of *IL* and *ILP*. The completeness of *ilm* is established using the method of de Jongh and Veltman [1990].

2.1 Definition. Let Γ, Δ be two maximal *ils*-consistent sets. Then Δ is called a *C-critical* successor of Γ if

(i) $\Gamma \prec \Delta$;
(ii) $\mathbf{I}C \notin \Gamma$;
(iii) $\neg C, \Box\neg C \in \Delta$.

2.2 Definition. A set of formulas Φ is *adequate* if

(i) if $B \in \Phi$, and C is a subformula of B, then $C \in \Phi$,
(ii) if $B \in \Phi$, and B is no negation, than $\neg B \in \Phi$.

Let Φ be a closed set. Then we say that a formula $\Diamond B$ is *almost in* Φ, if $\Diamond B \in \Phi$ or if $\mathbf{I}B \in \Phi$ or $B \equiv \top$.

Assuming that negations $\neg A$ are written as $A \to \bot$, any non-empty adequate set contains $\bot$ and $\top$.

2.3 Proposition. $il \vdash \neg\mathbf{I}B \to \Diamond\top$.

In our construction of countermodels for non-derivable formulas, we need two Propositions. The first one is analogous to Proposition 2.1.3, the second one is analogous to Proposition 2.1.4. Both of them hold for each of the systems under consideration here. However, to prove the converse of Proposition 1.2 for *ilm* we will need a special version of the second Proposition. This version will be proved in Section 4.

2.4 Proposition. *Let Γ be a maximal ils-consistent set such that $\neg\mathbf{I}C \in \Gamma$. Then there is a maximal ils-consistent C-critical successor Δ of Γ.*

Proof. Let Δ be a maximal consistent extension of

$$\{ D, \Box D \mid \Box D \in \Gamma \} \cup \{ \neg C, \Box\neg C \} \cup \{ \Box\bot \}.$$

Notice that if such a Δ exists, it must be a C-critical successor of Γ. Since

$$\{\, D, \Box D \mid \Box D \in \Gamma \,\} \cup \{\, \Box\bot \,\} \subseteq \Delta$$

it is a successor of Γ. (Note that $\Box\bot \notin \Gamma$, since otherwise $\Box C$ and $\mathbf{I}C \in \Gamma$.) And because $\{\neg C, \Box\neg C\,\} \subseteq \Delta$ it is also C-critical.

We only have to prove $\{\, D \mid \Box D \in \Gamma \,\} \cup \{\, \neg C \,\} \cup \{\, \Box\bot \,\}$ consistent, since $\Box\bot$ implies $\Box E$ for all E. Now, suppose that this set is inconsistent. Then there are $D_1, \ldots, D_m$ such that $D_1, \ldots, D_m,\ \neg C,\ \Box\bot \vdash \bot$. Then

$$\begin{aligned} D_1, \ldots, D_m &\vdash \Box\bot \to C, \\ \Rightarrow \Box D_1, \ldots, \Box D_m &\vdash \Box(\Box\bot \to C), \text{ by } R2, \\ \Rightarrow \Box D_1, \ldots, \Box D_m &\vdash \mathbf{I}C, \text{ by } I1 \text{ and } I3, \end{aligned}$$

So $\Gamma \vdash \mathbf{I}C$. This contradicts the consistency of Γ. ■

2.5 Proposition. *Assume that* $\mathbf{I}C \in \Gamma$, *and that* Δ *is a maximal ils-consistent* E*-critical successor of* Γ. *Then there is a maximal ils-consistent* E*-critical successor* Δ' *of* Γ *such that* $C \in \Delta'$.

Proof. Assume that there is no such Δ', then there are $\Box D_1, \ldots, \Box D_n \in \Gamma$ such that

$$D_1, \ldots, D_n,\ \Box D_1, \ldots, \Box D_n,\ \neg E,\ \Box\neg E,\ C \vdash \bot,$$

so

$$\begin{aligned} D_1, \ldots, D_n,\ \Box D_1, \ldots, \Box D_n &\vdash C \to E \vee \Diamond E, \\ \Rightarrow \Box D_1, \ldots, \Box D_n &\vdash \Box(C \to E \vee \Diamond E), \quad \text{by } R2. \end{aligned}$$

Then $\Gamma \vdash \Box(C \to E \vee \Diamond E)$. Since $\mathbf{I}C \in \Gamma$, axiom $I2$ yields $\Gamma \vdash \mathbf{I}(E \vee \Diamond E)$, and by axiom $I3$ we also have $\Gamma \vdash \mathbf{I}E$. It follows that $\mathbf{I}E \in \Gamma$ — which contradicts the fact that $\mathbf{I}E \notin \Gamma$. ■

We are now ready to prove the converse of Proposition 1.2. As we pointed out before, if for some $A \in \mathcal{L}(\Box, \mathbf{I})$ we have *ils* $\nvdash A$, then to show that *ILS* $\nvdash A$, it suffices to produce an *ILS*-model refuting A — this will be done separately for each of the systems under consideration here.

3 The Modal Completeness of *il* and *ilp*

Let us proceed without delay. We use the notation from Chapter 2.

3.1 Definition. Define W_Γ to be a minimal set of pairs $\langle \Delta, \tau \rangle$ such that

- (i) $\langle \Gamma, \langle\langle\rangle\rangle \rangle \in W_\Gamma$;
- (ii) if $\langle \Delta, \tau \rangle \in W_\Gamma$, $\Diamond B \in \Delta$ is almost in Φ and $C \in \Phi$, and if there is a maximal *ils*-consistent C-critical successor Δ' of Δ with $B, \Box\neg B \in \Delta'$, then $\langle \Delta', \tau ^\frown \langle\langle B, C\rangle\rangle \rangle \in W_\Gamma$ for *one* such Δ'.

Define R on W_Γ by putting $\bar{w} R \bar{v}$ iff $(\bar{w})_1 \subset (\bar{v})_1$.

Recall the following key properties of $\langle W_\Gamma, R \rangle$ from Chapter 2:

3.2 Facts. (i) W_Γ is finite;
(ii) if $\langle \Delta, \tau \rangle \in W_\Gamma$ and E occurs as the second component in some pair in τ, then $\neg E \in \Delta$;
(iii) if $\bar{w}$, $\bar{v} \in W_\Gamma$ and $(\bar{w})_1 \subseteq (\bar{v})_1$, then $(\bar{w})_0 \preceq (\bar{v})_0$;
(iv) $\langle W_\Gamma, R \rangle$ is a tree.

3.3 Theorem. *Let $A \in \mathcal{L}(\Box, \mathbf{I})$. Then il $\vdash A$ iff for all finite IL-models $\mathcal{M}$ we have $\mathcal{M} \models A$.*

Proof. As usual proving soundness is left to the reader. To prove completeness, assume that *il* $\nvdash A$. We want to produce an **IL**-model refuting A. Let Φ be a *finite* adequate set containing $\neg A$, and let Γ be a maximal *il*-consistent set containing $\neg A$. Construct $\langle W_\Gamma, R\rangle$ as above — using infinite maximal *il*-consistent sets. Then R has all the desired properties. Define S on W by putting $\bar{v} S_{\bar{w}} \bar{u}$ iff

$$(\bar{v})_1 = (\bar{w})_1{}^\frown\langle\langle B, C\rangle\rangle{}^\frown\tau \text{ and } (\bar{u})_1 = (\bar{w})_1{}^\frown\langle\langle B', C\rangle\rangle{}^\frown\sigma, \text{ for some } B,\ B',\ C,\ \tau \text{ and } \sigma.$$

Then S has all the properties required. We complete the proof by putting $\bar{w} \Vdash p$ iff $p \in (\bar{w})_0$, and proving that for all $F \in \Phi$, $\bar{w} \in W$ we have $\bar{w} \Vdash F$ iff $F \in (\bar{w})_0$. The proof is by induction on F. We only consider the cases $F \equiv \Diamond B$ and $F \equiv \mathbf{I}C$.

If $F \equiv \Diamond B \in (\bar{w})_0$ we have to show that $\exists \bar{v}\, (\bar{w} R \bar{v} \wedge B \in (\bar{v})_0)$. Note first that $\Diamond B$ is almost in Φ, and that $\bot \in \Phi$. By a well-known argument there is a successor Δ' of $(\bar{w})_0$ with B, $\Box\neg B \in \Delta'$. Moreover, Δ' is a $\bot$-*critical* successor of $(\bar{w})_0$. For, $\Diamond B \in (\bar{w})_0$ implies $\Diamond\top \in (\bar{w})_0$, so $\mathbf{I}\bot \in (\bar{w})_0$ would imply $\Diamond\bot \in (\bar{w})_0$, by axiom *I4* — which is impossible; therefore, $\mathbf{I}\bot \notin (\bar{w})_0$. Furthermore, it is clear that $\neg\bot$, $\Box\neg\bot \in \Delta'$. Put $\bar{v} := \langle \Delta', (\bar{w})_1{}^\frown\langle\langle B, \bot\rangle\rangle\rangle$. Then we may assume that $\bar{v} \in W_\Gamma$. It is clear that $\bar{w} R \bar{v}$ and $B \in \Delta'$ as required.

If $F \equiv \Diamond B \notin (\bar{w})_0$ then $\Box\neg B \in (\bar{w})_0$, and we have to show that $\forall \bar{v}\, (\bar{w} R \bar{v} \to \neg B \in (\bar{v})_0)$. But this is obvious from the definitions.

Assume $\mathbf{I}C \notin (\bar{w})_0$. Then $\neg\mathbf{I}C \in (\bar{w})_0$, and by Proposition 2.3 $\Diamond\top \in (\bar{w})_0$. By the induction hypothesis we have to show that $\exists \bar{v}\, (\bar{w} R \bar{v} \wedge \forall \bar{u}\, (\bar{v} S_{\bar{w}} \bar{u} \to \neg C \in (\bar{u})_0))$. Apply Proposition 2.4, with $\Gamma = (\bar{w})_0$, to obtain a C-critical successor Δ of Γ, and define $\bar{v} := \langle \Delta, (\bar{w})_1{}^\frown\langle\langle \top, C\rangle\rangle\rangle$. Since $\Diamond\top \in (\bar{w})_0$ is almost in Φ, we may assume that $\bar{v} \in W_\Gamma$. Furthermore, if $\bar{u} \in \bar{v} S_{\bar{w}}$ then C occurs in $(\bar{u})_1$, hence $\neg C \in (\bar{u})_0$, by Fact 3.2.(ii).

Assume $\mathbf{I}C \in (\bar{w})_0$. By the induction hypothesis we have to show that $\forall \bar{v}\, (\bar{w} R \bar{v} \to \exists \bar{u}\, (\bar{v} S_{\bar{w}} \bar{u} \wedge C \in (\bar{u})_0))$. So let $\bar{v} \in \bar{w} R$. Then $(\bar{v})_0 \succ (\bar{w})_0$, so $\Diamond\top \in (\bar{w})_0$, and therefore $\Diamond C \in (\bar{w})_0$ by axiom *I4*. By construction $\bar{v}$ is E-critical for some $E \in \Phi$. Now, apply Proposition 2.5, with $\Gamma = (\bar{w})_0, \Delta = (\bar{v})_0$, to obtain an E-critical successor Δ' of Γ that contains C, $\Box\neg C$. Since $\Diamond C$ is almost in Φ, we may assume that $\bar{u} = \langle \Delta', (\bar{w})_1{}^\frown\langle\langle C, E\rangle\rangle\rangle \in W_\Gamma$. Clearly, $\bar{u}$ does the job. ∎

3.4 Proposition. *Let $A \in \mathcal{L}(\Box, \mathbf{I})$. Then il $\vdash A$ iff IL $\vdash A$.*

Let il^ω denote *il* without the rule *R*2, but with all instances of the schema $\Box A \to A$. Propositions 3.4 and 3.1.6 imply:

3.5 Proposition. *Let $A \in \mathcal{L}(\Box, \mathbf{I})$. Then the following are equivalent:*

(i) $il^\omega \vdash A$

(ii) $il \vdash \left(\bigwedge_{\Box B \in \mathrm{Sub}(A)}(\Box B \to B) \wedge \bigwedge_{\mathbf{I}C \in \mathrm{Sub}(A)} \Diamond\top\right) \to A.$

Let us not waste any time, and prove the modal completeness of *ilp* right away.

3.6 Theorem. *Let $A \in \mathcal{L}(\Box, \mathbf{I})$. Then ilp $\vdash A$ iff for all finite ILP-models $\mathcal{M}$ we have $\mathcal{M} \models A$.*

Proof. Soundness is immediate. To prove completeness, assume that *ILP* $\nvdash A$. Let Φ be a *finite* adequate set containing $\neg A$, and let Γ be a maximal *ILP*-consistent set with $\neg A \in \Gamma$. Construct $\langle W_\Gamma, R\rangle$ as above — this time using infinite maximal *ILP*-consistent sets. Then R has all the properties required. Moreover, every $\bar{w} \in W_\Gamma$ differing from $\langle \Gamma, \langle\langle\rangle\rangle\rangle$ has exactly one immediate predecessor. S is defined on W_Γ by putting $\bar{v} S_{\bar{w}} \bar{u}$ iff

$$(\bar{v})_1 = (\bar{w})_1{}^\frown\tau{}^\frown\langle\langle B, C\rangle\rangle \text{ and } (\bar{u})_1 = (\bar{w})_1{}^\frown\tau{}^\frown\langle\langle B', C\rangle\rangle{}^\frown\sigma, \text{ for some } B,\ B',\ C,\ \tau \text{ and } \sigma.$$

One easily verifies that S has all the required properties. We complete the proof by putting $\bar{w} \Vdash p$ iff $p \in (\bar{w})_0$, and by proving that for all $F \in \Phi$, $\bar{w} \in W_\Gamma$, we have $\bar{w} \Vdash F$ iff $F \in (\bar{w})_0$. Once again the proof is by induction on F. The case $F \equiv \Diamond B$ is entirely analogous to the corresponding case in the completeness proof for *il*. So we only consider the case $F \equiv \mathbf{I}C$.

