THE FORMALIZATION OF INTERPRETABILITY

Albert Visser
Department of Philosophy, Rijksuniversiteit te Utrecht

Logic Group
Preprint Series
No. 47
August 1989

Department of Philosophy
University of Utrecht
Heidelberglaan 2
3584 CS Utrecht
The Netherlands

THE FORMALIZATION OF INTERPRETABILITY

Albert Visser
Department of Philosophy
University of Utrecht

August 1989

Department of Philosophy
University of Utrecht
Heidelberglaan 2
3584CS Utrecht
The Netherlands

MSC-1980 classification 03B45/03F30
Key words and phrases: Provability Logic,
interpretability, fragments of arithmetic

THE FORMALIZATION OF INTERPRETABILITY

Albert Visser

ABSTRACT: This paper contains a careful derivation of principles of Interpretability Logic valid in extensions of $I\Delta_0+\Omega_1$.

1 Introduction

Interpretability Logic is a generalization of Provability Logic in two senses. First its subject, interpretability, simply *is* a generalization of provability. Interpretability is provability combined with change of perspective. Secondly well known facts of Provability Logic have natural counterparts in Interpretability Logic. For example we have arithmetical completeness theorems in the style of Solovay for two systems of interpretability logic (see Berarducci[88], Shavrukov[88], Hájek & Montagna[89], Visser[88b]). The modal theorems on uniqueness and expliciteness of modalized fixed points have an immediate generalization (see Smorynski[87] for uniqueness, De Jongh & Visser[89] for expliciteness). Finally the closed fragment of interpretability logic 'collapses' to that of provability logic (see Hájek & Svejdar[89], Visser[89]).

The richer language of Interpretability Logic has several advantages: first a major metamathematical insight, the Model Existence Lemma, can in some sense be formulated in the logic: viz. in the form of the Interpretation Existence Lemma (J5, see below). Secondly Solovays Completeness Theorem for *Provability Logic* has an amazing stability: we find that Löb's logic is the provability logic of all Σ_1-sound RE theories U that extend $I\Delta_0$+EXP.(1) Thus Arithmetical Completeness gives no specific information distinguishing various theories. In the case of Interpretability Logic the situation is a bit better. Two important classes, viz. Σ_1-sound Essentially Reflexive Theories and Σ_1-sound Finitely Axiomatized Sequential Theories extending $I\Delta_0$+SUPEXP, have their own distinctive interpretability logic (resp. ILM and ILP). Moreover many individual theories not falling in one of these classes (like $I\Delta_0+\Omega_1$. $I\Delta_0$+EXP, PRA) have their own interpretability logics. (Not much is known about these, except in the case of $I\Delta_0$+EXP.) A third point is the possibility of applications to ordinary Provability Logic. (An example of this -the solution of Guaspari's problem by Dick de Jongh (building on work of Visser, Montagna, Pianigiani (in temporal order))- is forthcoming.)

This paper aims to be a careful presentation of the principles of Interpretability Logic valid in any theory extending $I\Delta_0+\Omega_1$. It turns out that there are more such principles than was conjectured in Viser[88b] (see section 8).

One basic obstacle in reasoning about interpretability in weak theories is the absence of the Σ_1-collection Principle. Tł is difficulty is illustrated (sections 4,5) and a strategy is developped to circumvent the difficulty (sections 4,6). This strategy differs from the one followed in Visser[88b]. (In an appendix, section 11, a metamathematical result is proved on the comparison of reasoning in $I\Delta_0+\Omega_1$ using the strategy and reasoning in $B\Sigma_1+\Omega_1$ without it.)

The plan of the paper is as follows: section 3 provides the necessary preliminaries. Secting 4 introduces interpretability and contains a discussion of the problems one meets when working without Σ_1-collection. Section 5 gives a more extended treatment of $B\Sigma_1$. Section 6 is my latest attempt to give a good presentation of the formalization of the Henkin Construction in $I\Delta_0+\Omega_1$. It supersedes the treatment in Visser[89b]. Section 7 is the heart of the paper: it contains the careful derivation of a number of insights essential to interpretability logic in $I\Delta_0+\Omega_1$. In section 8 the consequences of the work in 7 in terms of arithmetically valid modal principles is spelled out. Sections 9,10,11 are appendices. In section 9 I show how to derive versions of the Orey-Hájek characterization using the framework of the paper. Section 10 contains an alternative proof of the theorem by Hájek and Montagna that ILM is the logic of Π_1-conservativity for extensions of $I\Sigma_1$. In section 11 I describe a different strategy to prove the results of section 7: work in $B\Sigma_1+\Omega_1$ and then use a conservation result to show that what is proved is also provable in $I\Delta_0+\Omega_1$.

2 Prerequisites

The reader should know either the discussion of systems and arithmetization in Paris & Wilkie[87] or in Buss[85]. Moreover the reader should have some knowledge of cuts: see e.g. Paris & Wilkie[87].

3 Conventions, Notions & Elementary Facts.

3.1 $I\Delta_0+\Omega_1$ and the arithmetization of Syntax

$I\Delta_0+\Omega_1$ is the basic theory of this paper. For an introduction see Paris & Wilkie[87]. Here we briefly mention a few relevant facts.

$I\Delta_0$ is PA with induction restricted to Δ_0-formulas. J.H. Bennett shows that there is a Δ_0-formula exp(x)=y, such that $I\Delta_0$ verifies $((\exp(x)=y\wedge\exp(x)=z)\rightarrow y=z)$, $\exp(\underline{0})=\underline{1}$ and $\exp(Sx)=2.\exp(x)$. It is easy to see that $I\Delta_0$ verifies such familiar facts as $((x<y\wedge\exp(y)=z)\rightarrow\exists u\ \exp(x)=u)$, $((\exp(x)=u \wedge \exp(y)=v) \rightarrow \exp(x+y)=u.v$. (Similar remarks hold for x^y.)

Define $|x|$:=the smallest y such that exp(Sy)>Sx. Obviously the graph of $|.|$ is Δ_0. $I\Delta_0$ shows that $|.|$ is a total function, which is weakly monotonically increasing. If we consider the numbers as coding strings of a's and b's, where 0 codes the empty string, 1 codes a, 2 codes b, 3 codes aa, 4 codes ab, 5 codes ba, 6 codes bb, 7 codes aaa, then $|x|$ is the length of the string coded by x. Note (in $I\Delta_0$): $((x\neq\underline{0}\wedge \exp(x)=y)\to|y|=x)$

Define $x*y:=x.\exp(|y|)+y$. $I\Delta_0$ proves that * is total and weakly monotonically increasing in both arguments. $x*y$ is the code of the concatenation of the strings coded by x and y. Note: $I\Delta_0$ proves $|x*y|=|x|+|y|$. Moreover $I\Delta_0$ proves various elementary properties of * like associativity and $x*z=y*z\to x=y$.

Define $\omega_1(x):=\exp(|x|^2)$. Note (in $I\Delta_0$): $((x\neq\underline{0}\wedge \exp(x)=y\wedge\omega_1(y)=z)\to z=\exp(x^2))$ and "$x\neq\underline{0}\to \omega_1{}^n(\exp(x))=\exp(x^{\exp(n)})$ (if one of these exists)".

Let Ω_1 be the axiom "ω_1 is total". As is easily seen $I\Delta_0$ does not prove Ω_1. $I\Delta_0+\Omega_1$ is just right for treating syntax: e.g. Ω_1 guarantees that substitution of a term in a formula is possible. Sometimes it is pleasant to work with Nelson's #, which is defined by $x\#y:=\exp(|x|.|y|)$. As is easily seen $I\Delta_0+\Omega_1$ proves that # is total.

Theorem (Gaifman & Dimitracopoulos[82]): If f has Δ_0-graph than $I\Delta_0$+"f is total and weakly monotonically increasing"$\vdash I\Delta_0(f)$.

Here $\Delta_0(f)$ is the class of (translations of) formulas with only bounded quantifiers, where f is allowed to occur in the bounding terms.

It follows that $I\Delta_0+\Omega_1\vdash I\Delta_0(\omega_1)$, so in $I\Delta_0+\Omega_1$ we can work as if ω_1 were a function symbol in the language.

We code in $I\Delta_0+\Omega_1$ by first translating our syntactical objects into strings of a' and b's and then translate e.g. aabab into $\underline{1}*\underline{1}*\underline{2}*\underline{1}*\underline{2}$. Here * is a definable function that simulates concatenation. To do the usual formalization of syntax it is imperative that the function num(x) that assigns to x the code of the numeral of x is total. However it is easy to see that if we use as numeral for x: $S...S\underline{0}$ (S x times), then the code of this numeral will be exponential in x. Hence we use the following system of numerals: assign to 0 and 1 $\underline{0}$ and $S\underline{0}$; if we have assigned to $x\neq0$ numeral t, assign to $2.x$: $SS\underline{0}.t$, and to $2x+1$: $(SS\underline{0}.t+S\underline{0})$. Num(x) can be proved total even in $I\Delta_0$.

In the sequel we will often use that every term in x of the language of arithmetic extended with ω_1 can be estimated by $\omega_1{}^n(x$ for some standard n, provided that $x>2$. Moreover for every standard polynomial P(x), we have: for some standard n $\exp(P(|x|)<\omega_1{}^n(x)$, again provided that $x>2$. I find it rather tiresome to always mention the proviso $x>2$, so I will omit it. The reader could easily imagine a slightly adapted definition of ω_1 that would make the proviso superfluous.

3.2 Languages

In this paper we consider only relational languages, i.e. languages without function symbols and constants. So for example in the case of arithmetic, instead of + we have a ternary relation symbol, etc. . After this is said officially we will of course often *pretend* that we are working in a language with function symbols. Here one has to be careful: for example at a certain point we are working in $I\Delta_0+\Omega_1$ and we consider a function from n to the Gödelnumber of $\exists y\ y=\underline{n}$, where $\underline{n}$ is the numeral in the sense of section 3.1 corresponding to n. For the functional language it is easy to see that this function is total (in $I\Delta_0+\Omega_1$). Inspection of the translation procedure into the corresponding relational language shows that the formulas become only polynomially longer, so the function is also total for the relational language.

In our languages there are only finitely many relation symbols which include identity.

3.3 Special Classes of Formulas

We refer the reader to the discussion of special classes of formulas in Buss[1985]. We will use mainly $\Sigma_1{}^b$.

3.4 Theories and Provability

We consider, unless explicitely stated otherwise, only theories with identity for which a fixed list of formulas of their language is specified defining a set of natural numbers, 0, successor, addition and multiplication. We assume in most cases that $I\Delta_0+\Omega_1$ is provable for these natural numbers. Variables x,y,z,u,v,... will be taken to range over the designated numbers. So $\forall x A(x)$ means $\forall x(N(x)\rightarrow A(x))$ if N is the formula specifying the natural numbers of our theory. Syntactical notions will always be formalized in the designated natural numbers.

We consider a theory T as given by a formula $\alpha_T(x)$ having just x free plus the relevant information on what the set of natural numbers of the theory is. α_T gives the set of codes of the (non-predicate-logical) axioms of the theory. Different α or different natural numbers different theories; same α and same natural numbers same theory. Unless explicitely stated otherwise we will assume that α is a $\Sigma_1{}^b$-formula.

Example: Consider GB. Different definitions of the natural numbers in GB are possible. Under one such choice $GB\vdash PA$ and $GB\nvdash Con(ZF)$. Under another such choice: $GB\vdash I\Delta_0+\Omega_1+Con(ZF)$. We take the two different choices of the natural numbers to give us two different GB's.

The theory T+A is always axiomatized by α_{T+A} with: $x\in\alpha_{T+A} :\leftrightarrow x\in \alpha_T \vee x=\ulcorner \underline{A}\urcorner$.

Let $\text{Proof}_T(x,y)$ be the $\Sigma_1{}^b$-formula representing the relation: x is the Gödelnumber of a T-proof of the formula with Gödelnumber y. Proof_T will be built in some standard way from α_T. The precise choice of the system on which Proof_T is based is immaterial: any Hilbert style system or Natural Deduction system or Genzen style sequent system will do. If we want to stress that we are looking at the Proof-relation based on a certain specific formula β we write: Proof_β.

We assume for convenience that: $I\Delta_0+\Omega_1\vdash\forall x\exists!y\ \text{Proof}_T(x,y)$. Let $\text{Prov}_T(y) := \exists x\text{Proof}_T(x,y)$.

We write par abus de langage '$\text{Proof}_T(x,\ A(x_1,...,x_n)\ )$' for: $\text{Proof}_T(x,\ulcorner A(\dot{x}_1,...,\dot{x}_n)\urcorner)$, here:

i) all free variables of A are among those shown.

ii) $\ulcorner A(\dot{x}_1,...,\dot{x}_n)\urcorner$ is the "Gödelterm" for $A(x_1,...,x_n)$ as defined in Smoryński [1985], p43. Here we use instead of the usual numerals the efficient numerals of 3.1, so that: $I\Delta_0+\Omega_1\vdash\forall x_1,...,x_n\exists y\ \ulcorner A(\dot{x}_1,...,\dot{x}_n)\urcorner = y$.

$\Box_T A(x_1,...,x_n)$ will stand for: $\text{Prov}_T(\ulcorner A(\dot{x}_1,...,\dot{x}_n)\urcorner)$.

