Notre Dame Journal of Formal Logic
Volume 33, Number 2, Spring 1992

Unary Interpretability Logic

MAARTEN de RIJKE*

Abstract Let T be an arithmetical theory. We introduce a unary modal operator '**I**' to be interpreted arithmetically as the unary interpretability predicate over T. We present complete axiomatizations of the (unary) interpretability principles underlying two important classes of theories. We also prove some basic modal results about these new axiomatizations.

1 Introduction The language $\mathcal{L}(\Box)$ of propositional modal logic consists of a countable set of proposition letters $p_0, p_1, \ldots$, and connectives $\neg$, $\wedge$, and $\Box$. $\mathcal{L}(\Box, \rhd)$ is the langauge of (binary) interpretability logic, and extends $\mathcal{L}(\Box)$ with a binary operator '$\rhd$'. ('$A \rhd B$' is read: 'A interprets B'.) The *provability logic L* is propositional logic plus the axiom schemas $\Box(A \to B) \to (\Box A \to \Box B)$, $\Box A \to \Box\Box A$, and $\Box(\Box A \to A) \to \Box A$, and the rules Modus Ponens ($\vdash A, \vdash A \to B \Rightarrow \vdash B$) and Necessitation ($\vdash A \Rightarrow \vdash \Box A$). The *binary interpretability logic IL* is obtained from L by adding the axioms

(*J*1) $\Box(A \to B) \to A \rhd B$
(*J*2) $(A \rhd B) \wedge (B \rhd C) \to (A \rhd C)$
(*J*3) $(A \rhd C) \wedge (B \rhd C) \to (A \vee B) \rhd C$
(*J*4) $A \rhd B \to (\Diamond A \to \Diamond B)$
(*J*5) $\Diamond A \rhd A$,

where $\Diamond \equiv \neg\Box\neg$. *IL* is taken as the base system; extensions of *IL* with one or more of the following schemas have also been studied:

(*F*) $A \rhd \Diamond A \to \Box\neg A$
(*W*) $A \rhd B \to A \rhd (B \wedge \Box\neg A)$
(*M*$_0$) $A \rhd B \to (\Diamond A \wedge \Box C) \rhd (B \wedge \Box C)$
(*P*) $A \rhd B \to \Box(A \rhd B)$
(*M*) $A \rhd B \to (A \wedge \Box C) \rhd (B \wedge \Box C)$.

*Research supported by the Netherlands Organization for Scientific Research (NWO).

Received September 17, 1990; revised March 4, 1991

We use ILX to denote the system $IL + X$, where X is the name of some axiom schema. $ILMP$ denotes the system $IL + M + P$ plus the additional axiom $A \rhd B \rightarrow A \wedge (C \rhd D) \rhd B \wedge (C \rhd D)$. Let ILS be one of the systems introduced above; the system ILS^{ω} has as axioms all theorems of ILS plus all instances of the schema of *reflection*: $\Box A \rightarrow A$. Its sole rule of inference is Modus Ponens.

Recall that an L-frame is a pair $\langle W, R\rangle$ with $R \subseteq W^2$ transitive and conversely well-founded, and that an L-model is given by an L-frame $\mathfrak{F}$ together with a forcing relation $\Vdash$ that satisfies the usual clauses for $\neg$ and $\wedge$, while $u \Vdash \Box A$ iff $\forall v(uRv \Rightarrow v \Vdash A)$. A (*Veltman-*) *frame* for IL is a triple $\langle W, R, S\rangle$, where $\langle W, R\rangle$ is an L-frame, and $S = \{S_w : w \in W\}$ is a collection of binary relations on W satisfying

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

An *IL-model* is given by a Veltman-frame $\mathfrak{F}$ for IL together with a forcing relation $\Vdash$ that satisfies the above clauses for $\neg$, $\wedge$, and $\Box$, where

$$u \Vdash A \rhd B \Leftrightarrow \forall v(uRv \text{ and } v \Vdash A \Rightarrow \exists w(vS_u w \text{ and } w \Vdash B)).$$

An ILP-model is an IL-model that satisfies the extra condition: if $wRw'RuS_w v$ then $uS_{w'} v$. An ILM-model is an IL-model satisfying the extra condition: if $uS_w vRz$ then uRz. A model is an $ILMP$-model if it is both an ILM- and an ILP-model, and it also satisfies the condition: if $xRyS_x zRuS_y v$ then $uS_z v$.

In the sequel, T denotes a theory which has a reasonable notion of natural numbers and finite sequences. The theories we consider are either Σ_1^0-sound essentially reflexive theories (like PA), or Σ_1^0-sound finitely axiomatized sequential theories (like GB).

An *arithmetical interpretation* $(\cdot)^T$ of $\mathcal{L}(\Box, \rhd)$ in the language of T is a map which assigns to every proposition letter p a sentence p^T in the language of T, and which is defined on other modal formulas as follows:

1. $(\bot)^T$ is '$0 = 1$';
2. $(\cdot)^T$ commutes with $\neg$ and $\wedge$;
3. $(\Box A)^T$ is a formalization of '$T \vdash (A)^T$';
4. $(A \rhd B)^T$ is a formalization of '$T + (A)^T$ interprets $T + (B)^T$'.

So the operator $\rhd$ is interpreted arithmetically as the *binary* interpretability predicate over T. Interpretability over T may also be studied as a *unary* predicate on finite extensions of T. Obviously, the modal analysis of the unary interpretability predicate in the spirit of Solovay's analysis of provability has to be undertaken using a *unary* modal operator. It was Craig Smoryński who first introduced an operator to be interpreted as the unary interpretability predicate. (The present investigations were inspired by questions of his.) Švejdar was subsequently the first one to introduce a binary operator to be interpreted as the binary interpretability relation.

It is clear that interpretability as a binary relation is the more basic notion, since unary interpretability is reducible to it; moreover, in the sequel it will become clear that the modal language with $\rhd$ is more expressive in important ways than its unary reduct. On the modal side the reduction of unary to binary interpretability leads to the following definition:

Definition 1.1 Define in $\mathcal{L}(\Box, \rhd)$ the unary interpretability operator '**I**' by $\mathbf{I}A := \top \rhd A$, and let $\mathcal{L}(\Box, \mathbf{I})$ extend $\mathcal{L}(\Box)$ with **I**.

So $x \Vdash \mathbf{I}A$ iff $\forall y(xRy \rightarrow \exists z(yS_x z \wedge z \Vdash A))$. And given a theory T, it follows from the definition of an arithmetical interpretation that $(\mathbf{I}A)^T$ is a formalization of '$T + (A)^T$ is interpretable in T'.

Definition 1.2 The *unary interpretability logic il* is obtained from the provability logic L by adding the axioms

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

Several axioms have special names:

(f) $\mathbf{I}\Diamond\top \rightarrow \Box\bot$
(m) $\mathbf{I}A \rightarrow \mathbf{I}(A \wedge \Box\bot)$
(p) $\mathbf{I}A \rightarrow \Box\mathbf{I}A$.

We use *ilm* to denote the system $il + m$, and *ilp* to denote $il + p$. For other axiom schemas S we will simply refer to $ILS \cap \mathcal{L}(\Box, \mathbf{I})$ as *ils*. Let *ils* be one of the systems *il*, *ilm* or *ilp*. The system ils^ω has as axioms all theorems of *ils* plus all instances of the schema of *reflection*: $\Box A \rightarrow A$. Its sole rule of inference is Modus Ponens.

In Section 2 we prove that $il = IL \cap \mathcal{L}(\Box, \mathbf{I})$, $ilm = ILM \cap \mathcal{L}(\Box, \mathbf{I})$, and $ilp = ILP \cap \mathcal{L}(\Box, \mathbf{I})$ – thereby establishing that *ilp* is the unary interpretability logic of all finitely axiomatized sequential theories that extend $I\Delta_0$ + SupExp, and that *ilm* is the unary interpretability logic of all essentially reflexive theories. It will turn out that *ilm* is in fact the unary interpretability logic of all "reasonable" arithmetical theories. We end Section 2 with some remarks on the hierarchy of extensions of *il*.

Next, in Section 3 we study the closed fragment of $\mathcal{L}(\Box, \mathbf{I})$ and investigate the modalities in this language. We then state and prove Interpolation Theorems for *il*, *ilm*, and *ilp*. From these we obtain Fixed Point Theorems for these logics in a standard way.

We end this section with two useful propositions. Let *ils* be one of the systems *il*, *ilm*, or *ilp*, and let *ILS* be the corresponding binary system. We first show that $ils \subseteq ILS \cap \mathcal{L}(\Box, \mathbf{I})$:

Proposition 1.3 *Let* $A \in \mathcal{L}(\Box, \mathbf{I})$. *If* $ils \vdash A$ *then* $ILS \vdash A$.

Proof: It suffices to show that for $S = \top, P, M$, we have $ILS \vdash ils$. We only show that $IL \vdash I1$ and that $ILM \vdash m$.

By $J1$, $J5$, and $J3$ we have

$$IL \vdash (\Box\bot \vee \Diamond\Box\bot) \rhd \Box\bot. \quad (1)$$

Furthermore

$$IL \vdash \Box(\top \to (\top \wedge \Box\bot) \vee \Diamond(\top \wedge \Box\bot)) \Rightarrow IL \vdash \Box(\top \to \Box\bot \vee \Diamond\Box\bot)$$
$$\Rightarrow IL \vdash \top \rhd (\Box\bot \vee \Diamond\Box\bot), \text{ by } J1$$
$$\Rightarrow IL \vdash \top \rhd \Box\bot, \text{ by } J2 \text{ and } (1).$$

To prove that $ILM \vdash m$, we use the fact that in ILM we can derive $A \rhd B \to A \rhd B \wedge \Box\neg A$ (cf. Visser [11]). Therefore $ILM \vdash m$.

Here are some theorems and a derived rule of the unary systems:

Proposition 1.4
(a) *If* $il \vdash A$ *then* $il \vdash \mathbf{I}A$. *In particular,* $il \vdash \mathbf{I}\top$.
(b) $il \vdash \Box A \to \mathbf{I}A$.
(c) $il \vdash \mathbf{I}A \to \mathbf{I}(A \wedge \Box\neg A)$.
(d) $il + f \subseteq ilm \subseteq ilp$.

Proof: Items (a), (b), and (c) are left to the reader. To prove (d), note

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

That is, $ilp \vdash m$. This establishes the inclusion $ilm \subseteq ilp$. The inclusion $il + f \subseteq ilm$ is immediate.

Assuming that *il* does indeed axiomatize $IL \cap \mathcal{L}(\Box, \mathbf{I})$, we find that $\vdash \mathbf{I}A \Rightarrow \vdash A$ 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 **Completeness*** In this section we prove *il* to be modally complete with respect to finite *IL*-models. We also prove modal and arithmetical completeness results for *ilm* and *ilp*. To prove the arithmetical completeness of *ilm* (*ilp*) we first show that *ilm* (*ilp*) is modally complete with respect to *ILM*- (*ILP*)-models; after that we appeal to the existing arithmetical completeness results for *ILM* (*ILP*).

*2.1 **Preliminaries*** Our modal completeness proofs use *infinite* maximal consistent sets instead of the finite ones used, for example, to prove *L* or *IL* complete (in Smoryński [8] and de Jongh and Veltman [2], respectively). Our approach has the advantage that it can do without the large adequate sets employed there. In this subsection we establish some results that will provide us with the building blocks for constructing countermodels in our modal completeness proofs.

We start with some definitions. For the remainder of this subsection let *ils* denote *il*, *ilm*, or *ilp*.

Definition 2.1 Let Γ, Δ be two maximal *ils*-consistent sets:

1. Δ is called a *successor* of $\Gamma\,(\Gamma < \Delta)$ if
   (a) $A \in \Delta$ for each $\Box A \in \Gamma$
   (b) $\Box A \in \Delta$ for some $\Box A \notin \Gamma$
2. Δ is called a *C-critical successor* of $\Gamma\,(\Gamma <_C \Delta)$ if
   (a) $\Gamma < \Delta$
   (b) $\mathbf{I}C \notin \Gamma$
   (c) $\neg C, \Box\neg C \in \Delta$.

