Interpretability Logic

Albert Visser
University of Utrecht

1 Introduction

Interpretations are much used in metamathematics. The first application that comes to mind is their use in reductive Hilbert-style programs. Think of the kind of program proposed by Simpson, Feferman or Nelson (see Simpson[1988], Feferman[1988], Nelson[1986]). Here they serve to compare the strength of theories, or better to prove conservation results within a properly weak theory. An advantage of using interpretations is that even if their use should -perhaps- be classified as a proof-theoretical method, it is often possible to employ a model-theoretical heuristics. An example is given in section 7.2 where a conservation result due to Paris & Wilkie, which is proven by a model-theoretical argument, is formalized in a weak theory. For more discussion of and perspective on the use of interpretability in reductive programs the reader is referred to Feferman[1988].

A second application is the use of an interpretation of Elementary Syntax e.g in proving Gödel's Second Incompleteness Theorem: here the interpretation is essential both for the significance of the result and for the heuristics of the argument.

The notion of relative interpretability was made explicit in Tarski, Mostowski, Robinson [1953]; it was systematically studied in the twin pioneering papers Feferman[1960] and Orey[1961]. Lattices of interpretability types were considered in much detail e.g in Montague[1958], Mycielski[1962, 1977], Svejdar[1978], Lindström[1979], Pudlák[1983a]. The interest in these lattices is clearly motivated by the view that interpretability is an adequate means for comparison of strength of theories. Characterizations of relative interpretability for various kinds of theories were obtained by Hájek applying the Orey Compactness Theorem (for essentially reflexive theories) and by Friedman and Pudlák independently (for finitely axiomatized sequential theories; see respectively Smorynski[1985b] and Pudlák[1985]; a presentation of part of Friedman's result is given in sections 7.2, 7.3).

Both Solovay and Lindström proved that relative interpretability over essentially reflexive theories like PA or ZFC is complete Π_2 (see Lindström[1979], Solovay[?]). To be more specific for example the set of Σ_1-sentences S such that ZFC interprets

ZFC+S is complete Π_2. This awesome complexity has suggested to some that the usual notion allows too many interpretations. I'm not quite convinced: nobody said we have to *use* all of them. Another response is to study restrictions of the usual notion: there is still room for a lot of experimentation here.

Modal logics for interpretability were first studied by Hájek and then by Svejdar (see Hájek [1981], Svejdar[1983]). They studied logics with modal operators for provability and interpretability and with witness comparison relations. In Svejdar's system a number of important arguments can be formulated. Moreover Svejdar provides a number of different interpretations of his system. What one seeks in a Svejdar-type approach (which is analogous to Smorynski's approach in his "Ubiquitous Fixed Point Calculation") is a system that is as weak as possible, but still codifies the relevant class of arguments, the point being unification and simplication of a number of specific arguments from the literature. There is no need for the system to be complete w.r.t. any set of interpretations.

The approach in this paper is somewhat different: the focus of interest is to find logics that are sound and complete for interpretations in a given theory (or class of theories). If we know that a logic is sound and complete for interpretations in a given theory and a modal formula ϕ is consistent with the logic, then we know that we can find an interpretation of ϕ that is consistent with the given theory. Typically this interpretation is explicitly given by the proof of the Completeness Theorem.

Solovay's Completeness Theorem for *provability logic* is remarkably general: we have the same logic, viz. Löb's Logic L, for all theories T with the following properties: (i) they have a Σ_1-pro- vability predicate, (ii) they extend $I\Delta_0$+EXP, and (iii) they do not prove their own n-iterated inconsistency (i.e. $\Box_T^n\bot$) for any n. (If a theory T satisfies (i) and (ii), but not (iii) let n* be the least n such that $T\vdash\Box_T^n\bot$, then the provability logic of T is L+ $\Box^{n^*}\bot$. Suppose T has an R_1^+-provability predicate, extends $I\Delta_0+\Omega_1$ and has property (iii), then we know that L is sound for interpretations in T, but we do not know in general whether L is complete for interpretations in T. Specifically it is an open question what the provability logic of $I\Delta_0+\Omega_1$ is.) From one point of view the generality of Solovay's theorem is a disadvantage: one cannot expect information from it connected with specific properties of the theory considered. In this respect interpretability fares better: it turns out, for example, that properties like finiteness and essential reflexivity induce essentially different interpretability principles.

We study two kinds of questions. Let some property Ξ of theories be given: (i) which interpretability principles are valid in all theories satisfying Ξ?; and (ii) does Ξ determine the interpretability principles valid for interpretations in any given theory T satisfying Ξ? In this paper the following specific instances of questions (i) and (ii) are considered: (a) which interpretability principles are valid in all R_1^+-axiomatized theories extending $I\Delta_0+\Omega_1$?; (b) what is the interpretability logic of a given verifiably essentially reflexive theory U?; (c) what is the interpretability logic of a given finitely axiomatized sequential theory U extending $I\Delta_0+\Omega_1$? For questions (a), (b) conjectures are formulated. An answer is available for (c) in case U extends $I\Delta_0$+SUPEXP.