Occurrences of terms inside $\Box_T$ should be treated with some care. Is $\Box_T(A[t/x])$ intended $(\Box_T A(x))[t/x]$? We will always use the first, i.e. the small scope reading. In cases where t defin provably in U a total function and $U\vdash t=x\rightarrow\Box_V t=x$, the scope distinction may be ignored within U w.r.t. $\Box_V$. We have: $U\vdash\ (\Box_V A(x))[t/x]\ \leftrightarrow\ \Box_V(A[t/x])$.

$\Diamond_T$ will stand for: $\neg\Box_T\neg$.

Let the axiom set of T be given by $\alpha(x)$ then $\Box_T\restriction y$ stands for provability in the theory whos axiom set is given by $(\alpha(x)\wedge x\leq y)$.

4 Interpretations and interpretability

4.1 Interpretations

Interpretations are in this paper: one dimensional global relative interpretations without parameters. Consider two languages L_U and L_V. An interpretation M of L_V in L_U is given by (i) a function F from the relation symbols of L_V to formulas of the language of L_U and (ii) a formula $\delta(a)$ of L_U having just a free. The image of a relation symbol has precisely $a_1,...,a_n$ free, where n is the arity of the relation symbol. The image of = need not be $a_1=a_2$. The function F is canonically extended in the following way: $(R(b_1,...,b_n))^M:=A(b_1,...,b_n)$, where A=F(R). (To make substitution of the b's possible we rename bound variables in A if necessary. In fact it would be neater to set apart bound variables for the F(R) and for δ that do not occur in the original L_V.) $(.)^M$ commutes with the propositional connectives. $(\forall bB)^M:=\forall b(\delta(b)\rightarrow B^M)$. Similarly for $\exists$.

We can easily extend $(.)^M$ again to map proofs π (from assumptions) in L_V to proofs π^M from the translated assumptions in L_U in the obvious way. As is easily seen for a given interpretation M the lengths of the translated objects are given by a fixed polynomial in the lengths of the originals. The graphs of B^M (considered as a function in B and M) and of π^M (considered as a function in π and M) can be arithmetized by $\Delta_1{}^b$-formulas in such a way that the recursive clauses are verifiable in $I\Delta_0+\Omega_1$. Using the polynomial bound on the lengths of the values it is easy to verify that $I\Delta_0+\Omega_1$ proves that these functions are total. (This is verified in detail in Kalsbeek[89].)

4.2 Interpretability

Consider theories U (with language L_U) and V (with language L_V). What does it mean to say that V is interpretable in U via M? I think the obvious definition is this: for every $B\in \alpha_V$ there is a proof in U of B^M. (I assume in this discussion that we are dealing with sentences, in the case of formulas one should consider: $(\delta[B]\rightarrow B^M)$, where $\delta[B]$ is the conjunction of $\delta(b)$'s, for all free variables b of B.) Given this definition the next step is to show: if V is interpretable in U via M and if V proves C, say by π, then there is a proof π^* in U of C^M. Roughly π^* is π^M with proofs of the translated T'-axioms plugged in at the relevant places. Now here is the problem: in a theory like $I\Delta_0+\Omega_1$ we cannot exclude that the proofs of the translated V-axioms are cofinal in the natural numbers. In other words we cannot prove that there is a bound for these proofs. The axiom that would provide such bounds is Σ_1-collection.

Σ_1-collection $\quad \vdash \forall x<u\exists yA(x,y) \rightarrow \exists v\forall x<u\exists y<vA(x,y) \qquad A\in \Sigma_1$

(Note that we could equivalenty state the principle demanding: $A\in \Delta_0$.)

So we would get this basic property in $B\Sigma_1+\Omega_1$, where $B\Sigma_1:= I\Delta_0+\Sigma_1$-collection. In section 5 we elaborate the consequences of the lack of Σ_1-collection a bit more.

One way to evade the problem at hand is to make a definitional move. We change the definition of interpretability in such a way that the basic properties we want are guaranteed even in $I\Delta_0+\Omega_1$, but also in such a way that our definition and the usual one collapse in the presence of $B\Sigma_1+\Omega_1$. In my paper Visser[88b] I used the notion of theorems interpretability, by now I convinced myself that the notion of smooth interpretability introduced below is a better choice.

Define $(\forall x\exists y)^*A(x,y)$ by $\forall u\exists v\forall x<u\exists y<vA(x,y)$. Similarly for more variables. We also write: $(\forall x\in \alpha\exists y\in \beta)^*A(x,y)$ for: $\forall u\exists v\forall x<u(x\in \alpha\rightarrow\exists y<v(y\in \beta\wedge A(x,y)))$

Note that if $(\forall x\exists y)^*A(x,y)$ and $(\forall y\exists z)^*B(y,z)$, then: $(\forall x\exists y,z)^*(A(x,y)\wedge B(y,z))$.

Define:

$K{:}U \rhd_a V \quad :\Leftrightarrow \forall x \in \alpha_V \mathrm{Prov}_U(x^K).$
$K{:}U \rhd_s V \quad :\Leftrightarrow (\forall x \in \alpha_V \exists p)^* \mathrm{Proof}_U(p,x^K).$
$K{:}U \rhd_t V \quad :\Leftrightarrow \forall x \in \mathrm{Sent}_V(\mathrm{Prov}_V(x) \rightarrow \mathrm{Prov}_U(x^K)).$

Our first notion is *axioms interpretability*; our second notion is *smooth interpretability*, our third notion is *theorems interpretability*.

Note that if V is finitely axiomatized and α_V is the obvious formula representing the axioms of V, then these notions collapse in $I\Delta_0+\Omega_1$.

4.2.1 Fact

i) $I\Delta_0+\Omega_1 \vdash K{:}U \rhd_s V \rightarrow K{:}U \rhd_t V$
ii) $I\Delta_0+\Omega_1 \vdash K{:}U \rhd_t V \rightarrow K{:}U \rhd_a V$
iii) $I\Delta_0+\mathrm{EXP} \vdash K{:}U \rhd_t V \rightarrow K{:}U \rhd_s V$
iv) $B\Sigma_1+\Omega_1 \vdash K{:}U \rhd_t V \rightarrow K{:}U \rhd_s V$

Proof: The proof of (i) is simple, but is postponed till section 7 (7.4) where (i) is an immediate consequence of more general facts. The proofs of (ii) and (iv) are trivial.

We turn to the proof of (iii). Reason in $I\Delta_0$+EXP: suppose $K{:}U \rhd_t V$. Fix a bound u. Consider $B{:=}\bigwedge\{A<u \mid A \in \alpha_V\}$. (We need EXP to guarantee the existence of this formula!) As is easily seen B is provable in V, hence B^K is provable in U, say the proof is p. We can construct proofs q of the A^K by appending proofs of A^K from B^K in U to p. In the worst case the number of steps proceeding from B^K is u, the formulas ocurring in each such step are smaller than or equal to B^K. So $|q|<u.(|B^K|+m)+|p|$, for some standard m. So we may take our bound v for the q: $\exp(u.(|B^K|+m)+|p|)$. (Note that $|B^K|$ is about $|K|.u.|u|$.) □

There are two arguments to prefer smooth interpretability over theorems interpretability. First it is conceptually better to retain the distinction between axioms and theorems: the whole point of the fact that interpretations preserve logical structure becomes obscured when one uses theorems interpretability. Secondly the Orey-Hájek characterization is more naturally formulated using smooth interpretability (see section 9).

A somewhat different perspective on the use of $\rhd_s$ instead of $\rhd_a$ will be given in an appendix (section 11).

From now on we write: $M{:}\,U \rhd V$ for $M{:}U \rhd_s V$. Define:

$U \rhd V \quad \Rightarrow \exists M\, M{:}\,U \rhd V$
$M{:}A \rhd_U B \Leftrightarrow M{:}(U+A) \rhd (U+B)$

$$
\begin{array}{ll}
A\rhd_U B & :\Leftrightarrow (U+A)\rhd(U+B)\\
U\equiv V & :\Leftrightarrow U\rhd V \wedge V\rhd U\\
A\equiv_U B & :\Leftrightarrow (U+A)\equiv(U+B)
\end{array}
$$

4.3 Some special interpretations and some operations on interpretations

We can interpret any language in itself by ID:=$<\delta,F>$, where $\delta(x):=(x=x)$ and $F(R):=R$.

Let U be theory containing arithmetic. A U-cut I is given by a fomula $I(x)$ (often written: $x\in I$) such that U proves that $0\in I$ and that I is closed under Successor, Addition, Multiplication and ω_1. (We make the assumption of closure under ω_1 for convenience.)

We refer the reader to the discussion of cuts in Paris & Wilkie[87]. Some further information on cuts can be found in Pudlák[83a].

A cut I induces an interpretation $<\delta,F>$ of the language of arithmetic (par abus de langage we call this interpretation again I) by taking: $\delta:=I$ and $F(R)=R$.

Suppose K is an interpretation of L_1 in L_2, M is an interpretation of L_2 in L_3. Then P:=K∘M, the composition of K and M, is an interpretation of L_1 in L_3, with $F_P(R):=(F_K(R))^M$ and $\delta_P(x):=(\delta_M(x)\wedge(\delta_K(x))^M)$. It is easy to show that for any sentence A: $\vdash A^P\leftrightarrow(A^K)^M$.

Suppose K and M are interpretations of L_1 in L_2. Suppose A is a sentence of L_2. Then P:=K[A]M, the A-join of K and M, is an interpretation of L_1 in L_2, with $F_P(R):=((A\wedge F_K(R))\vee(\neg A\wedge F_K(R)))$ and $\delta_P(x):=((A\wedge\delta_K(x))\vee(\neg A\wedge\delta_M(x)))$. It is easy to show that for any sentence B:

$\vdash B^P\leftrightarrow((A\wedge B^K)\vee(A\wedge B^M))$.

5 Notes on $B\Sigma_1$

In this section we show that the arrow in 4.2.1.(ii) cannot be reversed. Before doing this we briefly mention some well known facts about $B\Sigma_1$.

i) If a model M of $I\Delta_0$ has an endextension N satisfying $I\Delta_0$ then M satisfies $B\Sigma_1$. Suppose $M\models\forall x<a\exists yA(x,y)$, where $A\in\Delta_0$. Let $b\in N\backslash M$. Then $N\models\forall x<a\exists y<bA(x,y)$. By applying the Δ_0-minimum principle in N we find the smallest b^* such that $N\models\forall x<a\exists y<b^*A(x,y)$. It follows that $N\models\exists x<a\forall y<b^*{-}1\neg A(x,y)$. On the other hand if b^* were in $N\backslash M$ we would have: $N\models\forall x<a\exists y<b^*{-}1A(x,y)$. Ergo $b^*\in M$ and thus: $M\models\forall x<a\exists y<b^*A(x,y)$.

ii) Every cut in $B\Sigma_1$ satisfies $B\Sigma_1$. This is immediate from (i).

iii) $B\Sigma_1$ is interpretable in $I\Delta_0$ on a cut. An easy argument to establish this is the following. It is shown in Visser[88b] that $I\Delta_0$ interprets $I\Delta_0+\neg EXP$ say by M. Let N be a model of $I\Delta_0$. M induces an internally definable model of $I\Delta_0+\neg EXP$. Let's call this model again M. There is a definable cut I in N that is isomorphic by an N-definable isomorphism F to an N-definable external cut I* of M (see Pudlák[85] or Visser[88b]). Let J* be a (provably) M-definable (and hence N-definable) cut of M such that for all x in M exp(x) exists in M. Then there is an element of M above $Y^*:=I^*\cap J^*$. So Y* satisfies $B\Sigma_1$. $Y:=F^{-1}(Y^*)$ is an N-definable N-cut isomorphic to Y*. Hence Y satifies $B\Sigma_1$. Since clearly the choice of the definitions of the cuts involved is independent of the specific model N, it follows that $Y{:}I\Delta_0 \rhd B\Sigma_1$.

Let U be the theory axiomatized by the Π_2-consequences of PA. So U extends PRA. We give a construction due to Jeff Paris of a model N of U that does not satisfy $B\Sigma_1$.

Let M be any model of PA+Incon(PA). Let N be the submodel given by the set of Σ_1-definable elements. Clearly N is non-standard and closed under S, + and . . We show that N and M satisfy the same Σ_1-formulas (with parameters in N). We first show: for A(x,...) in Δ_0 and a,... in N: if $M\models A(a,\ldots)$ then $N\models A(a,\ldots)$. This is clear for atoms and negations of atoms. The cases of conjunction and disjunction are easy. Suppose A(x,...) is $\forall y<t(x,\ldots)B(y,x,\ldots)$ and suppose $M\models \forall y<t(a,\ldots)\, B(y,a,\ldots)$. Clearly for every b in N with $b<t(a,\ldots)$: $M\models B(b,a,\ldots)$. It is easy to see that < has the same meaning in N, so: $N\models\forall y<t(a,\ldots)\, B(y,a,\ldots)$. Suppose A(x,...) is $\exists y<t(x,\ldots)\, B(y,x,\ldots)$ and suppose $M\models\exists y<t(a,\ldots)\, B(y,a,\ldots)$. Clearly $M\models\exists y<t(a,\ldots)(B(y,a,\ldots)\wedge\forall z<y\, \neg B(z,a,\ldots))$. Suppose: $M\models b<t(a,\ldots)\wedge B(b,a,\ldots)\wedge\forall z<b\neg B(z,a,\ldots)$. Clearly b is in N, so by the IH we find: $N\models b<t(a,\ldots)\wedge B(b,a,\ldots)$, hence $N\models\exists y<t(a,\ldots)\, B(y,a,\ldots)$.