Note that if $\Gamma <_C \Delta < \Delta'$ then $\Gamma <_C \Delta'$; and if $\Gamma < \Delta$ then $\Gamma <_\bot \Delta$.

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

1. if $B \in \Phi$, and C is a subformula of B, then $C \in \Phi$
2. if $B \in \Phi$, and B is no negation, then $\neg B \in \Phi$.

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

Proposition 2.3 *Let Γ be a maximal ils-consistent set such that $\Diamond C \in \Gamma$. Then there is a maximal ils-consistent $\Delta > \Gamma$ with $C, \Box\neg C \in \Delta$.*

Proof: Well-known (or cf. [8]).

Proposition 2.4 *Let Γ be a maximal ils-consistent set with $\neg\mathbf{I}C \in \Gamma$. Then there is a maximal ils-consistent Δ with $\Gamma <_C \Delta$ and $\Box\bot \in \Delta$.*

Proof: Let Δ be a maximal consistent extension of

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

Note that if such a Δ exists, it must be a C-critical successor of Γ: since

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

it is a successor of Γ; and because $\{\neg C, \Box\neg C\} \subseteq \Delta$ it is also C-critical.

We have only to prove $\{D : \Box D \in \Gamma\} \cup \{\neg C\} \neg \{\Box\bot\}$ consistent, since $\Box\bot$ implies $\Box\neg C$. 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

$$D_1, \ldots, D_m \vdash \Box\bot \to C \Rightarrow \Box D_1, \ldots, \Box D_m \vdash \Box(\Box\bot \to C)$$
$$\Rightarrow \Box D_1, \ldots, \Box D_m \vdash \mathbf{I}C, \text{ by } I1 \text{ and } I3.$$

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

Proposition 2.5 *Assume that $\mathbf{I}C \in \Gamma$, and that Δ is a maximal ils-consistent set with $\Gamma <_E \Delta$. Then there is a maximal ils-consistent set Δ' with $\Gamma <_E \Delta'$ such that $C, \Box\neg C \in \Delta'$.*

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

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

so

$$D_1, \dots, D_n \vdash C \wedge \Box\neg C \to E \vee \Diamond E$$
$$\Box D_1, \dots, \Box D_n \vdash \Box(C \wedge \Box\neg C \to E \vee \Diamond E)$$
$$\Gamma \vdash \Box(C \wedge \Box\neg C \to E \vee \Diamond E). \quad (2)$$

Since $\mathbf{I}C \in \Gamma$, it follows from 1.4 that $\mathbf{I}(C \wedge \Box\neg C) \in \Gamma$. By (2) and *I2* it follows that $\Gamma \vdash \mathbf{I}(E \vee \Diamond E)$, which, by *I3*, implies $\Gamma \vdash \mathbf{I}E$ and $\mathbf{I}E \in \Gamma$—but this contradicts the fact that $\mathbf{I}E \notin \Gamma$ by the existence of an E-critical successor of Γ.

2.2 Modal completeness of il Given some (infinite) maximal *il*-consistent set Γ and a finite adequate set Φ, we define the structure $\langle W_\Gamma, R\rangle$, which consists of pairs $\langle \Delta, \tau\rangle$. Here, the maximal consistent sets Δ are needed to handle the truth definition for formulas in $\Gamma \cap \Phi$. And the sequences of (pairs of) formulas τ are used to carefully index the pairs we add to W_Γ. In this way we make sure that $\langle W_\Gamma, R\rangle$ will be a finite tree.

For the time being, let Γ be an infinite maximal *il*-consistent set, and let Φ be a finite adequate set. We use $\overline{w}, \overline{v}, \dots$ to denote pairs $\langle \Delta, \tau\rangle$. If $\overline{w} = \langle \Delta, \tau\rangle$, then $(\overline{w})_0 = \Delta$, $(\overline{w})_1 = \tau$. 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. Finally, $(\overline{w})_1{}^\frown(\overline{v})_1$ denotes the concatenation of $(\overline{w})_1$ and $(\overline{v})_1$.

Definition 2.6 Define W_Γ to be a minimal set of pairs $\langle \Delta, \tau\rangle$ such that:

1. $\langle \Gamma, \langle\langle\ \rangle\rangle\rangle \in W_\Gamma$
2. 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 *il*-consistent set Δ' with $\Delta <_C \Delta'$ and $B, \Box\neg B \in \Delta'$, 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 $\overline{w}R\overline{v}$ iff $(\overline{w})_1 \subset (\overline{v})_1$. Define S on W_Γ by putting $\overline{v}S_{\overline{w}}\overline{u}$ iff for some B, B', C, τ, and σ:

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

Remark 2.7 In 2.6 the pairs $\langle B, C\rangle$ code the following: if $\langle \Delta', \tau^\frown\langle\langle B, C\rangle\rangle\rangle \in W_\Gamma$, then for some $\langle \Delta, \tau\rangle \in W_\Gamma$, Δ' is a C-critical successor of Δ, and $\langle \Delta', \tau^\frown\langle\langle B, C\rangle\rangle\rangle$ was added to W_Γ because $\Diamond B \in \Delta$ is almost in Φ.

Proposition 2.8
(a) *W_Γ is finite.*
(b) *If $(\overline{w})_1 = (\overline{v})_1$ then $\overline{w} = \overline{v}$.*
(c) *If $\overline{w}R\overline{v}$ then $(\overline{w})_0 < (\overline{v})_0$.*
(d) *$\langle W_\Gamma, R\rangle$ is a tree.*
(e) *$\langle W_\Gamma, R, S\rangle$ is an IL-frame.*
(f) *If $\langle \Delta, \tau\rangle \in W_\Gamma$ and E occurs as the second component in some pair in τ, then $\neg E$, $\Box\neg E \in \Delta$.*

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

(b) Induction on $\mathrm{lh}((\overline{w})_1) = \mathrm{lh}((\overline{v})_1)$.

(c) Fix w arbitrarily and prove the claim by induction on $\mathrm{lh}(\overline{v})$ where $\overline{w}R\overline{v}$.

(d) To prove that $\langle W_\Gamma, R\rangle$ is a tree, note first that transitivity and asymmetry are straightforward, so we prove only that for each $\overline{w} \in W_\Gamma$ the set of its R-predecessors is finite and linear. Finiteness is immediate by (a). To prove linearity, assume that $\overline{u}R\overline{w}$ and $\overline{v}R\overline{w}$. Then $(\overline{u})_1 \subset (\overline{w})_1$ and $(\overline{v})_1 \subset (\overline{w})_1$, so $(\overline{u})_1 \subseteq (\overline{v})_1$ or $(\overline{v})_1 \subseteq (\overline{u})_1$. If $(\overline{u})_1 = (\overline{v})_1$ then $\overline{u} = \overline{v}$ by (b), and we are done. If $(\overline{u})_1 \neq (\overline{v})_1$ then either $(\overline{u})_1 \subseteq (\overline{v})_1$ or $(\overline{v})_1 \subset (\overline{u})_1$, that is: $\overline{u}R\overline{v}$ or $\overline{v}R\overline{u}$.

(e) Left to the reader.

(f) Induction on the construction of W_Γ.

Theorem 2.9 *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: Proving soundness is left to the reader. To prove completeness, assume that $il \nvdash A$. We want to produce an IL-model that refutes A. Let Φ be a finite adequate set containing $\neg A$, and let Γ be a maximal il-consistent set containing $\neg A$. Construct $\langle W_\Gamma, R, S\rangle$ as in 2.6. We complete the proof by putting $\overline{w} \Vdash p$ iff $p \in (\overline{w})_0$ and by proving that for all $F \in \Phi$ and $\overline{w} \in W_\Gamma$ we have $\overline{w} \Vdash F$ iff $F \in (\overline{w})_0$. The proof is by induction on F. We only consider the cases $F \equiv \Diamond B$ and $F \equiv \mathbf{I}C$.

If $F \equiv \Diamond B \in (\overline{w})_0$ we have to show that $\exists \overline{v}(\overline{w}R\overline{v} \wedge B \in (\overline{v})_0)$. Note first that $\Diamond B$ is almost in Φ, and that $\bot \in \Phi$. By 2.3 there is a successor Δ of $(\overline{w})_0$ with B, $\Box\neg B \in \Delta$. Moreover, Δ is a $\bot$-critical successor of $(\overline{w})_0$. Put $\overline{v} := \langle \Delta, (\overline{w})_1{}^\frown\langle\langle B, \bot\rangle\rangle\rangle$. Then we may assume that $\overline{v} \in W_\Gamma$. It is clear that $\overline{w}R\overline{v}$ and $B \in (\overline{v})_0$ as required.

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

Assume $\mathbf{I}C \notin (\overline{w})_0$. Then $\neg\mathbf{I}C \in (\overline{w})_0$, and $\Diamond\top \in (\overline{w})_0$. By the induction hypothesis we have to show that $\exists \overline{v}(\overline{w}R\overline{v} \wedge \forall \overline{u}(\overline{v}S_{\overline{w}}\overline{u} \to \neg C \in (\overline{u})_0))$. Apply 2.4, with $\Gamma = (\overline{w})_0$, to obtain a Δ with $(\overline{w})_0 <_C \Delta$, and define $\overline{v} := \langle \Delta, (\overline{w})_1{}^\frown\langle\langle \top, C\rangle\rangle\rangle$. Since $\Diamond\top \in (\overline{w})_0$ is almost in Φ, we may assume that $\overline{v} \in W_\Gamma$. Furthermore, if $\overline{v}S_{\overline{w}}\overline{u}$ then C occurs as the second component in some pair in $(\overline{u})_1$, hence $\neg C \in (\overline{u})_0$, by 2.8(e).

Assume $\mathbf{I}C \in (\overline{w})_0$. By the induction hypothesis we have to show that $\forall \overline{v}(\overline{w}R\overline{v} \to \exists \overline{u}(\overline{v}S_{\overline{w}}\overline{u} \wedge C \in (\overline{u})_0))$. So let $\overline{v} \in \overline{w}R$. Then $(\overline{v})_0 \succ (\overline{w})_0$ by 2.8(c), so $\Diamond\top \in (\overline{w})_0$, and therefore $\Diamond C \in (\overline{w})_0$ by Axiom $I4$. By construction $(\overline{v})_0$ is E-critical for some $E \in \Phi$. Now, apply 2.5, with $\Gamma = (\overline{w})_0$, $\Delta = (\overline{v})_0$, to obtain a Δ' with $(\overline{w})_0 <_E \Delta'$ that contains C, $\Box\neg C$. Since $\Diamond C$ is almost in Φ, we may assume that $\overline{u} = \langle \Delta', (\overline{w})_1{}^\frown\langle\langle C, E\rangle\rangle\rangle \in W_\Gamma$. Clearly, $\overline{u}$ does the job.

Proposition 2.10 *Let $A \in \mathcal{L}(\Box, \mathbf{I})$. Then $IL \vdash A$ iff $il \vdash A$.*

Proof: By [2] we have for all $A \in \mathcal{L}(\Box, \rhd)$, $IL \vdash A$ iff for all finite IL-models $\mathcal{M}$, $\mathcal{M} \models A$. From this and 2.9 the proposition follows.

2.3 Modal and arithmetical completeness of ilm To prove the modal completeness of *ilm* we need to adapt the construction used in proving *il* complete somewhat. The countermodel we will construct in the completeness proof will consist of pairs $\langle \Delta, \tau \rangle$, where Δ is a maximal *ilm*-consistent set, and τ is a sequence of triples of formulas.

For the time being we fix a maximal *ilm*-consistent set Γ and a finite adequate set Φ.