2 Contents

Section 5 contains the necessary preliminaries. In section 6 the systems of interpretability logic IL, ILW, ILP and ILM are introduced. We take a brief look at their consequences and discuss their Kripke semantics and arithmetical significance. In section 7 the form of Friedman's characterization of interpretability for finitely axiomatized sequential theories that is needed to prove our arithmetical completeness result is derived. It turns out that it is convenient to prove this result from a technical lemma (7.2). This lemma is the formalized version of a result of Paris & Wilkie which provides a connection between $I\Delta_0+\Omega_1$ and $I\Delta_0$+EXP. I think this lemma is of some independent interest. Finally in section 8 it is shown that ILP is a complete axiomatization of the interpretability logic of finitely axiomatized sequential theories extending $I\Delta_0$+SUPEXP

3 Acknowledgements

The research on which this paper reports is part of a project together with Dick de Jongh, Craig Smorynski and Frank Veltman. Discussions with them were very important for me. Correspon- dence with George Kreisel and with Franco Montagna has been invariably stimulating.

4 Prerequisites

We presuppose some knowledge of Smorynski[1985a], Paris & Wilkie[1987], Pudlák[1985, 1986].

5 Conventions, Notions & Elementary Facts.

5.1 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. . Of course this is a severe and unjustifiable restriction. I am convinced that the restriction can be dropped almost everywhere. My only excuse is that at some places -especially where tableaux provability is involved- the use of a language with function symbols asks for some extra work: work I have not yet done.

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 assigning to n the Gödelnumber of $\exists y\ y=\underline{n}$, where $\underline{n}$ is the numeral in the sense of Paris and Wilkie[1987] 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 fomulas become only polynomially longer, so the function is also total for the relational language.

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

5.2 Special Classes of Formulas

We refer the reader to the discussion of special classes of formulas in Paris & Wilkie[1987].

Δ_0-formulas are formulas where all quantifiers are bounded by terms in 0, S, + and . (or rather the translations of such formulas in the relational language), where the variable of quantification does not occur in the bounding term. If the theory we are working in proves that some function f with Δ_0-graph is total, we may want to consider $\Delta_0(f)$-formulas, where the bounding terms also involve f. In Gaifman & Dimitracopoulos[1982] it is shown that if f is reasonable -roughly: if it doesn't jump up an down wildly- then $I\Delta_0$+"f is total" implies $I\Delta_0(f)$. For our purposes it is sufficient to know that ω_1 and exp are reasonable; here: $\exp(x):=2^x$.

5.3 Theories and Provability

We consider only theories with identity for which a fixed formulas of their language are specified giving us 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. As variables for general objects of the theory we will use a,b,... . 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 different theories; same α same theory. Unless explicitely stated otherwise we will always assume that α is an R_1^+-formula.

Let $Proof_T(x,y)$ be the R_1^+-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: $Proof_\beta$.

We assume for convenience that: $I\Delta_0+\Omega_1 \vdash \forall x \exists! y\ Proof_T(x,y)$. Let $Prov_T(y) := \exists x Proof_T(x,y)$.

We write par abus de langage '$Proof_T(u, \phi(x_1,...,x_n))$' for: $Proof_T(u, \ulcorner\phi(\dot{x}_1,\dots,\dot{x}_n)\urcorner)$, here:

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

ii) $\ulcorner\phi(\dot{x}_1,\dots,\dot{x}_n)\urcorner$ is the "Gödelterm" for $\phi(x_1,...,x_n)$ as defined in Smoryński [1985], p43. Here we use instead of the usual numerals the efficient numerals of Paris & Wilkie[1987], so that: $I\Delta_0+\Omega_1 \vdash \forall x_1,...,x_n \exists y\ \ulcorner\phi(\dot{x}_1,\dots,\dot{x}_n)\urcorner = y$.

$\Box_T\phi(x_1,...,x_n)$ will stand for: $Prov_T(\ulcorner\phi(\dot{x}_1,\dots,\dot{x}_n)\urcorner)$.

Occurrences of terms inside $\Box_T$ should be treated with some care. Is $\Box_T(\phi[t/x])$ intended or $(\Box_T\phi(x))[t/x]$? We will always use the first, i.e. the small scope reading. In cases where: U proves that t is total 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: $U \vdash (\Box_V\phi(x))[t/x] \leftrightarrow \Box_V(\phi[t/x])$.

We will also need normalized or cut-free provability: here we could choose Herbrand provability (as used in Pudlák[1985]) or cut-free provability in a sequent system or tableaux provability. We use tableaux provability as in Paris & Wilkie[1987]: we write $\mathrm{Tproof}_T(x,y)$ for: Tabincon- proof(U,x), where U is T plus the negation of the formula coded by y. $\mathrm{Tproof}_T(x,y)$ is given by an R_1^+-formula. $\mathrm{Tprov}_T(y)$ is $\exists x\,\mathrm{Tproof}(x,y)$. Tcon(T) is $\forall x \neg \mathrm{Tproof}(x,\ulcorner\underline{\bot}\urcorner)$. $\Delta_T\phi(x_1,...,x_n)$ will stand for: $\mathrm{Tprov}_T(\ulcorner\phi(\mathbf{x}\dot{\mathbf{C}}_1,...,\mathbf{x}\dot{\mathbf{C}}_n)\urcorner)$. Of course our remarks about scope of terms carry over to Δ.

