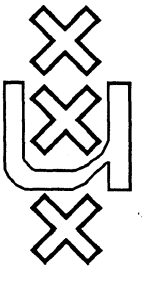
Institute for Language, Logic and Information

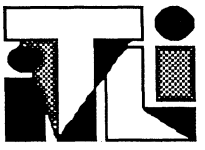
EXPLICIT FIXED POINTS FOR INTERPRETABILITY LOGIC

Dick de Jongh
Albert Visser

ITLI Prepublication Series
for Mathematical Logic and Foundations ML-89-01



University of Amsterdam



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

EXPLICIT FIXED POINTS FOR INTERPRETABILITY LOGIC

Dick de Jongh
Department of Mathematics and Computer Science
University of Amsterdam

Albert Visser
Department of Philosophy
University of Utercht

Received March 1989

Correspondence to:

Faculteit der Wiskunde en Informatica
(Department of Mathematics and Computer Science)
Roetersstraat 15
1018WB Amsterdam

or

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

1 Introduction

The basic theorems of *Provability Logic* are three in number. First is the Arithmetical Completeness Theorem. The second place is shared by the theorems affirming the Uniqueness of Fixed Points and the Explicit Definability of Fixed Points. In this paper we consider the problem of Uniqueness and Explicit Definability of Fixed Points for *Interpretability Logic*. It turns out that Uniqueness is an immediate corollary of a theorem of Smoryński, so most of the paper is devoted to proving Explicit Definability. More sketchy proofs of this Explicit Definability Theorem were given in Visser[88P] and, model-theoretically, in De Jongh & Veltman[88].

Interpretability Logic results from Provability Logic by adding a Binary Modal Operator $\rhd$. If T is a given theory containing enough Arithmetic, we can interpret the modal language into the language of T in the usual way. We interpret $A \rhd B$ as: (the formalization of) T+B is relatively interpretable in T+A. Interpretations of a modal language of this kind were first considered in Hájek[81] and Švejdar[83]. For a more extensive introduction to the various systems of Interpretability Logic see Visser[88].

The system **IL**, the basic system of Interpretability Logic considered in this paper, is a system of arithmetically valid principles. **IL** is definitely arithmetically incomplete, but very natural from the modal point of view. The language of **IL** is the usual language of Modal Propositional Logic with an extra binary connective $\rhd$. The theory **IL** is given as Propositional Logic plus:

L1 $\quad \vdash A \Rightarrow \vdash \Box A$

L2 $\quad \vdash \Box(A \to B) \to (\Box A \to \Box B)$

L3 $\quad \vdash \Box A \to \Box\Box A$

L4 $\quad \vdash \Box(\Box A \to A) \to \Box A$

J1 $\quad \vdash \Box(A \to B) \to A \rhd B$

J2 $\quad \vdash (A \rhd B) \wedge (B \rhd C) \to A \rhd C$

J3 $\quad \vdash (A \rhd C) \wedge (B \rhd C) \to A \vee B \rhd C$

J4 $\quad \vdash A \rhd B \to (\Diamond A \to \Diamond B)$

J5 $\quad \vdash \Diamond A \rhd A$

In the conventions for leaving out parentheses $\rhd$ binds stronger than $\to$, but less strong than the other connectives. The principle J5 is the Interpretation Existence Lemma: it is a syntactic form of the Model Existence Lemma.

L3 is doubly superfluous: as is well-known it can be derived from L4, but in **IL** it can also be derived from J4 and J5. (Interestingly, on the arithmetical side the alternative proof leads in some cases to better estimates on the length of proofs of provability.)

IL is valid for arithmetical interpretations in *adequate* theories T, i.e. theories into which $I\Delta_0+\Omega_1$ is translatable and whose axiom sets can be represented by a Δ_1^b-formula (see Buss[85] for a definition of the bounded hierarchy). It is surely arithmetically incomplete: the principle W introduced immediately below and some other principles discussed in section 4 are not provable in **IL**, but valid in every adequate theory.

Kripke models for **IL** were invented by Frank Veltman and a Kripke model completeness theorem was proved by De Jongh & Veltman (see De Jongh & Veltman[88]).