Definition 2.11 Define W_Γ to be a minimal set of pairs $\langle \Delta, \tau \rangle$ such that

1. $\langle \Gamma, \langle\langle\ \rangle\rangle\rangle \in W_\Gamma$.
2. If $\langle \Delta, \tau \rangle \in W_\Gamma$, $\Diamond B \in \Delta \cap (\Phi \cup \{\Diamond \top\})$, $C \in \Phi$ and if there exists a maximal *ilm*-consistent set Δ' with $\Delta <_C \Delta'$ and $B, \Box\neg B \in \Delta'$, then for *one* such Δ', $\langle \Delta', \tau^\frown\langle\langle B, \bot, C\rangle\rangle\rangle \in W_\Gamma$.
3. If $\langle \Delta, \tau \rangle \in W_\Gamma$, $\mathbf{I}B \in \Delta \cap \Phi$, $C \in \Phi$ and if there exists a maximal *ilm*-consistent set Δ' with $\Delta <_C \Delta'$ and $B, \Box\bot \in \Delta'$, then $\langle \Delta', \tau^\frown\langle\langle \bot, B, C\rangle\rangle\rangle \in W_\Gamma$, for *one* such Δ'.

Define R on W_Γ by putting $\bar{w}R\bar{v}$ if $(\bar{w})_1 \subset (\bar{v})_1$. Define S on W_Γ by putting $\bar{v}S_{\bar{w}}\bar{u}$ iff for some B, B', E, E', C, σ and σ'

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

and

$$\text{if } B \equiv \bot \text{ then } B' \equiv \bot,$$

and

$$\text{if } E' \equiv \bot \text{ then } B' \equiv B,\ E' \equiv E \text{ and } \sigma \subseteq \sigma'.$$

Remark 2.12 In 2.11 the triples $\langle B, E, C\rangle$ code the following: if $\langle \Delta', \tau^\frown\langle\langle B, E, C\rangle\rangle\rangle \in W_\Gamma$, then there is some $\langle \Delta, \tau \rangle \in W_\Gamma$ such that Δ' is a C-critical successor of Δ, and if $B \not\equiv \bot$ then $\langle \Delta', \tau^\frown\langle\langle B, E, C\rangle\rangle\rangle$ was added to W_Γ because $\Diamond B \in \Delta \cap (\Phi \cup \{\Diamond\top\})$; if $B \equiv \bot$ then $E \not\equiv \bot$ and $\langle \Delta', \tau^\frown\langle\langle B, E, C\rangle\rangle\rangle$ was added to W_Γ because $\mathbf{I}E \in \Delta \cap \Phi$.

Proposition 2.13
(a) *W_Γ is finite.*
(b) *If $(\bar{v})_1 = (\bar{w})_1{}^\frown\langle\langle B, E, C\rangle\rangle{}^\frown\sigma$ then either $B \equiv \bot$ or $E \equiv \bot$ (but not both); and if $B \equiv \bot$ then $\Box\bot \in (\bar{v})_0$ and $\sigma = \langle\ \rangle$.*
(c) *If $(\bar{w})_1 = (\bar{v})_1$ then $\bar{w} = \bar{v}$.*
(d) *If $\bar{w}R\bar{v}$ then $(\bar{w})_0 < (\bar{v})_0$.*
(e) *$\langle W_\Gamma, R, S\rangle$ is an ILM-frame.*
(f) *If $\bar{v} = \langle \Delta, \tau \rangle \in W_\Gamma$ and C occurs as the third component in some triple in τ then $\neg C$, $\Box\neg C \in \Delta$.*

Proof: Items (a), (b), (c), (d), and (f) are left to the reader. Let us check that $\langle W_\Gamma, R, S\rangle$ satisfies all the conditions to be an *ILM*-frame:

- it is easily seen that R is transitive and irreflexive—so by (a) it is also conversely well-founded;
- $S_{\bar{w}} \subseteq \bar{w}R \times \bar{w}R$ is immediate;
- to show that $S_{\bar{w}}$ is reflexive and transitive, use (b);
- to show that $\bar{w}R\bar{v}R\bar{u}$ implies $\bar{v}S_{\bar{w}}\bar{u}$, use (b);

- finally, we have to show that $\bar{v}S_{\bar{w}}\bar{u}R\bar{z}$ implies $\bar{v}R\bar{z}$; so assume that $\bar{v}S_{\bar{w}}\bar{u}$. By definition there are B, B', B'', E, E', E'', C, C'', σ, σ', and σ'' such that

$$(\bar{v})_1 = (\bar{w})_1{}^\frown\langle\langle B, E, C\rangle\rangle{}^\frown\sigma$$
$$(\bar{u})_1 = (\bar{w})_1{}^\frown\langle\langle B', E', C\rangle\rangle{}^\frown\sigma'$$
$$(\bar{z})_1 = (\bar{w})_1{}^\frown\langle\langle B', E', C\rangle\rangle{}^\frown\sigma'{}^\frown\langle\langle B'', E'', C''\rangle\rangle{}^\frown\sigma''.$$

Obviously, $B' \not\equiv \bot$, for otherwise, by (b), $\Box\bot \in (\bar{u})_0$, and, by (d) $\bot \in (\bar{z})_0$. Therefore, by (b), $E' \equiv \bot$ — but then $B \equiv B'$, $E \equiv E'$, and $\sigma \subseteq \sigma'$. In other words: $(\bar{v})_1 \subset (\bar{z})_1$, which means that $\bar{v}R\bar{z}$.

Theorem 2.14 *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: As before we prove only completeness. Assume *ilm* $\nvdash A$. Let Φ be a finite adequate set that contains $\neg A$, and let Γ be a maximal *ilm*-consistent set with $\neg A \in \Gamma$. Construct $\langle W_\Gamma, R, S\rangle$ as in 2.11. Define a forcing relation $\Vdash$ on $\langle W_\Gamma, R, S\rangle$ by putting $\bar{w} \Vdash p$ iff $p \in (\bar{w})_0$. As before, we prove by induction on F that for all $F \in (\Phi \cup \{\Diamond\top\})$ and $\bar{w} \in W_\Gamma$ we have $\bar{w} \Vdash F$ iff $F \in (\bar{w})_0$. We consider only the case $F \equiv \mathbf{I}B$. (The case $F \equiv \Diamond B$ is similar to the corresponding case in the proof of 2.9.)

The case that $F \equiv \mathbf{I}B \notin (\bar{w})_0$ is entirely analogous to the corresponding case in the proof of 2.9.

Assume that $F \equiv \mathbf{I}B \in (\bar{w})_0$. By the induction hypothesis we have to show that $\forall\bar{v}(\bar{w}R\bar{v} \to \exists\bar{u}(\bar{v}S_{\bar{w}}\bar{u} \wedge B \in (\bar{u})_0))$. So assume that $\bar{v} \in \bar{w}R$. Then $(\bar{v})_1 = (\bar{w})_1{}^\frown\langle\langle B', E', C\rangle\rangle{}^\frown\sigma$ for some B', E', C, and σ. By 2.13(f), $(\bar{v})_0$ is C-critical. Now $\mathbf{I}B \in (\bar{w})_0$ implies $\mathbf{I}(B \wedge \Box\bot) \in (\bar{w})_0$, by Axiom m. Apply 2.5 to find a Δ' with $(\bar{w})_0 <_C \Delta'$ and $B, \Box\bot \in \Delta'$. Since $\mathbf{I}B \in (\bar{w})_0 \cap \Phi$, we may assume that $\bar{u} := \langle\Delta', (\bar{w})_1{}^\frown\langle\langle\bot, B, C\rangle\rangle\rangle \in W_\Gamma$. Obviously, we have $\bar{v}S_{\bar{w}}\bar{u}$ and $B \in (\bar{u})_0$ as required.

Proposition 2.15 *Let $A \in \mathcal{L}(\Box, \mathbf{I})$. Then ILM $\vdash A$ iff ilm $\vdash A$.*

Proof: By [2] we have for all $A \in \mathcal{L}(\Box, \rhd)$, *ILM* $\vdash A$ iff for all finite *ILM*-models $\mathcal{M}$, $\mathcal{M} \models A$. From this and 2.14 the result follows.

Theorem 2.16 *Let $A \in \mathcal{L}(\Box, \mathbf{I})$, and let T be a Σ^0_1-sound essentially reflexive theory. Then ilm $\vdash A$ iff for all interpretations $(\cdot)^T$ of $\mathcal{L}(\Box, \mathbf{I})$ in the language of T, $T \vdash (A)^T$.*

Proof: By Berarducci ([1], Theorem 3.8) we have for all $A \in \mathcal{L}(\Box, \rhd)$, *ILM* $\vdash A$ iff for all interpretations $(\cdot)^T$ of $\mathcal{L}(\Box, \rhd)$ in the language of T, $T \vdash (A)^T$. From this and 2.15 the result follows.

Proposition 2.17 *Let $A \in \mathcal{L}(\Box, \mathbf{I})$. Then the following are equivalent:*
(a) *$ilm^\omega \vdash A$*
(b) *$ILM^\omega \vdash A$*
(c) *ilm* $\vdash (\bigwedge_{\Box B \in \mathrm{Sub}(A)}(\Box B \to B) \wedge \Diamond\top) \to A$.

Proof: The implication (a) $\Rightarrow$ (b) is trivial. By the proof of [1], Theorem 6.5, $ILM^\omega \vdash A$ implies

$$ILM \vdash \left(\bigwedge_{\Box B \in \mathrm{Sub}(A)} (\Box B \to B) \wedge \bigwedge_{C \rhd D \in \mathrm{Sub}(A)} (C \to \Diamond C) \right) \to A.$$

Since $A \in \mathcal{L}(\Box, \mathbf{I})$ this implies

$$ILM \vdash \left(\bigwedge_{\Box B \in \mathrm{Sub}(A)} (\Box B \to B) \wedge \Diamond \top \right) \to A.$$

Together with 2.15 this yields the implication (b) ⇒ (c). The implication (c) ⇒ (a) is straightforward since $ilm^{\omega} \vdash \Box B \to B$ for all $B \in \mathcal{L}(\Box, \mathbf{I})$, so in particular $ilm^{\omega} \vdash \Box \bot \to \bot$, i.e., $ilm^{\omega} \vdash \Diamond \top$.

Theorem 2.18 *Let $A \in \mathcal{L}(\Box, \mathbf{I})$, and let T be a Σ^0_1 sound essentially reflexive theory. Then $ilm^{\omega} \vdash A$ iff for all interpretations $(\cdot)^T$ of $\mathcal{L}(\Box, \mathbf{I})$ in the language of T, $(A)^T$ is true in the standard model.*

Proof: By [1], Theorem 6.5, we have $ILM^{\omega} \vdash A$ iff for all interpretations $(\cdot)^T$ of $\mathcal{L}(\Box, \rhd)$ in the language of T, $(A)^T$ is true in the standard model. By 2.17 this implies the theorem.

Proposition 2.19 *Let $A \in \mathcal{L}(\Box, \mathbf{I})$. Then $ILW \vdash A$ iff $ilm \vdash A$.*

Proof: Since m is a substitution instance of the axiom W, the direction from right to left is immediate from 2.10. Conversely, if $ilm \nvdash A$, then $ILM \nvdash A$ by 2.15. Recall from the proof of 1.3 that $ILM \vdash W$, i.e. that $ILM \supseteq ILW$. It follows that $ILW \nvdash A$.

Let us call an arithmetical theory a *reasonable* theory if it is sequential, Σ^0_1-sound, R^+_1-axiomatized, and its natural numbers satisfy $I\Delta_0 + \Omega_1$ (cf. Visser [10] for details and motivation).

Theorem 2.20 *The system ilm is the unary interpretability logic of all reasonable arithmetical theories.*

Proof: In Visser [10], Section 6.2, it is shown that ILW is valid for arithmetic interpretations in all reasonable arithmetical theories, hence by 2.19 the same holds for *ilm*. Therefore, the unary interpretability logic of all reasonable arithmetics contains *ilm*. Since, by 2.16, *ilm* is the unary interpretability logic of PA, the converse inclusion holds as well.

2.4 Modal and arithmetical completeness of ilp Instead of proving *ilp* modally complete with respect to *ILP*-models we prove a stronger result, notably the modal completeness of *ilp* with respect to *ILMP*-models. The proof of this result is a slight variation on the modal completeness proof for *ilm*.