$\Diamond_T$ will stand for: $\neg\Box_T\neg$, and ∇_T for: $\neg\Delta_T\neg$.

Let the axiom set of T be given by $\alpha(x)$ then $\Box_T\ulcorner y$ stands for provability in the theory whose axiom set is given by $(\alpha(x)\wedge x<y)$. $\Box_{T,x}$ will stand for restricted provability in the sense of Paris & Wilkie[1987].

For convenience we write $\Box_\Omega$ for provability in $I\Delta_0+\Omega_1$ and $\Box_{EXP}$ for provability in $I\Delta_0+EXP$.

5.4 Special Properties of Theories

A theory T, with designated natural numbers satisfying $I\Delta_0+\Omega_1$, is *sequential* if in it one can form sequences of any of its objects i.e. there is a relation $(s)_x=a$ such that T proves:

(i) $\forall s,x,a,b\ (\ ((s)_x=a\wedge(s)_x=b) \rightarrow a=b\ )$,

(ii) $\forall s\exists x\forall y\ (\ \exists b\ (s)_y=b \leftrightarrow y<x\ )$

(iii) $\exists s\forall x,a \neg (s)_x=a$,

(iv) $\forall s,a,x(\ \forall y<x\exists b\ (s)_y=b \rightarrow$
$\exists s'\forall b\forall y\leq x\ (\ (s')_y=b \leftrightarrow ((y<x\wedge(s)_y=b)\vee(y=x\wedge a=b))\ )\ )$

Our notion of sequentiality is only seemingly more restrictive than those in the literature: for any theory that is sequential e.g. in the sense of Pudlák[1983] one can define set of natural numbers satisfying $I\Delta_0+\Omega_1$ and a relation $(s)_x=a$ making the theory sequential in our sense. The notion of sequentiality is due to Pudlák. We will describe several important properties of sequential theories later.

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

A theory T is *essentially reflexive* if for all fomulas $\phi(x,...)$ of its language and for all natural numbers n: $T \vdash \forall x,...(\Box \ulcorner \underline{n}\ \phi(x,...) \rightarrow \phi(x,...))$. T is *verifiably* essentially reflexive if T is essentially reflexive and T proves the formalization of "T is essentially reflexive".

5.5 Interpretability

Interpretations are in this paper: one dimensional global relative interpretations without parameters. Consider two languages *L* and *L'*. An interpretation M of *L'* in *L* is given by (i) a function F from the relation symbols of *L'* to formulas of the language of *L* and (ii) a formula $\delta(a)$ of *L* 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: $(R(b_1,...,b_n))^M:=\phi(b_1,...,b_n)$, where $\phi=F(R)$. (To make substitution of the b's possible we rename bound variables in ϕ 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'*.) $(.)^M$ commutes with the propositional connectives. $(\forall b\psi)^M:=\forall b(\delta(b)\rightarrow\psi^M)$. We can easily extend $(.)^M$ again to map proofs π (from assumptions) in *L'* to proofs π^M from the translated assumptions in *L* in the obvious way (for free variables b one adds $\delta(b)$ as a hypothesis). 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 ψ^M (considered as a function in ψ and M) and of π^M (considered as a function in π and M) can be arithmetized by R_1^+-formulas in such a way that the recursive clauses are verifiable in $I\Delta_0+\Omega_1$. Because of the bound on the lengths of the values $I\Delta_0+\Omega_1$ proves that these functions are total.

Consider theories T (with language *L*) and T' (with language *L'*). What could it mean to say that T' is interpretable in T via M? I think the obvious interpretation is this: for every axiom ψ of T' there is a proof in T of ψ^M. (I assume in this discussion that we are dealing with sentences, in the case of formulas one should consider: $(\delta[\psi]\rightarrow \psi^M)$, where $\delta[\psi]$ is the conjunction of $\delta(b)$'s, for all free variables b of ψ.) Given the definition the next step is to show: if T' is interpretable in T via M and if T' proves χ, say by π, then there is a proof π^* in T of ψ^M. Roughly π^* is π^M with proofs of the translated T'-axioms plugged in at the relevant places. Now here is the problem: the verification of the existence of π^* requires (prima facie) $I\Sigma_1$, so in weak theories we don't have this step available. On the other hand what is the point of interpretability if we don't have the π^*?

Let us say that:

T' is *a-interpretable* via M in T if for every axiom ψ of T'
there is a proof in T of ψ^M.

T' is *t-interpretable* via M in T if for every theorem χ of T'
there is a proof in T of χ^M.

The proof π^* as described above could be said to *simulate* π.