Assume $\mathbf{I}C \notin (\bar{w})_0$. Then we may copy the proof for the corresponding case in the completeness proof for *il*.

Assume $\mathbf{I}C \in (\bar{w})_0$. By the induction hypothesis we have to show that $\forall \bar{v}\,(\bar{w}R\bar{v} \rightarrow \exists \bar{u}\,(\bar{v}S_{\bar{w}}\bar{u} \wedge C \in (\bar{u})_0))$. So assume $\bar{v} \in \bar{w}R$. Since $\langle W_\Gamma, R\rangle$ is a tree, we can find a unique immediate predecessor $\bar{w}'$ of $\bar{v}$. Then $\mathbf{I}C \in (\bar{w}')_0$ by axiom p. Obviously, $\Diamond\top \in (\bar{w}')_0$ and therefore $\Diamond C \in (\bar{w}')_0$, by axiom *I4*. By construction $(\bar{v})_0$ is an E-critical successor of $(\bar{w}')_0$ for some $E \in \Phi$, and $(\bar{v})_1 = (\bar{w}')_1{}^\frown\langle\langle D, E\rangle\rangle$ for some $D \in \Phi$. Now Proposition 2.5 yields an E-critical successor Δ' of $(\bar{w})_0$ with $C, \Box\neg C \in \Delta'$. Since $\Diamond C$ is almost in Φ we may assume that $\bar{u} := \langle \Delta', (\bar{w}')_1{}^\frown\langle\langle C, E\rangle\rangle\rangle \in W_\Gamma$. Clearly, $\bar{v}S_{\bar{w}}\bar{u}$ and $C \in (\bar{u})_0$ as required. ■

3.7 Proposition. *Let $A \in \mathcal{L}(\Box, \mathbf{I})$. Then $ilp \vdash A$ iff $ILP \vdash A$.*

Let U be a Σ_1-sound finitely axiomatized sequential theory that extends $\mathrm{I}\Delta_0 + \mathrm{SupExp}$, and let $(\cdot)^*$ range over the kind of arithmetic interpretations described in Chapter 3. Then:

3.8 Proposition. *Let $A \in \mathcal{L}(\Box, \mathbf{I})$. Then $ilp \vdash A$ iff for all $(\cdot)^*$, $U \vdash A^*$.*

Proof. By Visser [1990] we have for any $A \in \mathcal{L}(\Box, \rhd)$, $ILP \vdash A$ iff for all $(\cdot)^*$, $U \vdash A^*$. So, in particular, for $A \in \mathcal{L}(\Box, \mathbf{I})$ we have $ILP \vdash A$ iff for all $(\cdot)^*$, $U \vdash A^*$. The result now follows from Proposition 3.7. ■

Let ilp^ω denote *ilp* without the rule *R2*, but with all instances of the schema $\Box A \rightarrow A$. Propositions 3.7 and 3.1.8 imply:

3.9 Proposition. *Let $A \in \mathcal{L}(\Box, \mathbf{I})$. Then the following are equivalent:*

(i) $ilp^\omega \vdash A$
(ii) $ilp \vdash \left(\bigwedge_{\Box B \in \mathrm{Sub}(A)}(\Box B \rightarrow B) \wedge \bigwedge_{\mathbf{I}C \in \mathrm{Sub}(A)} \Diamond\top\right) \rightarrow A.$

Let U be a Δ_2-sound finitely axiomatized theory that extends $\mathrm{I}\Delta_0 + \mathrm{SupExp}$, and let $(\cdot)^*$ be as in Proposition 3.8. Then:

3.10 Proposition. *Let $A \in \mathcal{L}(\Box, \mathbf{I})$. Then $ilp^\omega \vdash A$ iff A^* is true for all $(\cdot)^*$.*

Proof. Combine the preceding Proposition with Proposition 3.8 and Theorem 3.3.2. ■

4 The Modal Completeness of *ilm* and *ilw*

The most elegant proof for the modal completeness of *ilm* we have been able to find is rather similar to the completeness proof of *ILM* as given in de Jongh and Veltman [1990]. In our construction of a model we can use sets and sequences of formulas in stead of the sets and sequences of *equivalence classes* of formulas employed in the proof in that paper.

4.1 Definition. A set of formulas Φ is called an *ilm-adequate* set if

(i) Φ is closed under single negations;
(ii) Φ is closed under subformulas;
(iii) $\mathbf{I}\Box\bot, \mathbf{I}\bot, \mathbf{I}\top, \Box\top \in \Phi$;

(iv) if $\mathbf{I}A \in \Phi$ then $\Diamond A \in \Phi$.

4.2 Definition. Let Γ, Δ be maximal *ilm*-consistent subsets of some given adequate set Φ. Let $\mathbf{I}C \in \Phi$. Then Δ is called a *C-critical successor* of Γ if

(i) $\Gamma \prec \Delta$;
(ii) $\mathbf{I}C \notin \Gamma$;
(iii) $\neg C, \Box\neg C \in \Delta$.

Note that successors of C-critical successors of Γ are again C-critical successors of Γ.
We restate Proposition 2.4:

4.3 Proposition. *Let Γ be a maximal ilm-consistent set in Φ such that $\neg\mathbf{I}C \in \Gamma$. Then there is a C-critical successor Δ of Γ with Δ maximal consistent in Φ.*

Proof. Since the adequacy conditions ensure that the relevant formulas are in Φ, we can copy the proof of Proposition 2.4. ∎

We need a special version of Proposition 2.5:

4.4 Proposition. *Assume that $\mathbf{I}C \in \Gamma$, and that Δ is a maximal ilm-consistent E-critical successor of Γ. Then there is a maximal ilm-consistent E-critical successor Δ' of Γ such that $C \in \Delta'$ and such that for all $\Box B \in \Phi$, if $\Box B \in \Delta$ then $\Box B \in \Delta'$.*

Proof. Assume that no such Δ' exists. Then there are $\Box D_1, \ldots, \Box D_n \in \Gamma$, $\Box B_1, \ldots, \Box B_m \in \Delta$ such that

$$D_1, \ldots, D_n,\ \Box D_1, \ldots, \Box D_n,\ \Box B_1, \ldots, \Box B_m,\ \neg E,\ \Box\neg E,\ C \vdash \bot,$$

so

$$D_1, \ldots, D_n,\ \Box D_1, \ldots, \Box D_n \vdash C \wedge \Box(B_1 \wedge \ldots \wedge B_m) \rightarrow E \vee \Diamond E,$$
$$\Rightarrow \Box D_1, \ldots, \Box D_n \vdash \Box(C \wedge \Box(B_1 \wedge \ldots \wedge B_m) \rightarrow E \vee \Diamond E).$$

Then $\Gamma \vdash \Box(C \wedge \Box(B_1 \wedge \ldots \wedge B_m) \rightarrow E \vee \Diamond E)$. Now, $\Gamma \vdash \mathbf{I}C$, so by axiom m also $\Gamma \vdash \mathbf{I}(C \wedge \Box(B_1 \wedge \ldots \wedge B_m))$. An application of axiom $I2$ now yields $\Gamma \vdash \mathbf{I}(E \vee \Diamond E)$, from which it follows by $I3$ that $\Gamma \vdash \mathbf{I}E$. By the adequacy conditions this implies $\mathbf{I}E \in \Gamma$ — this, however, contradicts the consistency of Γ, because $\mathbf{I}E \notin \Gamma$ by the fact that Δ is an E-critical successor of Γ. ∎

Let Γ be a given maximal *ilm*-consistent set. The E-critical successors Δ' produced by the previous Proposition will be called *special E-critical successors* of Γ.

4.5 Definition. Let Φ be an (*ilm*-)adequate set, and let Γ be a maximal *ilm*-consistent subset of Φ. Then Γ is said to have *depth n* if the maximal length of a complete chain $\Gamma = \Gamma_0 \prec \ldots \prec \Gamma_m$ in Φ is $n+1$ — where all Γ_is $(0 \leq i \leq m)$ are maximal *ilm*-consistent subsets of Φ.

4.6 Theorem. *Let $A \in \mathcal{L}(\Box, \mathbf{I})$. Then ilm $\vdash A$ iff for all finite ILM-models $\mathcal{M}$ we have $\mathcal{M} \models A$.*

Proof. Proving soundness is left to the reader. To prove completeness, assume that *ilm* $\nvdash A$. We construct a finite *ILM*-model refuting A.

Let Φ be a finite adequate set containing $\neg A$, and let Γ be a maximal *ilm*-consistent subset of Φ that contains $\neg A$. Define W to be the smallest set of pairs $\langle \Delta, \tau \rangle$ such that

(i) $\Gamma = \Delta$ or $\Gamma \prec \Delta$;

(ii) τ is a finite sequence of formulas from Φ, the length of which does not exceed the depth of Γ minus the depth of Δ.

So W is finite. Unfortunately, the sequence τ no longer provides sufficient information on the 'C-critical' status of Δ. Define R on W by putting $\bar{w}R\bar{v}$ iff $(\bar{w})_0 \prec (\bar{v})_0$ and $(\bar{w})_1 \subseteq (\bar{v})_1$. Then R is transitive and conversely well-founded. Before defining S we need a definition: let $\bar{w}, \bar{v} \in W$, then $\bar{v}$ is called a *C-critical R-successor* of $\bar{w}$ if

(i) $(\bar{v})_0$ is a (plain) C-critical successor of $(\bar{w})_0$;
(ii) $(\bar{v})_1 = (\bar{w})_1 {}^\frown \langle C \rangle {}^\frown \tau$, for some τ.

We can now define S on W. Put $\bar{v}S_{\bar{w}}\bar{u}$ iff

(1) $\bar{v}, \bar{u} \in \bar{w}R$;
(2) $(\bar{u})_1 \subseteq (\bar{v})_1$;
(3) for every $B \in \Phi$, if $\Box B \in (\bar{v})_0$, then $\Box B \in (\bar{u})_0$;
(4) if $\bar{v}$ is a C-critical R-successor of $\bar{w}$, then so is $\bar{u}$.

This time it is less obvious than before that S has all the properties required, so we will spell out the details:

- $S_{\bar{w}} \subseteq \bar{w}R \times \bar{w}R$ is obvious;
- so are reflexivity and transitivity;
- if $\bar{v}, \bar{u} \in \bar{w}R$ and $\bar{v}R\bar{u}$ then we clearly have (1), (2) and (3). And since successors of C-critical successors are C-critical successors as well, we also have (4). So $\bar{v}S_{\bar{w}}\bar{u}$;
- assume that $\bar{v}S_{\bar{w}}\bar{u}R\bar{u}'$ — we have to show that $\bar{v}R\bar{u}'$ holds. By (2) we obviously have $(\bar{v})_1 \subseteq (\bar{u}')_1$, and by (3) and the fact that $(\bar{u})_0 \prec (\bar{u}')_0$ it follows that $(\bar{v})_0 \prec (\bar{u}')_0$. Therefore, $\bar{v}R\bar{u}'$.

Now, to complete the proof we define $\Vdash$ on W by putting $\bar{w} \Vdash p$ iff $p \in (\bar{w})_0$, and show that for all $F \in \Phi$ we have $\bar{w} \Vdash F$ iff $F \in (\bar{w})_0$. The only interesting case in the inductive proof is the case $F \equiv \mathbf{I}C$.

Assume that $\mathbf{I}C \notin (\bar{w})_0$. Then $\neg\mathbf{I}C \in (\bar{w})_0$. By the induction hypothesis we have to show that $\exists \bar{v}\,(\bar{w}R\bar{v} \wedge \forall \bar{u}\,(\bar{v}S_{\bar{w}}\bar{u} \to \neg C \in (\bar{u})_0))$. Apply Proposition 4.3 with $\Gamma = (\bar{w})_0$ to obtain a C-critical successor Δ of Γ. Define $\bar{v} := \big\langle \Delta, (\bar{w})_1 {}^\frown \langle C \rangle \big\rangle$. Then $\bar{v} \in W$. Let $\bar{u} \in \bar{v}S_{\bar{w}}$, then $\bar{u}$ is a C-critical R-successor of $\bar{w}$, because $\bar{v}$ is one. Therefore, $\neg C \in (\bar{u})_0$.

Assume $\mathbf{I}C \in (\bar{w})_0$. By the induction hypothesis we have to show that $\forall \bar{v}\,(\bar{w}R\bar{v} \to \exists \bar{u}\,(\bar{v}S_{\bar{w}}\bar{u} \wedge C \in (\bar{u})_0))$. Suppose first that $\bar{v}$ is an E-critical R-successor of $\bar{w}$. Apply Proposition 4.4 with $\Gamma = (\bar{w})_0$ and $\Delta = (\bar{v})_0$ to obtain a *special* E-critical successor Δ' of Γ that contains C. Put $\bar{u} := \langle \Delta', (\bar{v})_1 \rangle$. Now the depth of Δ' cannot be larger than the depth of Δ, because all formulas of the form $\Box B$ that are in Δ, are also in Δ'. So $\bar{u} \in W$. Clearly, $\bar{v}S_{\bar{w}}\bar{u}$ holds. If, on the other hand, $\bar{v} \in \bar{w}R$ is not an E-critical R-successor of $\bar{w}$, then $(\bar{w})_0 \prec (\bar{v})_0$ is all we know. But $(\bar{v})_0$ is a $\bot$-critical successor of $(\bar{w})_0$. For, then we have

- $(\bar{w})_0 \prec (\bar{v})_0$;
- if $\Diamond\top \notin (\bar{w})_0$, then $\Box\bot \in (\bar{w})_0$ — which is impossible; so $\Diamond\top \in (\bar{w})_0$. Now, if $\mathbf{I}\bot \in (\bar{w})_0$, then axiom $I4$ implies that $\Diamond\bot \in (\bar{w})_0$, and this too is impossible, therefore we must have $\mathbf{I}\bot \notin (\bar{w})_0$;
- $\neg\bot, \Box\neg\bot \in (\bar{v})_0$ is immediate.