As before, we fix a maximal *ilp*-consistent set Γ and a finite adequate set Φ.

Definition 2.21 Define W_Γ to be a minimal set of pairs $\langle \Delta, \tau \rangle$ such that

1. $\langle \Gamma, \langle\langle\ \rangle\rangle \rangle \in W_\Gamma$.
2. If $\langle \Delta, \tau \rangle \in W_\Gamma$, $\Diamond B \in \Delta \cap (\Phi \cup \{\Diamond \top\})$, $C \in \Phi$, and if there exists a maximal *ilp*-consistent set Δ' with $\Delta \prec_C \Delta'$ and B, $\Box \neg B \in \Delta'$, then for *one* such Δ', $\langle \Delta', \tau^\frown \langle\langle B, \bot, C \rangle\rangle \rangle \in W_\Gamma$.

3. If $\langle \Delta, \tau \rangle \in W_\Gamma$, $\mathbf{I}B \in \Delta \cap \Phi$, $C \in \Phi$, and if there exists a maximal *ilp*-consistent set Δ' with $\Delta <_C \Delta'$ and B, $\Box\bot \in \Delta'$, then $\langle \Delta', \tau ^\frown \langle\langle \bot, B, C\rangle\rangle\rangle \in W_\Gamma$, for *one* such Δ'.

Define R on W_Γ by putting $\bar{w}R\bar{v}$ iff $(\bar{w})_0 \subset (\bar{v})_0$. Define S on W_Γ by putting $\bar{v}S_{\bar{w}}\bar{u}$ iff for some B, B', E, E', C, τ, and σ

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

and

$$\text{if } B \equiv \bot \text{ then } B' \equiv \bot$$

and

$$\text{if } E' \equiv \bot \text{ then } B' \equiv B,\ E' \equiv E.$$

Proposition 2.22
(a) W_Γ *is finite.*
(b) *If* $(\bar{v})_1 = (\bar{w})_1{}^\frown\tau^\frown\langle\langle B,E,C\rangle\rangle^\frown\sigma$ *then either* $B \equiv \bot$ *or* $E \equiv \bot$ *(but not both); and if* $B \equiv \bot$ *then* $\Box\bot \in (\bar{v})_0$ *and* $\sigma = \langle\ \rangle$.
(c) *If* $(\bar{w})_1 = (\bar{v})_1$ *then* $\bar{w} = \bar{v}$.
(d) *If* $\bar{w}R\bar{v}$ *then* $(\bar{w})_0 < (\bar{v})_0$.
(e) $\langle W_\Gamma, R\rangle$ *is a tree.*
(f) $\langle W_\Gamma, R, S\rangle$ *is an ILMP-frame.*
(g) *If* $\bar{v} = \langle \Delta, \tau\rangle \in W_\Gamma$ *and C occurs as the third component in some triple in* τ, *then* $\neg C$, $\Box\neg C \in \Delta$.

Proof: We prove only (f). The proof that $\langle W_\Gamma, R, S\rangle$ is an *ILM*-frame is similar to the proof of 2.13(e); to prove that $\langle W_\Gamma, R, S\rangle$ is also an *ILP*-frame, we have to show that $\bar{w}R\bar{w}'R\bar{u}S_{\bar{w}}\bar{v}$ implies $\bar{u}S_{\bar{w}'}\bar{v}$—but this is immediate. So it remains to be proved that $xRyS_xzRuS_yv$ implies uS_zv. Reasoning as in 2.13(e) we find that $xRyS_xzRu$ implies $xRy\underline{R}zRu$. Now, if $y = z$ then we trivially have uS_zv, and if yRz then we have uS_zv because $\langle W_\Gamma, R, S\rangle$ is an *ILP*-frame.

Theorem 2.23 *Let* $A \in \mathcal{L}(\Box,\mathbf{I})$. *Then ilp* $\vdash A$ *iff for all finite ILMP-models* $\mathcal{M}$ *we have* $\mathcal{M} \models A$.

Proof: As before we prove only completeness. Assume that *ilp* $\nvdash A$. Let Φ be a finite adequate set that contains $\neg A$, and let Γ be a maximal *ilp*-consistent set with $\neg A \in \Gamma$. Construct $\langle W_\Gamma, R, S\rangle$ as in 2.21. Define a forcing relation $\Vdash$ on $\langle W_\Gamma, R, S\rangle$ by putting $\bar{w} \Vdash p$ iff $p \in (\bar{w})_0$. As before, we prove by induction on F that for all $F \in \Phi \cup \{\Diamond\top\}$ and $\bar{w} \in W_\Gamma$ we have $\bar{w} \Vdash F$ iff $F \in (\bar{w})_0$. The case $F \equiv \Diamond B$ is similar to the corresponding case in the proof of 2.9. So we consider only the case $F \equiv \mathbf{I}B$.

The case that $F \equiv \mathbf{I}B \notin (\bar{w})_0$ is entirely analogous to the corresponding case in the proof of 2.9.

Assume that $F \equiv \mathbf{I}B \in (\bar{w})_0$. By the induction hypothesis we have to show that $\forall\bar{v}(\bar{w}R\bar{v} \to \exists\bar{u}(\bar{v}S_{\bar{w}}\bar{u} \wedge B \in (\bar{u})_0))$. So assume that $\bar{v} \in \bar{w}R$. Since $\langle W_\Gamma, R\rangle$ is a tree, we can find a unique immediate R-predecessor $\bar{w}'$ of $\bar{v}$. By Axiom p we must have $\mathbf{I}B \in (\bar{w}')_0$, and so by Axiom m, $\mathbf{I}(B \wedge \Box\bot) \in (\bar{w}')_0$. By construction there are $B', E', C' \in \Phi$ such that $(\bar{v})_1 = (\bar{w}')_1{}^\frown\langle\langle B',E',C'\rangle\rangle$, that is: $(\bar{w}')_0 <_{C'} (\bar{v})_0$. By 2.5 there exists a Δ with $(\bar{w}')_0 <_{C'} \Delta$ and B, $\Box\bot \in \Delta$. Since

$\mathbf{I}B \in (\bar{w}')_0 \cap \Phi$, and $C' \in \Phi$ we may assume that $\bar{u} := \langle \Delta, (\bar{w}')_1 {}^\frown \langle\langle \bot, B, C' \rangle\rangle\rangle \in W_\Gamma$. Obviously, we have $\bar{v} S_{\bar{w}} \bar{u}$ and $B \in (\bar{u})_0$ as required.

Proposition 2.24 *Let* $A \in \mathcal{L}(\Box, \mathbf{I})$. *The* $ILMP \vdash A$ *iff* $ilp \vdash A$.

Proof: If $ilp \vdash A$ then by 1.3 $ILP \vdash A$, and hence $ILMP \vdash A$. Conversely, if $ILMP \vdash A$, then for all (finite) *ILMP*-models $\mathcal{M}$, $\mathcal{M} \models A$. So by 2.23, $ilp \vdash A$.

Proposition 2.25 *Let* $A \in \mathcal{L}(\Box, \mathbf{I})$. *Then* $ILP \vdash A$ *iff* $ilp \vdash A$.

Proof: The direction from right to left follows from 1.3. To prove the other direction, note that $ilp \nvdash A$ implies $ILMP \nvdash A$, by the previous proposition, and this in turn implies $ILP \nvdash A$.

Theorem 2.26 *Let* $A \in \mathcal{L}(\Box, \mathbf{I})$ *and let* T *be a* Σ_1^0*-sound finitely axiomatized sequential theory that extends* $I\Delta_0$ + SupExp. *Then* $ilp \vdash A$ *iff for all interpretations* $(\cdot)^T$ *of* $\mathcal{L}(\Box, \mathbf{I})$ *in the language of* T, $T \vdash (A)^T$.

Proof: By 2.25 we have $ilp \vdash A$ iff $ILP \vdash A$, for all $A \in \mathcal{L}(\Box, \mathbf{I})$. By [11], Theorem 8.2, this is equivalent to: for all interpretations $(\cdot)^T$ of $\mathcal{L}(\Box, \rhd)$ in the language of T, $T \vdash (A)^T$. This implies the theorem.

Proposition 2.27 *Let* $A \in \mathcal{L}(\Box, \mathbf{I})$. *Then the following are equivalent:*
(a) $ilp^\omega \vdash A$
(b) $ILP^\omega \vdash A$
(c) $ilp \vdash (\bigwedge_{\Box B \in \mathrm{Sub}(A)} (\Box B \to B) \land \Diamond \top) \to A$.

Proof: The implications (a) $\Rightarrow$ (b) and (c) $\Rightarrow$ (a) are trivial. The implication (b) $\Rightarrow$ (c) follows from 2.25.

Theorem 2.28 *Let* $A \in \mathcal{L}(\Box, \mathbf{I})$, *and let* T *be a* Δ_2*-sound finitely axiomatized sequential theory that extends* $I\Delta_0$ + SupExp. *Then* $ilp^\omega \vdash A$ *iff for all interpretations* $(\cdot)^T$ *of* $\mathcal{L}(\Box, \mathbf{I})$ *in the language of* T, $(A)^T$ *is true in the standard model.*

Proof: By de Rijke [5], Theorem 3.2, we have $ILP^\omega \vdash A$ iff for all interpretations $(\cdot)^T$ of $\mathcal{L}(\Box, \rhd)$ in the language of T, $(A)^T$ is true in the standard model. By 2.27 this yields the theorem.

2.5 On the hierarchy of extensions of il In [2], [10], and [11] the following extensions of *IL* in $\mathcal{L}(\Box, \rhd)$ are considered:

$$IL \subset ILF \subset ILW \subset ILWM_0 \subset \begin{matrix} ILP \\ ILM \end{matrix} \subset ILMP.$$

(All inclusions are proper.)

As a corollary to 2.19 and 2.24 we find that this hierarchy partly collapses when we only consider formulas $A \in \mathcal{L}(\Box, \mathbf{I})$:

$$il \subset ilf \subset ilw = ilwm_0 = ilm \subset ilp = ilmp.$$

(Recall that $ilx = ILX \cap \mathcal{L}(\Box, \mathbf{I})$.) To see that there is no total collapse we prove the following result:

Proposition 2.29
(a) *ilm* $\neq$ *ilp*
(b) *ilf* $\neq$ *ilm*
(c) *il* $\neq$ *ilf*.

Proof: Parts (a) and (c) may be proved using two simple models. To prove (b) we use a construction due to Švejdar (cf. [9]). It suffices to show that $ILF \nvdash m$. Consider Figure 1 below. We claim that $w \Vdash F$, i.e., that $w \Vdash A \rhd \Diamond A \to \Box\neg A$, for all $A \in \mathcal{L}(\Box, \mathbf{I})$. Suppose that $w \Vdash A \rhd \Diamond A$. Then

(a) if $b \Vdash A$ then $a \Vdash A$
(b) $d \nVdash A$ — otherwise $d \Vdash \Diamond A$, which is impossible
(c) for each B, $a \Vdash B$ iff $c \Vdash B$
(d) $c \nVdash A$ — otherwise $c \Vdash \Diamond A$, which is impossible
(e) $a \nVdash A$, by (c) and (d)
(f) $b \nVdash A$, by (a) and (e)
(g) $w \Vdash \Box\neg A$, by (b), (d), (e), and (f).

On the other hand, $w \nVdash \mathbf{I}A \to I(A \wedge \Box\bot)$, for we have $w \Vdash \mathbf{I}p$ while $w \nVdash \mathbf{I}(p \wedge \Box\bot)$, since b has no S_w-successor at which $p \wedge \Box\bot$ holds.

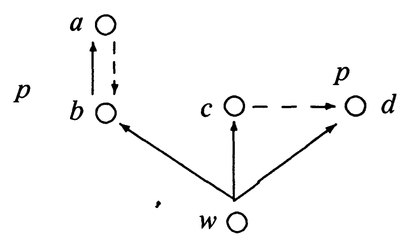

Figure 1.