An immediate consequence is that for A(x,...) in Σ_1 and a,... in N: $M\models A(a,\ldots) \Leftrightarrow N\models A(a,\ldots)$. It follows for B(x,...) in Π_2 and a,... in N: $M\models B(a,\ldots) \Rightarrow N\models B(a,\ldots)$.

So $N\models U$+Incon(PA). Let $SAT_0(x,y,z)$ be the predicate that expresses in U: y is a Σ_1-formula with one free variable and x is a sequence witnessing that z satisfies y. In U (even in $I\Delta_0$+EXP) one can prove that $SAT(y,z):=\exists x SAT_0(x,y,z)$ has the Tarskian properties of a Σ_1-satisfaction predicate. Let DEF(y,z) be $\exists w[(SAT_0((w)_0,y,(w)_1)\wedge(w)_1=z)\wedge\forall v<w\neg(SAT_0((v)_0,y,(v)_1)\wedge(v)_1=z)]$. Let a be non-standard in N. We have: $N\models\forall z\exists y<a DEF(y,z)$ and hence $N\models\forall z<a+1\exists y<a DEF(y,z)$ (*). Suppose to get a contradiction that N satisfies $B\Sigma_1$. Then (*) is equivalent to a Σ_1-formula, hence $M\models\forall z<a+1\exists y<a DEF(y,z)$. M is a model of PA, so we can conclude in M by the Pigeonhole principle that two z's below a+1 share a definition. This leads immediately to a contradiction.

It follows that there is a model of U without an end-extension satisfying $I\Delta_0$.

Now we have the means available to show that U does not prove: $K{:}W\rhd_a V\rightarrow K{:}W\rhd_t V$

Let $Mincon(p){:}\leftrightarrow(Proof_{PA}(p,\perp)\wedge\forall z<p\neg Proof_{PA}(z,\perp))$. Take K:=ID, $W:=I\Delta_0$+EXP and let V be axiomatized by: $x\in\alpha_V :\leftrightarrow x\in\alpha_W\vee\exists z<x\ x=B(z)$, where $B(z)=\ulcorner\exists p(Mincon(p)\wedge\exists y<p DEF(y,\check{z}))\urcorner$.

Finally let $A:=\exists p(\mathrm{Mincon}(p)\wedge\forall z<p+1\exists y<p\mathrm{DEF}(y,z))$.

We show: $N\models \mathrm{ID}:W\rhd_a V$. Reason in N: for some p we have $\mathrm{Mincon}(p)\wedge\forall z\exists y<p\mathrm{DEF}(y,z))$. Hence by Σ_1-completeness $\forall z\Box_W(\mathrm{Mincon}(p)\wedge\exists y<p\mathrm{DEF}(y,z))$. So $\forall z\Box_W B(z)$. $\Box$(N)

We show: $N\models\Box_V A$. Reason in N: consider p satisfying Mincon(p). Clearly by Σ_1-completeness we have some V-proof q of Mincon(p). We can find W-proofs r, with length polynomially bounded in $|p|$ and $|z|$ of $((\mathrm{Mincon}(p)\wedge B(z))\rightarrow\exists y<p\mathrm{DEF}(y,z))$. Ergo we can find V-proofs s of length polynomially bounded in $|q|,|p|,|z|$ of $\exists y<p\mathrm{DEF}(y,z)$. Hence $\forall z<p+1\exists s<\exp(P(|q|,|p|,|z|))$ $\mathrm{Proof}_V(s,\exists y<p\mathrm{DEF}(y,z))$ for some standard polynomial P. Now it is easy to construct a V-proof of $\forall z<p+1\exists y<p\mathrm{DEF}(y,z)$. Thus $\Box_V(\mathrm{Mincon}(p)\wedge\forall z<p+1\exists y<p\mathrm{DEF}(y,z))$. It follows that $\Box_V A$. $\Box$(N)

We show $N\not\models\Box_W A$. Suppose it did. Then we would have $M\models\Box_W A$. But M is a model of PA, and PA proves the reflection principle for W. Hence $M\models A$. Quod non!

So in N the axioms of V are theorems of W, but W has theorems that are not theorems of V. This shows that it is -at least in N- not always a good idea to prove things from first principles. $\Box$

Open Problems

i) Show that $I\Delta_0+\Omega_1\nvdash K:U\rhd_t V\rightarrow K:U\rhd_s V$.
ii) Show that $I\Delta_0+\mathrm{EXP}\nvdash K:(U+A)\rhd_a(U+B)\rightarrow K:(U+A)\rhd_t(U+B)$.
iii) Show that $I\Delta_0+\Omega_1\nvdash U\rhd_a V\rightarrow U\rhd_t V$.

6 The Henkin Construction

We reason in $I\Delta_0+\Omega_1$. Let β be a (standard) formula that provably codes a set of sentences of a language L. We do not place any constraints on the complexity of β, nor do we demand that L contains the language of arithmetic. We assume that L is $\Sigma_1{}^b$. Let V be the theory axiomatized by β.

Let U be any extension of $I\Delta_0+\Omega_1+\mathrm{con}V$. We construct formulas K, D, an extension L^+ of L with new constants C and a substitution function σ (defined on pairs of a formula of L^+ and a sequence of elements of D to sentences of L^+) satisfying the following claims.

Let $(.)^K$ be the interpretation given by D and: $R(x,...)^K:=K(\sigma(\ulcorner\underline{R(x,...)}\urcorner,<x,...>))$. Let $\mathrm{Cl}^K(A(x,...)):=\forall x,...\in D\ A(x,...)$.

Claim 1: there is a standard k such that:

$\forall A\in\mathrm{Sent}(L)\exists p<\omega_1{}^k(A)\ \mathrm{Proof}_U(p,\Box_V A\rightarrow K(A))$

Claim 2: there is a standard n such that:

$$\forall A\in L^{+}\ \exists p<\omega_1{}^{n}(A)\ \mathrm{Proof}_U(p,\mathrm{Cl}^K(A^K(x,\ldots)\leftrightarrow K(\sigma(\ulcorner\underline{A(x,\ldots)}\urcorner,<x,\ldots>)).$$

Claim 3: there is a standard r such that:

$$\forall A\in \mathrm{Sent}(L)\ \exists p<\omega_1{}^{r}(A)\ \mathrm{Proof}_U(p,\Box_V A\rightarrow A^K).$$

Claim 4: Let α provably define a set of sentences of L. Let W be the theory axiomatized by α, then: $(\forall x\in \alpha\exists p)^*\mathrm{Proof}_U(p,x\in \beta) \rightarrow U\rhd W$

Define L^+ as follows: L^+ is the smallest extension of L such that if A is in L^+, containing at most x free, then there are constants $c[\exists xA]$ and $c[\forall xA]$ in L^+. L^+ is again $\Sigma_1{}^b$. We use the coding of section 3.1, but consider it as coding strings of 0 and 1's. Let $\prec$ be the 'initial sequence' ordering. Par abus de langage we write x0 for the concatenation of x with 0 etc. $|x|$ is the length of sequence x. Define:

$u\in T[x] :\Leftrightarrow$ u is a T^+-sentence; $(x)_u=0$ or ($u=\mathrm{NEG}(v)$ and $(x)_v=1$) or (there is a w of the form $\exists zA(z)$ such that $(x)_w=0$ and u codes $A(c[\exists zA])$) or (there is a w of the form $\forall zA(z)$ such that $(x)_w=1$ and u codes $\neg A(c[\forall zA])$).

Note that $u\in T[x]$ is $\Sigma_1{}^b$. Moreover: $\Box_U\forall x,y(x\prec y \rightarrow T[x]\subseteq T[y])$ and $\Box_U T[0]=\emptyset$.

Define further: $x\in \mathrm{TREE} :\Leftrightarrow \mathrm{Con}(V+T[x])$. Clearly $\Box_U\forall x,y((x\prec y \wedge y\in \mathrm{TREE}) \rightarrow x\in \mathrm{TREE})$. Moreover: $\Box_U 0\in \mathrm{TREE}$. We show that $\Box_U\forall x(x\in \mathrm{TREE} \rightarrow(x0\in \mathrm{TREE} \vee x1\in \mathrm{TREE}))$. Reason in U: Suppose $x\in \mathrm{TREE}$, i.e. $\mathrm{Con}(V+T[x])$. Let $u:=|x|+1$. In case u does not code an L^+-sentence we have: $T[x0]=T[x1]=T[x]$, so we are done. We treat the case that u codes a sentence of the form $\exists zA(z)$, the other cases are analogous or easier. So suppose u codes $\exists zA(z)$. Then $T[x0] = T[x]+ \{\exists zA(z),A(c[\exists zA(z)])\}$ (note that the existence of $A(c[\exists zA(z)])$ requires Ω_1) and $T[x1] = T[x]+ \{\neg\exists zA(z)\}$. The constant $c[\exists zA(z)]$ does not occur in $V+T[x]$ (because we used a natural Gödelnumbering), hence we can convert a proof of falsity in $V+T[x]+\{\exists zA(z),A(c[\exists zA(z)])\}$ into a proof of falsity in $V+T[x]+\{\exists zA(z)\}$. Thus if both $V+T[x0]$ and $V+T[x1]$ were inconsistent, we could convert the proofs of inconsistency in a proof of inconsistency of $V+T[x]$ in the usual way. (All these conversions are available in $I\Delta_0+\Omega_1$.) □(U)

Define $\mathrm{PATH}:=\{x\in \mathrm{TREE}$ there is no y in TREE to the left of x$\}$. As is easily seen: $\Box_U\forall x\in \mathrm{PATH}$ $(x0\in \mathrm{PATH} \vee x1\in \mathrm{PATH})$ and $\Box_U 0\in \mathrm{PATH}$. Also $\Box_U\forall x,y\in \mathrm{PATH}(x\prec y \vee y\prec x \vee x=y)$.

Let $X:=\{x|$ for some y in PATH $x=|y|\}$. By the above U proves that 0 is in X and that 0 is closed under successor. By Solovay's methods we can shorten X to a U-cut I. For purposes of presentation we will define our interpretation for L with just one unary relation symbol R. The general case is, of course, precisely the same. Define:

$x \in L^0$:⇔ $x \in I$ and x is a code of an L-sentence.

$x \in L^1$:⇔ $x \in I$ and x is a code of an L^+-sentence.

$x \in F^1(y)$:⇔ $x,v \in I$ and y is a code of a variable, x is a code of an L^+-formula with at most the variable coded by y free.

$x \in D$:⇔ x is of the form c[A] for $A \in L^1$ and A is a sentence of the form $\exists uB(u)$ or $\forall uB(u)$.

$K(x)$:⇔ $x \in L^1$ and there is an $y \in$ PATH with $|y| \leq x$ and $x \in T[y]$.

$\sigma(A,<x,...>)$:= the result of substituting x,... for the variables corresponding to the places in the sequence <x,...> in A. Any remaining variables in A are to be replaced by, say, $c[\exists u\ u=u]$. It is intended that $A \in L^+$ and $x,... \in D$.

(So e.g. $\sigma(\ulcorner \underline{R(x_0)} \urcorner,<x>)=\ulcorner \underline{R(c[A])} \urcorner$ if $x=\ulcorner \underline{c[A]} \urcorner$.)

$R^K(x)$:⇔ $x \in D \wedge K(\sigma(\ulcorner \underline{R(x_0)} \urcorner,<x>))$.

i) We have: $\Box_U \forall x \in L^C\ (Prov_V(x) \rightarrow K(x))$.

Reason in U: Suppose $x \in L^0$ and $Prov_V(x)$. Since x is in I there is a y in PATH with $|y|=x$. Say x codes B. V+T[y] is consistent, and either B or $\neg B$ is in T[y]. Clearly $\neg B$ cannot be in T[y], so B is. □(U)

Note that given the fact that β is standard the U-proof constructed above is standard. Moreover for some standard s we find: $\forall x \exists p<\omega_1{}^s(A)\ Proof_U(p,x \in I)$, also, L being $\Sigma_1{}^b$, for some standard r: $\forall x \in L \exists p<\omega_1{}^r(x)\ Proof_U(p,x \in L)$. We find that $\forall x \in L \exists p<\omega_1{}^q(x)\ Proof_U(p,x \in L^0)$ for some standard q. Combining we get claim 1: $\forall A \in Sent(L)\ \exists p<\omega_1{}^k(A)\ Proof_U(p,\Box_V A \rightarrow K(A))$, for some standard k.

ii) K 'commutes' provably in U with the logical constants on L^1.

We first show (a): $\Box_U \forall x \in L^1\ K(x) \vee K(NEG(x))$ and (b) $\Box_U \forall x \in L^1 \neg(K(x) \wedge K(NEG(x)))$. Reason in U:

a) Consider x in L^1. x is in I so there is an y in PATH with $|y|=x$. In case $(y)_x=0$ we have $x \in T[y]$, hence K(x). In case $(y)_x=1$ we have $NEG(x) \in T[y]$, hence K(NEG(x)).

b) Suppose K(x) and K(NEG(x)). There are y and y' in PATH with x in T[y] and NEG(x) in T[y']. We have y=y' or $y \prec y'$ or $y' \prec y$. Let z be the $\prec$-maximum of y, y'. Clearly both x and NEG(x) are in T[z]. But T[z] is consistent. Contradiction. □(U)

We treat the cases of negation, conjunction and universal quantification: we show

(c) $\Box_U \forall x \in L^1\ K(NEG(x)) \leftrightarrow \neg K(x)$

(d) $\Box_U \forall x,y \in L^1\ K(\mathrm{CONJ}(x,y)) \leftrightarrow (K(x) \wedge K(y))$

(e) $\Box_U \forall y \in I\ (\mathrm{VAR}(y) \rightarrow \forall x \in F^1(y)\ (\ K(\mathrm{UQ}(y,x)) \leftrightarrow \forall z \in D\ K(\sigma(x,y,z))\ )$

Here if z codes c[B], x codes A(u) and y codes u: $\sigma(x,y,z)=\ulcorner A(c[B])\urcorner$. Note that by Ω_1 both UQ(y,x) and σ(x,y,z) are in L^1.