Applying Proposition 4.4 once again, with $\Gamma = (\bar{w})_0$, $\Delta = (\bar{v})_0$ and $E = \bot$, we get a *special* $\bot$-critical successor Δ' of Γ with $C \in \Delta'$. Finally, $\bar{u} := \langle \Delta', (\bar{v})_1 \rangle$ does the job. ∎

4.7 Proposition. *Let $A \in \mathcal{L}(\Box, \mathbf{I})$. Then $ilm \vdash A$ iff $ILM \vdash A$.*

Let U be an extension of $I\Sigma_1$ that is Σ_1-sound. Define:

$$\varphi \rhd_U^* \psi \iff \forall \chi \in \Pi_1^0 \, (\Box_{U+\varphi}\chi \to \Box_{U+\psi}\chi).$$

$(\cdot)^*$ is called a U–ΠCons interpretation of $\mathcal{L}(\Box, \rhd)$ if

(i) $(\cdot)^*$ maps proposition letters to sentences of the language of U;
(ii) $(\cdot)^*$ commutes with the connectives;
(iii) $(\Box A)^*$ is $\Box_U(A^*)$;
(iv) $(A \rhd B)^*$ is $(A^*) \rhd_U^* (B^*)$.

4.8 Proposition. *Let $A \in \mathcal{L}(\Box, \mathbf{I})$. Then ilm $\vdash A$ iff for all U–*ΠCons *interpretations* $(\cdot)^*$, $U \vdash A^*$.

Proof. This follows from the corresponding result for $A \in \mathcal{L}(\Box, \rhd)$ and *ILM* (cf. Montagna & Hájek [1989] and Visser [1989], Theorem 10.1), together with Proposition 4.7. ■

Let ilm^ω denote *ilm* without the rule $R2$, but with all instances of the schema $\Box A \to A$. Propositions 4.7 and 3.1.8 imply:

4.9 Proposition. *Let $A \in \mathcal{L}(\Box, \mathbf{I})$. Then the following are equivalent:*

(i) $ilm^\omega \vdash A$
(ii) $ilm \vdash \left(\bigwedge_{\Box B \in \mathrm{Sub}(A)}(\Box B \to B) \wedge \bigwedge_{\mathbf{I}C \in \mathrm{Sub}(A)} \Diamond\top\right) \to A.$

Define a formula $A \in \mathcal{L}(\Box, \rhd)$ to be ***ω-valid*** if A^* is true in the standard model for all 'normal' arithmetic interpretations $(\cdot)^*$ with $(\Box A)^*$ is (a formalization of) '$PA \vdash A^*$', and $(A \rhd B)^*$ is (a formalization of) '$PA + A$ interprets $PA + B$'.

4.10 Proposition. *ilm^ω precisely axiomatizes all ω-valid formulas $A \in \mathcal{L}(\Box, \mathbf{I})$.*

Proof. This follows from the corresponding result for $A \in \mathcal{L}(\Box, \rhd)$ and ILM^ω (cf. Berarducci [1989]), together with the preceding Proposition and Proposition 4.7. ■

We end this section with some remarks on the hierarchy of extensions of *il*.

4.11 Proposition. *('ilw = ilm') Let $A \in \mathcal{L}(\Box, \mathbf{I})$. Then ilm $\vdash A$ iff ILW $\vdash A$.*

Proof. Since m is a substitution instance of the axiom W, the direction from left to right is immediate. Conversely, if *ilm* $\nvdash A$ then *ILM* $\nvdash A$ by Proposition 4.7. Now, recall that *ILM* $\vdash W$ by Proposition 1.2, therefore *ILW* $\nvdash A$. ■

4.12 Proposition. *ilm is the unary interpretability logic of all reasonable arithmetics.*

Proof. Let *ul* be the unary interpretability logic of all reasonable arithmetics. By Visser [1988], Section 9, *ILW* is valid for arithmetic interpretations in all reasonable arithmetics, hence the same holds for *ilw* = *ilm*. It follows that *ilm* $\subseteq$ *ul*. On the other hand, in essentially reflexive theories U like *PA* binary interpretability and $\rhd_U^*$ are provably extensionally equal (cf. Visser [1989], Section 10). Hence by Proposition 4.8 *ilm* is the unary interpretability logic of *PA*. It follows that *ul* must be contained in *ilm*. Therefore, *ul* = *ilm*, and *ilm* is indeed the unary interpretability logic of all reasonable arithmetics. ■

As another corollary to Proposition 4.11 we find that the (real) hierarchy

$$\begin{array}{ccccccccc} & & & & & & ILP & & \\ & & & & & \nearrow & & \searrow & \\ IL & \longrightarrow & ILF & \longrightarrow & ILW & \longrightarrow\ ILWM_0 & & & ILMP \\ & & & & & \searrow & & \nearrow & \\ & & & & & & ILM & & \end{array}$$

partly collapses when we only consider formulas $A \in \mathcal{L}(\Box, \mathbf{I})$:

$$il \longrightarrow ilf \longrightarrow ilw = ilwm_0 = ilm \longrightarrow ilp \;---\; ilmp.$$

There is no total collapse, however, since $ilm \neq ilp$, $ilf \neq ilw$ and $il \neq ilf$. We conjecture that $ilp = ilmp$.

4.13 Proposition. (i) $ilm \neq ilp$;
(ii) $ilf \neq ilw$;
(iii) $il \neq ilf$.

Proof. (i) It suffices to show that $ilm \nvdash \mathbf{I}C \to \Box\mathbf{I}C$. Consider Figure 1 below. We clearly have $w \Vdash \mathbf{I}p$, since every R-successor of w has an S_w-successor at which p holds. However, v does not force $\mathbf{I}p$, for it has an R-successor (u) that is not S_v-succeeded by a point at which p holds — so $w \Vdash \neg\Box\mathbf{I}p$.

(ii) It suffices to show that $ILF \nvdash m$. Consider Figure 2. We claim that $w \Vdash F$, i.e., $w \Vdash A \rhd \Diamond A \to \Box\neg A$ for all $A \in \mathcal{L}(\Box, \rhd)$. Suppose that $w \Vdash A \rhd \Diamond A$. Then

(i) if $b \Vdash A$ then $a \Vdash A$
(ii) $d \nVdash A$ — otherwise $d \Vdash \Diamond A$, which is impossible
(iii) for each B, $a \Vdash B \iff c \Vdash B$
(iv) $c \nVdash A$ — otherwise $c \Vdash \Diamond A$, which is impossible
(v) $a \Vdash A$, by (iii) and (iv)
(vi) $b \nVdash A$, by (v) and (i)
(vii) $w \Vdash \Box\neg A$, by (ii), (iv), (v) and (vi).

On the other hand $w \nVdash m$, for we have $w \Vdash \top \rhd p$, whereas $w \nVdash \top \rhd (p \wedge \Box\bot)$, since b has no S_w-successor at which $p \wedge \Box\bot$ holds.

(iii) We have $ilf \vdash \mathbf{I}\Diamond\top \to \Box\bot$ or, equivalently, $ilf \vdash \Diamond\top \to \neg\mathbf{I}\Diamond\top$, by axiom F. On the other hand, we have $il \nvdash \Diamond\top \to \neg\mathbf{I}\Diamond\top$, as is clear from the model in Figure 3. ■

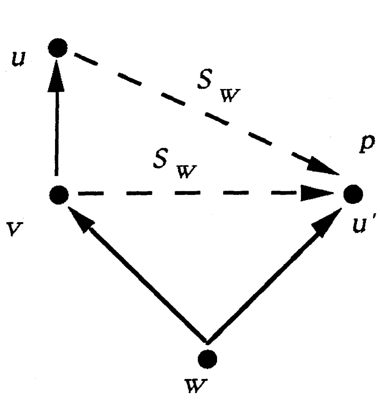


Figure 1.

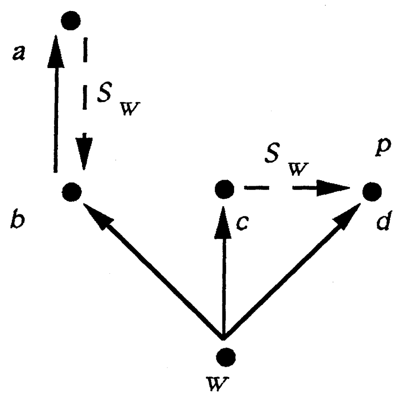


Figure 2.

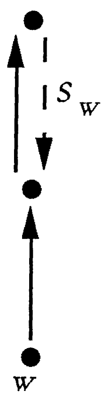


Figure 3.

(Plain arrows denote R-links. Reflexive S-links and S-links induced by R-links have been left out.)

Part (ii) of the Proposition may be used to answer a question of Dick de Jongh:

4.14 Proposition. *ILW is not conservative over ILF for formulas in one proposition letter.*

5 Unique and Explicit Fixed Points

We present a model-theoretic proof for the Explicit Definability of Fixed Points in *il*.

First we introduce some notation. We use $A(p)$ for a formula in which p possibly occurs; p is said to occur *modalized* in $A(p)$ if p occurs only in the scope of a $\Box$ or $\mathbf{I}$. $A(C)$ denotes the result of substituting C for p in $A(p)$. From the *il*-axioms we can deduce the following extensionality principle:

$$(E)\ \Box(A \leftrightarrow B) \rightarrow (\mathbf{I}A \leftrightarrow \mathbf{I}B).$$

Let sr_0 denote $R1$, $R2$, $L1$, $L2$, $L3$ and E. Then sr_0 proves

$$(S1)\ sr_0 \vdash B \leftrightarrow C \Rightarrow sr_0 \vdash A(B) \leftrightarrow A(C)$$
$$(S2)\ sr_0 \vdash \Box^+ (B \leftrightarrow C) \rightarrow (A(B) \leftrightarrow A(C))$$

If p is modalized in $A(p)$, then

$$(S3)\ sr_0 \vdash \Box(B \leftrightarrow C) \rightarrow (A(B) \leftrightarrow A(C))$$
$$(LR)\ sr_0 \vdash B \rightarrow (\Box A \rightarrow A) \Rightarrow sr_0 \vdash B \rightarrow A$$

5.1 Theorem. (Uniqueness Theorem) *Assume that p occurs modalized in A, then*

$$sr_0 \vdash (\Box^+ (p \leftrightarrow A(p)) \wedge \Box^+ (q \leftrightarrow A(q))) \rightarrow (p \leftrightarrow q).$$

Proof. By $S3$:

$$\vdash (\Box^+ (p \leftrightarrow A(p)) \wedge \Box^+ (q \leftrightarrow A(q))) \rightarrow (\Box(p \leftrightarrow q) \rightarrow (p \leftrightarrow q)),$$

so by LR

$$\vdash (\Box^+ (p \leftrightarrow A(p)) \wedge \Box^+ (q \leftrightarrow A(q))) \rightarrow (p \leftrightarrow q).$$ ∎

From the fact that *il* extends L it is obvious that to prove the existence of explicit fixed points in *il* it suffices to find a fixed point for $\mathbf{I}B(p)$, i.e., to find a formula C such that $il \vdash C \leftrightarrow \mathbf{I}B(C)$. After that we can proceed as in the standard proof for L — cf. Smoryński [1985a].

We need the following result:

5.2 Proposition. *Let $\mathcal{M}$ be an il-model, and $w \in \mathcal{M}$. Then we have $w \Vdash \mathbf{I}B(\Box\bot)$ iff $w \Vdash \mathbf{I}B(\mathbf{I}B(\Box\bot))$.*

Proof. We list some simple facts about arbitrary w. We write $u \Vdash_{max} A$ iff $u \Vdash A$ and for all $v \in uR$, $v \nVdash A$ and we write $w\underline{R}u$ for wRu or $w = u$.

(i) if $w \Vdash_{max} B(\Box\bot)$ then $w \Vdash \Box\neg B(\Box\bot)$;
(ii) if $w \Vdash \Box\neg B(\Box\bot)$ then, if $w\underline{R}u$, then $u \Vdash \Box\bot \iff u \Vdash \mathbf{I}B(\Box\bot)$;
(iii) if $w \Vdash_{max} B(\Box\bot)$, then, if $w\underline{R}u$, then $u \Vdash B(\Box\bot) \iff u \Vdash B(\mathbf{I}B(\Box\bot))$;
(iv) if $w \Vdash_{max} B(\Box\bot)$, then, if $w\underline{R}u$, then $u \Vdash_{max} B(\Box\bot) \iff u \Vdash_{max} B(\mathbf{I}B(\Box\bot))$;
(v) if $w \Vdash_{max} B(\Box\bot)$, then $w \Vdash_{max} B(\mathbf{I}B(\Box\bot))$;
(vi) if $w \Vdash_{max} B(\mathbf{I}B(\Box\bot))$, then $w \Vdash_{max} B(\Box\bot)$, for assume that $w \Vdash_{max} B(\mathbf{I}B(\Box\bot))$, then $w \Vdash \Box\neg B(\Box\bot)$ by (v), so by (ii) we have for all u with $w\underline{R}u$, $u \Vdash \mathbf{I}B(\Box\bot) \iff u \Vdash \Box\bot$, hence $w \Vdash_{max} B(\Box\bot)$.