(Plain arrows denote R-links; dashed arrows denote S_w-links; reflexive S-links and S-links induced by R-links have been left out.)

3 Answers to some standard questions In this section we answer some questions that come naturally with any extension of L. Notably, what are the closed formulas and the modalities in $\mathcal{L}(\mathbf{I})$ and $\mathcal{L}(\Box, \mathbf{I})$? We also prove interpolation and fixed point theorems for *il*, *ilm*, and *ilp*.

3.1 Closed formulas and modalities As usual we start with some definitions. A formula C is called *closed* if it does not contain any proposition letters. Let $\mathcal{F}$ be a frame. Define the *depth* $d(w)$ of $w \in \mathcal{F}$ by $d(w) = \sup\{d(v) + 1 : wRv\}$.

Proposition 3.1 *Let w, v be two points (not necessarily in the same model). If $d(w) = d(v)$ then $w \Vdash C$ iff $v \Vdash C$ for all closed formulas $C \in \mathcal{L}(\Box)$.*

Proof: This is by induction on $d(w) = d(v)$.

Proposition 3.2 *Let ω^* be the natural numbers with the ordering reversed, i.e., $\omega^* = \langle \mathbf{N}, > \rangle$. Let C be a closed formula in $\mathcal{L}(\Box)$. Then $L \vdash C$ iff C is valid on ω^*. (I.e., iff for every $w \in \omega^*$ and every $\Vdash$ on ω^*, $w \Vdash C$.)*

Proof: The direction from left to right is obvious. To prove the other one, assume that $L \nvdash C$; then for some finite L-model $\mathcal{M}$ with root w, $w \nVdash C$. Let $n = d(w)$, and let $\Vdash'$ be any forcing relation on ω^*. It is clear that, in ω^*, the element n has depth n. So by the previous proposition, $n \nVdash' C$.

Proposition 3.3 *Let C be a closed formula in $\mathcal{L}(\Box)$. Then $L \vdash (C \vee \Diamond C) \leftrightarrow \Diamond^k \top$, for some $k \in \omega \cup \{\omega\}$. (Here, $\Diamond^\omega \top \equiv \bot$.)*

Proof: By the previous proposition it suffices to show that for all closed formulas C in $\mathcal{L}(\Box)$, there is some $k \in \omega \cup \{\omega\}$ such that $(C \vee \Diamond C) \leftrightarrow \Diamond^k \top$ is valid on ω^*. This is left to the reader.

Proposition 3.4 *Let X be a logic that extends $il + f$. Then every closed formula in $\mathcal{L}(\mathbf{I})$ is, provably in X, equivalent to one of $\Diamond\top$, $\Box\bot$, $\bot$, or $\top$. Hence, every closed formula in $\mathcal{L}(\Box,\mathbf{I})$ is equivalent, over X, to a closed formula in $\mathcal{L}(\Box)$.*

Proof: This is by induction on the closed formula C. The only nontrivial case is $C \equiv \mathbf{I}B$, where B is a closed formula in $\mathcal{L}(\mathbf{I})$. Now by the induction hypothesis, B is a closed formula in $\mathcal{L}(\Box)$. Furthermore, $il \vdash \mathbf{I}B \leftrightarrow \mathbf{I}(B \vee \Diamond B)$. So, $il \vdash \mathbf{I}B \leftrightarrow \mathbf{I}\Diamond^k\top$, for some $k \in \omega \cup \{\omega\}$. If $k = 0$, then $\mathbf{I}\Diamond^k\top \equiv \mathbf{I}\top$, and $X \vdash \mathbf{I}B \leftrightarrow \top$. If $k = \omega$, then $\mathbf{I}\Diamond^k\top \equiv \mathbf{I}\bot$, and

$$\begin{aligned} il \vdash \mathbf{I}\bot &\to (\neg\Diamond\top \vee \Diamond\bot), \text{ by } I3 \\ &\to (\Box\bot \vee \Diamond\bot) \\ &\to \Box\bot \\ &\to \mathbf{I}\bot, \text{ by } 1.4. \end{aligned}$$

So $X \vdash C \leftrightarrow \Box\bot$. If $0 < k < \omega$, then

$$\begin{aligned} X \vdash \mathbf{I}\Diamond^k\top &\to \mathbf{I}\Diamond\top, \text{ by Axiom } \Box A \to \Box\Box A \\ &\to \Box\bot, \text{ by Axiom } f \\ &\to \Box\Diamond^k\top \\ &\to \mathbf{I}\Diamond^k\top, \text{ by } 1.4 \end{aligned}$$

So $X \vdash C \leftrightarrow \Box\bot$.

By the Normal Form Theorem for closed formulas in $\mathcal{L}(\Box)$, it follows from 3.4 that in extensions of $il + f$ every closed formula in $\mathcal{L}(\Box,\mathbf{I})$ is equivalent to a Boolean combination of formulas of the form $\Box^n\bot$, for some $n \in \omega \cup \{\omega\}$.

Below $il + f$ the situation is more complicated. Note for example that there are infinitely many pairwise nonequivalent closed $\mathcal{L}(\Box,\mathbf{I})$-formulas, none of which is equivalent to a (closed) formula in $\mathcal{L}(\Box)$. To see this, let $A_1 := \mathbf{I}\Diamond\top$, $A_{n+1} := \Diamond(A_n \wedge \Diamond^{n+1}\top)$, and consider the Veltman-frame $\mathcal{F}$ depicted in Figure 2. Let $\Vdash$ be any forcing relation on $\mathcal{F}$ with, for all $i \in \omega \cup \{-1\}$, $a_i \Vdash p$ iff $b_i \Vdash p$; then for all $B \in \mathcal{L}(\Box)$, $a_i \Vdash B$ iff $b_i \Vdash B$. On the other hand, we have for all $i \in \omega \setminus \{0\}$, $a_i \nVdash A_i$ and $b_i \Vdash A_i$. This shows that none of the A_i is equivalent to an $\mathcal{L}(\Box)$-formula. To see that $il \nvdash A_i \leftrightarrow A_j$, if $i \neq j$, note that for all i, and all $j > i$, $b_i \Vdash A_i \wedge \neg A_j$.

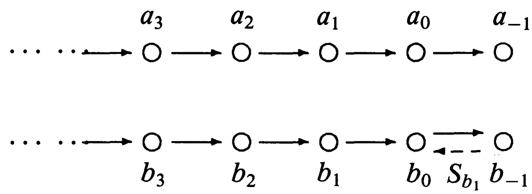

Figure 2.

It is still open whether there exist reasonable normal forms for closed formulas in subsystems of $il + f$.

We now examine the modalities in $\mathcal{L}(\mathbf{I})$ and $\mathcal{L}(\Box,\mathbf{I})$. (Recall that a modality is nothing but a sequence consisting of modal operators and/or dual versions of these operators.) We say that two modalities α and β are *equivalent over ils* if for all $A \in \mathcal{L}(\Box,\mathbf{I})$, $ils \vdash \alpha A \leftrightarrow \beta A$. A modality α is called a *constant* modality (over *ils*) if there is a closed formula C such that for all A, $ils \vdash \alpha A \leftrightarrow C$ (i.e., if for all A, B, $ils \vdash \alpha A \leftrightarrow \alpha B$). We use $\bar{\mathbf{I}}$ as an abbreviation for $\neg\mathbf{I}\neg$.

We start with the modalities over extensions of *il*. Unlike modalities in more traditional modal languages, almost all modalities in $\mathcal{L}(\mathbf{I})$ are constant. For example:

Proposition 3.5 *Let* $A \in \mathcal{L}(\Box,\mathbf{I})$. *Then*
(a) $il \vdash \mathbf{I}\mathbf{I}A \leftrightarrow \top$
(b) $il \vdash \bar{\mathbf{I}}\bar{\mathbf{I}}A \leftrightarrow \bot$
(c) $il \vdash \mathbf{I}\Box A \leftrightarrow \top$
(d) $il \vdash \bar{\mathbf{I}}\Diamond A \leftrightarrow \bot$.

Proposition 3.6 *Let* $A \in \mathcal{L}(\Box,\mathbf{I})$. *Then* $il \vdash \mathbf{I}\bar{\mathbf{I}}\mathbf{I}A \leftrightarrow \mathbf{I}\bar{\mathbf{I}}\top \; (\leftrightarrow \mathbf{I}\Diamond\top)$.

Proof: One direction is almost immediate:

$$\begin{aligned} il \vdash \mathbf{I}\bar{\mathbf{I}}\mathbf{I}A &\rightarrow \mathbf{I}\bar{\mathbf{I}}\mathbf{I}\top \\ &\rightarrow \mathbf{I}\bar{\mathbf{I}}\top, \text{ since } il \vdash \Box(\mathbf{I}\top \leftrightarrow \top). \end{aligned}$$

To prove the other one, we show that $il \vdash \bar{\mathbf{I}}\mathbf{I}\bar{\mathbf{I}}A \rightarrow \bar{\mathbf{I}}\mathbf{I}\top$:

$$\begin{aligned} il \vdash \bar{\mathbf{I}}\mathbf{I}\bar{\mathbf{I}}A \wedge \neg\bar{\mathbf{I}}\Box\bot &\rightarrow \mathbf{I}\Diamond\top \wedge \bar{\mathbf{I}}\mathbf{I}\bar{\mathbf{I}}\top \\ &\rightarrow \mathbf{I}\Diamond\top \wedge \bar{\mathbf{I}}\mathbf{I}\Diamond\top, \text{ since } il \vdash \Box(\bar{\mathbf{I}}\top \leftrightarrow \Diamond\top) \\ &\rightarrow \mathbf{I}\Diamond\top \wedge \bar{\mathbf{I}}(\Diamond\top \rightarrow \Diamond\Diamond\top), \quad \text{by Axiom } I4. \end{aligned}$$

Now $il \vdash \mathbf{I}\Diamond\top \wedge \bar{\mathbf{I}}(\Diamond\top \rightarrow \Diamond\Diamond\top) \rightarrow \bot$, by 1.4, and $il \vdash \bar{\mathbf{I}}\Box\bot \leftrightarrow \bar{\mathbf{I}}\mathbf{I}\bot$. Therefore $il \vdash \bar{\mathbf{I}}\mathbf{I}\bar{\mathbf{I}}A \rightarrow \bar{\mathbf{I}}\mathbf{I}\bot$.

As a corollary we find the following result:

Proposition 3.7 *Let X be a logic that extends il. Then*
(a) *every modality in* $\mathcal{L}(\mathbf{I})$ *is equivalent* (*over* X) *to one of* $\langle\ \rangle$, $\mathbf{I}$, $\bar{\mathbf{I}}$, $\mathbf{II}$, $\mathbf{I}\bar{\mathbf{I}}$, $\bar{\mathbf{I}}\mathbf{I}$, $\bar{\mathbf{I}}\bar{\mathbf{I}}$, $\mathbf{I}\bar{\mathbf{I}}\mathbf{I}$, $\mathbf{I}\bar{\mathbf{I}}\bar{\mathbf{I}}$, $\bar{\mathbf{I}}\mathbf{II}$, *or* $\bar{\mathbf{I}}\mathbf{I}\bar{\mathbf{I}}$;
(b) *if X is il then the only nonconstant modalities in* $\mathcal{L}(\mathbf{I})$ *are* $\langle\ \rangle$, $\mathbf{I}$, $\bar{\mathbf{I}}$, $\mathbf{I}\bar{\mathbf{I}}$, *and* $\bar{\mathbf{I}}\mathbf{I}$.

Proof: Note first that if α, β are one of the modalities mentioned in (a), and if $\alpha \not\equiv \beta$, then α and β are not equivalent over *il*. Let α be a modality in $\mathcal{L}(\mathbf{I})$. Then either $\alpha \in \{\langle\ \rangle, \mathbf{I}, \bar{\mathbf{I}}, \mathbf{I}\bar{\mathbf{I}}, \bar{\mathbf{I}}\mathbf{I}\}$, and we are done, or for some α' we have $\alpha \in \{\mathbf{II}\alpha', \mathbf{I}\bar{\mathbf{I}}\mathbf{I}\alpha', \mathbf{I}\bar{\mathbf{I}}\bar{\mathbf{I}}\alpha', \bar{\mathbf{I}}\bar{\mathbf{I}}\alpha', \bar{\mathbf{I}}\mathbf{II}\alpha', \bar{\mathbf{I}}\mathbf{I}\bar{\mathbf{I}}\alpha'\}$. In the latter case an application of 3.5 or 3.6 yields (a).