(c) is immediate from (a) and (b). For (d) and (e) reason in U:

d) Consider x, y in L^1 and suppose K(x) and K(y). Let z:=CONJ(x,y). As is easily seen z is in I and hence in L^1. There is a w in PATH with |w|=z. Either z or NEG(z) are in T[w]. As is easily seen x and y are in T[w], so by the consistency of T[w] z must be in T[w], so K[z]. In case e.g. ¬K(x) we have K(NEG(x)) and reasoning as before we find K(NEG(CONJ(x,y))), so ¬K(CONJ(x,y)).

e) Consider $y \in I$ with VAR(y) and $x \in F^1(y)$. First suppose K(UQ(y,x)). Clearly UQ(u,x) is in L^1. Consider z in D. As is easily seen σ(x,y,z) is in L^1. Let v be the maximum of UQ(y,x) and σ(x,y,z). There is a w in PATH with |w|=v. We have UQ(y,x) in T[w] and either σ(x,y,z) or NEG(σ(x,y,z)). By the consistency of T[w] we must have σ(x,y,z) in T[w] and hence K(σ(x,y,z)). Suppose for the converse that ¬K(UQ(y,x)). Let v:=UQ(y,x) and let w be in PATH with |w|=v. Reasoning as before we find that $(v)_w=1$ and thus that $\mathrm{NEG}(\sigma(x,y,z)) \in T[v]$. Clearly $a:=\ulcorner \underline{c}[\urcorner * v * \ulcorner]\urcorner$ is in D and we have ¬K(σ(x,y,a)). □

Note that all the proofs we provided are standard.

At this point we know enough to employ a slightly more convenient notation. We use variables d,e to range over D and write 'K(A(d,...))' for: $K(\sigma(\ulcorner \underline{A(x_0,...)} \urcorner, \langle d,...\rangle))$.

iii) We prove claim 2: for some standard n:

$\forall A \in \mathrm{Sent}(L^+) \exists p < \omega_1^n(A) \mathrm{Proof}_U(p, \forall d,... \in D\ (K(A(d,...)) \leftrightarrow A(d,...)^K))$.

Let's call the statement following $\Box_U$ in claim 2: E{A}. To prove claim 2 we use $\Delta_0(\omega_1)$-induction on A, which is available in $I\Delta_0+\Omega_1$. This induction is trivial using (ii). It is sufficient to provide the bound on the proofs. Equivalently we must provide a standard polynomial P such that the length |p| of p is bounded by P(|A|). Let's call the length of the proof of E{A}: λ(A). I consider a specific example: say A=(B∧C) and suppose we have proofs of E{B} and E{C}. To construct a proof of E{A} we give proofs of: A=CONJ(B,C), and $\forall x\ K(\mathrm{CONJ}(x,y)) \leftrightarrow (K(x) \wedge K(y))$. The length of the first proof is polynomially bounded in |A| and the length of the second one is standard. Now the proofs of E{B}, E{C}, A=CONJ(B,C), and $\forall x\ K(\mathrm{CONJ}(x,y)) \leftrightarrow (K(x) \wedge K(y))$ can be combined to a proof of E{A} of length bounded by: λ(B)+λ(C)+Q(|A|), where Q is a suitable standard polynomial.. For each connective we find such a polynomial. Let Q* be a polynomial that majorizes all polonomials corresponding to the connectives. Noting that |B|+|C|<|A| it is now easy to show that: λ(A)≤|A|.Q*(|A|), e.g. in the case considered we have e.g:

$$\lambda(A)\leq\lambda(B)+\lambda(C)+Q(|A|)\leq|B|Q^*(|B|)+|C|Q^*(|C|)+Q^*(|A|)\leq(|B|+|C|+1)Q^*(|A|)\leq|A|Q^*(|A|).\ \square$$

NOTE: If β is Σ_1 then by a result of Wilkie $I\Delta_0+\Omega_1+\mathrm{con}(V)$ is interpretable on a cut in $Q+\mathrm{con}(V)$. So in this case we can reduce our assumption that $I\Delta_0+\Omega_1+\mathrm{con}(V)$ is contained in U to the assumption that Q+conV is contained in U. In fact we may assume that U contains Q and proves con(V) on a cut.

iv) Claim 3 is a direct consequence of claims 1 and 2.

v) Suppose α provably defines a set of sentences of L. Let W be the theory axiomatized by α. Suppose $(\forall x\in\alpha\exists p)^*\mathrm{Proof}_U(p,x\in\beta)$. We show: $U\rhd W$ (claim 4).

Fix u. Let $A<u$, $A\in\alpha$. For some w depending only on u there is a U-proof $p<w$ of $A\in\beta$. We have for some standard r: $\forall B\in\mathrm{Sent}(L)\ \exists q<\omega_1{}^r(B)\ \mathrm{Proof}_U(q,\square_V B\rightarrow B^K)$. Hence, as is easily seen for siome standard r': $\forall B\in\mathrm{Sent}(L)\ \exists q'<\omega_1{}^{r'}(B)\ \mathrm{Proof}_U(q',\beta(B)\rightarrow B^K)$. Ergo for some standard r*: $\exists q^*<\omega_1{}^{r^*}(\max(u,w))\ \mathrm{Proof}_U(q,A^K)$.Take $v:=\omega_1{}^{r^*}(\max(u,w))$. □

6.1 Corollary: (in $I\Delta_0+\Omega_1$) let β,U,V be as before. Suppose β is $\Sigma_1{}^b$, then $U\rhd V$.

Proof: take in claim 4 $\alpha:=\beta$ (and thus W:=V). As is well known $\forall x\in\beta\exists p<\omega_1{}^r(x)\ \mathrm{Proof}_U(p,x\in\beta)$ for some standard r. So for given u, we may take $v:=\omega_1{}^r(u)$. □

6.2 Corollary: (in $I\Delta_0+\Omega_1$) suppose W is axiomatized by α and α is $\Sigma_1{}^b$. We have:
$\forall x\square_U\mathrm{Con}\restriction x(W)\rightarrow U\rhd W$

Proof: Suppose $\forall x\square_U\mathrm{Con}\restriction x(W)$. Let $\beta(x):\leftrightarrow(\alpha(x)\wedge\mathrm{Con}\restriction x(V))$. Let V be axiomatized by β. We have: $\square_U\mathrm{Con}(V)$. To apply claim 4 we need only show: $(\forall x\in\alpha\exists p)^*\mathrm{Proof}_U(p,x\in\beta)$. Fix u. We have for some standard n: $\forall x\in\alpha\exists p<\omega_1{}^n(x)\ \mathrm{Proof}_U(p,x\in\alpha)$. Moreover $\square_U\mathrm{Con}\restriction u(W)$ and $\square_U(\mathrm{Con}\restriction u(W)\rightarrow\forall y<u\ \mathrm{Con}\restriction y(W))$. Hence for some q: $\mathrm{Proof}_U(q,\forall y<u\ \mathrm{Con}\restriction y(W))$. For some standard k: $\forall x<u\exists p<\omega_1{}^k(x)\ \mathrm{Proof}_U(p,x<u)$. Hence we can construct a U-proof r of $\mathrm{Con}\restriction x(W)$ with $|r|<|q|+|x|^{\exp(k)}+m|x|+s$, with m and s standard. So for some standard a: $r<\omega_1{}^a(\max(x,q))$. Combining we find a U-proof d of $x\in\alpha^*{}_V$ and standard b such that $d<\omega_1{}^b(\max(x,q))$. Take $v:=\omega_1{}^b(\max(u,q))$. □

7 Facta Selecta

In this section we verify various interpretability principles in $I\Delta_0+\Omega_1$.

7.1 Weakening

We have in $I\Delta_0+\Omega_1$: if $\alpha_V \subseteq \alpha_W$, $\alpha_X \subseteq \alpha_U$ and $X \rhd W$, then $U \rhd V$.

7.2 Addition

We verify in $I\Delta_0+\Omega_1$: $(K{:}U \rhd V \wedge \Box_U A^K) \rightarrow U \rhd (V+A)$. (Here $\alpha_{V+A}(x):=(\alpha_V(x) \vee x=\ulcorner \underline{A} \urcorner)$.)

Suppose $K{:}U \rhd V$ and $\mathrm{Proof}_U(p,A^K)$. Fix u. We have a w such that for all $x<u$, $x \in \alpha_V$ there is a $q<w$ $\mathrm{Proof}_U(q,x^K)$. Take $v:=\max(w,p+1)$. As is easily seen for any x with $x<u$, $x \in \alpha_{V+A}$ there is a $q<v$ $\mathrm{Proof}_U(q,x^K)$.

An immediate consequence of 7.2 is: $\Box_U A \rightarrow U \rhd U+A$. (Take $V:=U$, $K:=ID$.)

7.3 Transitivity

We verify in $I\Delta_0+\Omega_1$: $(U \rhd V \wedge V \rhd W) \rightarrow U \rhd W$.

Suppose $K{:}U \rhd V$ and $M{:}V \rhd W$. We show that $M \circ K{:}U \rhd W$. Fix u. Let v be such that for any $x<u$ in α_W there is a $p<v$ with $\mathrm{Proof}_V(p,x^M)$. Let w be such that for any $y<v$ in α_V there is a $q<w$ with $\mathrm{Proof}_U(q,y^K)$. Consider any $x<u$ in α_W. We have a $p<v$ with $\mathrm{Proof}_V(p,x^M)$. Now we can produce a proof p^* from the axioms $\{y^K | y$ is a V-axiom ocurring in $p\}$ of $x^{M \circ K}$. $|p^*|$, the length of p^*, is linear in $|p|$ and $|K|$. Now add U-proofs r of the y^K to p^*. Call the result q. Clearly the V-axioms y ocurring in p satisfy $y<p<v$. So the proofs r satisfy $r<w$. It follows that $|q|<|p^*|.|w|<c|K|.|p|.|w|< c|K|.|v|.|w|$, where c is standard. So for a sufficiently large standard n: $q<\omega^n(\max(K,v,w))$. We may conclude $(\forall x \in \alpha_W \exists q)^*\mathrm{Proof}_U(q,x^{M \circ K})$.

7.4 Smooth interpretability implies theorems interpretability

We verify in $I\Delta_0+\Omega_1$: $(K{:}U \rhd V \wedge \Box_V A) \rightarrow \Box_U A^K$.

Suppose $K{:}U \rhd V$ and $\Box_V A$. We have $ID{:}V \rhd (V+A)$, and hence $ID \circ K{:}U \rhd (V+A)$, so $\Box_U A^K$.

7.5 The principle M_0

By 6.1 we have in $I\Delta_0+\Omega_1$ $(U+\mathrm{Con}(V)) \rhd V$. We can strengthen this to (in $I\Delta_0+\Omega_1$): for S a $\exists\Sigma_1{}^b$-sentence: $V \rhd W \rightarrow (U+\mathrm{Con}(V)+S) \rhd (W+S)$.

Suppose $M{:}V \rhd W$. Let Q be the single axiom of Robinson's Arithmetic. We have $\Box_V Q^M$ and hence $\Box_U \Box_V Q^M$. Also $\Box_U(S \rightarrow \Box_Q S)$ and hence: $\Box_U(S \rightarrow \Box_V S^M)$. Ergo: $\Box_{U+\mathrm{Con}(V)+S} \Box_V S^M$ and hence: $\Box_{U+Con(V)+S}Con(V+S^M)$. We may conclude:

$(U+Con(V)+S)\rhd(U+Con(V+S^M))\rhd(V+S^M)\rhd(W+S)$.

7.6 Disjunction Elimination Property

We verify in $I\Delta_0+\Omega_1$: $((U+A)\rhd V\wedge(U+B)\rhd V)\rightarrow(U+(A\vee B))\rhd V$.

Suppose $K:(U+A)\rhd V$ and $M:(U+B)\rhd V$. We leave it to the reader to check that:

$K[A]M:(U+(A\vee B))\rhd V$.

(K[A]M is introduced in 4.3.)

7.7 A theorem of Feferman

Let S be $\exists\Sigma_1{}^b$. We verify in $I\Delta_0+\Omega_1$ that: $(U+S)\rhd(U+S+Incon(U))$.

We have $\Box_{U+Con(U)}Con(U+Incon(U))$. Hence $(U+Con(U))\rhd(U+Con(U+Incon(U)))$, so by 7.3 and 7.4: $(U+Con(U))\rhd(U+Incon(U))$. Moreover trivially: $(U+Incon(U))\rhd(U+Incon(U))$. Hence by 7.6: $U\rhd(U+Incon(U))$.

We leave it to the reader to prove the following trivial sharpening of Feferman's result: let S be $\exists\Sigma_1{}^b$. Then (in $I\Delta_0+\Omega_1$): $(U+S)\rhd(U+S+Incon(U))$.