T' is *s-interpretable* via M in T if for every proof π of T' there is a simulating proof π^* in T.

Clearly (in $I\Delta_0+\Omega_1$) s-interpretability implies t-interpretability which in turn implies a-interpretabi- lity. My choice to solve the problem mentioned above is simply to take t-interpretability as my notion of interpretability. One could argue that from the philosophical point of view s-interpretability would be the best choice. However t-interpretability is somewhat easier to define and somewhat easier to work with. Moreover I am not aware of any point where the difference between the notions becomes important.

Note that our problem vanishes if T' is finitely axiomatized: it is easy to see that in this case $I\Delta_0+\Omega_1$ proves that a-interpretability implies t-interpretability. A further idea is to impose a bound on the proofs of the translated axioms of T':

T' is *e-interpretable* via M in T if there is a polynomial p such that for every axiom ψ of T' there is a proof in T of ψ^M that is shorter than p of the length of ψ.

Again it is not difficult to see that $I\Delta_0+\Omega_1$ proves that e-interpretability implies t-interpretability. Moreover by applying a well known result we find: if $I\Delta_0+\Omega_1$ proves that T' is a-interpretable in T via M, then $I\Delta_0+\Omega_1$ proves that T' is e-interpretable in T via M and hence that T' is t-interpretable in T via M. So if we verify in $I\Delta_0+\Omega_1$ that M is an interpretation of T' in T we need only worry about the axioms.

We write:

$M:U \rhd V$, for the arithmetization of: V is t-interpretable in U via M.

We can arrange it so that M occurs in the arithmetization as a number, so it is possible to quantify over M in the theory. Define:

$$U \rhd V \quad :\Leftrightarrow \exists M\ M: U \rhd V$$
$$M:\phi \rhd_U \psi :\Leftrightarrow M:(U+\phi) \rhd (U+\psi)$$
$$\phi \rhd_U \psi \quad :\Leftrightarrow (U+\phi) \rhd (U+\psi)$$
$$U \equiv V \quad :\Leftrightarrow U \rhd V \wedge V \rhd U$$
$$\phi \equiv_U \psi \quad :\Leftrightarrow (U+\phi) \equiv (U+\psi)$$

Finally let me mention an important fact (which is just a variation of the similar fact stated for Herbrand consistency, see e.g. Pudlák[1985]):

5.5.1 Fact: for T R_1^+-axiomatized: $I\Delta_0+\Omega_1 \vdash (I\Delta_0+\Omega_1+\nabla_T \top) \rhd T$.

Proof: The proof will be given in detail in Marianne Kalsbeek's Masters Thesis. It involves carefully constructing the systematic tableaux τ for T on a suitable cut I and then producing a path that is provably infinite on a cut J shortening I. □

5.6 Cuts

Consider a theory T with designated natural numbers satisfying $I\Delta_0+\Omega_1$. A T-cut is a definable set I of natural numbers such that T proves that: $\underline{0} \in I$, $((x<y \wedge y \in I) \rightarrow x \in I)$, "I is closed under $S,+,.,\omega_1$". This definition of cut is a bit stronger that usual, but because any cut in the weaker sense can be shortened to a cut in our sense the difference in definition does no harm. For an introduction to Solovay's method of shortening cuts the reader is refered to Paris & Wilkie[1987]. We collect a few facts to be used later.

5.6.1 Fact: Let I be a T-cut, then $I\Delta_0+\Omega_1 \vdash \forall x \Box_T x \in I$.

This fact is due indepently to Pudlák (see Pudlák[1985]) and Paris & Wilkie (see Paris & Wilkie[1987]). It depends crucially on the use of efficient numerals and is proved by carefully constructing the proof of $x \in I$ from x and from the proof in T that I is a cut. A slightly sharpend version (due to Paris and Wilkie) is:

5.6.2 Fact: Let I be a T-cut, then for some n $I\Delta_0+\Omega_1 \vdash \forall x \Box_{T,\underline{n}} x \in I$.

Let $\exp(x):=2^x$. Define: $\mathrm{itexp}(x,0):=x$, $\mathrm{itexp}(x,Sy):=\exp(\mathrm{itexp}(x,y))$, and $\mathrm{supexp}(y):=\mathrm{itexp}(1,y)$. One can find a Δ_0-formula representing the graph of itexp, such that the recursive clauses for itexp are verifiable in $I\Delta_0+\Omega_1$. We have:

5.6.3. Fact:
$I\Delta_0+\Omega_1 \vdash \forall y(\ (\exp(y) \text{ exists}) \rightarrow$
$\exists\ I\Delta_0+\Omega_1$-cut I such that $\Box_\Omega(\forall x \in I\ \mathrm{itexp}(x,y) \text{ exists})$).

Proof: This is an immediate consequence of the proof of lemma 2.2 of Pudlák[1986].
□

5.6.4 Consequence:
$I\Delta_0+\Omega_1 \vdash \forall x,y(\ (\exp(y)\ \text{exists}) \rightarrow \Box_\Omega(\text{itexp}(x,y)\ \text{exists})\ )$.