To prove (b), note first that $\langle\ \rangle$, $\mathbf{I}$, $\bar{\mathbf{I}}$, $\mathbf{I}\bar{\mathbf{I}}$, and $\bar{\mathbf{I}}\mathbf{I}$ are indeed nonconstant modalities; that they are the only such modalities in $\mathcal{L}(\mathbf{I})$ is immediate from 3.5, 3.6, and (a).

Proposition 3.8 *Let X be a logic that extends il. Then every modality in* $\mathcal{L}(\Box, \mathbf{I})$ *is equivalent* (*over* X) *to a modality of the form* $\alpha_1\beta_1 \ldots \alpha_n\beta_n$, *where the* α_i*s are modalities in* $\mathcal{L}(\Box)$ *and for* $1 \leq i < n$, $\beta_i \in \langle\langle\ \rangle, \mathbf{I}, \bar{\mathbf{I}}, \mathbf{I}\bar{\mathbf{I}}, \bar{\mathbf{I}}\mathbf{I}\}$, *while* $\beta_n \in \{\langle\ \rangle, \mathbf{I}, \bar{\mathbf{I}}, \mathbf{II}, \mathbf{I}\bar{\mathbf{I}}, \bar{\mathbf{I}}\mathbf{I}, \bar{\mathbf{I}}\bar{\mathbf{I}}, \mathbf{I}\bar{\mathbf{I}}\mathbf{I}, \mathbf{I}\bar{\mathbf{I}}\bar{\mathbf{I}}, \bar{\mathbf{I}}\mathbf{II}, \bar{\mathbf{I}}\mathbf{I}\bar{\mathbf{I}}\}$.

We continue with a somewhat simpler case: the modalities over extensions of *ilm*. Here there are even fewer nonconstant modalities in $\mathcal{L}(\mathbf{I})$. For a start, we have the following stronger version of 3.6.

Proposition 3.9 *Let* $A, B \in \mathcal{L}(\Box, \mathbf{I})$. *Then* $ilm \vdash \mathbf{I}\bar{\mathbf{I}}A \leftrightarrow \Box\bot$.

Proof: Since $ilm \vdash \Box\bot \to \Box\bar{\mathbf{I}}A$, we have $ilm \vdash \Box\bot \to \mathbf{I}\bar{\mathbf{I}}A$, by 1.4. To prove the converse, note that $ilm \vdash \Box(\bar{\mathbf{I}}A \wedge \Box\bot \to \bot)$. So since $ilm \vdash \mathbf{I}\bar{\mathbf{I}}A \to \mathbf{I}(\bar{\mathbf{I}}A \wedge \Box\bot)$, by Axiom m, we have $ilm \vdash \mathbf{I}\bar{\mathbf{I}}A \to \mathbf{I}\bot$, by Axiom $I2$. Thus $ilm \vdash \mathbf{I}\bar{\mathbf{I}}A \to \Box\bot$.

Proposition 3.10 *Let X be a logic that extends ilm. Then every modality in* $\mathcal{L}(\mathbf{I})$ *is equivalent* (*over* X) *to one of* $\langle\ \rangle$, $\mathbf{I}$, $\bar{\mathbf{I}}$, $\mathbf{II}$, $\mathbf{I}\bar{\mathbf{I}}$, $\bar{\mathbf{I}}\mathbf{I}$, *or* $\bar{\mathbf{I}}\bar{\mathbf{I}}$. *Moreover, if X is ilm or ilp then the only nonconstant modalities in* $\mathcal{L}(\mathbf{I})$ *are* $\langle\ \rangle$, $\mathbf{I}$, *and* $\bar{\mathbf{I}}$.

Proof: Immediate from 3.5 and 3.9.

Proposition 3.11 *Let* $A \in \mathcal{L}(\Box, \mathbf{I})$. *Then*
(a) $ilm \vdash \mathbf{I}\Diamond A \leftrightarrow \Box\bot$;
(b) $ilm \vdash \bar{\mathbf{I}}\Box A \leftrightarrow \Diamond\top$.

Proposition 3.12 *Let X be a logic that extends ilm. Then*
(a) *every modality in* $\mathcal{L}(\Box, \mathbf{I})$ *is equivalent* (*over* X) *to a modality of the form* $\alpha\beta$, *where* α *is a* (*possibly empty*) *modality in* $\mathcal{L}(\Box)$, *and* $\beta \in \{\langle\ \rangle, \mathbf{I}, \bar{\mathbf{I}}, \mathbf{I}\bar{\mathbf{I}}, \bar{\mathbf{I}}\mathbf{I}, \mathbf{II}, \bar{\mathbf{I}}\bar{\mathbf{I}}\}$;
(b) *if X is ilm or ilp, then the only nonconstant modalities in* $\mathcal{L}(\Box, \mathbf{I})$ *are* $\Diamond^k$, $\Box^k$, $\Box^k\mathbf{I}$, *and* $\Diamond^k\bar{\mathbf{I}}$.

Proof: Let γ be a nonempty modality in $\mathcal{L}(\Box, \mathbf{I})$. If γ is in fact a modality in $\mathcal{L}(\mathbf{I})$, then we are done by 3.10. So assume that $\gamma \equiv \alpha\beta$ where α is the largest prefix of γ that is still a modality in $\mathcal{L}(\Box)$ (so α may be empty). Then, again by 3.10, β is equivalent to one of $\langle\ \rangle$, $\mathbf{I}\beta'$, $\bar{\mathbf{I}}\beta'$, $\mathbf{II}\beta'$, $\mathbf{I}\bar{\mathbf{I}}\beta'$, $\bar{\mathbf{I}}\mathbf{I}\beta'$, or $\bar{\mathbf{I}}\bar{\mathbf{I}}\beta'$, where β' is either empty or a modality with a nonempty prefix in $\mathcal{L}(\Box)$. In the first case we are done; in the latter case we can use 3.5, 3.9, and 3.11 to check all cases, and see that β is (equivalent to) a modality of the desired form.

Next, to prove (b) let γ be a nonconstant modality in $\mathcal{L}(\Box,\mathbf{I})$; by (a) we may assume that $\gamma \equiv \alpha\beta$, with $\alpha\beta$ as described in (a). Since γ is assumed to be nonconstant $\beta \in \{\langle\ \rangle, \mathbf{I}, \bar{\mathbf{I}}\}$. Moreover since $\Box\Diamond$ and $\Diamond\Box$ are constant we may assume that $\alpha \equiv \Diamond^k$, or $\alpha \equiv \Box^k$, for some k.

If $\beta \equiv \langle\ \rangle$, then $\gamma \equiv \Diamond^k$ or $\gamma \equiv \Box^k$; in both cases γ is nonconstant for all k.

If $\beta \equiv \mathbf{I}$, then $\gamma \equiv \Diamond^k\mathbf{I}$ or $\gamma \equiv \Box^k\mathbf{I}$. Since $ilm \vdash \Diamond\mathbf{I}A \leftrightarrow \Diamond\top$, we have that $\Diamond^k\mathbf{I}$ is constant for all $k \geq 1$; on the other hand, for any k, $\Box^k\mathbf{I}$ is nonconstant, as the reader may verify.

Similarly, if $\beta \equiv \bar{\mathbf{I}}$, then γ is nonconstant iff $\gamma \equiv \Diamond^k\bar{\mathbf{I}}$.

For the remainder of this section let T be a Σ_1^0-sound essentially reflexive theory. (Modulo some obvious changes most of the remarks in the sequel hold equally well for Σ_1^0-sound finitely axiomatized sequential theories that extend $\mathrm{I}\Delta_0$ + SupExp.) Let $\Box_T$ be a formalization (in the language of T) of provability in T; $\Diamond_T\varphi$ is short for $\neg\Box_T\neg\varphi$; $\mathbf{I}_T$ is a formalization (in the language of T) of the unary interpretability predicate over T.

Assume that φ is a sentence in the language of T that is not of the form $(\neg)\mathbf{I}_T\psi$ or $(\neg)\Box_T\psi$. We want to know what the theory T can say about sentences of the form $\beta\varphi$, where β is (the arithmetical version of) a nonempty modality of the form $(\neg)\mathbf{I}\beta'$. By 3.12(a) we have to consider only 6 cases.

Note first that no formula of the form $\neg\mathbf{I}_T\varphi$ can be provable in T, for we have $ilm \vdash \Box\neg\mathbf{I}A \rightarrow \Box\bot$ for all $A \in \mathcal{L}(\Box,\mathbf{I})$. So $T \vdash \Box_T\neg\mathbf{I}_T\varphi \rightarrow \Box_T(0=1)$, for all sentences φ in the language of T. Therefore if $T \vdash \neg\mathbf{I}_T\varphi$ then $T \vdash \Box_T(0=1)$. Since T is assumed to be Σ_1^0-sound, this implies that for no φ, $T \vdash \neg\mathbf{I}_T\varphi$. Similarly, since $ilm \vdash \mathbf{I}\bar{\mathbf{I}}A \leftrightarrow \Box\bot$, we cannot have $T \vdash \mathbf{I}_T\bar{\mathbf{I}}_T\varphi$ for any sentence φ. Moreover, we do have for all sentences φ, $T \vdash \mathbf{I}_T\mathbf{I}_T\varphi$, because $ilm \vdash \mathbf{I}\mathbf{I}A$. The only remaining case, then, is $\beta \equiv \mathbf{I}$. Here we have the following possibilities:

1. $T \vdash \varphi$, and then $T \vdash \mathbf{I}_T\varphi$, $T \nvdash \mathbf{I}_T\neg\varphi$
2. $T \vdash \neg\varphi$, and then $T \nvdash \mathbf{I}_T\varphi$, $T \vdash \mathbf{I}_T\neg\varphi$
3. $T \nvdash \varphi$, $T \nvdash \neg\varphi$ and $T \vdash \mathbf{I}_T\varphi$, $T \vdash \mathbf{I}_T\neg\varphi$
4. $T \nvdash \varphi$, $T \nvdash \neg\varphi$ and $T \vdash \mathbf{I}_T\varphi$, $T \nvdash \mathbf{I}_T\neg\varphi$
5. $T \nvdash \varphi$, $T \nvdash \neg\varphi$ and $T \nvdash \mathbf{I}_T\varphi$, $T \vdash \mathbf{I}_T\neg\varphi$
6. $T \nvdash \varphi$, $T \nvdash \neg\varphi$ and $T \nvdash \mathbf{I}_T\varphi$, $T \nvdash \mathbf{I}_T\neg\varphi$.

By our previous remarks, no strengthening of this classification is possible by replacing '$T \nvdash$' by '$T \vdash \neg$' somewhere.

We leave it to the reader to supply examples of cases 1 and 2; the sentence $\Box_T(0=1)$ is a sentence that satisfies case 4, and its negation satisfies case 5; below we will provide examples of sentences that satisfy cases 3 and 6. Recall that an *Orey sentence* for T is a sentence ψ such that both ψ and $\neg\psi$ are interpretable in T. So a sentence satisfying case 3 is an example of a sentence that is provably in T an Orey sentence for T. Our example below of a sentence satisfying case 6 is an example of a sentence that is—unprovably in T—an Orey sentence for T.

Example 3.13 *There is a sentence φ that satisfies case* 3.

Proof: Put $A \equiv \neg\Box p \wedge \neg\Box\neg p \wedge \Box\mathbf{I}p \wedge \Box\mathbf{I}\neg p$. We prove that $ilm^{\omega} \nVdash \neg A$; then, by 2.18, there is an interpretation $(\cdot)^T$ of $\mathcal{L}(\Box,\mathbf{I})$ in the language of T such that $(\neg A)^T$ is false in the standard model. Hence $(A)^T$ is true. Put $\varphi = (p)^T$ and we are done.