Similarly we have for any U-cut I: $(U+S^I)\rhd(U+S^I+Incon^I(U))$. (Again this sharpening is really nothing but a different choice of the natural numbers of U.)

7.8 A generalization of Löb's Principle

We prove in $I\Delta_0+\Omega_1$: $K:U\rhd V\rightarrow\Box^+{}_{I\Delta 0+\Omega 1}(\Box_U(\Box_U{}^KA\rightarrow A)\rightarrow\Box_UA)$.

(Here $\Box_U{}^KA:=(\Box_UA)^K$, $\Box^+{}_WB:=(B\wedge\Box_WB)$.)

Suppose: $K:U\rhd V$.

Find λ such that $\Box^+{}_{I\Delta 0+\Omega 1}(\lambda\leftrightarrow(\Box_U{}^K\lambda\rightarrow A))$. We also have: $\Box_V\Box_U(\lambda\leftrightarrow(\Box_U{}^K\lambda\rightarrow A))$ and hence $\Box_V(\Box_U\lambda\leftrightarrow\Box_U(\Box_U{}^K\lambda\rightarrow A))$. We claim: $\Box_V(\Box_U\lambda\rightarrow\Box_U\Box_U{}^K\lambda)$. This is true because we have: $\Box_UQ^K$ (here Q is the single axiom of Robinson's Arithmetic) and hence $\Box_V\Box_UQ^K$. Also: $\Box_V(\Box_U\lambda\rightarrow\Box_Q\Box_U\lambda)$. It follows that $\Box_V(\Box_U\lambda\rightarrow\Box_U(Q^K\rightarrow\Box_U{}^K\lambda))$.

We find $\Box_V(\Box_U\lambda\rightarrow\Box_UA)$. We may conclude $\Box_U(\Box_U{}^K\lambda\rightarrow\Box_U{}^KA)$. Assume $\Box_U(\Box_U{}^KA\rightarrow A)$. Hence $\Box_U(\Box_U{}^K\lambda\rightarrow A)$ and thus $\Box_U\lambda$. By the same reasoning as before: $\Box_U\Box_U{}^K\lambda$. Hence $\Box_UA$.

This gives us Löb's Rule outside the $\Box$. Note however that we get by $\exists\Sigma_1{}^b$-completeness: $\Box_{I\Delta0+\Omega1}\Box_U(\Box_U{}^K\lambda\rightarrow\Box_U{}^K A)$ and hence $\Box_{I\Delta0+\Omega1}(\Box_U(\Box_U{}^K A\rightarrow A) \rightarrow \Box_U(\Box_U{}^K\lambda\rightarrow A))$, ergo: $\Box_{I\Delta0+\Omega1}(\Box_U(\Box_U{}^K A\rightarrow A) \rightarrow \Box_U\lambda)$. Also $\Box_{I\Delta0+\Omega1}(\Box_U\lambda \rightarrow \Box_U\Box_U{}^K\lambda)$, so:

$$\Box_{I\Delta0+\Omega1}(\Box_U(\Box_U{}^K A\rightarrow A) \rightarrow \Box_U A). \qquad \Box$$

7.9 The principle W*

Let $S\in\exists\Sigma_1{}^b$. We verify in $I\Delta_0+\Omega_1$: $U\rhd V \rightarrow (V+S)\rhd(V+S+\mathrm{Incon}(U))$.

We provide two different proofs.

First proof: Our first proof has two variants: one that uses sequentiality and one that does not.

First variant: Suppose U is sequential and $K{:}U\rhd V$. There is a U-cut I such that U proves that "there is an isomorphism between I and an external cut of the natural numbers of K". We find: $(V+S+\mathrm{Con}(U))\rhd(V+\mathrm{Con}(U+S^I))\rhd(U+S^I)\rhd(U+S^I+\mathrm{Incon}^I(U))$. Now $U+S^I+\mathrm{Incon}^I(U)$ proves $(S\wedge\mathrm{Incon}(U))^K$, so by 7.2: $K{:}(U+S^I+\mathrm{Incon}^I(U))\rhd(V+S+\mathrm{Incon}(U))$.

Ergo $(V+S+\mathrm{Con}(U)) \rhd (V+S+\mathrm{Incon}(U))$. Clearly $(V+S+\mathrm{Incon}(U)) \rhd (V+S+\mathrm{Incon}(U))$. Hence: $(V+S)\rhd(V+S+\mathrm{Incon}(U))$. $\Box$

Second variant: Suppose $K{:}U\rhd V$. We have by 7.8: $\Box_V(\mathrm{Con}(U)\rightarrow\mathrm{Con}(U+\mathrm{Incon}^K(U)))$. Hence: $\Box_V((S\wedge\mathrm{Con}(U))\rightarrow\mathrm{Con}(U+S^K+\mathrm{Incon}^K(U)))$. So we may conclude:

$$(V+S+\mathrm{Con}(U))\rhd(V+\mathrm{Con}(U+S^K+\mathrm{Incon}^K(U)))\rhd(U+S^K+\mathrm{Incon}^K(U))\rhd(V+S+\mathrm{Incon}(U)).$$

Also $(V+S+\mathrm{Incon}(U))\rhd(V+S+\mathrm{Incon}(U))$ and we are done. $\Box$

Second proof: We reason in $I\Delta_0+\Omega_1$: define $\mathrm{conj}(x,0){:=}\ulcorner\top\urcorner$, $\mathrm{conj}(x,y+1){:=}\ulcorner(\urcorner*\mathrm{conj}(x,y)*\ulcorner\wedge\urcorner*x*\ulcorner)\urcorner$. One can produce a $\Delta_1{}^b$-formula representing the graph of conj such that $I\Delta_0+\Omega_1$ proves the recursive clauses of the definition (assuming existence of the righthand side of the second clause). Moreover $I\Delta_0+\Omega_1$ proves: if exp(y) exists then conj(x,y) exists.

Let the interpretation K be given. To fit the proof into our framework we use a variant of Craig's Trick. Define $\alpha_{V*}{:=}\{y|\exists x,p<y\ (y=\mathrm{conj}(x,|p|)\wedge x\in\alpha_V\wedge\mathrm{Proof}_U(p,x^K))\}$. Clearly α_{V*} is $\Sigma_1{}^b$. We call V* the U,K-associate of V.

Step 1: we show $\mathrm{ID}{:}V\rhd V^*$.

Fix u. Consider any $y<u$ in α_{V*}. There are x and p below y, such that $y=\mathrm{conj}(x,|p|)$ and $x\in\alpha_V$. Let q be the obvious proof in propositional logic of y from x. Evidently q has $|p|$ steps in which at most two formulas occur of length smaller then $|y|$ (which is about $|p|.|x|$). So $|q|$ can be estimated

by $2.|y|.|p|+k|p|$, for some standard k. Moreover: $|p|<|y|<|u|$, so $|q|$ can be estimated by $2|u|^2+k|u|$. So for suitably large standard n: $q<\omega_1{}^n(u)$. Choose $v:= \omega_1{}^n(u)$.

Step 2: We show $K:U \rhd V^*$.

Fix u. Consider $y<u$, $y\in \alpha_{V^*}$. There are x and p below y, such that $y=conj(x,|p|)$ and $Proof_U(p,x^K)$. We transform p into a proof q of y^K by appending to p the proof in propositional logic of y^K from x^K. By reasoning essentially the same as in step 1 we find a standard n such that $q<\omega_1{}^n(u)$. Take $v:=\omega_1{}^n(u)$.

Step 3: we show: $K:U \rhd V \rightarrow ID:V^* \rhd V$.

Fix u. Suppose $K:U \rhd V$, so there is a w, such that for any x in α_V with $x<u$ there is a $p<w$ with $Proof_U(p,x^K)$. Consider any x in α_V with $x<u$. Let $p<w$ be a U-proof of x^K. Then $y:=conj(x,|p|)$ is a V^*-axiom. Let q be a proof of x from y. One easily sees that $|q|$ is estimated by $m.|x|.|p|+n.|p|$, for standard m and n. also. $m.|x|.|p|+n.|p|<m.|u|.|w|+n.|u|$. So for suitably large standard k: $q<\omega_1{}^k(max(u,w))$. Pick $v:=\omega_1{}^k(max(u,w))$.

Step 4: $\Box_{V^*}(\Box_{V^*}\bot\rightarrow\Box_U\bot)$

Step 2 gives: $\Box_{V^*}(U \rhd V^*)$, hence: $\Box_{V^*}(\Box_{V^*}\bot\rightarrow\Box_U\bot)$.

Step 5: $U \rhd V \rightarrow (V+S) \rhd\ V+S+\Box_U\bot)$.

Suppose $U \rhd V$. We find: $(V+S) \rhd (V^*+S) \rhd (V^*+S+\Box_{V^*}\bot) \rhd (V^*+S+\Box_U\bot) \rhd (V+S+\Box_U\bot)$. □

8 Modal Principles

The system IL is given by tne following principles:

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

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

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

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

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

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

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

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

J5 $\vdash \Diamond A \rhd A$

From the materials of section 7 we easily see that the following three principles are also arithmetically valid:

M_0 $\vdash A\rhd B \to (\Diamond A\wedge\Box C)\rhd(B\wedge\Box C)$
W $\vdash A\rhd B \to A\rhd(B\wedge\Box\neg A)$
W* $\vdash A\rhd B \to (B\wedge\Box C)\rhd(B\wedge\Box C\wedge\Box\neg A)$

Note that M_0 immediatetely implies J5 and that W* immediately implies W. We show that: $ILWM_0$=ILW*.

First we derive M_0 in ILW*. Suppose $A\rhd B$. It follows that $A\rhd(B\vee\Diamond A)$. Ergo by W*: $((B\vee\Diamond A)\wedge\Box C)\rhd((B\vee\Diamond A)\wedge\Box C\wedge\Box\neg A)\rhd(B\wedge\Box C)$. We may conclude: $(\Diamond A\wedge\Box C)\rhd(B\wedge\Box C)$.

We derive W* in $ILWM_0$ (the argument is due to Dick de Jongh). As is easily seen we have: $(B\wedge\Box C)\rhd((B\wedge\Box C\wedge\Diamond A)\vee(B\wedge\Box C\wedge\Box\neg A))$ (*). Suppose $A\rhd B$. We want to derive: $(B\wedge\Box C)\rhd(B\wedge\Box C\wedge\Box\neg A)$. By (*) it sufficient to show: $(B\wedge\Box C\wedge\Diamond A)\rhd(B\wedge\Box C\wedge\Box\neg A)$ (**). By W we have: $A\rhd(B\wedge\Box\neg A)$, so by M_0: $(\Diamond A\wedge\Box C)\rhd(B\wedge\Box\neg A\wedge\Box C)$. Reshuffling this a bit and strengthening the 'premiss' we find (**). □

Concluding we may say that the system $ILWM_0$ or equivalently ILW* is arithmetically valid in any $\Sigma_1{}^b$-axiomatized theory with designated natural numbers satisfying $I\Delta_0+\Omega_1$.

It is easy to see that $ILWM_0$ corresponds precisely to the ILW-frames with the extra property $RSR\subseteq R$. Hence ILW does not prove M_0. This refutes the conjecture of Visser[88b] that ILW is precisely the interpretability logic of all reasonable arithmetics. So it is time for a new conjecture!

Conjecture: The principles of $ILWM_0$ are precisely the principles valid in all $\Sigma_1{}^b$-axiomatized theories with designated natural numbers satisfying $I\Delta_0+\Omega_1$.

9 Appendix: the Orey-Hájek Characterization

It has often gone unnoticed that there are two quite different proofs of the Orey-Hájek characterization. When we restrict ourselves to, say, extensions of PA in the language of PA the difference between the proofs is immaterial. In our context, however, the two proofs lead to different statements and to a different range of validity.

Remember that by 6.2: $I\Delta_0+\Omega_1\vdash \forall x\Box_U \mathrm{Con}{\ulcorner}x(V) \to U\rhd V$.

9.1 Orey-Hájek 1

$I\Delta_0+EXP\vdash \forall x\Box_U Con{\restriction}x(U) \to (U\rhd V\leftrightarrow\forall x\Box_U Con{\restriction}x(V))$.

Proof: Reason in $I\Delta_0+EXP$: Suppose $\forall x\Box_U Con{\restriction}x(U)$ and $K:U\rhd V$. Consider any x. There is a v such that $\forall z<x(\alpha_V(z)\to\exists p<v Proof_V(p,x^K))$. By Σ_1-completeness it follows that:

$$\Box_U\forall z<x(\alpha_V(z)\to\exists p<v Proof_V(p,x^K)).$$

Also we have: $\Box_U Con{\restriction}v(U)$.

Reason inside $\Box_U$: suppose $Proof_V{\restriction}x(q,\bot)$. The V-axioms z used in q are all smaller than x, and hence their translations z^K have U-proofs p smaller than v. Consider the translated proof q^K. By plugging in the proofs p of the z^K we obtain a U-proof q* of $\bot$. q* will certainly exist, because its length can be bounded by $P(|v|,|q|,|K|)$ for some standard polynomial P. Clearly q* is a $U{\restriction}v$-proof. This contradicts $Con{\restriction}v(U)$ We may conclude: $Con{\restriction}x(V)$. □

9.2 Open Question: Is the dependence of 9.1 on EXP necessary?

9.3 Orey-Hájek 2

$I\Delta_0+\Omega_1\vdash \exists I\in \text{V-cuts}\forall x\Box_V Con^I{\restriction}x(V) \to (U\rhd V\leftrightarrow\exists J\in \text{U-cuts}\forall x\Box_U Con^J{\restriction}x(V))$.