Proof: By 5.6.2 and 5.6.3. □

5.6.5 Consequence: For T R_1^+-axiomatized: $I\Delta_0+\Omega_1 \vdash \Box_T\phi \rightarrow \Box_\Omega\Delta_T\phi$.

Proof: If a proof x is converted in a tableaux proof, the result is of order itexp(x,|x|), where |x| is the length of x in the sense of the number of symbols (as in Paris & Wilkie[1987]). So $|x| \approx \log(x)$. This estimate can be extracted from the one concerning cut-elimination on p876 of Schwichtenberg[1977], using the close connection between cut-free and tableaux proofs. We have:
$I\Delta_0+\Omega_1 \vdash \Box_T\phi \rightarrow \exists x \Box_\Omega \text{Proof}_T(x,\phi)$ and $I\Delta_0+\Omega_1 \vdash \forall x \Box_\Omega(\text{itexp}(x,|x|)\ \text{exists})$. So our result follows by induction inside $\Box_\Omega$ using itexp(x,|x|) as a bound. □

An important property of sequential theories is the presence of partial truthpredicates (see Pudlák[1986]). As a consequence of this a finitely axiomatized sequential theory T proves its own tableaux consistency on a T-cut I, i.e.:

5.6.6 Fact: $T \vdash \nabla_T^I \top$

It follows that $T \rhd (I\Delta_0+\Omega_1+\nabla_T\top)$ and hence by 5.5.1: $T \equiv (I\Delta_0+\Omega_1+\nabla_T\top)$.

At this point is is perhaps good to mention a possible source of confusion. $I\Delta_0$+EXP is infinitely axiomatized but finitely axiomatizable. In this paper we will use the results stated for finitely axiomatized sequential theories freely for $I\Delta_0$+EXP. The simplest way to justify this is simply to stipulate that by $I\Delta_0$+EXP we will understand the theory given by some fixed finite axiomatization. Another way is to check the results directly for $I\Delta_0$+EXP under its obvious axiomatization: this is possible because in $I\Delta_0$+EXP the usual truthpredicates for Σ_n-formulas are available and because of the agreeable form of the Δ_0-induction scheme. A third way is to prove in $I\Delta_0$+EXP the equivalence of tableaux provability in its finitely axiomatized form and tableaux provability in its infinitely axiomatized form. For simplicity I will opt for the first way out. Of course similar remarks hold for extensions of $I\Delta_0$+EXP with finitely many axioms and for $I\Delta_0$+SUPEXP.

5.7 Some Facts about $I\Delta_0+\Omega_1$ and $I\Delta_0$+EXP

5.7.1 **Fact:** For $\psi \in \Pi_2$: $I\Delta_0+EXP \vdash \forall x(\Delta_\Omega \psi(x) \rightarrow \psi(x))$.

Proof: This is the contraposition of Lemma 8.10 of Paris & Wilkie[1987] with a parameter added. The extra parameter doesn't require any significant changes in Paris & Wilkie's proof. □

5.7.2 **Fact:** For every $\psi(x,y) \in \Delta_0$, having only x,y free, there is an $I\Delta_0+\Omega_1$-cut I such that: $I\Delta_0+\Omega_1 \vdash \forall x \in I \exists y \psi(x,y) \Leftrightarrow I\Delta_0+EXP \vdash \forall x \exists y \psi(x,y)$.

Proof:"$\Leftarrow$" This is an entirely trivial variation of corollary 8.8 of Paris & Wilkie[1987]: the extra existential quantifier rides along for free. "$\Rightarrow$" Suppose I is an $I\Delta_0+\Omega_1$-cut and $I\Delta_0+\Omega_1 \vdash \forall x \in I \exists y \psi(x,y)$. It follows that for some m: $I\Delta_0+EXP \vdash \Box_{\Omega,\underline{m}} \forall x \in I \exists y \psi(x,y)$. On the other hand for some k: $I\Delta_0+EXP \vdash \forall x \Box_{\Omega,\underline{k}} x \in I$, so it follows that for some n: $I\Delta_0+EXP \vdash \forall x \Box_{\Omega,\underline{n}} \exists y \psi(x,y)$. By the estimate in Paris & Wilkie[1987], p293, we can prove cut-elimination for restricted provability in $I\Delta_0+EXP$, so $I\Delta_0+EXP \vdash \forall x \Delta_\Omega \exists y \psi(x,y)$. By 5.7.1 we may conclude that: $I\Delta_0+EXP \vdash \forall x \exists y \psi(x,y)$. □

5.7.3 **Consequence:** for S,S' in Σ_1:

$$I\Delta_0+EXP \vdash S \rightarrow S' \Rightarrow I\Delta_0+\Omega_1 \vdash \Box_\Omega S \rightarrow \Box_\Omega S'.$$

Proof: Suppose $I\Delta_0+EXP \vdash S \rightarrow S'$, then for some $I\Delta_0+\Omega_1$-cut I $I\Delta_0+\Omega_1 \vdash S^I \rightarrow S'$, so $I\Delta_0+\Omega_1 \vdash \Box_\Omega S^I \rightarrow \Box_\Omega S'$. On the other hand: $I\Delta_0+\Omega_1 \vdash \Box_\Omega S \rightarrow \Box_\Omega S^I$. □

5.8 Π_1-cut-conservativity

Define: $T \vdash^c \phi :\Leftrightarrow$ there is a T-cut I such that $T \vdash \phi^I$.