Now, assume that $w \Vdash \mathbf{I}B(\Box\bot)$ and that wRv holds. We have to find a u such that vS_wu and $u \Vdash B(\mathbf{I}B(\Box\bot))$. wRv implies the existence of a u with vS_wu and $u \Vdash B(\Box\bot)$. Since vS_wuRu' implies vS_wu', we may safely assume that $u \Vdash_{max} B(\Box\bot)$. But then $u \Vdash B(\mathbf{I}B(\Box\bot))$, by (v).

Conversely, let $w \Vdash \mathbf{I}B(\mathbf{I}B(\Box\bot))$, and assume that wRv. We have to find a u with vS_wu and $u \Vdash B(\Box\bot)$. Since wRv holds we find a u such that vS_wu and $u \Vdash B(\mathbf{I}B(\Box\bot))$. Again, we may safely assume that $u \Vdash_{max} B(\mathbf{I}B(\Box\bot))$. Then (vi) yields $u \Vdash_{max} B(\Box\bot)$. ∎

By our previous remarks and the completeness results from Section 3 it follows that

5.3 Theorem. *For each $A(p, q_1, \ldots, q_n) \in \mathcal{L}(\Box, \mathbf{I})$ in which p occurs modalized there is a provably unique $B(q_1, \ldots, q_n) \in \mathcal{L}(\Box, \mathbf{I})$ such that il $\vdash B(q_1, \ldots, q_n) \leftrightarrow A(B(q_1, \ldots, q_n))$.*

5.4 Remark. Another way to prove the fixed point theorem for *il* would be to use the existing fixed point theorem for *IL* (cf. de Jongh and Visser [1989]) together with the conservation results obtained above. To see this we repeat the key lemma de Jongh and Visser use to obtain the explicit definability of fixed points in *IL* (without, however, explaining the notions not used in this thesis):

> Let U be any extension of SR_0 satisfying:
>
> > Every formula $A(p)$ of the form $\Box B(p)$ or $B(p)\#C(p)$ has a fixed point such that $A(p) \leq D$.
>
> Then:
>
> > For every formula $A(p)$ with p modalized, there is a formula D such that p does not occur in D, $A(p) \leq D$ and $U \vdash D \leftrightarrow A(D)$.

An inspection of the inductive proof of this lemma shows that if the formula $A(p)$ of the form $\Box B(p)$ or $\top \# C(p)$ has its fixed point in $\mathcal{L}(\Box, \mathbf{I})$, then so does every $A(p) \in \mathcal{L}(\Box, \mathbf{I})$ with p modalized. An appeal to our conservation results then completes the explicit definability theorem for *il*, *ilp* and *ilm*.

5. Internal Definitions

We study several axioms of the form $f(p,q) \leftrightarrow p \rhd q$, where $f(p,q) \in \mathcal{L}(\Box)$. Such a biconditional will be called a defining schema. Its left-hand side is called an *(internal) definition.*

After having presented some examples of definitions in the first section, we use several of these in Section 2 to find out more about the hierarchy of extensions of *IL*.

1 Introduction

Before considering some examples, we define a new notion. Every definition $f(p,q)$ of $p \rhd q$ gives rise to a translation $(\cdot)^f : \mathcal{L}(\Box, \rhd) \longrightarrow \mathcal{L}(\Box)$ in a canonical way:

$$\begin{aligned} (p)^f &\equiv p, \quad \text{for proposition letters } p, \\ (A \circ B)^f &\equiv A^f \circ B^f, \quad \circ = \vee, \wedge, \rightarrow, \\ (\neg A)^f &\equiv \neg(A^f), \\ (\Box A)^f &\equiv \Box(A^f), \\ (A \rhd B)^f &\equiv f(A^f, B^f). \end{aligned}$$

1.1 Definition. Let $f(p,q)$ be a definition, and $(\cdot)^f$ its associated translation. Then $f(p,q)$ is called a *good* definition for a formula A if $L \vdash A^f$; $f(p,q)$ is called a good definition for a set of formulas X, if it is good for all $A \in X$.

Definitions for IL

Here is a minimal good definition for *IL*:

1.2 Proposition. *$\Diamond p \rightarrow \Diamond q$ is a good definition for IL; it is implied by every good definition for IL.*

Proof. The first part is left to the reader; the second part follows from axiom $J4$. ∎

Some variations on $\Diamond p \rightarrow \Diamond q$ also yield good definitions for *IL*. Let I be a finite subset of $\omega \setminus \{0\}$. Then ZI denotes the definition:

$$(\Diamond p \rightarrow \Diamond q) \wedge \bigwedge_{i \in I} (\Diamond^i p \rightarrow \Diamond^i q).$$

It is a simple exercise to check that each ZI is a good definition for *IL*. Let $J \subseteq ((\omega \setminus \{0\}) \times \omega)$ be finite. Then

$$(\Diamond p \rightarrow \Diamond q) \wedge \bigwedge_{\langle i,j \rangle \in J} (\Diamond^i (p \wedge \Box^j \bot) \rightarrow \Diamond^i (q \wedge \Box^j \bot))$$

is a good definition for *IL*, as the reader may verify.

Let $MAX \equiv \Box(p \rightarrow q \vee \Diamond q)$. Since it is not entirely trivial to check that MAX is good for *IL*, we will supply some details:

1.3 Proposition. *MAX is a good definition for IL.*

Proof. Let $(\cdot)^m$ denote the translation associated with MAX. We only show that $L \vdash (J2)^m$, i.e.,

$$L \vdash \Box(A^m \to B^m \vee \Diamond B^m) \wedge \Box(B^m \to C^m \vee \Diamond C^m) \to \Box(A^m \to C^m \vee \Diamond C^m).$$

Omitting the superscripts we have:

$$\begin{aligned} L \vdash \Box(A \to B \vee \Diamond B) \wedge \Box(B \to C \vee \Diamond C) &\to \Box(\Diamond B \to \Diamond(C \vee \Diamond C)) \\ &\to \Box(\Diamond B \to \Diamond C \vee \Diamond\Diamond C) \\ &\to \Box(\Diamond B \to \Diamond C), \text{ by } L2 \\ &\to \Box(\Diamond B \to C \vee \Diamond C) \\ &\to \Box(B \vee \Diamond B \to C \vee \Diamond C) \\ &\to \Box(A \to C \vee \Diamond C). \end{aligned}$$

■

1.4 Proposition. *Let $f(p,q)$ be a good definition for IL. Then $L \vdash \Box(A \to B \vee \Diamond B) \to f(A,B)$.*

Proof. By $J1$ we have

$$\begin{aligned} L \vdash \Box(A \to B \vee \Diamond B) &\to f(A, B \vee \Diamond B) \\ &\to f(A,B), \text{ by } J5 \text{ and } J3. \end{aligned}$$

■

By Proposition 1.2 it follows that we have, for any good definition $f(p,q)$ for IL, that $L \vdash \Box(A \to B \vee \Diamond B) \to f(A,B)$ and $L \vdash f(A,B) \to (\Diamond A \to \Diamond B)$. For definitions of the form $\Box f(p,q)$ this may be strengthened to:

1.5 Proposition. *Let $\Box f(p,q)$ be a good definition for IL. Then*

$$L \vdash \Box f(A,B) \leftrightarrow \Box(A \to B \vee \Diamond B).$$

Proof. We only have to prove the implication from left to right. This may done using Smoryński's trick. Assume that the implication is not provable, then we find an L-model $\mathcal{M}_1$ with root w_1 such that $w_1 \nVdash \Box f(A,B) \to \Box(A \to B \vee \Diamond B)$. Then there must be a w_2 with $w_1 R w_2$ and $w_2 \Vdash f(A,B) \wedge \Box f(A,B) \wedge A \wedge \neg B \wedge \neg\Box B$. Let $\mathcal{M}_2$ be the submodel generated by w_2, and let $\mathcal{M}_3$ be the result of appending a new root w_3 to $\mathcal{M}_2$ (cf. Sections 1.2 and 3.1). Then $w_3 \Vdash \Box f(A,B)$, but $w_3 \nVdash (\Diamond A \to \Diamond B)$ — which contradicts the assumption that $f(p,q)$ is good for IL. ■

Definitions for ILP

It is left to the reader to check that neither $(\Diamond p \to \Diamond q)$ nor any of the ZI is good for ILP. On the other hand, axiom $L2$ implies that MAX is a good definition for ILP.

Smoryńsky [1989] introduces the schema $\Box(\Diamond p \to \Diamond q) \leftrightarrow p \rhd q$. Let $PRC \equiv \Box(\Diamond p \to \Diamond q)$. (Here PRC stands for 'Provable Relative Consistency'.) We leave it to the reader to check that PRC is good for $J1$, $J2$, $J3$, $J5$ and P, but not for $J4$.

There is an obvious way to remedy this defect: simply add $\Diamond p \to \Diamond q$ to PRC. By checking all axioms, one easily verifies that the resulting definition, PRC^+, is good for ILP. One can generalize this procedure to the following Proposition:

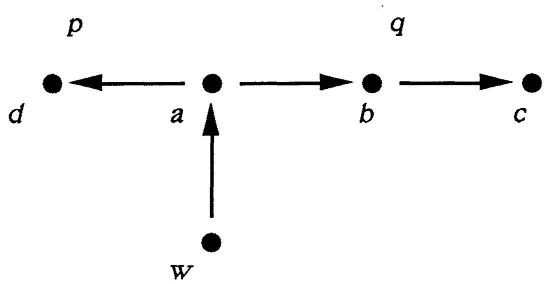
1.6 Proposition. *Assume that $f(p,q)$ is a good definition for $J1$, $J2$, $J3$ and $J5$. Then $f(p,q) \wedge (\Diamond p \to \Diamond q)$ is good for IL.*

Note, by the way, that $\Box^+(\Diamond p \to \Diamond q)$ is a minimal good definition for ILP: it is entailed by every other good definition for ILP. For, let $f(p,q)$ be such a definition, then $L \vdash f(p,q) \to$

$(\Diamond p \to \Diamond q)$ by $J4$. So $L \vdash \Box f(p,q) \to \Box(\Diamond p \to \Diamond q)$, by $R2$ and $L1$. Now, $L \vdash f(p,q) \to \Box f(p,q)$ by axiom P, therefore $L \vdash f(p,q) \to \Box(\Diamond p \to \Diamond q) \wedge (\Diamond p \to \Diamond q)$.

Definitions for ILM

It is easily verified that neither $(\Diamond p \to \Diamond q)$ nor any of the ZI is good for *ILM*. To see that PRC^+ is not a good definition for the axiom M, consider the following L-model (transitive R-arrows have been left out):



It is a routine matter to check that $w \Vdash \Box^+(\Diamond p \to \Diamond q)$, but $w \nVdash \Box^+(\Diamond(p \wedge \Box r) \to \Diamond(q \wedge \Box r))$.

To see that, on the other hand, MAX is a good definition for *ILM*, we only have to show that $L \vdash (M)^m$. Note that

$$L \vdash \Box(A \to B \vee \Diamond B) \to \Box(A \wedge \Box C \to (B \vee \Diamond B) \wedge \Box C)$$
$$\to \Box(A \wedge \Box C \to (B \wedge \Box C) \vee (\Diamond B \wedge \Box C)).$$

Now, $L \vdash \Diamond B \wedge \Box C \to \Diamond(B \wedge \Box C)$. Therefore, $L \vdash \Box(A \to B \vee \Diamond B) \to \Box(A \wedge \Box C \to (B \wedge \Box C) \vee \Diamond(B \wedge \Box C))$, as required.

For some time it was thought that MAX was the only good definition for *ILM*; however, there are in fact infinitely many good definitions for *ILM*:

1.7 Proposition. *Let $n \in \omega$. Then $\Box(p \to q \vee \Diamond q) \vee \Diamond^n(q \wedge \Box\bot)$ is a good defintion for ILM.*

Proof. The case $n = 0$ is proved above. Here we will prove the case $n = 1$. Let $(\cdot)^t$ be the translation associated with $\Box(p \to q \vee \Diamond q) \vee \Diamond(q \wedge \Box\bot)$. We only check that $L \vdash (J2)^t$, and leave the other cases to the reader. To prove that $L \vdash (J2)^t$ we have to show that

$$L \vdash (\Box(A \to B \vee \Diamond B) \vee \Diamond(B \wedge \Box\bot)) \wedge (\Box(B \to C \vee \Diamond C) \vee \Diamond(C \wedge \Box\bot))$$
$$\to \Diamond(A \to C \vee \Diamond C) \vee \Diamond(C \wedge \Diamond\bot).$$

Assume the antecedent. If we have $\Diamond(C \wedge \Box\bot)$ then there is nothing to prove. So assume $\neg\Diamond(C \wedge \Box\bot)$. Then $\Box(B \to C \vee \Diamond C)$. If we have $\Box(A \to B \vee \Diamond B)$, then we can use the proof of Proposition 1.3 to get $\Box(A \to C \vee \Diamond C)$. So suppose we have $\neg\Box(A \to B \vee \Diamond B)$, then we must have $\Diamond(B \wedge \Box\bot)$. Together with $\Box(B \to C \vee \Diamond C)$ this yields $\Diamond(C \wedge \Box\bot)$, and we are done. ■

By Proposition 1.4 any good definition $f(p,q)$ for *ILM* is implied by $\Box(p \to q \vee \Diamond q)$. The following result, which is due to Dick de Jongh, states that any good definition $f(p,q)$ for *ILM* implies $\Box(p \to q \vee \Diamond q) \vee \Diamond(q \wedge \Box\bot)$:

1.8 Proposition. *Let $f(p,q)$ be a good definition for ILM. Then $L \vdash f(p,q) \to \Box(p \to q \vee \Diamond q) \vee \Diamond(q \wedge \Box\bot)$.*

Proof. Suppose L does not prove the implication. Then we find a Kripke model $\mathcal{M}$ for L with root w such that $w \Vdash f(p,q)$, $\Diamond(p \wedge \Box^+ q)$, $\Box(q \to \neg\Box\bot)$. It follows that there is a v with wRv and $v \Vdash p \wedge \Box^+\neg q$, and that q is not forced on the endpoints of $\mathcal{M}$.