Proof: Reason in $I\Delta_0+\Omega_1$: suppose for some V-cut I: $\forall x\Box_V Con^I{\restriction}x(V)$. We show:

$$(U\rhd V\leftrightarrow\exists J\in \text{U-cuts}\forall x\Box_U Con^J{\restriction}x(V)).$$

The "←"part is just 6.2 for a different choice of the natural numbers in U. We treat the "→" part. Suppose $K:U\rhd V$. We can find a U-cut J* such that U proves: J* is isomorphic by, say F, to an external cut of the natural numbers of K. Suppose the isomorphic image of J* on the K-side is I*. 'In K' take the intersection H of I and I* and let J be the set of F-originals of H. So $J=J^*\cap F^{-1}(I^K)$. As is easily seen J is a U-cut and (using that $\forall x\Box_U x\in J$): $\forall x\Box_U Con^J{\restriction}x(V)$. □

9.4 Remark: The difference between Orey Hájek 1 and 2 becomes nearly invisible if U and V are both essentially reflexive.

There is a characterization parallel to the Orey-Hájek characterization for Π_1-conservativity. Define:

$$U\rhd^*V :\leftrightarrow \forall P\in\Pi_1\text{-sentences } (\Box_V P \to \Box_U P).$$

We have:

9.5 Orey-Hájek for Π_1-conservativity

Suppose U extends $I\Delta_0+EXP$. Then:

$$I\Delta_0+\Omega_1\vdash \forall x\Box_V Con{\restriction}x(V) \to (U\rhd^*V\leftrightarrow\forall x\Box_U Con{\restriction}x(V)).$$

Proof: Reason in $I\Delta_0+\Omega_1$: Suppose $\forall x \Box_V \mathrm{Con}{\upharpoonright}x(V)$. "$\rightarrow$" Trivial. "$\leftarrow$" Consider $P \in \Pi_1$-sentences. Suppose $\Box_V P$. Then for some x $\Box_U \Box_V {\upharpoonright}x P$. Ergo: $\Box_U \mathrm{Con}{\upharpoonright}x(V+P)$. We may assume that x is large enough, so that the axioms of Robinson's Arithmetic occur below x. So we have by Σ_1-completeness: $\Box_U(\neg P \rightarrow \Box_V {\upharpoonright}x(\neg P))$. In other words $\Box_U(\mathrm{Con}{\upharpoonright}x(V+P) \rightarrow P)$. We may conclude $\Box_U P$. □

The dependence on EXP can be avoided if we consider $\forall \Pi_1{}^b$-conservativity instead of Π_1-conservativity.

10 Appendix: on a result of Hájek & Montagna

We sketch an alternative proof of a beautiful result of Hájek and Montagna: suppose U is an extension of $I\Sigma_1$ that is Σ_1-sound$^{(2)}$. Define:

$$A \rhd^*_U B :\leftrightarrow \forall P \in \Pi_1\text{-sentences } (\Box_{U+B} P \rightarrow \Box_{U+A} P).$$

(.)* is an U-∏Con-interpretation of the language of interpretability logic if:

i) (.)* maps propositional atoms to sentences of the language of U,

ii) (.)* commutes with the propositional connectives,

iii) $(\Box A)^* := \Box_U A^*$,

iv) $(A \rhd B)^* := A^* \rhd^*_U B^*$.

10.1 Theorem (Hájek & Montagna)

$\mathrm{ILM} \vdash A \Leftrightarrow$ for all U-∏Con-interpretations (.)*: $U \vdash A^*$.

10.1 generalizes the result of Berarducci-Shavrukov, because, as is well known, in essentially reflexive theories U $\rhd_U$ and $\rhd^*_U$ are provably extensionally equal. Our proof of Hájek-Montagna follows Berarducci's proof of Berarducci-Shavrukov as closely as possible. For the details on the model-theoretic side (and its formalization) the reader is referred to the papers Berarducci[88] and Hájek-Montagna[89].

We work towards the proof of Completeness via a series of lemmas and definitions. We start with a theorem of Hájek.

10.2 Theorem (Hájek)

$I\Sigma_1 \vdash \forall A \in \Sigma_3 \forall x\, \Box_{I\Sigma_1}(A(x) \rightarrow \mathrm{Con}(Q+A(x)))$

Proof (sketch): Reason in $I\Sigma_1$: let A be given. We allow free variables in A, so a moment's reflection will convince the reader that it is sufficient to prove the result for A in Π_2. Note that inside the $\square_{I\Sigma_1}$ x occurs as a (coded) numeral. Fix x. Let $A(\dot{x})$ be $\forall u\exists v A_0(u,v)$, where A_0 is Δ_0. The assumption of $A(\dot{x})$ in $I\Sigma_1$ can be replaced by the introduction of a new function symbol F with defining equation $F(u)=v :\leftrightarrow (A_0(u,v)\wedge\forall w<v\neg A_0(u,w))$. Let's call $I\Sigma_1$ in the extended language plus the defining equation of F: $I\Sigma_1{}^+$. Clearly it is sufficient to prove: $\square_{I\Sigma_1{}^+}$ $\mathrm{Con}(Q+\forall u A_0(u,F(u)))$. Note that $\forall u A_0(u,F(u))$ is $\Pi_1(F)$. As is well known $I\Sigma_1{}^+$ proves $I\Sigma_1(F)$ (*). Moreover in $I\Sigma_1(F)$ we have a $\Sigma_1(F)$-truthpredicate TR (**). Finally $I\Sigma_1$ proves cut-elimination for predicate logic (***). Using (*), (**), (***) one easily shows:

$$\square_{I\Sigma_1{}^+}\forall B\in\Sigma_1(F)(\square_Q B\rightarrow TR(B)).$$

From this the desired result is immediate. □

10.3 Definition

Let X be the set of Boole(Σ_2)-sentences. Let conj(y,v) be the result of taking the v-fold conjunction of y. Clearly: if 2^v exists, then conj(y,v) exists. Define:

$$\beta(x):=\exists p,y<x(\ (y\in X\wedge x=\mathrm{conj}(y,|p|)\wedge \mathrm{Proof}_U(p,y)).$$

Evidently β is $\Sigma_1{}^b$. Let U* be the theory axiomatized by β. $\mathrm{Proof}_{U*}(x,y)$ will be $\Sigma_1{}^b$.

$I\Sigma_1{}^*$ is similarly defined.

The intended analogy here is: U is to U* as GB is to ZF.

10.4 Lemma

$I\Delta_0+\Omega_1\vdash\ \forall y\in X(\mathrm{Prov}_U(y)\leftrightarrow \mathrm{Prov}_{U*}(y))$.

Proof: Reason in $I\Delta_0+\Omega_1$ First suppose $\mathrm{Proof}_{U*}(p,y)$. Let x be a β-axiom used in p. There are y,q<x such that x=conj(y,|q|) and $\mathrm{Proof}_U(q,y)$. So insert into p before the x's the proofs q of y followed by the obvious proofs w of x from y. Call the result p*. It is easy to see that |w| will be estimated by P(|q|,|x|) for some standard polynomial P. The number of insertions will be at most |p|. Note |x|<|p|, |q|<|p|. Hence |p*|<|p|.(|p|+P(|p|,|p|)).

Next suppose $y\in X$ and $\mathrm{Proof}_U(p,y)$. It follows that conj(y,|p|) is in β. We leave it to the reader to show that the length of the proof of y from conj(y,|p|) is estimated by Q(|p|) for some standard polynomial Q. □

10.5 Lemma

$I\Sigma_1{}^*\vdash\forall A\in X\ \square_{I\Sigma_1{}^*}(A\rightarrow\mathrm{Con}(Q+A))$

Proof: immediate from 10.2 and 10.4. . □

10.6 Lemma

$I\Sigma_1^* \vdash \forall A \in X \forall x \Box_{U^*+A} \mathrm{Con}{\restriction}x(U^*+A)$.

Proof: Reason in $I\Sigma_1^*$: let A and x be given. Take the conjunction B of the axioms of U^*+A below x. (B exists because we have EXP.) Clearly B is in X. By 10.5 $\Box_{I\Sigma 1^*}(B \rightarrow \mathrm{Con}(Q+B))$. By elementary reasoning it follows that $\Box_{I\Sigma 1^*}(B \rightarrow \mathrm{Con}{\restriction}x(U^*+A))$. So $\Box_{U^*}(B \rightarrow \mathrm{Con}{\restriction}x(U^*+A))$. Also $\Box_{U^*+A}B$. Hence $\Box_{U^*+A}\mathrm{Con}{\restriction}x(U^*+A)$. □

10.7 Lemma

$I\Sigma_1^* \vdash \forall A,B \in X\ (\ A \rhd^*{}_U B \leftrightarrow A \rhd^*{}_{U^*} B \leftrightarrow A \rhd_{U^*} B\ )$.

Proof: Reason in $I\Sigma_1^*$. The first equivalence is easy: for $A,B \in X$ and $P \in \Pi_1$-sentences, we have: $(A \rightarrow P),(B \rightarrow P) \in X$. Hence $\Box_U(A \rightarrow P) \leftrightarrow \Box_{U^*}(A \rightarrow P)$, and $\Box_U(B \rightarrow P) \leftrightarrow \Box_{U^*}(B \rightarrow P)$.

For the second equivalence note that by 10.5 and respectively 9.5 and 9.1 both $A \rhd^*{}_{U^*} B$ and $A \rhd_{U^*} B$ are equivalent to $\forall x \Box_{U^*+A} \mathrm{Con}{\restriction}x(U^*+B)$. □

10.8 Definition

We call $(.)^*$ an U^*,X-OH-interpretation if:

i) $(.)^*$ maps propositional atoms to X,

ii) $(.)^*$ commutes with the propositional connectives,

iii) $(\Box A)^* := \Box_{U^*} A^*$,

iv) $(A \rhd B)^* := \forall x \Box_{U^*+A} \mathrm{Con}{\restriction}x(U^*+B)$.

Note that if $(.)^*$ is an U^*,X-OH-interpretation, then $A^* \in X$.

10.9 Theorem

$ILM \vdash A \Leftrightarrow$ for all U^*,X-OH-interpretations $(.)^*$: $U^* \vdash A^*$.

Before proving 10.9, we show that 10.9 implies 10.1.

Proof of 10.1 from 10.9: Suppose $ILM \nvdash A$, then there is an U^*,X-OH-interpretation $(.)^*$ such that $U^* \nvdash A^*$. Define an U-Πcon-interpretation $(.)^\circ$ by stipulating that for any atom p: $p^\circ := p^*$. By induction on A one easily shows using 10.4 and 10.6: $U^* \vdash A^\circ \leftrightarrow A^*$. Hence $U^* \nvdash A^\circ$. We may conclude by 10.4: $U\nvdash A^{\circ}$. □

The proof of 10.9 is an adaptation of Berarducci's proof. The trick here is to use only the Orey-Hájek equivalent in the argument. So we must eliminate all 'model-theoretical' reasoning in favour of syntactical arguments.

Sketch of the proof of 10.9: The Soundness side is routine. Suppose, to prove Completeness, that $ILM\nvdash A$. Then there is a simplified ILM-model K with bottom node b, such that $b\nVdash A$. Say the domain of K is V. We can arrange it so that (provably in U*) there is a k>0 such that every x in V forces $\Box^{k}\bot$. We attach a new R-bottom 0 below K.

We define a primitive recursive function F satisfying the following conditions. Let L:=Lim(F), i.e. $L=z :\leftrightarrow \exists x\,(F(x)=z \wedge \forall y>x\ F(y)=z)$. Note that L=z is Σ_2. (One can show that L=z is even Δ_2.) As we will see: $U^*\vdash\exists z\ L=z$, for the moment we will simply assume this fact. We will use L as a term: it should always been given the small scope reading.

10.9.1 Berarducci's conditions

In U* we have:

R) $\forall x,y\in V\cup\{0\}\ (L=x\wedge xRy) \rightarrow \Diamond_{U^*}L=y$

¬R) $\forall x\in V\cap \mathrm{Range}(F)\ \sqsupset_{U^*}xRL$

S) $\forall x\in V\ L=x \rightarrow \forall u\Box_{U^*}\forall y,z\in V((L=y\wedge xRz\wedge ySz)\rightarrow\Diamond_{U^*}\upharpoonright u L=z)$

¬S) $x<y \rightarrow F(x)SF(y)$

We will verify these conditions later on.

Define a U*,X-OH-interpretation (.)* by: $p^*=\exists z\,(L=z\wedge z\Vdash p)$. We show first that (.)* is the counterexample we are looking for.

10.9.2 The proof from the conditions

We show in U*: for all x in V:

(i) $(x\Vdash C\wedge L=x) \rightarrow C^*$

(ii) $(x\nVdash C\wedge L=x) \rightarrow \neg C^*$

We treat the case of $C=E\rhd G$. Reason in U*: Suppose (i) and (ii) hold for E an G.

(i) Suppose $x\in V$, $x\Vdash E\rhd\cdot G$, L=x. To show: $\forall x\Box_{U^*+E^*}\mathrm{Con}\upharpoonright x(U^*+G^*)$. Let u be any large enough number. Reason in □:

Suppose E*. Let y bc as guaranteed by condition ¬R: xRy and L=y. By the IH: $y\Vdash E$. We have: $x \Vdash E \rhd G$ (by Σ-completeness), xRy, $y \Vdash E$. So for some z: ySz, xRz, $z \Vdash G$. By condition S: $\Diamond_{U^*}\ulcorner uL=z$. So by IH: $\Diamond_{U^*}\ulcorner uG^*$.