Other important interpretability logics which have been studied are the extensions **ILW**, **ILP** and **ILM** of **IL** obtained by adding to **IL** respectively the principles W, P, M:

W	$\vdash A \rhd B \rightarrow A \rhd B \wedge \Box \neg A$
P	$\vdash A \rhd B \rightarrow \Box(A \rhd B)$
M	$\vdash A \rhd B \rightarrow A \wedge \Box C \rhd B \wedge \Box C$

Kripke model completeness theorems for **IL**, **ILP** and **ILM** were proved by De Jongh & Veltman ([88]), arithmetic completeness was proved for **ILP** by Visser ([88]) with respect to all sequential finitely axiomatizable theories extending $I\Delta_0$+SUPEREXP, and for **ILM** arithmetic completeness with respect to **PA** and other essentially reflexive theories has been established indepedently by Berarducci and Shavrukov. **ILW**, which is contained in both **ILP** and **ILM**, is still arithmetically valid in any adequate theory T. It is conjectured that **ILW** contains precisely the principles valid in every reasonable theory T, i.e.:

$$\mathbf{ILW} \vdash A \Leftrightarrow \text{for all adequate T, for all interpretations } * \text{ in T, T} \vdash (A)^*.$$

The restriction to **IL** is for our purpose in this paper no limitation: theories that are arithmetically complete are evidently extensions of **IL** and every extension of **IL** inherits Uniqueness and Explicit Definability of Fixed Points from **IL**. In one respect restriction to **IL** does make a difference however: in a stronger theory the explicit fixed points could take a simpler form. We show that this indeed happens for **ILW**.

Although the Explicit Definability of Fixed Points is a beautiful property for a system to have, the other side of the coin is that fixed points of formulas expressible in a system satisfying it can never give anything new. Thus, one cannot expect in pure interpretability logic interesting fixed points like the Rosser fixed points featuring in provability logic extended with witness comparison symbols.

2 Unique & Explicit Fixed Points in general

For our purposes we need the careful discussion of bi-modal self-reference in Smoryński[85] (p.172-176) in a slightly adapted form. Let $\mathbf{SR_0}$ be the following system in the the language of modal propositional logic extended with a binary operator #:

L1 $\quad \vdash A \Rightarrow \vdash \Box A$

L2 $\quad \vdash \Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$

L3 $\quad \vdash \Box A \rightarrow \Box\Box A$

L4 $\quad \vdash \Box(\Box A \rightarrow A) \rightarrow \Box A$

E $\quad \vdash \Box(A \leftrightarrow B) \rightarrow (A\#C \leftrightarrow B\#C)$

$\quad \vdash \Box(A \leftrightarrow B) \rightarrow (C\#A \leftrightarrow C\#B)$

Here E stands for Extensionality.

Define $\Box^+A := (A \wedge \Box A)$. We write Ap for a formula A in which p possibly occurs, in which case, e.g., AB stands for the result of the substitution of B for p in Ap and AAB for the result of substituting AB for p in Ap. We say that p occurs *modalized* in Ap, if p occurs in Ap only in the scope of $\Box$ and #. Two immediate consequences of our theory are the Substitution Principles S_1, S_2, S_3 and Löb's Rule LR:

S_1 $\quad \vdash B \leftrightarrow C \Rightarrow \vdash AB \leftrightarrow AC$

S_2 $\quad \vdash \Box^+(B \leftrightarrow C) \rightarrow (AB \leftrightarrow AC)$

S_3 $\quad$ Suppose p is modalized in Ap, then:

$\quad \vdash \Box(B \leftrightarrow C) \rightarrow (AB \leftrightarrow AC)$

LR $\quad$ Let B be a conjunction of formulas of the form $\Box C$ or $\Box^+C$, then:

$\quad \vdash B \rightarrow (\Box A \rightarrow A) \Rightarrow \vdash B \rightarrow A$

2.1 Uniqueness Theorem

Suppose p occurs modalized in A, then: $\mathbf{SR_0} \vdash (\Box^+(p \leftrightarrow Ap) \wedge \Box^+(q \leftrightarrow Aq)) \rightarrow (p \leftrightarrow q)$.