Now reduce the model to p and q, turn it into a tree (cf. Smoryński [1985a] for the details), and define $y \Vdash r$ iff vRy, where $r \notin \{p, q\}$. Using the fact that q is not forced on endpoints, one easily verifies that $w \nVdash f(p,q) \to (\Diamond(p \wedge \Box r) \to \Diamond(q \wedge \Box r))$, and so $L \nvdash f(p,q) \to (\Diamond(p \wedge \Box r) \to \Diamond(q \wedge \Box r))$. However, since $f(p,q)$ is a good definition for *ILM*, we have

$$\begin{aligned} L \vdash f(p,q) &\to f(p \wedge \Box r, q \wedge \Box r), \text{ by axiom } M \\ &\to (\Diamond(p \wedge \Box r) \to \Diamond(q \wedge \Box r)), \text{ by axiom } J4. \end{aligned}$$

So we have reached a contradiction. ∎

How to Get New Good Definitions from Old Ones

Let X be some set of axioms in $\mathcal{L}(\Box, \rhd)$. Given a good definition for X, there are several ways of making new ones out of it. One way was indicated above:

(i) If $f(p,q)$ is good for $J1$, $J2$, $J3$ and $J5$, then $f(p,q) \wedge (\Diamond p \to \Diamond q)$ is good for *IL*.

Here are some other ways:

(ii) If $f_1(p,q)$ and $f_2(p,q)$ are good for *IL* (*ILP*, *ILM*), then so are $f_1(p,q) \wedge f_2(p,q)$ and $\Box^+ f_1(p,q)$.
(iii) If $f(p,q)$ is good for *IL*, then $\Box^+ f(p,q)$ is good for *ILP*.

2 On Extensions of *IL*

We first establish a simple result to make sure that substitution will work properly.

One easily verifies that if $f(p,q)$ is good for $J1$ and $J2$, then

$$EXT \qquad \begin{aligned} L + f(A,B) \leftrightarrow A \rhd B \vdash \Box(D \leftrightarrow E) \to (D \rhd F \leftrightarrow E \rhd F), \\ L + f(A,B) \leftrightarrow A \rhd B \vdash \Box(D \leftrightarrow E) \to (F \rhd D \leftrightarrow F \rhd E). \end{aligned}$$

Let $F(p)$ denote a formula in which p possibly occurs, and let $F(D) \equiv F(p)[p := D]$. Using EXT it is a routine matter to show that for $f(p,q)$ as above we have

$$L + f(A,B) \leftrightarrow A \rhd B \vdash D \leftrightarrow E \quad \Rightarrow \quad L + f(A,B) \leftrightarrow A \rhd B \vdash F(D) \leftrightarrow F(E).$$

This substitution principle may then be used to prove the following Proposition:

2.1 Proposition. *Let $f(p,q)$ be good definition for $J1$ and $J2$, and let $(\cdot)^f$ be its associated translation. Then for all $D \in \mathcal{L}(\Box, \rhd)$, $L + f(A,B) \leftrightarrow A \rhd B \vdash D \leftrightarrow D^f$.*

Proof. A simple induction on D. ∎

2.2 Proposition. *Let $f(p,q)$ be a good definition for $X \subseteq \mathcal{L}(\Box, \rhd)$, and $(\cdot)^f$ its associated translation. If $L + X + f(A,B) \leftrightarrow A \rhd B \vdash D$, then $L \vdash D^f$.*

Proof. Induction on the length of the derivation of D from $L + X + f(A,B) \leftrightarrow A \rhd B$. ∎

Propositions 2.1 and 2.2 may be used to give an alternative characterization of some good definitions.

2.3 Proposition. *Let $f(p,q) \in \mathcal{L}(\Box)$, and let $X \subseteq \mathcal{L}(\Box, \rhd)$ contain $J1$ and $J2$. Then $f(p,q)$ is a good definition for X iff $L + X + f(A,B) \leftrightarrow A \rhd B$ is conservative over L.*

Proof. $\Rightarrow$: If $D \in \mathcal{L}(\Box)$, then

$$\begin{aligned} L + X + f(A,B) \leftrightarrow A \rhd B \vdash D & \\ & \Rightarrow L \vdash D^f, \text{ by Proposition 2.2,} \\ & \Rightarrow L \vdash D, \text{ since } D \in \mathcal{L}(\Box) \text{ implies } D \equiv D^f. \end{aligned}$$

$\Leftarrow$: Suppose that $L + X + f(A,B) \leftrightarrow A \rhd B$ is conservative over L. If $f(p,q)$ is not good for X, then $L \nvdash D^f$ for some $D \in X$, so by conservativity $L + X + f(A,B) \leftrightarrow A \rhd B \nvdash D^f$. But by Proposition 2.1 $L + X + f(A,B) \leftrightarrow A \rhd B \vdash D \leftrightarrow D^f$, whence $L + X + f(A,B) \leftrightarrow A \rhd B \nvdash D$ — which contradicts $D \in X$. ∎

2.4 Proposition. *Assume that $f(p,q)$ is a good definition for J1 and J2. If $f(p,q)$ is a good definition for $X \subseteq \mathcal{L}(\Box, \rhd)$, then $L + f(A,B) \leftrightarrow A \rhd B \vdash X$.*

Proof. Suppose $L + f(A,B) \leftrightarrow A \rhd B \nvdash D$, for some $D \in X$. By Proposition 2.1 it follows that $L + f(A,B) \leftrightarrow A \rhd B \nvdash D^f$ and $L \nvdash D^f$ — but then $f(p,q)$ can't be good for X. ∎

It follows that if $f(p,q)$ is good for *IL*, then $L + f(A,B) \leftrightarrow A \rhd B = IL + f(A,B) \leftrightarrow A \rhd B$.

2.5 Proposition. *Assume that $f(p,q)$ is a good definition for J1 and J2. If $f(p,q)$ is a good definition for $X \subseteq \mathcal{L}(\Box, \rhd)$, then $L + X \vdash f(A,B) \leftrightarrow A \rhd B$ iff $L + f(A,B) \leftrightarrow A \rhd B$ axiomatizes $L + X$.*

Proof. $\Rightarrow$: We have to prove that for all D

$$L + f(A,B) \leftrightarrow A \rhd B \vdash D \text{ iff } L + X \vdash D.$$

The direction from left to right is easy since $L + X \vdash f(A,B) \leftrightarrow A \rhd B$, while the other one follows from the previous Proposition.

$\Leftarrow$: If $L + f(A,B) \leftrightarrow A \rhd B$ axiomatizes $L + X$, then we have in particular $L + X \vdash f(A,B) \leftrightarrow A \rhd B$. ∎

2.6 Definition. A logic $X \supseteq IL$ is called a *maximal* L-conservative extension of *IL*, if X is an L-conservative extension of *IL*, while $Y \supsetneq X$ implies that Y is not L-conservative.

2.7 Proposition. *Let $f(p,q)$ be a good definition for IL. Then $L + f(A,B) \leftrightarrow A \rhd B$ is a maximal L-conservative extension of IL.*

Proof. Suppose $f(p,q)$ is good for *IL*. By Proposition 2.3 we have that $L + f(A,B) \leftrightarrow A \rhd B$ is an L-conservative extension of *IL*. To show it is maximal, assume that $Y \supsetneq L + f(A,B) \leftrightarrow A \rhd B$, say $Y \vdash D$, while $L + f(A,B) \leftrightarrow A \rhd B \nvdash D$. Let $(\cdot)^f$ be the translation associated with $f(p,q)$. Then by Proposition 2.1 we have that $L + f(A,B) \leftrightarrow A \rhd B \vdash D \leftrightarrow D^f$, and $Y \vdash D \leftrightarrow D^f$, therefore $Y \vdash D^f$. On the other hand $L + f(A,B) \leftrightarrow A \rhd B \nvdash D^f$, so $L \nvdash D^f$. Since $D^f \in \mathcal{L}(\Box)$, this proves that Y is not L-conservative. ∎

This last result may be used to show that no good definition for *IL*, *ILP* or *ILM* can be derivable in *IL*, *ILP* or *ILM* respectively. For suppose the contrary, say $IL \vdash f(A,B) \leftrightarrow A \rhd B$, then $L + f(A,B) \leftrightarrow A \rhd B$ axiomatizes *IL*, by Proposition 2.5. By Proposition 2.7, then, it would be a maximal L-conservative system — which is certainly not the case, since $IL \subsetneq ILP$. To show that neither *ILP* nor *ILM* proves a good definition for itself, one uses the fact that both are contained in the L-conservative system *ILMP*.

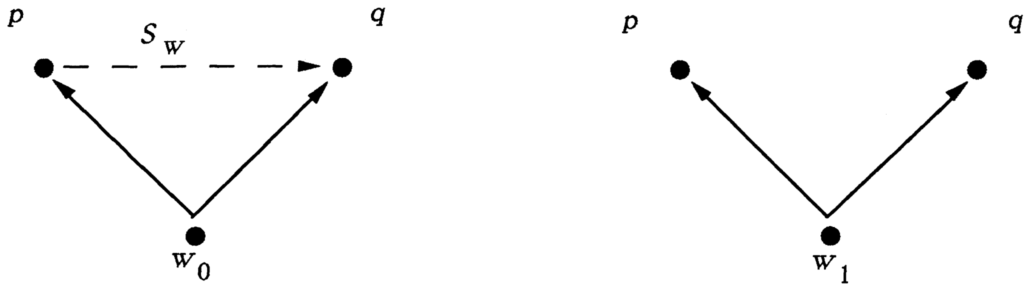
Johan van Benthem pointed out to us that these non-definability results may also be obtained using a so-called Padoa counterexample. The idea is to give two models with identical W, R and $\Vdash$ (on the proposition letters p and q), but with different S, in such a way that

- both of these models are *IL*-models;

- their extensions of $p \rhd q$ are different.

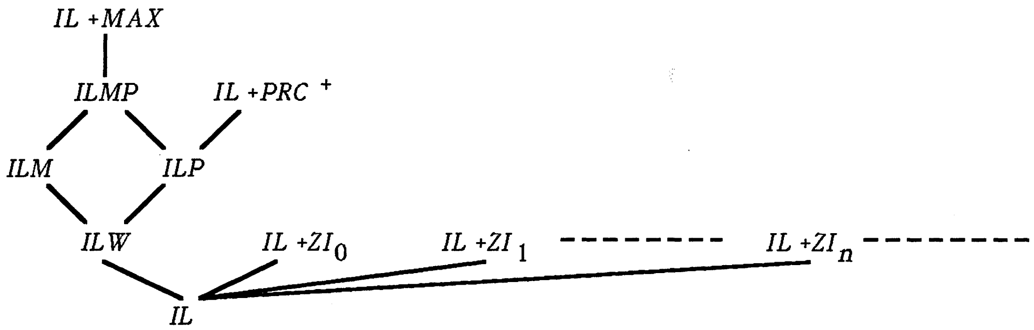
Clearly, if two such models exist, we can not have $IL \vdash p \rhd q \leftrightarrow f(p,q)$ for any $f(p,q) \in \mathcal{L}(\Box)$.

One easily verifies that the following two models satisfy the conditions mentioned above:



(Note that both models are *ILM*- and *ILP*-models, so the above proof also shows the non-definability of '$\rhd$' in *ILM* and *ILP*.)

Recall that in section 1 we found $\aleph_0$ good definitions for *IL*: each of the ZI (where I is a finite subset of $\omega \setminus \{0\}$) is good for *IL*. So the picture that emerges frome Proposition 2.7 is the following. We have one 'basic' *L*-conservative system, *IL*, that can be extended in (at least) $\aleph_0$ essentially different ways without destroying *L*-conservativity:



(Here, we write *IL*+X, where X is some definition, to denote *IL*+$X([p := A],[q := B]) \leftrightarrow A \rhd B$; $\{ I_j \}_{j \in \omega}$ is some enumeration of the finite subsets of $\omega \setminus \{0\}$.)

Several natural questions arise at this point:

- Are all maximal *L*-conservative extensions of *IL* of the form $L + f(A,B) \leftrightarrow A \rhd B$ for some good definition for *IL*?
- What is going on high up in the picture? E.g., is there one big *L*-conservative extension of *IL* that contains all other *L*-conservative extensions of *IL*?