(ii) Suppose $x \in V$, $x \nVdash E \rhd G$, $L=x$. To show: $\neg\forall x\Box_{U^*+E^*}\mathrm{Con}\ulcorner x(U^*+G^*)$. Let y be such that xRy, $y \Vdash E$, for all z with xRz, ySz: $z \nVdash G$. By R: $\Diamond_{U^*}L=y$. Assume to get a contradiction: $\forall x\Box_{U^*+E^*}\mathrm{Con}\ulcorner x(U^*+G^*)$. Let u be big enough. We have by $\neg$R: $\Box_{U^*}\exists z\, xRz \wedge L=z$, so provided that u is sufficiently large: $\Box_{U^*}\Box_{U^*}\ulcorner u\exists z\, xRz \wedge L=z$. Reason inside $\Box_{U^*}$:

> Suppose $L=y$. By Σ-completeness: $y \Vdash E$, so by IH: E^*. Our assumption gives: $\Diamond_{U^*}\ulcorner uG^*$. Suppose $F(v)=y$, then $\Box_{U^*}\ulcorner uF(v)=y$, so $\Box_{U^*}\ulcorner u\ ySL$. Moreover we have: $\Box_{U^*}\ulcorner uxRL$ By applying Σ-completeness twice we see: $\Box_{U^*}\ulcorner u(x \nVdash E \rhd G)$, $\Box_{U^*}\ulcorner u\ xRy$. Conclude $\Box_{U^*}\ulcorner uL \nVdash G$, so by IH: $\Box_{U^*}\ulcorner u\neg G^*$. Contradiction: so $L \neq y$.

Contradiction, so $\neg\forall x\Box_{U^*+E^*}\mathrm{Con}\ulcorner x(U^*+G^*)$. □

Now $b \in V$ and by assumption $b \nVdash A$. Hence $U^* \vdash L=b \rightarrow \neg A^*$. Suppose $U^* \vdash A^*$, then $U^* \vdash L \neq b$ and hence by the definition of F: $U^* \vdash L \neq 0$. It follows $U^* \vdash \Box_{U^*}{}^{k}\bot$, quod non by Σ_1-soundness.

10.9.3 Definition of F

Let $\lambda(x)$ be the largest U^*-axiom ocurring in x, if there is such. $\lambda(x):=0$ otherwise. We define F simultaneously with an auxiliary function primitive recursive G.

Stage 0:
$F(0):=0$, $G(0):=\infty$.

Stage x+1:
$F(x+1):=u$, $G(x+1):=\lambda(x)$ if

$$[\mathrm{Proof}_{U^*}(x,L\neq u), F(x)Ru] \text{ or } [\mathrm{Proof}_{U^*}(x,L\neq u), F(x)Su, F(\lambda(x))Ru, \lambda(x)<G(x)]\ ;$$

$F(x+1):=F(x)$, $G(x+1):=G(x)$ otherwise.

L is lim(F). One can show in $I\Sigma_1$ that L exists. An immediate consequence is that U^* proves that L exists, the statement "L exists" being Σ_2. Define on $V \cup \{0\}$:

$$\text{R-rank}(x):=\sup\{1+\text{R-rank}(y) \mid xRy\}$$

We can arrange it so that $\lambda x.\text{R-rank}(x)$ is primitive recursive and that for some K $I\Sigma_1$ proves that for all $x \in V \cup \{0\}\text{R-rank}(x)<K$. One can also show in $I\Sigma_1$: $x<y \rightarrow \text{R-rank}(F(x)) \leq \text{R-rank}(F(y))$. It is now easy to show (even without induction because K is standard!) that $\lambda x.\text{R-rank}(F(x))$ will assume a minimum m. Say at u this minimum is assumed. It is easily seen that from u on only the second clause in the definition of F is operative, so whenever the value of F changes (after u) G will decrease. So it is sufficient to show that G assumes a minimum. This uses the Σ_1 Least Number Principle. It is well known that the Σ_1 Least Number Principle is derivable in $I\Sigma_1$.

10.9.4 Proof of Berarducci's conditions

R and $\neg$S are trivial. To prove $\neg$R reason in U*:

Let $x \in V$, $F(u)=x$. Clearly $\Box_{U^*}F(u)=x$, hence $\Box_{U^*}xSL$. We must have: $\Box_{U^*}L \neq x$ by the definition of F. Reason inside $\Box_{U^*}$:

> Say $L=y$. Suppose that not xRy. For some $v \geq u$ and some w: $w \neq y$, $F(v)=w$, $F(v+1)=y$. (This uses the Δ_1 least number principle: $v+1$ is the smallest number above u such that $F(v+1)=y$.) Evidently not wRy. Hence: $Proof_{U^*}(v, L \neq y)$, $\lambda(v)<G(v)$ and $F(\lambda(v))Ry$. Is it possible that $xSF(\lambda(v))$? No, or else $xSF(\lambda(v))Ry$ and thus xRy; quod non. So $\lambda(v)<u$. Ergo $\Box_{U^*}\upharpoonright uL \neq y$. u is an "external" number, so by reflection $L \neq y$, contradiction. Conclude xRy, i.e. xRL. □

To prove S: reason in U*:

Suppose $x \in V$ and $L=x$. Consider any number u that is large enough. Say $F(u)=a$. Clearly aSx, and hence $\Box_{U^*}aSx$. Reason inside $\Box_{U^*}$:

> Suppose: $y,z \in V$, $L=y$, xRz, ySz and (to get a contradiction:) $\Box_{U^*}\upharpoonright u\ L \neq z$. u is "external" so by reflection $L \neq z$, so $z \neq y$. Suppose for all $v \geq w$ $F(v)=y$. It is easy to see that for all $v \geq w$ $G(v)>u$: suppose not, then it would follow that $\Box_{U^*}\upharpoonright u\ L \neq y$, and thus (u being "external") $L \neq y$, quod non. Let p be a U*-proof of $L \neq z$ with $\lambda(p)=u$ and $p>w$. (It is easily seen that such a p should exist!) We have: $Proof_{U^*}(p, L \neq z)$, $F(p)=y$, ySz, $\lambda(p)=u<G(p)$, $F(\lambda(p))SF(u)=aSxRz$ and thus $F(\lambda(p))Rz$. Conclude: $F(p+1)=z$. Contradiction! Ergo $\Diamond_{U^*}\upharpoonright u$ $L \neq z$. □

As it were accidentally we proved two extra theorems.

10.10 Definition

We call (.)* a U*-interpretation if:

i) (.)* maps propositional atoms to sentences of the language of U,
ii) (.)* commutes with the propositional connectives,
iii) $(\Box A)^* := \Box_U A^*$,
iv) $(A \rhd B)^* := A^* \rhd_{U^*} B^*$.

We call (.)* an U*,X-interpretation if (.) is a U*-interpretation and (.)* maps propositional atoms to elements of X.

10.11 Theorem

i) $ILM \vdash A \Leftrightarrow$ for all U*,X-interpretations (.)*: $U^* \vdash A^*$.

ii) (For all U*-interpretations (.)*: $U^* \vdash A^*$) $\Rightarrow$ ILM$\vdash$A.

Proof: Left to the industrious reader. □

10.12 Examples

The following example shows that the interpretability logic of $I\Sigma_1{}^*$ is strictly weaker than ILM. From the arithmetical completeness of ILP (see Visser[88b]) for interpretations in $I\Sigma_1$, we know that there are sentences A,B,C such that $I\Sigma_1 \nvdash A \rhd_{I\Sigma_1} B \rightarrow (A\wedge\Box_{I\Sigma_1}C) \rhd_{I\Sigma_1} (B\wedge\Box_{I\Sigma_1}C)$. $I\Sigma_1$ is finitely axiomatizable (and this fact is verifiable in $I\Sigma_1$). Let D be a single axiom for $I\Sigma_1$. It is easily seen that $I\Sigma_1{}^* \nvdash (D\wedge A) \rhd_{I\Sigma_1*} (D\wedge B) \rightarrow ((D\wedge A)\wedge\Box_{I\Sigma_1*}(D\rightarrow C)) \rhd_{I\Sigma_1*} ((D\wedge B)\wedge\Box_{I\Sigma_1*}(D\rightarrow C))$. Hence we have found A',B',C' such that $I\Sigma_1{}^* \nvdash A' \rhd_{I\Sigma_1*} B' \rightarrow (A'\wedge\Box_{I\Sigma_1*}C') \rhd_{I\Sigma_1*} (B'\wedge\Box_{I\Sigma_1*}C')$.

The following example shows that the interpretability logic of U* for any U extending $I\Sigma_1$ is not a sublogic of ILP: by 10.11(i): $\vdash \Diamond A \rhd \Diamond B \rightarrow (\Diamond A\wedge\Box C) \rhd (\Diamond B\wedge\Box C)$, is valid for U*-interpretations. On the other hand one shows by an easy Kripke model argument that ILP does not imply this principle.

11 Appendix: conservation results for $B\Sigma_1$ over $I\Delta_0$

I think the reader will agree that working with 'smoothened' notions to compensate the absence of Σ_1-collection is rather tiresome. Also, perhaps, comparison of certain arguments in $B\Sigma_1+\Omega_1$ about axioms interpretability with their counterparts in $I\Delta_0+\Omega_1$ about smooth interpretability will have suggested to the reader that there is a systematical relation between these arguments. Ideally what one would like is a method to convert $B\Sigma_1+\Omega_1$-proofs (of some interesting class) leading to a conclusion about axioms interpretability into $I\Delta_0+\Omega_1$-proofs leading to similar conclusions about smooth interpretability.

In this section I will formulate a result that brings us halfway to the ideal: namely a conservation result proved by model theoretical methods. So we will just know that there is an $I\Delta_0+\Omega_1$-proof of the sort we are looking for, but we have no interesting method to find it.

To find our result we just have to take a closer look at a model construction that is well known from the literature.

Let M be any model of $I\Delta_0$. We can, by Compactness, always find an extension N of M such that $(M,\{m|m\in M\})$ is elementary equivalent to $(N,\{m|m\in M\})$ and such that there is an $n^*\in N$ with $M<n^*$ (i.e. for all $m\in M$ $m<_N n^*$). Consider any such model N. Let M* be the model given by $\{n\in N|\exists m\in M\ n<m\}$. Let $A(x,...)$ be any arithmetical formula. We say that $A(x,...)$ is

(M,N)-preserved if for all m,... in M: $M\models A(m,...) \Rightarrow M^*\models A(m,...)$.

11.1 Lemma

i) The Π_1 formulas are (M,N)-preserved. Moreover if M* is closed under ω_1 or EXP then this can be strengthened to $\Pi_1(\omega_1)$, resp. $\Pi_1(\mathrm{EXP})$.

ii) The (M,N)-preserved formulas are closed under conjunction, disjunction and existential quantification.

iii) Suppose A(x,y,...) is (M,N)-preserved and for all $m,k,r... \in M^*$:

$$M^*\models (A(m,r,...)\wedge n<m) \rightarrow A(n,r,...),$$

then $\forall x A(x,y,...)$ is (M,N)-preserved.

Proof: completely trivial. □

From 11.1 it is immediate that M* is a model of $I\Delta_0$ and hence of $B\Sigma_1$. Moreover as is easily seen Ω_1 and EXP are (M,N)-preserved. So if M is e.g. a model of $I\Delta_0+\Omega_1$, then so is M*.

Remember that $(\forall x\exists y)^*A(x,y,z,...)$ means $\forall u\exists v\forall x<u\exists y<vA(x,y,z,...)$. If A is Δ_0 it is easily seen by 11.1(i),(ii) that $B(u,z,...):=\exists v\forall x<u\exists y<vA(x,y,z,...)$ is (M,N)-preserved. B(u,z,...) satisfies the condition of 11.1(iii) in u. So it follows that $(\forall x\exists y)^*A(x,y,z,...)$ is (M,N)-preserved.

A is a Σ_3*-formula if A is of the form $\exists z(\forall x\exists y)^*A_0(x,y,z)$, where A_0 is Δ_0. Note that Σ_2 is (modulo provable equivalence) a subclass of Σ_3*. Note also that Σ_3* is closed (modulo provable equivalence) under conjunction. We similarly define $\Sigma_3^*(\omega_1)$. Clearly Σ_3*-formulas are (M,N)-preserved; if M satisfies Ω then $\Sigma_3^*(\omega_1)$-formulas are (M,N)-preserved.

The interpretability principles P_s we have been considering in section 7 are all of the form: $\forall u(A\rightarrow B)$, where A is a (possibly empty) conjunction of $\Sigma_3^*(\omega_1)$-formulas and B is a $\Sigma_3^*(\omega_1)$-formula. Clearly we may assume that A is a single $\Sigma_3^*(\omega_1)$-formula.

Let P_a be the axioms-interpretability variant of P_s. We want to show: $B\Sigma_1+\Omega_1\vdash P_a \Rightarrow I\Delta_0+\Omega_1\vdash P_s$. Evidently it is sufficient to prove: $B\Sigma_1+\Omega_1\vdash P_s \Rightarrow I\Delta_0+\Omega_1\vdash P_s$, since $B\Sigma_1+\Omega_1\vdash P_a\rightarrow P_s$. In this form, however, the problem does not seem to be solvable: we need an extra observation. The observation is this: in proving a principle P_a in $B\Sigma_1+\Omega_1$, we accomplish a bit more than stated: we explicitely provide the transformation of interpretations involved. This means that we really prove something of the form: $\forall u,z$(z-wit-A$\rightarrow$f(z)-wit-B), where f is an $I\Delta_0+\Omega_1$-provably recursive function. (If C is $\exists zC_0(z)$, then z-wit-C is just $C_0(z)$.) Let's call our principle in this stronger form P^+_s.