Proof: By S_3: $\vdash (\Box^+(p \leftrightarrow Ap) \wedge \Box^+(q \leftrightarrow Aq)) \rightarrow (\Box(p \leftrightarrow q) \rightarrow (p \leftrightarrow q))$. So LR gives us the desired conclusion. $\Box$

The Uniqueness Theorem was in its original form due to Bernardi, De Jongh and Sambin. In its present form it is due to Smoryński. Assuming the modal completeness theorem an alternative model-theoretic proof along the lines of the implicit definability theorem (see theorem 3.1, p.109, Smoryński[85]) is easily given.

Let $\mathbf{SR_1}$ be $\mathbf{SR_0}$ plus the following axiom:

L3' $\quad \vdash A\#B \rightarrow \Box(A\#B)$.

An immediate consequence of $\mathbf{SR_1}$ is LR^+:

LR^+ Let B be a conjunction of formulas of the form $\Box C$ or $\Box^+ C$ or C#D, then:
$\vdash B \rightarrow (\Box A \rightarrow A) \Rightarrow \vdash B \rightarrow A$

In this general setting the Explicit Definability Theorem is split up into two parts, from which the theorem itself can then be deduced as a Corollary.

2.2 Explicit Definability Theorem, part 1

Let Ap be either of the form $\Box Bp$ or Bp#Cp, then there is a formula D such that: $\mathbf{SR_1} \vdash D \leftrightarrow AD$.

Proof: Suppose Ap is $\Box Bp$ or Bp#Cp. Take $D := A\top$. We have from L3': $\vdash A\top \rightarrow \Box^+(A\top \leftrightarrow \top)$, and hence by S_2: $\vdash A\top \rightarrow AA\top$. On the other hand by S_3: $\vdash AA\top \rightarrow (\Box A\top \rightarrow A\top)$. So LR^+ gives us: $\vdash AA\top \rightarrow A\top$. □

To state the second part of the Explicit Definability Theorem we introduce a simple notion. Fix for the moment a propositional variable p. We write:

$Ap \leq Bp :\Leftrightarrow$ whenever Ap can be written as $A^*(p,E_1q,...,E_nq)$, where q does not occur in $A^*(p,r_1,...,r_n)$ and p does not occur in the E_kq, then Bp can be written as $B^*(p,E_1q,...,E_nq)$, where q does not occur in $B^*(p,r_1,...,r_n)$. (Not all r_k need actually occur in $B^*(p,r_1,...,r_n)$, and neither need p.)

The intuitive content of $Ap \leq Bp$ is that propositional letters q different from p occur in Bp in no other context than they occur in Ap. Clearly $\leq$ is transitive. We allow that the sequence $E_1q,...,E_nq$ is empty; this means that $Ap \leq Bp$ implies that if q occurs in Bp, then q occurs in Ap. We have:

2.3 Lemma

i) Suppose $Ap \leq Bp$ and $Ap \leq Cp$, then $Ap \leq BCp$.

ii) Suppose $Ap \leq B(p,p)$, $Ap \leq Cp$ and $Ap \leq Dp$, then $Ap \leq B(Cp,Dp)$.

iii) Suppose that Ap is of the form BCp, that p really occurs in Cp and that p does not occur in Cq, then $Ap \leq Bp$ and $Ap \leq Cp$.

iv) If at most the propositional variable p occurs in Bp, then $Ap \leq BAp$

v) Suppose $A(p,q) \leq B(p,q)$, then $A(p,p) \leq B(p,p)$.

vi) If Ap=Bp#Cp and p really occurs in Ap, then $Ap \leq Bp$.