We say that U is Π_1-cut-conservative over V if for all Π_1-sentences P:

$$V \vdash^c P \Rightarrow U \vdash^c P.$$

We show that for sequential U: U interprets V $\Rightarrow$ U is Π_1-cut-conservative over V. The proof will be verifiable in $I\Delta_0+\Omega_1$.

Proof: (The proof is really just a proof of lemma 3.3 of Pudlák[1985]) Suppose U interprets V. We will use outline for variables ranging over the domain assigned to V

in the translation, for translated constants and predicates of V. Suppose I is a V-cut, P a Π_1-sentence and $V \vdash P^I$. Reason in U:

The idea is to try to map the numbers of U into the 'translated numbers' of V. A small complication is that translated identity need only be an equivalence relation. So the 'function' we define will be multivalued.

Define for $x \in \omega$:

$F(x,y) :\leftrightarrow$ there is a sequence σ of elements of $\mathbb{\omega}$ such that $(\sigma)_0 = \mathbb{0}$, for $u<x$ $(\sigma)_{u+1} = \mathbb{S}((\sigma)_u)$, $(\sigma)_x = y$.

Let I_0 be the set of x's such that: $\exists y \in \mathbb{\omega}\ (F(x,y) \wedge \forall z \in \mathbb{\omega}\ (F(x,z) \rightarrow y=z)$. As is easily seen I_0 contains 0 and is closed under successor. Clearly F behaves like a function w.r.t. = on I_0, so we will write $f(x)=y$ instead of $F(x,y)$ for $x \in I_0$.

Define $I_1 := \{x \in I_0 | \forall y \in I_0\ (y+x \in I_0 \wedge f(y)+f(x)=f(y+x))\}$. It is easily seen that I_1 contains 0 and is closed under sucessor and addition. $I_2 := \{x \in I_1 | \forall y \in I_1\ (y.x \in I_1 \wedge f(y).f(x)=f(y.x))\}$. Again it is easily seen that I_2 contains 0 and is closed under successor, addition and multiplication. Clearly on I_2 f commutes with 0,S, + and . .

Let $I_3 := \{x \in I_2 | \forall y \leq x\ y \in I_2 \wedge \forall y \leq f(x)\ \exists z \leq x\ f(z)=y\}$. I_3 contains 0 and is closed under succes- sor. Finally let I* be the result of shortening I_3 to a cut that is closed under S, +, . and ω_1. Let $\mathbb{I}^*$ be the image of I* under f. Both I* and $\mathbb{I}^*$ are initial segments of their respective natural numbers, which are isomorphic w.r.t. 0, S, +, and . . Note that $\mathbb{I}^*$ need not be definable in V: for V it is an "external cut". We find for Δ_0-formulas $\phi(x_1,...,x_n)$:

$\forall x_1,...,x_n \in I^*\ \phi(x_1,...,x_n) \leftrightarrow \phi(f(x_1),...,f(x_n))$,

and thus for Π_1-sentences ψ: $\psi^{I^*} \leftrightarrow \psi^{\mathbb{I}^*}$.

By assumption we had $\mathbb{P}^{\mathbb{I}}$ where $\mathbb{I}$ is a translated V-cut. Let $\mathbb{J} := \mathbb{I} \cap \mathbb{I}^*$ and let $J := f^{-1}(\mathbb{J})$. We find that J is an U-cut isomorphic to $\mathbb{J}$ and thus P^J. □

Suppose V is also sequential and suppose U is Π_1-cut-conservative over V. We show that in this case V is locally interpretable in U.

Proof: Consider a finite subtheory V_0 of V. We have for some V-cut I: $V \vdash \mathrm{Tcon}^I(V_0)$. So for some U-cut J $U \vdash \mathrm{Tcon}^J(V_0)$. Ergo U interprets V_0. □

5.8.1 Application: We show for finite sequential U and V:
$I\Delta_0+\Omega_1 \vdash U \rhd V \leftrightarrow \exists\, I\Delta_0+\Omega_1\text{-cut } I\ \Box_\Omega(\mathrm{Tcon}(U)\rightarrow \mathrm{Tcon}^I(V))$.