We have no answer to the first question, but we conjecture that it has a negative answer. As to the second question, any system that contains all *L*-conservative extensions of *IL* cannot itself be a conservative extension of *L*. For let X be such a system. Then X extends both $L + \Box^+(\Diamond A \rightarrow \Diamond B) \leftrightarrow A \rhd B$ and $L + \Box(A \rightarrow B \vee \Diamond B) \leftrightarrow A \rhd B$. Since $L \nvdash \Box^+(\Diamond A \rightarrow \Diamond B) \rightarrow \Box(A \rightarrow B \vee \Diamond B)$, X properly extends both of these systems, therefore it cannot be *L*-conservative by Proposition 2.7.

6. On the Modal Theory of $\mathcal{L}(\Box, \rhd)$

The first section of this chapter has a less formal character than the rest of this thesis; it contains some general remarks on the modal theory of $\mathcal{L}(\Box, \rhd)$. In Section 2 we characterize the first-order formulas (in some appropriate first-order language) that are equivalent — on Veltman models — to a formula in a natural extension of our modal language.

1 Some General Remarks

We make the following change in notation. Let $\mathcal{F}$ be an $\mathcal{L}(\Box, \rhd)$-frame, i.e., a triple $\langle W, R, S \rangle$ with $W \neq \emptyset$, $R \subseteq W^2$ and $S \subseteq W^3$. From now on we write $\mathcal{M} = \langle \mathcal{F}, V \rangle$ in stead of $\mathcal{M} = \langle \mathcal{F}, \Vdash \rangle$ to denote a model on $\mathcal{F}$. Here, V is a function $V : PROP \longrightarrow 2^W$, where $PROP$ is the set of proposition letters. '$\mathcal{M} \models A[w]$' (A is true at w in $\mathcal{M}$) is defined inductively by

$$\begin{array}{rll} \mathcal{M} \models p[w] & \text{iff} & w \in V(p) \\ \mathcal{M} \models \neg A[w] & \text{iff} & \mathcal{M} \not\models A[w] \\ \mathcal{M} \models A \wedge B[w] & \text{iff} & \mathcal{M} \models A[w] \text{ and } \mathcal{M} \models B[w] \\ \mathcal{M} \models \Diamond A[w] & \text{iff} & \exists v\,(wRv \wedge \mathcal{M} \models A[v]) \\ \mathcal{M} \models A \rhd B & \text{iff} & \forall v\,(wRv \wedge \mathcal{M} \models A[v] \Rightarrow \exists u\,(vS_w u \wedge \mathcal{M} \models B[u])). \end{array}$$

Furthermore, $\mathcal{M} \models A$ if $\mathcal{M} \models A[w]$ for all $w \in W$; $\mathcal{F} \models A[w]$ if, for all valuations V on $\mathcal{F}$, $\langle \mathcal{F}, V \rangle \models A[w]$, and $\mathcal{F} \models A$ if, for all $w \in W$, $\mathcal{F} \models A[w]$.

As is well-known, $\mathcal{L}(\Box)$-formulas can be translated into first-order ones in a language $\mathcal{L}_1$ that contains one binary predicate symbol R, and unary predicate symbols P corresponding to the proposition letters in $\mathcal{L}(\Box)$:

(1) $\tau(p) = Px$
(2) $\tau(\neg B) = \neg\tau(B)$
(3) $\tau(B \wedge C) = \tau(B) \wedge \tau(C)$
(4) $\tau(\Box B) = \forall y\,(xRy \rightarrow \tau(B)[x := y])$

where y is a variable not occurring in $\tau(B)$. As a matter of fact, all $\mathcal{L}(\Box)$-formulas translate into a *two-variable* fragment of the above first-order language, e.g., the two-variable transcription of $\Box\Diamond(p \rightarrow \Diamond q)$ is

$$\forall y\,(xRy \rightarrow \exists x\,(yRx \wedge (Px \rightarrow \exists y\,(xRy \wedge Qy)))).$$

Note that τ is not surjective, even on equivalence classes of formulas. Not all formulas in $\mathcal{L}_1$ — or even in the appropriate 2-variable fragment thereof — are the τ-transcription of an $\mathcal{L}(\Box)$-formula. E.g., $\forall y\,(xRy)$ is not equivalent to an $\mathcal{L}(\Box)$-formula: unlike $\mathcal{L}(\Box)$-formulas it is not preserved under disjoint unions.

It is clear that τ may be extended to $\mathcal{L}(\Box, \rhd)$ by adding a clause for '$\rhd$':

(5) $\tau(A \rhd B) = \forall y\,(xRy \wedge \tau(A)[x := y] \rightarrow \exists z\,(yS_x z \wedge \tau(B)[x := z])).$

Here, y and z are variables not occurring in $\tau(A)$ and $\tau(B)$. Note that this translation takes $\mathcal{L}(\Box, \rhd)$-formulas into a *three-variable* fragment of the appropriate first-order language. Here too τ is not surjective, not even with respect to the 2-variable fragment. This is witnessed by the same formula as the one given above.

In van Benthem [1989] a *modality* is defined to be a function on sets defined by some first-order schema

$$\lambda x.\varphi(x, A_1, \ldots, A_n)$$

which is continuous in the sense of commuting with arbitrary unions of its arguments A_i ($1 \leq i \leq n$). (Obviously, for the "dual" version of the operator, one has to substitute 'intersections' for 'unions'.) A typical example may be read off from the above clauses defining '$\mathcal{M} \models A[w]$':

$$\Diamond A \equiv \lambda x.\exists y\,(xRy \wedge A(y)).$$

Now, the definition of '$\rhd$' in this λ-notation is

$$\lambda x.F(A, B) \equiv \lambda x.\forall y\,(xRy \wedge A(y) \rightarrow \exists z\,(yS_x z \wedge B(z))),$$

or, equivalently,

$$\lambda x.F(A, B) \equiv \lambda x.\forall y \exists z\,(xRy \wedge A(y) \rightarrow (yS_x z \wedge B(z))).$$

One easily verifies that $\lambda x.F(A, B)$ is not continuous:

$$\lambda x.F(A, B \cup C) \neq \lambda x.F(A, B) \cup \lambda x.F(A, C).$$

Neither does it commute with intersections of its arguments:

$$\lambda x.F(A \cap B, C) \neq \lambda x.F(A, C) \cap \lambda x.F(B, C).$$

So '$\rhd$' does not qualify as a modality according to van Benthem's definition. One obvious way out is to drop this definition, and to try and find a more liberal one. There are, however, several good reasons for excluding operators like '$\rhd$' from a general theory of modal operators in the spirit of van Benthem [1989]. One such reason is that we have a good duality theory for modalities in the sense of van Benthem, while no such theory is available for operators like '$\rhd$'. Let us elaborate a bit.

An alternative to the Kripke-like structures for ordinary modal logics is offered by so-called 'modal algebras'. These are Boolean algebras with operations (that is, functions from A^n to A, for any n, where A is the carrier of the Boolean algebra). According to the famous representation theorem in Jónsson and Tarski [1951] there is a unique way of passing from a Kripke frame to a modal algebra, and back. This duality between Kripke structures and modal algebras has proved to be very useful in classical modal logic. (Cf. van Benthem [1985].) Now, passing from an $\mathcal{L}(\Box, \rhd)$-frame to some sort of algebra is easy — one merely has to use the clauses for '$\Diamond$' and '$\rhd$' used to define '$\mathcal{M} \models A$'. Unfortunately, the Jónsson and Tarski representation result is only concerned with operators that are definable by a first-order formula of the form $\exists x_1 \ldots \exists x_n\, \psi$ (or dually: $\forall x_1 \ldots \forall x_n\, \psi$), where ψ is quantifier free. By a result in van Benthem [1989] all van Benthem's modalities are indeed of this form. But clearly, any operator corresponding to '$\rhd$' has to have a $\forall\exists$-definition; therefore the Jónsson and Tarski representation result cannot be applied here to get back into the realm of frames.

When interpreted in frames, $\mathcal{L}(\Box, \rhd)$-formulas get *second-order* transcriptions, with equivalences

$$\mathcal{F} \models A[w] \iff \mathcal{F} \models \forall P_1 \ldots \forall P_n\, \tau(A)[w]$$
$$\mathcal{F} \models A \iff \mathcal{F} \models \forall P_1 \ldots \forall P_n\, \tau(A),$$

proposition letters occurring in A. One of the rare general results in classical modal logic, the well-known Sahlqvist Theorem, is, among other things, concerned with the following question:

which modal formulas define first-order conditions on frames? Here, we will not attempt to obtain any general answers or results concerning this issue. We only want to point out that even syntactically very simple $\mathcal{L}(\Box, \rhd)$-formulas may already define second-order conditions.

As an example we consider the formula $\top \rhd p \to \neg(\top \rhd \neg p)$. Now, ordinary modal formulas in which a quantifier combination $\forall\exists$ occurs in the antecedent of an implication are known to lead us outside of the realm of first-order definability (cf. van Benthem [1985]) — so one might expect that the formula $\top \rhd p \to \neg(\top \rhd \neg p)$ is not first-order definable. And indeed, it is not.

1.1 Proposition. *Let $\mathcal{M}$ be an $\mathcal{L}(\Box, \rhd)$-model such that $\mathcal{M} \models \forall xyz\,(yS_xz \leftrightarrow xRyRz)$. Then, for all $w \in W$,*

(i) $\mathcal{M} \models \Box\Diamond A[w]$ iff $\mathcal{M} \models \top \rhd A[w]$;
(ii) $\mathcal{M} \models \Diamond\Box A[w]$ iff $\mathcal{M} \models \neg(\top \rhd \neg A)[w]$.

1.2 Proposition. *Let $\mathcal{F}$ be an $\mathcal{L}(\Box, \rhd)$-frame such that $\mathcal{F} \models \forall xyz\,(yS_xz \leftrightarrow xRyRz)$. Then, for all $w \in W$, $\mathcal{F} \models \Box\Diamond p \to \Diamond\Box p[w]$ iff $\mathcal{F} \models \top \rhd p \to \neg(\top \rhd \neg p)$.*

1.3 Proposition. $\Box\Diamond p \to \Diamond\Box p$ *is not first-order definable.*

Sketch of the Proof. In van Benthem [1985] an uncountable Kripke frame $\mathcal{F}$ is defined such that $\mathcal{F} \models \Box\Diamond p \to \Diamond\Box p$. It is then shown that this formula fails on some countable elementary substructure $\mathcal{F}'$ of $\mathcal{F}$. From this the Proposition follows. ∎

1.4 Proposition. $\top \rhd p \to \neg(\top \rhd \neg p)$ *is not first-order definable.*

Proof. Consider the 'proof' of the preceding Proposition. The frame $\mathcal{F}$ in that proof may be expanded to an $\mathcal{L}(\Box, \rhd)$-frame by putting $\forall xyz\,(yS_xz \leftrightarrow xRyRz)$. By Proposition 1.2 it follows that $\mathcal{F} \models \top \rhd p \to \neg(\top \rhd \neg p)$. Being a first-order sentence, $\forall xyz\,(yS_xz \leftrightarrow xRyRz)$ holds on $\mathcal{F}'$. But then, by another application of Proposition 1.2, $\mathcal{F}' \not\models \top \rhd p \to \neg(\top \rhd \neg p)$. We may conclude that $\top \rhd p \to \neg(\top \rhd \neg p)$ is not first-order definable. ∎

2 Relations between Models

The first-order formulas $\varphi \equiv \varphi(x)$ that are (equivalent to) a translation of an $\mathcal{L}(\Box)$-formula can be characterized using the following relation between $\mathcal{L}(\Box)$-models:

2.1 Definition. A relation $\mathcal{Z}$ between two $\mathcal{L}(\Box)$-models $\mathcal{M}_1$ and $\mathcal{M}_2$ is called a *bisimulation* if it is non-empty and if it satisfies

(i) if $x_1\mathcal{Z}x_2$, then x_1, x_2 satisfy the same proposition letters (or unary predicates);
(ii) if $x_1\mathcal{Z}x_2$ and $x_1R_1y_1$, then there exists a $y_2 \in W_2$ such that $x_2R_2y_2$ and $y_1\mathcal{Z}y_2$;
(iii) similar to (ii) — but in the opposite direction.

Now $\mathcal{L}(\Box)$-formulas are invariant for bisimulations, i.e., if $\mathcal{Z}$ is a bisimulation between $\mathcal{M}_1$ and $\mathcal{M}_2$ then

$$\text{if } w_1\mathcal{Z}w_2, \text{ then } \mathcal{M}_1 \models A[w_1] \text{ iff } \mathcal{M}_2 \models A[w_2],$$

for all $A \in \mathcal{L}(\Box)$. Conversely, if a first-order formula $\varphi \equiv \varphi(x)$ is invariant for bisimulations, then it is equivalent to a translation of an $\mathcal{L}(\Box)$-formula (cf. van Benthem [1985], de Rijke [1989]).

Given this model-theoretic characterization of the translations of $\mathcal{L}(\Box)$-formulas, it seems natural to try and find a characterization of the translations of $\mathcal{L}(\Box, \rhd)$-formulas in terms of their invariance under some relation between $\mathcal{L}(\Box, \rhd)$-models. The minimal conditions a bisimulation has to fulfil to preserve true formulas of the form $A \rhd B$ are

(iv) if $x_1 \mathcal{Z} x_2$ and $x_2 R_2 y_2$, then there is a y_1 with $x_1 R_1 y_1$, $y_1 \mathcal{Z} y_2$ such that for all z_1 with $y_1 S_{1x_1} z_1$ there is a z_2 with $z_1 \mathcal{Z} z_2$ and $y_2 S_{2x_2} z_2$;

(v) similar to (iv) — but in the opposite direction.