Proofs: The proofs of (i) and (ii) are trivial. For (iii), it is sufficient to note that $A^*(p,E_1q,...,E_nq)$ must be of the form $B^*(C^*(p,E_1q,...,E_nq),E_1q,...,E_nq)$. (The occurrence of p in Cp must be real, to

make sure that Cp cannot be a subformula of one of the E_kq.) (iv) is easy. Ad (v): suppose A(p,p) is of the form $A^*(p,p,E_1r,...,E_nr)$. This means that A(p,q) is of the form $A^*(p,q,E_1r,...,E_nr)$. So B(p,q) must be of the form $B^*(p,q,E_1r,...,E_nr)$. Clearly q does not occur in the E_kr, so the form for B(p,p) we are looking for is $B^*(p,p,E_1r,...,E_nr)$. For (vi), note that $A^*(p,E_1q,...,E_nq)$ must be of the form $B^*(p,E_1q,...,E_nq)\#C^*(p,E_1q,...,E_nq)$. □

2.4 Explicit Definability Theorem, part 2

Let U be any extension of $\mathbf{SR_0}$ satisfying:

FIX Every formula Ap of the form □Bp or Bp#Cp has a fixed point D such that Ap≤D.

For every formula Ap with p modalized, there is a formula D such that: p does not occur in D, Ap≤D and $U\vdash D\leftrightarrow AD$.

Proof: Let p be modalized in Ap. Let $Ap=B(C_1p,...,C_np)$, where the C_kp are either of the form □Ep or of the form Ep#Fp and where p does not occur in $B(q_1,...,q_n)$.

Our proof is by induction on n. First suppose n=1. Suppose Ap is of the form BCp, where p does not occur in Bq and Cp is either of the form □Dp or Dp#Ep. We may assume that p really occurs in Cp. Let D be the fixed point of CBp guaranteed by FIX. We show that $\vdash BD\leftrightarrow ABD$. We have $\vdash D\leftrightarrow CBD$. So by S_1: $\vdash BD \leftrightarrow BCBD$, and clearly BCBD = ABD. Trivially p does not occur in BD. We have: Ap≤Bp, Ap≤Cp, hence Ap≤CBp. Because CBp≤D, it follows that Ap ≤D and thus Ap≤BD.

For the induction step we have to show how to reduce the number of 'components' in Ap. Suppose q does not occur in Ap. Define $A^*(p,q)$ by $B(C_1p,...,C_{n-1}p,C_nq)$. $A^*(p,q)$ has n-1 components in which p occurs, so we may apply the induction hypothesis to get Dq with $A^*(p,q)\leq Dq$ and $\vdash Dq\leftrightarrow A^*(Dq,q)$. Clearly Dq can be written as FC_nq, where q does not occur in Fr. Applying the basis step of our induction to FC_np we find an E with: $\vdash E\leftrightarrow DE$, and thus $\vdash E\leftrightarrow A^*(DE,E)$. By S_1 it follows that $\vdash E\leftrightarrow A^*(E,E)$. Clearly $A^*(E,E)=AE$. Evidently p does not occur in E. Finally: $Ap=A^*(p,p)\leq Dp\leq E$. □

2.5 Corollary

(a) For every formula Ap with p modalized, there is a formula D such that p does not occur in D and $\mathbf{SR_1}\vdash D\leftrightarrow AD$.

(b) For every formula Ap in the language of interpretbility logic with p modalized, there is a formula D such that p does not occur in D and $\mathbf{ILP}\vdash D\leftrightarrow AD$.

Proof: (a) The fixed points D for formulas Ap of the form □Bp or Bp#Cp which $\mathbf{SR_1}$ has by the Explicit Definability Theorem, part 1, are □B⊤ and B⊤#C⊤ respectively. Since, by lemma 2.3(i) and (iv), □Bp≤□B⊤ and Bp#Cp≤B⊤#C⊤, $\mathbf{SR_1}$ satisfies FIX.

(b) Follows immediately from (a). □

Corollary 2.5(a) is Smoryński's version of the Explicit Definability Theorem with a proof along the lines of his "slightly easier proof" (see Smoryński[85], p.81). The original theorem was due to De Jongh and Sambin. Our proof differs only in two minor details from Smoryński's. First, for our purpose of proving the theorem for **IL**, it is essential that 2.4 is not proven in $\mathbf{SR_1}$, as $\mathbf{SR_1}$ is valid for **ILP**, but not for **IL**, or even for **ILM**. Secondly, the artifice of using $\leq$ was added, because the generality of theorem 2.4 forced us to be more explicit than usual about the property of the fixed points needed to get the proof to work. Surely our choice of the property 'Ap$\leq$D' is not the most parsimonious one, but we submit that it is fairly natural.