11.2 Theorem

$B\Sigma_1+\Omega_1\vdash P^+{}_s \Rightarrow I\Delta_0+\Omega_1\vdash P^+{}_s$

Proof: We reason by contraposition. Suppose M is a model of $I\Delta_0+\Omega_1+\neg P^+{}_s$. Construct N and M* as above. Clearly M* satisfies $B\Sigma_1+\Omega_1$. Hence it is sufficient to show $M^*\models\neg P^+{}_s$. $\neg P^+{}_s$ has the form: $\exists z((\forall x\exists y)^*A_0(x,y,z)\wedge\exists u\forall v\exists a<u\forall b<v\neg B_0(a,b,f(z)))$. By 11.1 it is immediate that $\neg P^+{}_s$ is (M,N)-preserved (using that M* is closed under ω_1), so we are done. □

11.3 Open Problem

Let's write i-$I\Delta_0+\Omega_1$ for the constructivistic version of $I\Delta_0+\Omega_1$, etc. Show by purely syntactical methods: for A,B in Σ_3*: i-$B\Sigma_1+\Omega_1\vdash\forall u(A\rightarrow B) \Rightarrow$ i-$I\Delta_0+\Omega_1\vdash\forall u(A\rightarrow B)$.

Footnotes:

1) The restriction to Σ_1-sound theories is a convenient understatement of the stability phenomenon. For a more accurate treatment see Artemov[86] or Visser[84]
2) The attentive reader will note that what we really need is the much weaker condition: for all $k>0$ $U\nvdash\Box_U{}^k\bot$.

References:

Artemov, 1986, *On Modal Logics axiomatizing Provability,* Math. USSR Izvestia, Vol 27, No.3, 401-429.

Bennet, J.H., 1962, *On spectra,* Thesis, Princeton University, Princeton.

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

Berarducci, A., *The interpretability Logic of Peano Arithmetic,* Manuscript, 1988.

Feferman, S., 1960, *Arithmetization of metamathematics in a general setting,* Fund. Math. 49, 33-92.

Friedman, H., ?, *Translatability and relative consistency II.*

Gaifman, H. & Dimitracopoulos, C., 1982, *Fragments of Peano's Arithmetic and the MRDP theorem,* in: *Logic and Algorithmic,* Monography 30 de l'Enseignement Mathematique, Genève, 187-206.

Guaspari D., 1979, *Partially conservative extensions of arithmetic,* Transactions of the AMS 254, 47-68.

Hájek, P., 1971, *On interpretability in set theories I,* Commentationes Mathematicae Universitatis Carolinae 12, 73-79.

Hájek, P., 1972, *On interpretability in set theories II,* Commentationes Mathematicae Universitatis

Carolinae 13, 445-455.

Hájek, P., 1979, *On partially conservative extensions of arithmetic,* in: Barwise, J. &al eds.,*Logic Colloquium '78,* North Holland, Amsterdam, 225-234.

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

Hájek, P., ?, *On partially conservative extensions of arithmetic II.*

Hájek, P., ?, *Positive results on fragments of arithmetic.*

Hájek, P., ?, *On logic in fragments of arithmetic.*

Hájek, P., & Hájková, M., 1972, *On interpretability in theories containing arithmetic,* Fundamenta Mathematica 76, 131-137.

Hájek, P., & Montagna, M., 1989, *ILM is the logic of Π_1-conservativity,* manuscript.

Hájek, P., & Svejdar, V., 1989, *A note on the normal form of closed formulas of Interpretability Logic,* manuscript.

Jongh, D.H.J. de & Veltman F., 1988, *Provability logics for relative interpretability,* ITLI Prepublication series, ML-88-03; to appear in the Proceedings of the Heyting Conference, Bulgaria, 1988.

Jongh, D.H.J. de & Visser A., 1989, *Explicit Fixed Points in Interpretability Logic,* Logic Group Preprint Series 44, Department of Philosophy, University of Utrecht.

Kalsbeek, M.B., 1988, *An Orey Sentence for Predicative Arithmetic,* ITLI Prepublication Series X-8-01.

Lindström, P., 1979, *Some results on interpretability,* in: Proceedings of the 5th Scandinavian Logic Symposium, Aalborg, 329-361.

Lindström, P., 1980, *Notes on partially conservative sentences and interpretability,* Philosophical Communications, Red Series no 13, Göteborg.

Lindström, P., 1981, *Remarks on provability and interpretability,* Philosophical Communications, Red Series no 15, Göteborg.

Lindström, P., 1982, *More on partially conservative sentences and interpretability,* Philosophical Communications, Red Series no 17, Göteborg.

Lindström, P., 1984a, *On faithful interpretability, in:* Richter, M.M. &al eds, *Computation and Proof Theory,* Logic Colloquium '83, Lecture Notes in Mathematics 1104, Springer Verlag, Berlin, 279-288.

Lindström, P., 1984b, *On certain lattices of degrees of interpretability,* Notre Dame Journal of Formal Logic 25, 127-140.

Lindström, P., 1984c, *On partially conservative sentences and interpretability,* Proceedings of the AMS 91, 436-443.

Lindström, P., 1984d, *Provability and interpretability in theories containing arithmetic,* Atti degli incontri di logica matematica 2, 431-451.

Nelson E., 1986, *Predicative Arithmetic,* Math Notes 32, Princeton University Press, Princeton.

Orey, S., 1961, *Relative Interpretations,* Zeitschrift für Mathematische Logik und Grundlagen der Mathematik 7, 146-153

Parikh, R., 1971, *Existence and feasibility in arithmetic,* JSL 36, 494-508.

Paris, J.B., Dimitracopoulos, C., 1983, *A note on the undefinability of cuts,* JSL 48, 564-569.

Paris, J., Wilkie, A., 1987, *On the scheme of induction for bounded arithmetic formulas,* Annals for Pure and Applied Logic 35, 261-302.

Pudlák, P., 1983a, *Some prime elements in the lattice of interpretability types,* Transactions of the AMS 280, 255-275.

Pudlák, P., 1983b, *A definition of exponentiation by a bounded arithmetical formula,* Commentationes Mathematicae Universitatis Carolinae 24, 667-671.

Pudlák, P., 1985, *Cuts, consistency statements and interpretability,* JSL 50, 423-441.

Shavrukov, V. Yu., 1988, *The Logic of Relative Interpretability over Peano Arithmetic,* Preprint 5 of the Steklov Mathematical Institute.

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

Smorynski, C., 1985b, *Nonstandard models and related developments in the work of Harvey Friedman,* in: Harrington, L.A. &alii eds., *Harvey Friedman's Research on the Foundations of Mathematics,* North Holland, 212-229.

Solovay, R., ?, *On interpretability in set theories.*

Svejdar, V., 1978, *Degrees of interpretability,* Commentationes Mathematicae Universitatis Carolinae 19, 789-813.

Svejdar, V., 1981, *A sentence that is difficult to interpret,* Commentationes Mathematicae Universitatis Carolinae 22, 661-666.

Svejdar, V., 1983, *Modal analysis of generalized Rosser sentences,* JSL 48, 986-999.

Svejdar, V., 1989, *Some independence results in Interpretability Logic,* manuscript.

Takeuti, G., 1988, *Bounded arithmetic and truth definition,* Annals of Pure & Applied Logic 36, 75-104.

Tarski, A., Mostowski, A., Robinson, R.M., 1953, *Undecidable theories,* North Holland, Amsterdam.

Visser, A., 1984, *The provability logics of recursively enumerable theories extending Peano Arithmetic at arbitrary theories extending Peano Arithmetic,* Journal of Philosophical Logic 13, 79-113.

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

Visser, A., 1988b, *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, Bulgaria, 1988.

Visser, A., 1989, *An inside view of EXP: the closed fragment of the provability logic of $I\Delta_0+\Omega_1$ with a propositional constant for EXP,* Logic Group Preprint Series nr 43, Dept. of Philosophy, University of Utrecht, Heidelberglaan 2, 3584CS Utrecht.

Logic Group Preprint Series

Department of Philosophy
University of Utrecht
Heidelberglaan 2
3584 CS Utrecht
The Netherlands

1 C.P.J. Koymans, J.L.M. Vrancken, *Extending Process Algebra with the empty process*, September 1985

2 J.A. Bergstra, *A process creation mechanism in Process Algebra*, September 1985

3 J.A. Bergstra, *Put and get, primitives for synchronous unreliable message passing*, October 1985

4 A. Visser, *Evaluation, provably deductive equivalence in Heyting's arithmetic of substitution instances of propositional formulas*, November 1985

5 G.R. Renardel de Lavalette, *Interpolation in a fragment of intuitionistic propositional logic*, January 1986

6 C.P.J. Koymans, J.C. Mulder, *A modular approach to protocol verification using Process Algebra*, April 1986

7 D. van Dalen, F.J. de Vries, *Intuitionistic free abelian groups*, April 1986

8 F. Voorbraak, *A simplification of the completeness proofs for Guaspari and Solovay's R*, May 1986

9 H.B.M. Jonkers, C.P.J. Koymans & G.R. Renardel de Lavalette, *A semantic framework for the COLD-family of languages*, May 1986

10 G.R. Renardel de Lavalette, *Strictheidsanalyse*, May 1986

11 A. Visser, *Kunnen wij elke machine verslaan? Beschouwingen rondom Lucas' argument*, July 1986

12 E.C.W. Krabbe, *Naess's dichotomy of tenability and relevance*, June 1986

13 H. van Ditmarsch, *Abstractie in wiskunde, expertsystemen en argumentatie*, Augustus 1986

14 A. Visser, *Peano's Smart Children, a provability logical study of systems with built-in consistency*, October 1986

15 G.R. Renardel de Lavalette, *Interpolation in natural fragments of intuitionistic propositional logic*, October 1986

16 J.A. Bergstra, *Module Algebra for relational specifications*, November 1986

17 F.P.J.M. Voorbraak, *Tensed Intuitionistic Logic*, January 1987

18 J.A. Bergstra, J. Tiuryn, *Process Algebra semantics for queues*, January 1987

19 F.J. de Vries, *A functional program for the fast Fourier transform*, March 1987

20 A. Visser, *A course in bimodal provability logic*, May 1987

21 F.P.J.M. Voorbraak, *The logic of actual obligation, an alternative approach to deontic logic*, May 1987

22 E.C.W. Krabbe, *Creative reasoning in formal discussion*, June 1987

23 F.J. de Vries, *A functional program for Gaussian elimination*, September 1987

24 G.R. Renardel de Lavalette, *Interpolation in fragments of intuitionistic propositional logic*, October 1987 (revised version of no. 15)

25 F.J. de Vries, *Applications of constructive logic to sheaf constructions in toposes*, October 1987

26 F.P.J.M. Voorbraak, *Reaeneren met onzekerheid in expertsystemen*, November 1987

27 P.H. Rodenburg, D.J. Hoekzema, *Specification of the fast Fourier transform algorithm as a term rewriting system*. December 1987

28 D. van Dalen, *The war of the frogs and the mice, or the crisis of the Mathematische Annalen,* December 1987

29 A. Visser, *Preliminary Notes on Interpretability Logic,* January 1988

30 D.J. Hoekzema, P.H. Rodenburg, *Gauß elimination as a term rewriting system*, January 1988

31 C. Smoryński, *Hilbert's Programme,* January 1988

32 G.R. Renardel de Lavalette, *Modularisation, Parameterisation, Interpolation,* January 1988

33 G.R. Renardel de Lavalette, *Strictness analysis for POLYREC, a language with polymorphic and recursive types*, March 1988

34 A. Visser, *A Descending Hierarchy of Reflection Principles,* April 1988

35 F.P.J.M. Voorbraak, *A computationally efficient approximation of Dempster-Shafer theory*, April 1988

36 C. Smoryński, *Arithmetic Analogues of McAloon's Unique Rosser Sentences,* April 1988

37 P.H. Rodenburg, F.J. van der Linden, *Manufacturing a cartesian closed category with exactly two objects*, May 1988

38 P.H. Rodenburg, J.L.M.Vrancken, *Parallel object-oriented term rewriting : The Booleans*, July 1988

39 D. de Jongh, L. Hendriks, G.R. Renardel de Lavalette, *Computations in fragments of intuitionistic propositional logic*, July 1988

40 A. Visser, *Interpretability Logic,* September 1988

41 M. Doorman, *The existence property in the presence of function symbols,* October 1988

42 F. Voorbraak, *On the justification of Dempster's rule of combination*, December 1988

43 A. Visser, *An inside view of EXP, or: The closed fragment of the provability logic of* $I\Delta_0+\Omega_1$, February 1989

44 D.H.J. de Jongh & A. Visser, *Explicit Fixed Points in Interpretability Logic,* March 1989

45 S. van Denneheuvel & G.R. Renardel de Lavalette, *Normalisation of database expressions involving calculations*, March 1989

46 M.F.J. Drossaers, *A Perceptron Network Theorem Prover for the Propositional Calculus*, July 1989

47 A. Visser, *The Formalization of Interpretability,* August 1989