Let an *extended bisimulation* be a bisimulation that satisfies the extra conditions (iv) and (v). Then a straightforward induction establishes that

2.2 Proposition. $\mathcal{L}(\Box, \rhd)$*-formulas are invariant for extended bisimulations.*

In his Vissser [1988] Albert Visser uses a notion of bisimulation that is similar to our extended bisimulations. He uses his notion to show that every Veltman model for *IL* (or *ILP* or *ILM*) is bisimular to a *simplified* Veltman model for *IL* (or *ILP*, *ILM* respectively). The simplification consists of having *one* binary relation S in stead of a number of binary relations S_w. The clause for '$\rhd$' in the definition of a forcing relation is then changed to

$$w \Vdash A \rhd B \iff \forall v\,(wRv \wedge v \Vdash A \Rightarrow \exists u\,(wRu \wedge vSu \wedge u \Vdash B)).$$

Some additional motivation for the notion of extended bisimulations may be found in the following result:

2.3 Proposition. *Let* $\mathcal{M}_1$, $\mathcal{M}_2$ *be two finite* $\mathcal{L}(\Box, \rhd)$*-models. If for all* $A \in \mathcal{L}(\Box, \rhd)$, $\mathcal{M}_1 \models A[w_1]$ *iff* $\mathcal{M}_2 \models A[w_2]$, *then there exists an extended bisimulation between* $\mathcal{M}_1$ *and* $\mathcal{M}_2$ *that connects* w_1 *and* w_2.

Proof. Define $w \mathcal{Z} w'$ iff for all $A \in \mathcal{L}(\Box, \rhd)$, $\mathcal{M}_1 \models A[w]$ iff $\mathcal{M}_2 \models A[w']$. We claim that $\mathcal{Z}$ is an extended bisimulation. We only check conditions (ii) and (iv) of the definition.

(ii) Assume that $x_1 \mathcal{Z} x_2$ and $x_1 R_1 y_1$, and suppose that there is no $y_2 \in W_2$ satisfying $x_2 R_2 y_2$ and $y_1 \mathcal{Z} y_2$. Let $x_2 R_2 := \{ u_1, \ldots, u_n \}$. Then for each u_i we have $\neg y_1 \mathcal{Z} u_i$. By the definition of $\mathcal{Z}$ this implies the existence of formulas A_i such that $\mathcal{M} \not\models A_i[y_1]$ and $\mathcal{M} \models A_i[u_i]$. Since $\mathcal{M}_2$ is finite, $A \equiv \bigvee_i A_i \in \mathcal{L}(\Box, \rhd)$. Then

$$\mathcal{M}_1 \models \neg A[y_1] \text{ and, for all } u_i,\ \mathcal{M}_2 \models A[u_i],$$

so

$$\mathcal{M}_1 \models \neg\Box A[x_1] \text{ and } \mathcal{M}_2 \models \Box A[x_2],$$

which contradicts $x_1 \mathcal{Z} x_2$.

(iv) Assume that $x_1 \mathcal{Z} x_2$ and $x_2 R_2 y_2$. Suppose that there is no y_1 satisfying condition (iv). Let $x_1 R_1 := \{ u_1, \ldots u_m \}$. Then for each u_i we have either

not $u_i \mathcal{Z} y_2$

or

there is a z_{1i} with $u_i S_{1x_1} z_{1i}$ without there being a z_{2i} with $z_{1i} \mathcal{Z} z_{2i}$ and $y_2 S_{2x_2} z_{2i}$.

By the definition of $\mathcal{Z}$ the first case yields formulas A_i such that $\mathcal{M}_1 \models A_i[u_i]$ and $\mathcal{M}_2 \models \neg A_i[y_2]$. Since $\mathcal{M}_1$ is finite we have $A \equiv \bigvee_i A_i \in \mathcal{L}(\Box, \rhd)$ and so

$$\mathcal{M}_1 \models A[u_i], \text{ for all } u_i \text{ with } \neg u_i \mathcal{Z} y_2, \text{ and } \mathcal{M}_2 \models \neg A[y_2].$$

Assume that $y_2 S_{2x_2} = \{ v_1, \ldots, v_k \}$. Then the second case implies that for each v_j there is a formula B_{ij} such that $\mathcal{M}_1 \models B_{ij}[z_{1i}]$ and $\mathcal{M}_2 \models \neg B_{ij}[v_j]$. So, letting $B_i \equiv \bigwedge_j B_{ij}$, we find

$$\mathcal{M}_1 \models B_i[z_{1i}], \text{ and for all } v_j,\ \mathcal{M}_2 \models \neg B_i[v_j].$$

But then $\mathcal{M}_1 \models \neg A \rhd \bigvee_i B_i[x_1]$ and $\mathcal{M}_2 \not\models \neg A \rhd \bigvee_i B_i[x_2]$, which contradicts $x_1 \mathcal{Z} x_2$. ∎

Although this notion seems a plausible candidate for a characterization of the transcriptions of $\mathcal{L}(\Box, \rhd)$-formulas, it does not fit. Without going into details here (but cf. Remark 2.8), and without claiming to have pinpointed the causes of these difficulties, we think that they have to do with

(i) the fact that our first-order formalism now essentially needs at least three variables — while we are trying to characterize this fragment by focusing on its formulas with one free variable;

(ii) the fact that we are working with a quantifier combination $\forall\exists$:

$$\forall y\,(xRy \wedge Py \rightarrow \exists z\,(yS_x z \wedge Qz)).$$

Given (i) it seems much more appropriate to work with a formalism allowing up to three free variables. This motivates the following definitions:

2.4 Definition. We define the first-order language $\mathcal{L}_{pi} \subseteq \mathcal{L}_1$. Fix three variables x, y, z, then

(i) $Px \in \mathcal{L}_{pi}$ for each unary $P \in \mathcal{L}_1$;

(ii) if φ, $\psi \in \mathcal{L}_{pi}$, then also $\neg\varphi$, $\varphi \wedge \psi$, $\varphi \vee \psi$, $\varphi \rightarrow \psi \in \mathcal{L}_{pi}$;

(iii) if $\varphi(x,y) \in \mathcal{L}_{pi}$, then $\exists y\,(xRy \wedge \varphi(x,y)) \in \mathcal{L}_{pi}$;

(iv) if $\varphi(x,y,z) \in \mathcal{L}_{pi}$, then $\exists z\,(yS_x z \wedge \varphi(x,y,z)) \in \mathcal{L}_{pi}$,

and similar patterns for the other variables.

Note that the formulas $\varphi \equiv \varphi(x) \in \mathcal{L}_{pi}$ that contain one free variable are a proper superset of the set of the τ-translations of $\mathcal{L}(\Box, \rhd)$-formulas. E.g., $\exists y\,(xRy \wedge \exists z\,(yS_x z \wedge Pz))$ is in $\mathcal{L}_{pi}$, but it certainly is not the transcription of some $A \in \mathcal{L}(\Box, \rhd)$. Nevertheless, we like to think that the set of $\varphi(x) \in \mathcal{L}_{pi}$ with one free variable, forms a natural extension of the set of translations of $\mathcal{L}(\Box, \rhd)$-formulas.

Ordinary bisimulations may be viewed as a relation between sequences of objects of length at most 2. In stead of imposing some extra condition on the relation between single objects, as with the above *extended* bisimulations, the right way to generalize the notion of bisimulation seems to be the extension to sequences of objects of length larger than 2:

2.5 Definition. A *trisimulation* between two models $\mathcal{M}_1$, $\mathcal{M}_2$ is a non-empty relation $\mathcal{Z}$ between objects in W_1, W_2, between ordered pairs $\langle x_1, y_1\rangle \in W_1^2$ and $\langle x_2, y_2\rangle \in W_2^2$, where $x_1R_1y_1$ and $x_2R_2y_2$, and between triples $\langle x_1, y_1, z_1\rangle \in W_1^3$ and $\langle x_2, y_2, z_2\rangle \in W_2^3$, where $x_1R_1y_1S_{1x_1}z_1$ and $x_2R_2y_2S_{2x_2}z_2$, satisfying:

(i) if $x_1\mathcal{Z}x_2$, then x_1, x_2 satisfy the same proposition letters (or unary predicates);

(ii) if $x_1\mathcal{Z}x_2$ and $x_1R_1y_1$, then there is a y_2 such that $\langle x_1, y_1\rangle\mathcal{Z}\langle x_2, y_2\rangle$;

(iii) similar to (ii) — but in the opposite direction;

(iv) if $\langle x_1, y_1\rangle\mathcal{Z}\langle x_2, y_2\rangle$ and $y_1S_{1x_1}z_1$, then there is a z_2 with $\langle x_1, y_1, z_1\rangle\mathcal{Z}\langle x_2, y_2, z_2\rangle$;

(v) similar to (iv) — but in the opposite direction;

(vi) if $\langle x_1, y_1\rangle\mathcal{Z}\langle x_2, y_2\rangle$, then both $x_1\mathcal{Z}x_2$ and $y_1\mathcal{Z}y_2$, and if $\langle x_1, y_1, z_1\rangle\mathcal{Z}\langle x_2, y_2, z_2\rangle$, then $x_1\mathcal{Z}x_2$, $y_1\mathcal{Z}y_2$, and $z_1\mathcal{Z}z_2$, and also $\langle x_1, y_1\rangle\mathcal{Z}\langle x_2, y_2\rangle$.

The first thing to note is that $\mathcal{L}(\Box, \rhd)$-formulas are invariant for trisimulations, i.e., if $\mathcal{Z}$ is a trisimulation between $\mathcal{M}_1$ and $\mathcal{M}_2$, and $w_1\mathcal{Z}w_2$, then $\mathcal{M}_1 \models A[w_1]$ iff $\mathcal{M}_2 \models A[w_2]$ for all $A \in \mathcal{L}(\Box, \rhd)$. And more generally:

2.6 Proposition. *Let $\varphi \equiv \varphi(x) \in \mathcal{L}_{pi}$. Then φ is invariant for trisimulations.*

We are now ready to prove an invariance result for $\mathcal{L}_{pi}$:

2.7 Theorem. *Let $\varphi \equiv \varphi(x) \in \mathcal{L}_1$. Then φ is equivalent to an $\mathcal{L}_{pi}$-formula iff it is invariant for trisimulations.*

Proof. One direction is the preceding Proposition. To prove the other one, assume that $\varphi(x) \in \mathcal{L}_1$ is invariant for trisimulations. Define $X_\varphi := \{ \psi(x) \in \mathcal{L}_{pi} \mid \varphi \models \psi \}$. We will prove that $X_\varphi \models \varphi$. By the compactness theorem it then follows that $\psi \models \varphi$, for some $\psi \in X_\varphi$. Hence $\models \varphi \leftrightarrow \psi$.

Assume that $\mathcal{M}_1 \models X_\varphi[w]$; we have to prove that $\mathcal{M}_1 \models \varphi[w]$. Introduce a new individual constant $\mathbf{w}$ to stand for the object w and define $\mathcal{L}^*_{pi} := \mathcal{L}_{pi} \cup \{ \mathbf{w} \}$. In the remainder of this proof we use the following notation: if $\psi \in \mathcal{L}_{pi}$ then $\psi^* \equiv \psi[x := \mathbf{w}]$, and if $T \subseteq \mathcal{L}_{pi}$ then $T^* := \{ \psi^* \mid \psi \in T \}$. Of course $\mathcal{M}_1$ is easily expanded to an $\mathcal{L}^*_{pi}$-model $\mathcal{M}^*_1$ by interpreting $\mathbf{w}$ as w.

Let $T_0 := \{ \psi \in \mathcal{L}_{pi} \mid \mathcal{M}_1 \models \psi[w] \}$, and assume that $T \subseteq T_0$ is finite, say $T = \{ \psi_1, \ldots, \psi_n \}$. We claim that there exists an $\mathcal{L}^*_{pi}$-model $(\mathcal{N}, \mathbf{w}) \models \varphi^* \wedge \bigwedge T^*$. For suppose that such a model does not exist, then

$$
\begin{aligned}
(\mathcal{N}, \mathbf{w}) \models \neg \bigwedge T^*, &\text{ for every } \mathcal{L}^*_{pi} - \text{model } (\mathcal{N}, \mathbf{w}) \text{ such that } (\mathcal{N}, \mathbf{w}) \models \varphi^*, \\
&\Rightarrow \models \varphi^* \to \neg \bigwedge T^* \\
&\Rightarrow \neg \bigwedge T \in X_\varphi \\
&\Rightarrow \mathcal{M}_1 \models \neg \bigwedge T[w], \text{ since } \mathcal{M}_1 \models X_\varphi[w].
\end{aligned}
$$

However, $\bigwedge T \in T_0$, and so $\mathcal{M}_1 \models \bigwedge T[w]$ — contradiction.