3 Explicit fixed points for IL

As is easily seen **IL** satisfies the principle E of the system $\mathbf{SR_0}$. So, the Uniqueness Theorem, 2.1, holds for **IL**. On the other hand, using **IL**-models, one can show that **IL** does not satisfy L3'. So, the proof of the Explicit Definability Theorem, part 1, is not available for **IL**. Thus we have to provide a different proof for Explicit Definability, part 1 for **IL**. This is the main aim of this section. Before giving the proof we list some theorems of **IL**.

Define: $A \equiv B :\Leftrightarrow (A \rhd B) \wedge (B \rhd A)$.

K1 $\vdash A \equiv (A \vee \Diamond A)$ J1,J5,J3

Let $\phi A := (A \vee \Diamond A)$, $\psi A := (A \wedge \Box\neg A)$, then by L1-L3:

K2 $\vdash \phi A \leftrightarrow \phi\phi A$
$\vdash \phi A \leftrightarrow \phi\psi A$
$\vdash \psi A \leftrightarrow \psi\psi A$
$\vdash \psi A \leftrightarrow \psi\phi A$

Immediate consequences of the above are:

K3 $\vdash A \rhd A \wedge \Box\neg A$

K4 $\vdash A \equiv A \wedge \Box\neg A$

Note that: K4 is an alternative for axiom J5.

K5 $\vdash A \rhd \bot \rightarrow \Box\neg A$ J4

Feferman's Principle is the following:

F $\vdash \Diamond A \rightarrow \neg(A \rhd \Diamond A)$

F is *not* derivable in **IL**. However, the following weakening of F is derivable:

K6 $\vdash \Diamond A \rhd \neg(A \rhd \Diamond A)$

Proof: By the above it is sufficient to show: **IL**$\vdash (\Diamond A \wedge \Box\neg\Diamond A) \rightarrow \neg(A \rhd \Diamond A)$. We have:

$$\begin{aligned} \vdash (\Diamond A \wedge \Box\neg\Diamond A \wedge (A \rhd \Diamond A)) &\rightarrow (\Diamond A \wedge \Box\Box\neg A \wedge (A \rhd \Diamond A)) \\ &\rightarrow (\Diamond A \wedge A \rhd \bot) \\ &\rightarrow (\Diamond A \wedge \Box\neg A) \\ &\rightarrow \bot \end{aligned}$$ □

Start of the proof of Explicit Definability, part 1.

E1 Suppose: $\vdash \Box\neg A\top \rightarrow C$, then $\vdash A\top \wedge \Box\neg A\top \leftrightarrow AC \wedge \Box\neg AC$.

Proof: The "$\rightarrow$" side is immediate, because $\Box\neg A\top \rightarrow \Box^+(C \leftrightarrow \top)$.
"$\leftarrow$" Suppose $\vdash \Box\neg A\top \rightarrow C$. Reason inside the "$\vdash$": Suppose AC and $\Box\neg AC$. We have: $\Box(\Box\neg A\top \rightarrow \Box^+(C \leftrightarrow \top))$. Combining this with $\Box\neg AC$ we get: $\Box(\Box\neg A\top \rightarrow \neg A\top)$. Hence by Löb's Principle: $\Box\neg A\top$. It follows that $\Box^+(C \leftrightarrow \top)$. Combining this with AC we find $A\top$. □

E2 Suppose: $\vdash \Box\neg A\top \rightarrow C$, then $\vdash A\top \equiv AC$. E1,K4

E3 $\vdash A\top \equiv A(A\top \rhd B\Box\neg A\top)$

Proof: We have $\vdash \Box\neg A\top \rightarrow A\top \rhd B\Box\neg A\top$. Apply E2. □