Proof: Reason in $I\Delta_0+\Omega_1$:

First suppose $U \rhd V$. Clearly $(I\Delta_0+\Omega_1+\mathrm{Tcon}(U)) \rhd U$ and hence $(I\Delta_0+\Omega_1+\mathrm{Tcon}(U)) \rhd V$. There is a V-cut J such that $\Box_V \mathrm{Tcon}^J(V)$, so by Π_1-cut-conservativity there is an $I\Delta_0+\Omega_1+$ Tcon(U)-cut J* such that $\Box_\Omega(\mathrm{Tcon}(U)\rightarrow \mathrm{Tcon}^{J^*}(V))$. Define: $x\in I :\leftrightarrow (x\in J^* \vee \neg\, \mathrm{Tcon}(U))$. As is easily seen I is an $I\Delta_0+\Omega_1$-cut and $\Box_\Omega(\mathrm{Tcon}(U)\rightarrow \mathrm{Tcon}^I(V))$.

Suppose $\exists\, I\Delta_0+\Omega_1$-cut $I\ \Box_\Omega(\mathrm{Tcon}(U)\rightarrow \mathrm{Tcon}^I(V))$. We have:
$U \rhd (I\Delta_0+\Omega_1+\mathrm{Tcon}(U)) \rhd (I\Delta_0+\Omega_1+\mathrm{Tcon}(V)) \rhd V$. □

5.8.2 Consequence: let T be a finitely axiomatized and sequential. Let $U:=T+\phi$, $V=T+\bot$, we find: $I\Delta_0+\Omega_1 \vdash \Box_T\phi \leftrightarrow \Box_\Omega \Delta_T\phi$.

5.8.3 Application: Let U be finitely axiomatized and sequential and let P be a Π_1-sentence. We have: $U \rhd P \Leftrightarrow I\Delta_0+\mathrm{EXP}+\mathrm{Tcon}(U) \vdash P$.

Proof: As is easy to see: $U \rhd P \Leftrightarrow$ for some $I\Delta_0+\Omega_1$-cut I: $I\Delta_0+\Omega_1+\mathrm{Tcon}(U) \vdash P^I$. Apply 5.7.2. □

6 Principles

The language of IL is the language of modal propositional logic with one extra binary operator $\rhd$. An interpretation of this language in a theory T with a designated set of natural numbers satisfying $I\Delta_0+\Omega_1$ is a function (.)* that maps the atoms of the modal language on arbitrary sentences of the language of T, commutes with the propositional connectives (including $\bot$) and satisfies: $(\Box\phi)^*=\Box_T\phi^*$ and $(\phi \rhd \psi)^*=\phi^* \rhd_T \psi^*$. Here $\Box_T$ and $\rhd_T$ are the arithmetizations in the designated set of natural numbers of respectively provability in T and interpretability over T.

6.1 IL, the basic logic

The theory IL is useful as a basic theory from the modal standpoint. From the point of view of arithmetical interpretations it is too weak: as we will see the principle W,

which is not derivable in IL, is valid for interpretations in all reasonable theories. The theory IL is given as Propositional Logic plus:

L1 $\vdash\phi \Rightarrow \vdash\Box\phi$
L2 $\vdash \Box(\phi\rightarrow\psi) \rightarrow (\Box\phi\rightarrow\Box\psi)$
L3 $\vdash \Box\phi \rightarrow \Box\Box\phi$
L4 $\vdash \Box(\Box\phi\rightarrow\phi)\rightarrow\Box\phi$
J1 $\vdash \Box(\phi\rightarrow\psi) \rightarrow \phi \triangleright \psi$
J2 $\vdash (\phi \triangleright \psi\wedge\psi \triangleright \chi) \rightarrow \phi \triangleright \chi$
J3 $\vdash (\phi \triangleright \chi\wedge\psi \triangleright \chi) \rightarrow (\phi\vee\psi) \triangleright \chi$
J4 $\vdash \phi \triangleright \psi \rightarrow (\Diamond\phi\rightarrow\Diamond\psi)$
J5 $\vdash \Diamond\phi \triangleright \phi$

Note that the principle L3 is doubly superfluous: it follows both from L1, L2, L4 (by a well known argument) and from L1, L2, J4, J5 (by a trivial argument).

6.1.1 Reasoning in IL

It is pleasant to get some feeling for reasoning in IL. This section aims to provide some examples.

K1 $\vdash \phi \equiv (\phi\vee\Diamond\phi)$

Proof: immediate by J1, J5, J3. □

Let $F\phi := (\phi\vee\Diamond\phi)$, $G\phi := (\phi\wedge\Box\neg\phi)$, then:
K2 $\vdash F\phi \leftrightarrow FF\phi$
$\vdash F\phi \leftrightarrow FG\phi$
$\vdash G\phi \leftrightarrow GG\phi$
$\vdash G\phi \leftrightarrow GF\phi$

Immediate consequences are:
K3 $\vdash \phi \triangleright (\phi\wedge\Box\neg\phi)$
K4 $\vdash \phi \equiv (\phi\wedge\Box\neg\phi)$

Note that: K3 is an alternative for axiom J5.