By compactness we obtain an $\mathcal{L}^*_{pi}$-model $\mathcal{M}^*_2 := (\mathcal{M}_2, \mathbf{w}) \models \varphi^* \wedge \bigwedge T^*_0$. Now let U be a countably incomplete ultrafilter on ω, and let $\mathcal{M}^*_3 := \Pi_U \mathcal{M}^*_1$ and $\mathcal{M}^*_4 := \Pi_U \mathcal{M}^*_2$. It follows that $\mathcal{M}^*_3$ and $\mathcal{M}^*_4$ are ω-saturated (cf. Chang & Keisler [1973], Theorem 6.1.1). By the Łoś Theorem we find that the interpretations w_3 and w_4 of $\mathbf{w}$ in $\mathcal{M}^*_3$ and $\mathcal{M}^*_4$, respectively, both realize T_0, and that $\mathcal{M}^*_4 \models \varphi^*$. Define a trisimulation $\mathcal{Z}$ between $\mathcal{M}^*_3$ and $\mathcal{M}^*_4$ by putting

$$
\begin{aligned}
x_1 \mathcal{Z} x_2 \iff & \text{ for all } \psi \in \mathcal{L}_{pi}, \\
& \mathcal{M}_3 \models \psi[x_1] \text{ iff } \mathcal{M}_4 \models \psi[x_2]; \\
\langle x_1, y_1 \rangle \mathcal{Z} \langle x_2, y_2 \rangle \iff & \; x_1 R_3 y_1 \text{ and } x_2 R_4 y_2 \text{ and for all } \psi \in \mathcal{L}_{pi}, \\
& \mathcal{M}_3 \models \psi[x_1 y_1] \text{ iff } \mathcal{M}_4 \models \psi[x_2 y_2]; \\
\langle x_1, y_1, z_1 \rangle \mathcal{Z} \langle x_2, y_2, z_2 \rangle \iff & \; x_1 R_3 y_1 S_{3x_1} z_1 \text{ and } x_2 R_4 y_2 S_{2x_2} z_2 \text{ and for all } \psi \in \mathcal{L}_{pi}, \\
& \mathcal{M}_3 \models \psi[x_1 y_1 z_1] \text{ iff } \mathcal{M}_4 \models \psi[x_2 y_2 z_2].
\end{aligned}
$$

Note that $\mathcal{Z}$ satisfies condition (vi) of Definition 2.5. Let us check the other conditions as well. $\mathcal{Z}$ is non-empty because we have $w_3 \mathcal{Z} w_4$:

$$
\begin{aligned}
\mathcal{M}_3 \models \psi[w_3] &\Rightarrow (\mathcal{M}_3, \mathbf{w}) \models \psi^* \\
&\Rightarrow \psi^* \in T^*_0, \text{ otherwise } (\mathcal{M}_1, \mathbf{w}) \models \neg\psi^*, \\
&\qquad \neg\psi^* \in T^*_0 \text{ and } (\mathcal{M}_3, \mathbf{w}) \models \neg\psi^*, \\
&\Rightarrow (\mathcal{M}_4, \mathbf{w}) \models \psi^* \\
&\Rightarrow \mathcal{M}_4 \models \psi[w_4],
\end{aligned}
$$

and the converse implication is proved similarly. Next we check conditions (i)–(v) in Definition 2.5.

(i) By definition.

(ii) Assume that $x_1 \mathcal{Z} x_2$ and $x_1 R_3 y_1$; we have to find a y_2 with $\langle x_1, y_1 \rangle \mathcal{Z} \langle x_2, y_2 \rangle$. Consider $\Psi = \{ \psi \in \mathcal{L}_{pi} \mid \mathcal{M}_3 \models \psi[x_1 y_1] \}$. Then $\Psi(\mathbf{x}_2) \cup \{ \mathbf{x}_2 R y \}$ is finitely satisfiable in $\mathcal{M}_4$. For suppose not, then there is a finite $\Psi_0 \subseteq \Psi$ with

$$
\begin{aligned}
\mathcal{M}_4 \models & \forall y \, (xRy \to \neg \bigwedge \Psi_0(x, y))[x_2] \\
& \Rightarrow \mathcal{M}_4 \models \neg \exists y \, (xRy \wedge \neg \bigwedge \Psi_0(x, y))[x_2]
\end{aligned}
$$

$$\Rightarrow \mathcal{M}_3 \models \neg\exists y\,(xRy \wedge \neg \bigwedge \Psi_0(x,y))[x_1], \text{ since } x_1\mathcal{Z}x_2.$$

But this contradicts the fact that $\mathcal{M}_3 \models xRy \wedge \bigwedge \Psi(x,y)[x_1,y_1]$. So $\Psi(\mathbf{x}_2,y) \cup \{\,\mathbf{x}_2R_y\,\}$ is a type of $(\mathcal{M}_4,\mathbf{x}_2)$. Because this model is ω-saturated we find a $y_2 \in W_4$ such that $(\mathcal{M}_4,\mathbf{x}_2) \models \mathbf{x}_2Ry \wedge \bigwedge \Psi(\mathbf{x}_2,y)[y_2]$, i.e., $\mathcal{M}_4 \models xRy \wedge \bigwedge \Psi(x,y)[x_2y_2]$. We conclude that $\langle x_1,y_1\rangle\mathcal{Z}\langle x_2,y_2\rangle$.

(iii) Similar to (ii).

(iv) Suppose that $\langle x_1,y_1\rangle\mathcal{Z}\langle x_2,y_2\rangle$ and $y_1S_{3x_1}z_1$. This time we are looking for a z_2 with $\langle x_1,y_1,z_1\rangle\mathcal{Z}\langle x_2,y_2,z_2\rangle$. We now consider $\Psi = \{\,\psi \in \mathcal{L}_{pi} \,|\, \mathcal{M}_3 \models \psi[x_1y_1z_1]\,\}$, and claim that $\Psi(\mathbf{x}_2,\mathbf{y}_2,z) \cup \{\,\mathbf{y}_2S_{\mathbf{x}_2}z_2\,\}$ is finitely satisfiable in $\mathcal{M}_4$. Again, if it is not, there must be a finite $\Psi_0 \subseteq \Psi$ with

$$\begin{aligned}\mathcal{M}_4 \models & \forall z\,(yS_xz \to \neg \bigwedge \Psi_0(x,y,z))[x_2y_2]\\ & \Rightarrow \mathcal{M}_4 \models \neg\exists z\,(yS_xz \wedge \neg \bigwedge \Psi_0(x,y,z))[x_2y_2]\\ & \Rightarrow \mathcal{M}_3 \models \neg\exists z\,(yS_xz \wedge \neg \bigwedge \Psi_0(x,y,z))[x_1y_1z_1], \text{ since } \langle x_1,y_1\rangle\mathcal{Z}\langle x_2,y_2\rangle. \qquad (\star)\end{aligned}$$

But this contradicts the fact that $\mathcal{M}_3 \models yS_xz \wedge \bigwedge \Psi(x,y,z)[x_1y_1z_1]$. So $\Psi(\mathbf{x}_2,\mathbf{y}_2,z)\cup$ is a type of $(\mathcal{M}_4,\mathbf{x}_2,\mathbf{y}_2)$. Because this model is ω-saturated we find a $z_2 \in W_4$ with $(\mathcal{M}_4,\mathbf{x}_2,\mathbf{y}_2) \models \mathbf{y}_2S_{\mathbf{x}_2}z_2 \wedge \bigwedge \Psi(\mathbf{x}_2,\mathbf{y}_2,z)[z_2]$, i.e., $\mathcal{M}_4 \models yS_xz \wedge \bigwedge \Psi(x,y,z)[x_2,y_2,z_2]$. Hence $\langle x_1,y_1,z_1\rangle\mathcal{Z}\langle x_2,y_2,z_2\rangle$.

(v) Similar to (iv).

So we have a trisimulation $\mathcal{Z}$ between $\mathcal{M}_3^*$ and $\mathcal{M}_4^*$. Now consider the following diagram:

$$\begin{array}{ccccccc}\mathcal{M}_1 & & & & & & \mathcal{M}_2\\ \vdots & & & & & & \vdots\\ \mathcal{M}_1^* & \prec & \Pi_U\mathcal{M}_1^* = \mathcal{M}_3^* & \text{———} & \mathcal{M}_4^* = \Pi_U\mathcal{M}_2^* & \succ & \mathcal{M}_2^*.\end{array}$$

We have

$$\begin{aligned}\mathcal{M}_2^* \models \varphi^* & \Rightarrow \mathcal{M}_4^* \models \varphi^*\\ & \Rightarrow \mathcal{M}_3^* \models \varphi^*, \text{ by invariance for trisimulations,}\\ & \Rightarrow \mathcal{M}_1^* \models \varphi^*\\ & \Rightarrow \mathcal{M}_1 \models \varphi[w].\end{aligned}$$

■

2.8 Remark. The starred line in the above proof is precisely where we use the fact that we are working in $\mathcal{L}_{pi}$. For this is where we use the fact that we have more than one free variable available, and the fact that $\exists z\,(yS_xz \wedge \varphi(x,y,z)) \in \mathcal{L}_{pi}$.

2.9 Remark. Since finite models $\mathcal{M}_1$, $\mathcal{M}_2$ are always saturated, we have, as a special case of the above proof, that the following are equivalent:

(i) $\mathcal{M}_1 \models A[w_1]$ iff $\mathcal{M}_2 \models A[w_2]$, for all $A \in \mathcal{L}_{pi}$;
(ii) there exists a trisimulation between $\mathcal{M}_1$ and $\mathcal{M}_2$ that connects w_1 and w_2.

We end this chapter with a corollary to Theorem 2.7:

2.10 Proposition. *Let* $\mathbf{M}$ *be a class of* $\mathcal{L}(\Box,\rhd)$*-models. Then* $\mathbf{M} = \{\,\mathcal{M} \,|\, \mathcal{M} \models A\,\}$ *for some* $A \in \mathcal{L}_{pi}$ *iff* $\mathbf{M}$ *is closed under trisimulations and ultraproducts, while its complement is closed under ultraproducts.*

Proof. If $\mathbf{M} = \{\,\mathcal{M} \,|\, \mathcal{M} \models A\,\}$ for some $A \in \mathcal{L}_{pi}$, then $\mathbf{M}$ is closed under trisimulations and ultraproducts. The complement of $\mathbf{M}$ is defined by $\neg\forall x\,\tau(A)$, hence closed under ultraproducts by the Łoś-theorem.

Conversely, suppose **M** and its complement satisfy the stated conditions. Since **M** is closed under trisimulations, it and its complement are closed under isomorphisms. So there is an $\mathcal{L}_1$-sentence φ such that for all models $\mathcal{M}$, $\mathcal{M} \in \mathbf{M}$ iff $\mathcal{M} \models \varphi$ (cf. Chang and Keisler [1973], Corollary 6.1.16). We may safely assume that $\varphi \equiv \forall x\, \psi$, for some ψ. Now, **M** is closed under trisimulations, so ψ is invariant for trisimulations. Therefore, the Theorem yields an $A \in \mathcal{L}_{pi}$ such that

$$\models \forall x\, (\tau(A) \leftrightarrow \psi)$$
$$\models \forall x\, \tau(A) \leftrightarrow \forall x\, \psi$$
$$\models \forall x\, \tau(A) \leftrightarrow \varphi.$$

So $\mathcal{M} \in \mathbf{M} \iff \mathcal{M} \models \varphi \iff \mathcal{M} \models \forall x\, \tau(A)$, and A defines **M**. ■

The End.

References

van Benthem, J.

[1985] *Modal Logic and Classical Logic*, Bibliopolis, Naples.

[1989] Modal Logic as a Theory of Information, *ITLI Prepublication Series LP-89-05*, University of Amsterdam.

Berarducci, A.

[1989] The Interpretability Logic of Peano Arithmetic, manuscript.

Chang, C. & J. Keisler

[1973] *Model Theory*, North-Holland, Amsterdam.

de Jongh, D. & F. Veltman

[1990] Provability Logics for Relative Interpretability, *ITLI Prepublication Series ML-88-03*, University of Amsterdam. To appear in the *Proceedings of the 1988 Heyting Colloquium*, Plenum Press, Boston.

de Jongh, D. & A. Visser

[1989] Explicit Fixed Points for Interpretability Logic, *ITLI Prepublication Series ML-89-01*, University of Amsterdam.

Jónsson, E. & A. Tarski

[1951] Boolean Algebras with Operators. Part I, *Am. J. Math.* **73**, 891-939.

Montagna, F. & P. Hájek

[1989] *ILM* is the Logic of Π_1^0-conservativity, preprint Siena.

Pudlák, P.

[1985] Cuts, Consistency Statements and Interpretations, *J. Symbolic Logic* **50**, 423-441.

de Rijke, M.

[1989] The Modal Theory of Inequality, *ITLI Prepublicaton Series X-89-05*, University of Amsterdam.

Smoryński, C.

[1985a] *Self-Reference and Modal Logic*, Springer-Verlag, New York.

[1985b] Nonstandard Models and Related Developments in the Work of Harvey Friedman, in: Harrington, L.A., M.D. Morley, A. Ščedrov & S.G. Simpson (eds.), *Harvey Friedman's Research on the Foundations of Mathematics*, North-Holland, Amsterdam, 212-229.

[1989] Letter to Dick de Jongh, April 15, 1989.

Švejdar, V.

[1988] Some Independence Results in Interpretability Logic, preprint.

Visser, A.

[1988] Preliminary Notes on Interpretability Logic, *Logic Group Preprint Series* No. 14, Department of Philosophy, University of Utrecht.

[1989] The Formalization of Interpretability, *Logic Group Preprint Series* No. 47, Department of Philosophy, University of Utrecht.

[1990] Interpretability Logic, to appear in the *Proceedings of the 1988 Heyting Colloquium*, Plenum Press, Boston.

Wilkie, A.J. & J.B. Paris

[1987] On the Scheme of Induction for Bounded Arithmetic Formulas, *Annals of Pure and Applied Logic* **35**, 261-302.

The ITLI Prepublication Series

1990

Logic, Semantics and Philosophy of Language

LP-90-01 Jaap van der Does — A Generalized Quantifier Logic for Naked Infinitives

Computation and Complexity Theory

CT-90-01 John Tromp, Peter van Emde Boas — Associative Storage Modification Machines

Other Prepublications

X-90-01 A.S. Troelstra — Remarks on Intuitionism and the Philosophy of Mathematics, Revised Version

X-90-02 Maarten de Rijke — Some Chapters on Interpretability Logic