E4 $\vdash \Box\neg B\Box\neg A\top \rightarrow (A\top \rhd B\Box\neg A\top \leftrightarrow \Box\neg A\top)$

Proof:
$$\begin{aligned} \vdash \Box\neg B\Box\neg A\top \rightarrow (A\top \rhd B\Box\neg A\top &\leftrightarrow A\Box\neg A\top \rhd \bot) \\ &\leftrightarrow \Box\neg A\top) \end{aligned}$$ □

E5 $\vdash \Box\neg B\Box\neg A\top \rightarrow \Box^+(A\top \rhd B\Box\neg A\top \leftrightarrow \Box\neg A\top)$

E6 $\vdash B\Box\neg A\top \wedge \Box\neg B\Box\neg A\top \leftrightarrow B(A\top \rhd B\Box\neg A\top) \wedge \Box\neg B(A\top \rhd B\Box\neg A\top)$

Proof: "$\rightarrow$" : immediate by E5 and S_2. For the "$\leftarrow$"-side it is clearly sufficient to show:

$$\vdash \Box\neg B(A\top \rhd B\Box\neg A\top) \rightarrow \Box\neg B\Box\neg A\top$$

This follows by:

$$\begin{aligned} \vdash \Box\neg B(A\top \rhd B\Box\neg A\top) &\rightarrow \Box(\Box\neg B\Box\neg A\top \rightarrow \neg B\Box\neg A\top) && (E5,S_2) \\ &\rightarrow \Box\neg B\Box\neg A\top && \Box \end{aligned}$$

E7 $\vdash B\Box\neg A\top \equiv B(A\top \rhd B\Box\neg A\top)$ E6,K4

E8 $\vdash A\top \rhd B\Box\neg A\top \leftrightarrow A(A\top \rhd B\Box\neg A\top) \rhd B(A\top \rhd B\Box\neg A\top)$ E3,E7

End of the proof of Explicit Definability, part 1.

It is easy to see that p does not occur in $A\top \rhd B\Box\neg A\top$. We have: $(Ap\rhd Bp)\leq(A\top \rhd B\Box\neg A\top)$. For assume that p really occurs in $Ap\rhd Bp$. By 2.3: $(Ap\rhd Bp)\leq Ap\leq A\top\leq\Box\neg A\top$. Also $(Ap\rhd Bp)\leq\top$. Combining by 2.3(ii) we find: $(Ap\rhd Bp)\leq(A\top \rhd B\Box\neg A\top)$. So, we can apply 2.4 and conclude Explicit Definability for **IL**:

for every formula Ap with p modalized, there is a formula D such that:

p does not occur in D, and $\mathbf{IL}\vdash D\leftrightarrow AD$.

4 The system ILW

The principle W is very powerful. It can be viewed (in our limited context) as a generalization both of Gödel's Second Incompleteness Theorem and of Gödel's Completeness Theorem (in the guise of the Interpretation Existence Lemma). To illustrate this we show that **ILW** can be axiomatized as follows:

L1 $\quad\vdash A \Rightarrow \vdash \Box A$

L2 $\quad\vdash \Box(A\rightarrow B)\rightarrow(\Box A\rightarrow\Box B)$

J1 $\quad\vdash \Box(A\rightarrow B)\rightarrow A\rhd B$

J2 $\quad\vdash (A\rhd B)\wedge(B\rhd C)\rightarrow A\rhd C$

J3 $\quad\vdash (A\rhd C)\wedge(B\rhd C)\rightarrow A\vee B\rhd C$

J4 $\quad\vdash A\rhd B\rightarrow(\Diamond A\rightarrow\Diamond B)$

W $\quad\vdash A\rhd B\rightarrow A\rhd B\wedge\Box\neg A$

First prove Feferman's principle F by substituting $\Diamond A$ for B in W (this uses L1, L2, J1, J2). Löb's Principle (L4) then follows from F:

$$\begin{aligned}\vdash\ \Box(\Box A\rightarrow A)&\rightarrow\Box(\neg A\rightarrow\Diamond\neg A)\\&\rightarrow\neg A\rhd\Diamond\neg A\\&\rightarrow\neg\Diamond\neg A\\&\rightarrow\Box A\end{aligned}$$

Using L4 one derives L3 by a well-known trick. Next we derive K2. Using K2 and $\vdash A\equiv A\wedge\Box\neg A$ which is immediate by W, we get: $\vdash A\equiv A\vee\Diamond A$ and hence, by J1, J5.