Now, to prove that $ilm^{\omega} \nVdash \neg A$ we show that

$$ilm \nVdash \left(\bigwedge_{\Box B \in \mathrm{Sub}(\neg A)} (\Box B \rightarrow B) \wedge \bigwedge_{\mathbf{I}D \in \mathrm{Sub}(\neg A)} \Diamond\top \right) \rightarrow \neg A. \qquad (3)$$

Define $\mathcal{M}$ as in Figure 3.

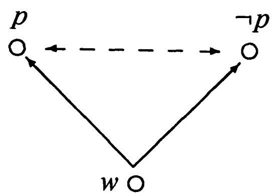

Figure 3.

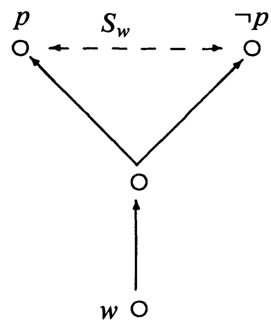

Figure 4.

We leave it to the reader to check that $w \Vdash \bigwedge_{\mathbf{I}D \in \mathrm{Sub}(\neg A)} \Diamond\top$ and that $w \Vdash \bigwedge_{\Box B \in \mathrm{Sub}(\neg A)} (\Box B \rightarrow B)$; from this and $w \Vdash A$ we obtain (3).

Example 3.14 *There is a sentence φ that satisfies case* 6 *and such that φ is, unprovably in T, an Orey sentence for T.*

Proof: Put $A \equiv \neg\Box p \wedge \neg\Box\neg p \wedge \neg\Box\mathbf{I}p \wedge \neg\Box\mathbf{I}\neg p \wedge \mathbf{I}p \wedge \mathbf{I}\neg p$. We have to show only that $ilm^{\omega} \nVdash \neg A$, then we find an interpretation $(\cdot)^T$ of $\mathcal{L}(\Box,\mathbf{I})$ in the language of T such that $(A)^T$ is true. Put $\varphi = (p)^T$ and we are done.

We leave it to the reader to check that the model depicted in Figure 4 shows that $ilm^{\omega} \nVdash \neg A$.

Note that the model used in 3.14 is not an *ILP*-model. Therefore the sentence φ given there works only for essentially reflexive theories T. We leave it to the reader to find a φ that satisfies case 6 if T is a Σ_1^0-sound finitely axiomatized theory that extends $\mathrm{I}\Delta_0 + \mathrm{SupExp}$. He or she will not be able to find a sentence φ that satisfies 3.14 for such T. For, let T be such a theory, and assume that $T \nvdash \mathbf{I}_T\varphi$ while $T + \varphi$ is interpretable in T. Then $\omega \models \mathbf{I}_T\varphi$. Hence, $\omega \models \Box_T\mathbf{I}_T\varphi$ (since $\omega \models \mathbf{I}_T\varphi \rightarrow \Box_T\mathbf{I}_T\varphi$), and so $T \vdash \mathbf{I}_T\varphi$—a contradiction.

An inspection of the arithmetical completeness proof of *ILM* shows that the sentences φ found in 3.13 and 3.14 may be taken to be Σ_2^0-sentences.

3.2 Interpolation and fixed point theorems Our proof of the interpolation theorem for *il*, *ilm*, and *ilp* extends Smoryński's proof of the interpolation theorem for L (cf. [7]).

Definition 3.15 Let $A \in \mathcal{L}(\Box, \mathbf{I})$. Then $\mathcal{L}_A$ is the sublanguage of $\mathcal{L}(\Box, \mathbf{I})$ consisting of all formulas having only proposition letters occurring in A. (So $\top, \bot \in \mathcal{L}_A$, for any A.) A set $X \subseteq \mathcal{L}_A$ is maximal *ils*-consistent in $\mathcal{L}_A$ if for all $C \in \mathcal{L}_A$, either $C \in \mathcal{L}_A$ or $\neg C \in \mathcal{L}_A$.

A pair $\langle X, Y \rangle$ with $X \subseteq \mathcal{L}_A$, $Y \subseteq \mathcal{L}_B$ is called *separable* if for some $C \in \mathcal{L}_A \cap \mathcal{L}_B$, $C \in X$ and $\neg C \in Y$. If $\langle X, Y \rangle$ is not separable it is *inseparable.*

A pair $\langle X, Y \rangle$ with $X \subseteq \mathcal{L}_A$, $Y \subseteq \mathcal{L}_B$ is called a *complete* pair if

1. $\langle X, Y \rangle$ is inseparable
2. X is maximal *ils*-consistent in $\mathcal{L}_A$
3. Y is maximal *ils*-consistent in $\mathcal{L}_B$.

Our proof of the interpolation theorem for *il* (*ilm, ilp*) is in fact nothing but another modal completeness proof for *il* (*ilm, ilp*)—using complete pairs instead of plain maximal *il* (*ilm, ilp*)-consistent sets. The construction of a countermodel is entirely analogous to the constructions in 2.6, 2.11, and 2.21. The main difference is the result that supplies us with the input for our construction. That is: 2.3, 2.4, and 2.5 have to be restated and reproved for complete pairs.

Definition 3.16 Let $\langle X, Y \rangle, \langle X', Y' \rangle$ be complete pairs.

1. $\langle X, Y \rangle < \langle X', Y' \rangle$ ($\langle X', Y' \rangle$ is a *successor* of $\langle X, Y \rangle$) if
   (a) $A \in X' \cup Y'$ for all $\Box A \in X \cup Y$
   (b) $\Box A \in X' \cup Y'$ for some $\Box A \notin X \cup Y$
2. $\langle X', Y' \rangle$ is called a *C-critical successor* of $\langle X, Y \rangle$ ($\langle X, Y \rangle <_C \langle X', Y' \rangle$) if
   (a) $\langle X, Y \rangle < \langle X', Y' \rangle$
   (b) $\mathbf{I}C \notin X \cup Y$
   (c) $\neg C, \Box\neg C \in X' \cup Y'$.

Proposition 3.17 *Let $X_0 \subseteq \mathcal{L}_A$, $Y_0 \subseteq \mathcal{L}_B$ be such that $\langle X_0, Y_0 \rangle$ is an inseparable pair. Then there exists a complete pair $\langle X, Y \rangle$ with $X_0 \subseteq X \subseteq \mathcal{L}_A$ and $Y_0 \subseteq Y \subseteq \mathcal{L}_B$.*

Proof: See [7], Lemma 1.1.

Proposition 3.18 *Let $\langle X, Y \rangle$ be a complete pair such that $\Diamond C \in X \cup Y$. Then there exists a complete pair $\langle X', Y' \rangle > \langle X, Y \rangle$ with $C, \Box\neg C \in X' \cup Y'$.*

Proof: See [7], Lemma 1.2.

Proposition 3.19 *Let $\langle X, Y \rangle$ be a complete pair such that $\mathbf{I}C \notin X \cup Y$. Then there exists a critical complete pair $\langle X', Y' \rangle$ with $\langle X, Y \rangle <_C \langle X', Y' \rangle$ and $\Box\bot \in X' \cup Y'$.*

Proof: Assume that no such $\langle X', Y' \rangle$ exists. We distinguish three cases. In each case we argue that X and Y are separable after all. We prove only one case in detail.

Case 1. $\mathbf{I}C \in \mathcal{L}_A \backslash \mathcal{L}_B$. Then by 3.17 and compactness there are $\Box F_1, \ldots, \Box F_m \in X$, $\Box G_1, \ldots, \Box G_n \in Y$, and $D \in \mathcal{L}_A \cap \mathcal{L}_B$ such that

$$F_1, \ldots, F_m, \neg C, \Box\neg C, \Box\bot \vdash D \tag{4}$$

$$G_1, \ldots, G_n, \Box\bot \vdash \neg D. \tag{5}$$

By (4) we have $\Box F_1, \ldots, \Box F_m \vdash \Box(\Box\bot \to (\neg D \to C \vee \Diamond C))$. Now

$$il \vdash \neg \mathbf{I}C \wedge \Box(\Box\bot \to (\neg D \to C \vee \Diamond C)) \to \neg\Box(\Box\bot \to \neg D).$$

So $X \vdash \neg\Box(\Box\bot \to \neg D)$. On the other hand, (5) yields $Y \vdash \Box(\Box\bot \to \neg D)$. So X and Y are separable—a contradiction.

Case 2. $\mathbf{I}C \in \mathcal{L}_B \backslash \mathcal{L}_A$. Similar to Case 1.

Case 3. $\mathbf{I}C \in \mathcal{L}_A \cap \mathcal{L}_B$. By 3.17 and compactness one can see that, using formulas $\Box F_i$, $\Box G_i$ as in Case 1, one obtains that for some $D \in \mathcal{L}_A \cap \mathcal{L}_B$ we have $X \vdash \neg\Box(\Box\bot \to (D \to C \vee \Diamond C))$ and $Y \vdash \Box(\Box\bot \to (D \to C \vee \Diamond C))$. Hence X and Y are separable—a contradiction.

Proposition 3.20 *Let $\langle X, Y\rangle$ be a complete pair with $\neg\mathbf{I}C \in X \cup Y$ and $\mathbf{I}E \in X \cup Y$. Then there exists a complete pair $\langle X', Y'\rangle$ with $\langle X, Y\rangle <_C \langle X', Y'\rangle$ and $E, \Box\neg E \in X' \cup Y'$.*

Proof: Assume that no such $\langle X', Y'\rangle$ exists. We distinguish nine cases. As before, in each case we argue that X and Y are separable after all; we consider only one case in some detail.

Case 1. $\mathbf{I}E \in \mathcal{L}_A \backslash \mathcal{L}_B$, $\mathbf{I}C \in \mathcal{L}_A \backslash \mathcal{L}_B$. By 3.17 and compactness there exist $\Box F_1, \ldots, \Box F_m \in X$, $\Box G_1, \ldots, \Box G_n \in Y$, and $D \in \mathcal{L}_A \cap \mathcal{L}_B$ such that

$$F_1, \ldots, F_m, \neg C, \Box\neg C, E, \Box\neg E \vdash D \tag{6}$$

$$G_1, \ldots, G_n \vdash \neg D. \tag{7}$$

Now (6) yields

$$\Box F_1, \ldots, \Box F_m, \Box\neg D \vdash \Box(E \wedge \Box E \to C \vee \Diamond C)$$

$$\Box F_1, \ldots, \Box F_m, \Box\neg D \vdash \mathbf{I}(E \wedge \Box\neg E) \to \mathbf{I}(C \vee \Diamond C), \quad \text{by Axiom } I2$$

$$\Box F_1, \ldots, \Box F_m, \Box\neg D \vdash \mathbf{I}E \to \mathbf{I}C, \quad \text{by 1.4(c) and Axiom } I3$$

$$\Box F_1, \ldots, \Box F_m \vdash \mathbf{I}E \wedge \neg\mathbf{I}C \to \neg\Box\neg D$$

$$X \vdash \neg\Box\neg D.$$

On the other hand (7) yields $Y \vdash \Box\neg D$. So X and Y are separable—a contradiction.

Case 2. $\mathbf{I}E \in \mathcal{L}_A \backslash \mathcal{L}_B$, $\mathbf{I}C \in \mathcal{L}_B \backslash \mathcal{L}_A$. Then by 3.17 and compactness one can see that, using formulas $\Box F_i$, $\Box G_i$ as in Case 1, one obtains that for some $D \in \mathcal{L}_A \cap \mathcal{L}_B$ we have $X \vdash \mathbf{I}D$ and $Y \vdash \neg\mathbf{I}D$, so X and Y are separable.

Case 3. $\mathbf{I}E \in \mathcal{L}_A \backslash \mathcal{L}_B$, $\mathbf{I}C \in \mathcal{L}_A \cap \mathcal{L}_B$. Reasoning as in Case 1, one finds a formula $D \in \mathcal{L}_A \cap \mathcal{L}_B$ such that $X \vdash \neg\Box(D \to C \vee \Diamond C)$ and $Y \vdash \Box(D \to C \vee \Diamond C)$. Again, this means that X and Y are separable—a contradiction.