W is not derivable in **IL**. To show this we need some model theory: we use Frank Veltman's **IL**-models. An **IL**-model M is of the form: $<K,R,S,\Vdash>$, where: K is non-empty; R is a binary relation on K, which is transitive, upwards well-founded; S is a ternary relation on K, which we treat as a K-indexed set of binary relations S_k on K; the S_k are reflexive, transitive; we have: $kRmS_kn\Rightarrow kRn$ and $kRmRn\Rightarrow mS_kn$; $\Vdash$ is a forcing relation on M, where R is the accessibility relation for $\Box$ and:

$k\Vdash A\rhd B :\Leftrightarrow$ for all m with kRm and $m\Vdash A$ there is an n with mS_kn and $n\Vdash B$.

It is easy to show that **IL** is valid in **IL**-models, and **IL** is complete w.r.t. (finite) **IL**-models (De Jongh & Veltman[88]).

Consider the **IL**-model on $\{\alpha,\beta,\gamma\}$ generated by $\alpha R\beta R\gamma$, $\gamma S_\alpha \beta$, $\gamma \Vdash p$. Clearly $\alpha \Vdash p \rhd \Diamond p$, but $\alpha \nVdash \Box\neg p$. Hence Feferman's Principle doesn't hold at α and so a fortiori W fails.

We show that the Fixed Point of $Ap \rhd Bp$ found in Section 3 simplifies in **ILW** to $AT \rhd BT$:

$$\vdash AT \rhd BT \leftrightarrow AT \rhd B\Box\neg AT.$$

Proof:
$$\begin{aligned} \vdash AT \rhd BT &\leftrightarrow AT \rhd BT \wedge \Box\neg AT \\ &\leftrightarrow AT \rhd B\Box\neg AT \wedge \Box\neg AT \\ &\leftrightarrow AT \rhd B\Box\neg AT \end{aligned}$$ □

Finally we show that the simplified fixed point doesn't work in **IL**. Consider $q \rhd \neg p$. The **ILW**-style fixed point in p for this formula is: $q \rhd \neg T$, i.e. modulo **IL** provable equivalence: $\Box\neg q$. If this were a fixed point in **IL**, we would have: $\mathbf{IL} \vdash \Box\neg q \leftrightarrow q \rhd \Diamond q$. We have already seen that this is not the case.

References:

Boolos, G., 1979, *The Unprovability of Consistency,* CUP, London.

Buss, S., 1985, *Bounded Arithmetic*, Thesis, Princeton University, Princeton. Reprinted: 1986, Bibliopolis, Napoli.

De Jongh, D.H.J. & Veltman, F., 1988, *Provability Logics for Relative Interpretability*. To appear in the Proceedings of the Heyting Conference, Chaika, Bulgaria, 1988.

Hájek, P., 1981, *Interpretability in Theories containing Arithmetic II,* Commentationes Mathematicae Universitatis Carolinae 22, 667-688.

Smoryński, C., 1985, *Self-Reference and Modal Logic,* Springer Verlag.

Švejdar, V., 1983, *Modal Analysis of Generalized Rosser Sentences,* JSL 48, 986-999.

Visser, A., 1988, *Interpretability Logic,* Logic Group Preprint Series nr 40, Dept. of Philosophy, University of Utrecht, Heidelberglaan 2, 3584CS Utrecht. To appear in the Proceedings of the Heyting Conference, Chaika, Bulgaria, 1988.

Visser, A., 1988P, *Preliminary Notes on Interpretability Logic,* Logic Group Preprint Series nr 29, Dept. of Philosophy, University of Utrecht, Heidelberglaan 2, 3584CS Utrecht.

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

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

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 Solutions

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