Case 4. $\mathbf{I}E \in \mathcal{L}_B \backslash \mathcal{L}_A$, $\mathbf{I}C \in \mathcal{L}_B \backslash \mathcal{L}_A$. Similar to Case 1.

Case 5. $\mathbf{I}E \in \mathcal{L}_B \backslash \mathcal{L}_A$, $\mathbf{I}C \in \mathcal{L}_A \backslash \mathcal{L}_B$. Similar to Case 2.

Case 6. $\mathbf{I}E \in \mathcal{L}_B \backslash \mathcal{L}_A$, $\mathbf{I}C \in \mathcal{L}_A \cap \mathcal{L}_B$. Similar to Case 3.

Case 7. $\mathbf{I}E \in \mathcal{L}_A \cap \mathcal{L}_B$, $\mathbf{I}C \in \mathcal{L}_A \backslash \mathcal{L}_B$. Reasoning as in Case 1, one can find a formula $D \in \mathcal{L}_A \cap \mathcal{L}_B$ such that $X \vdash \neg\Box(E \wedge \Box\neg E \to \neg D)$ and $Y \vdash \Box(E \wedge \Box\neg E \to \neg D)$. So X and Y are separable—a contradiction.

Case 8. $\mathbf{I}E \in \mathcal{L}_A \cap \mathcal{L}_B$, $\mathbf{I}C \in \mathcal{L}_B \backslash \mathcal{L}_A$. Similar to Case 7.

Case 9. $\mathbf{I}E \in \mathcal{L}_A \cap \mathcal{L}_B$, $\mathbf{I}C \in \mathcal{L}_A \cap \mathcal{L}_B$. Reasoning as in Case 1, one can find a formula $D \in \mathcal{L}_A \cap \mathcal{L}_B$ such that $X \vdash \neg\Box(E \wedge \Box\neg E \to (D \to C \vee \Diamond C))$ and $Y \vdash \Box(E \wedge \Box\neg E \to (D \to C \vee \Diamond C))$. So, again, X and Y are separable—a contradiction.

Theorem 3.21 (Interpolation Theorem) *Let ils be one of il, ilm, or ilp. If ils* $\vdash A \to B$, *then there is a formula C having only proposition letters occurring in both A and B such that ils* $\vdash A \to C$ *and ils* $\vdash C \to B$.

Proof: The proof is by contraposition. Fix A and B and assume that no interpolant exists. We will show that *ils* $\nvdash A \to B$ by constructing a countermodel to the implication.

Note that the assumption that no interpolant exists between A and B means: $\{A\}$ and $\{\neg B\}$ are separable. So by 3.17 there exists a complete pair $\langle X, Y\rangle$ with $\{A\} \subseteq X \subseteq \mathcal{L}_A$ and $\{\neg B\} \subseteq Y \subseteq \mathcal{L}_B$.

Put $\Gamma := \langle X, Y\rangle$ and construct W_Γ as in 2.6 (or 2.11 if *ils* = *ilm*, and 2.21 if *ils* = *ilp*)—starting with $\langle \Gamma, \langle\langle\ \rangle\rangle\rangle$ and adding pairs $\langle \Delta, \tau\rangle$ consisting of complete pairs Δ and sequences τ of pairs (or triples) of formulas. Using 3.18, 3.19, and 3.20 one can then mimic the proof of 2.9 (or 2.14 or 2.23) to find a countermodel to the implication $A \to B$.

To state Beth's Theorem and the Fixed Point Theorem for *il*, *ilm*, and *ilp*, we first introduce some notation and terminology. 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 a $\mathbf{I}$. $A(C)$ denotes the result of substituting C for p in $A(p)$.

Theorem 3.22 (Beth's Theorem) *Let* $A(r) \in \mathcal{L}(\Box, \mathbf{I})$ *contain neither proposition letter p nor q. If ils* $\vdash A(p) \wedge A(q) \to (p \leftrightarrow q)$ *then, for some* $C \in \mathcal{L}_{A(r)} \backslash \{r\}$, *ils* $\vdash A(p) \to (p \leftrightarrow C)$.

Proof: The theorem may be derived from 3.21 in a standard way (cf. [7]).

Proposition 3.23
(a) *il* $\vdash \Box(A \leftrightarrow B) \to (\mathbf{I}A \leftrightarrow \mathbf{I}B)$;
(b) *il* $\vdash \Box^+(B \leftrightarrow C) \to (A(B) \leftrightarrow A(C))$.

If p occurs modalized in A(p) and B is a conjunction of formulas of the form $\Box E$ *and* $\Box^+ E$ *then*
(c) *il* $\vdash \Box(C \leftrightarrow D) \to (A(C) \leftrightarrow A(D))$;
(d) *il* $\vdash B \to (\Box A \to A)$ *implies il* $\vdash B \to A$;
(e) *il* $\vdash \Box^+(p \leftrightarrow A(p)) \wedge \Box^+(q \leftrightarrow A(q)) \to (p \leftrightarrow q)$.

Theorem 3.24 (Explicit Definability of Fixed Points) *Let p occur modalized in $A(p)$. Then there is a formula B with only those proposition letters of A other than p and such that $il \vdash B \leftrightarrow A(B)$.*

Proof: The theorem may be derived from 3.22 and 3.23 in a standard way (cf. [7]).

Remark 3.25 Admittedly, our proof of the Fixed Point Theorem does not yield explicit information on what the fixed point of a given formula looks like. To find an explicit calculation of fixed points one may appeal to de Jongh and Visser's proof of the Fixed Point Theorem for *IL* and other binary interpretability logics in [3]. Using our conservation results 2.10, 2.15, and 2.25, their calculations can easily be carried over to the unary systems: the fixed point of a formula $\mathbf{I}A(p)$ turns out to be $\mathbf{I}A(\Box\bot)$.

4 Concluding remarks In [8] the bi-modal provability logic PRL_1 is defined in a modal language $\mathcal{L}(\Box_1, \Box_2)$ with two provability operators. Besides modus ponens it has as a rule of inference necessitation for $\Box_1$; its axioms are the usual L-axioms for $\Box_1$ plus $\Box_2(A \to B) \to (\Box_2 A \to \Box_2 B)$, $\Box_1 A \to \Box_2 A$, and $\Box_2 A \to \Box_1\Box_2 A$. Define a translation $(\cdot)^* : \mathcal{L}(\Box, \mathbf{I}) \to \mathcal{L}(\Box_1, \Box_2)$ by

$$\begin{aligned} p^t &:= p \\ (\neg A)^t &:= \neg A^t \\ (A \wedge B)^t &:= A^t \wedge B^t \\ (\Box A)^t &:= \Box_1\Box_2 A^t \\ (\mathbf{I}A)^t &:= \Box_1(\Diamond_2\top \to \Diamond_2 A^t). \end{aligned}$$

Using Visser's alternative semantics for *ILP* (cf. [11]) one may then show that for all $A \in \mathcal{L}(\Box, \mathbf{I})$, $ilp \vdash A$ iff $PRL_1 \vdash A^t$.

This much about a connection of (one of) our new logics with a previously known one. Let us look in the opposite direction now, and consider an extension of the language $\mathcal{L}(\Box, \mathbf{I})$. Montagna and Hájek [4] show that *ILM* is the logic of Π_1^0-conservativity in the following sense: given a Σ_1^0-sound extension T of $I\Sigma_1$, define the interpretation $(A \rhd B)^*$ of a formula $A \rhd B$ in the language of T to be '$T + B^*$ is Π_1^0-conservative over $T + A^*$'; then $ILM \vdash A$ iff for all such $(\cdot)^*$, $T \vdash A^*$. It is well known that in essentially reflexive theories like *PA*, relative interpretability and Π_1^0-conservativity (in the above sense) are provably extensionally equivalent. However in finitely axiomatized theories like $I\Sigma_1$ the two notions no longer coincide. So it is natural to extend $\mathcal{L}(\Box, \rhd)$ with an operator $\rhd_M$ to be interpreted arithmetically as Π_1^0-conservativity. (It is convenient in this context to write $\rhd_P$ instead of $\rhd$ for the 'old' operator $\rhd$.) As axioms we take the usual L-axioms and rules plus the *ILM*-axioms for $\rhd_M$, and the *ILP*-axioms for $\rhd_P$. In addition we have the following 'mixed' axiom: $A \rhd_M B \to A \wedge (C \rhd_P D) \rhd_M B \wedge (C \rhd_P D)$. The resulting system is called *ILM/P*. The relevant models are tuples $\langle W, R, S^M, S^P, \Vdash\rangle$ where $\langle W, R, S^M, \Vdash\rangle$ is an *ILM*-

model, $\langle W, R, S^P, \Vdash \rangle$ is an *ILP*-model, where the following extra condition holds: if $xRyS_x^M zRuS_y^P v$ then $uS_z^P v$. It is still open whether *ILM/P* is modally complete with respect to such *ILM/P*-models. The unary counterpart *ilm/p* of *ILM/P* is defined in a language $\mathcal{L}(\Box, \mathbf{I}_M, \mathbf{I}_P)$ with two unary interpretability operators; its axioms and rules are those of L plus the *ilm*-axioms for $\mathbf{I}_M$ and the *ilp*-axioms for $\mathbf{I}_P$; *ilm/p* has no 'mixed' axioms. It has been shown by the present author that *ilm/p* is modally complete with respect to *ILM/P*-models (cf. de Rijke [6]).

We end with a remark on the method used here to prove modal completeness results for the unary logics. Recall that it employs infinite maximal consistent sets and a 'small' adequate set instead of finite maximal consistent sets that are contained in a 'large' adequate set (as used, for example, in [8] and [2]). Our method has also been used to prove the modal completeness of several of the binary interpretability logics mentioned in this paper (cf. [5]).

Acknowledgment Some of the results listed in this paper were obtained as part of my Master's Thesis written at the Department of Mathematics and Computer Science of the University of Amsterdam under the supervision of Dick de Jongh; I want to thank him for his questions and suggestions.

REFERENCES

[1] Berarducci, A., "The interpretability logic of Peano Arithmetic," *The Journal of Symbolic Logic*, vol. 55 (1990), pp. 1059–1089.

[2] de Jongh, D. and F. Veltman, "Provability logics for interpretability," *ITLI Prepublication Series ML-88-03*, University of Amsterdam, pp. 31–42 in *Mathematical Logic, Proceedings of the 1988 Heyting Conference*, edited by P. Petkov, Plenum Press, New York, 1990.

[3] de Jongh, D. and A. Visser, "Explicit fixed points for interpretability logic," *Studia Logica*, vol. 50 (1991), pp. 39–50.

[4] Montagna, F. and P. Hájek, "*ILM* is the logic of Π_1^0-conservativity," unpublished.

[5] de Rijke, M., "Some chapters on interpretability logic," *ITLI Prepublication Series X-90-02*, University of Amsterdam, 1990.

[6] de Rijke, M., "Bi-unary interpretability logic," *ITLI Prepublication Series X-90-12*, University of Amsterdam, 1990.

[7] Smoryński, C., "Beth's Theorem and self-referential statements," pp. 253–261 in *Logic Colloquium '77*, edited by A. Macintyre et al., North-Holland, Amsterdam, 1978.

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

[9] Švejdar, V., "Some independence results in interpretability logic," *Studia Logica*, vol. 50 (1991), pp. 29–38.

[10] Visser, A., "Preliminary notes on interpretability logic," *Logic Group Preprint Series*, No. 14, Department of Philosophy, University of Utrecht, 1988.

[11] Visser, A., "Interpretability logic," pp. 175–210 in *Mathematical Logic, Proceedings of the 1988 Heyting Conference*, edited by P. Petkov, Plenum Press, New York, 1990.

Department of Mathematics and Computer Science
University of Amsterdam
Plantage Muidergracht 24
1018 TV Amsterdam
The Netherlands