K5 $\vdash \phi \triangleright \bot \rightarrow \Box\neg\phi$

Proof: by J4. □

Feferman's Principle is the following:
F $\vdash \Diamond\phi \rightarrow \neg(\phi \triangleright \Diamond\phi)$

A Kripke Model argument shows that F is *not* derivable in IL. However the following weakening is derivable:
K6 $\vdash \Diamond\phi \triangleright \neg(\phi \triangleright \Diamond\phi)$

Proof: It is sufficient to show: IL$\vdash (\Diamond\phi \wedge \Box\neg\Diamond\phi) \rightarrow \neg(\phi \triangleright \Diamond\phi)$. We have:

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

In IL one can already derive the existence of unique and explicit fixed points for modalized formulas. For a (model-theoretical) proof the reader is referred to de Jongh & Veltman[?], this volume.

6.2 The logic ILW

ILW is IL plus the principle W:
W $\vdash \phi \triangleright \psi \rightarrow \phi \triangleright (\psi \wedge \Box\neg\phi)$

It may amuse the reader to show that ILW can be more efficiently axiomatized using only L1, L2, J1, J2, J3, J4, W.

W characterizes the set of IL-frames such that $R \circ S_x$ is upwards wellfounded for all x in their do- main (see de Jongh & Veltman[?], this volume). I conjecture that ILW is complete for this set of structures. One can show that completeness w.r.t this set of frames implies completeness w.r.t a more restricted class of frames, namely those in which there are no infinite R,S-chains, where the index of S may vary. ILW is valid for interpretations in theories T with designated natural numbers satisfying $I\Delta_0+\Omega_1$, whose axiom sets can be represented by a R_1^+-formula. I conjecture that:

ILW$\vdash \phi \Leftrightarrow$ for all T with designated natural numbers
satisfying $I\Delta_0+\Omega_1$, with R_1^+ axiom sets,
for all interpretations (.)* into T: T$\vdash \phi^*$

6.2.1 Consequences of W

A first consequence of W is Feferman's Principle F:

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

This is immediate by substituting $\Diamond\phi$ for ψ. A second consequence is the Contraposition Principle:

KW1 $\vdash \phi \rhd \Diamond\top \rightarrow \top \rhd \neg\phi$

Proof: $\vdash \phi \rhd \Diamond\top \rightarrow \phi \rhd (\Diamond\top \wedge \Box\neg\phi)$

$\rightarrow \phi \rhd \Diamond\neg\phi$

$\rightarrow \phi \rhd \neg\phi$

Ergo by J1 and J3 we have KW1. □

Both F and KW1 characterize the same class of IL structures as W. However I do not know whether W is derivable either in ILF or in IL(KW1).

Given the arithmetical validity of ILW we have the following consequence: Paris & Wilkie show that $EXP \rhd_\Omega \Diamond_\Omega \top$, ergo by KW1: $\top \rhd_\Omega \neg EXP$, i.o.w.: $(I\Delta_0+\Omega_1) \rhd (I\Delta_0+\Omega_1+\neg EXP)$.

Autobiographical note: this proof of the interpretability of $I\Delta_0+\Omega_1+\neg EXP$ in $I\Delta_0+\Omega_1$ could have been a nice example of how the logic allows one to discover new interpretations. Alas, things did not go like that. First I sketched a proof in $I\Delta_0+EXP$ of the tableaux consistency of $I\Delta_0+\Omega_1+\neg EXP$ adapting a method from Paris & Wilkie[1987]. This gives us $(I\Delta_0+\Omega_1) \rhd (I\Delta_0+\Omega_1+\neg EXP)$. Then I constructed an interpretation of $I\Delta_0+\Omega_1+\neg EXP$ in $I\Delta_0+\Omega_1+con(I\Delta_0+\Omega_1)$ using the Henkin construction described in 7.2.2.1. This again gives us $(I\Delta_0+\Omega_1) \rhd (I\Delta_0+\Omega_1+ \neg EXP)$. Then I started to wonder about the connection of this fact and $(I\Delta_0+EXP) \rhd (I\Delta_0+\Omega_1+con\ (I\Delta_0+\Omega_1))$. This led me to prove the arithmetical validity of KW1 directly. Then I showed that KW1 is valid in all IL-structures with RoS_x upwards wellfounded. And *finally* I gave the simple modal proof of KW1.

6.2.2 Arithmetical validity of ILW

We verify that ILW is valid for arithmetical interpretations in theories T with designated natural numbers satisfying $I\Delta_0+\Omega_1$, whose axiom sets can be represented by a R_1^+-formula.

The axioms L1-4 are verified in Paris & Wilkie[1987]. J1, J2, J3 and J4 are trivial, given the fact that we opted for t-interpretability. We turn to J5: the Interpretation Existence Lemma. Before we proceed let me answer an obvious question: J5 follows from the stronger principle $\nabla\phi \rhd \phi$, which is assumed in this paper, so why bother to prove it? The answer is (i) to fix a number of concepts that we will use later on in the paper, and (ii) because the assumptions on provability in the proof are so weak that the argument also works for alternative notions like Feferman provability. The present construction is essentially Henkin's, refined by Feferman (see Feferman[1960]), with some twists due to Pudlák and Friedman.

6.2.2.1 The Henkin Construction

Let U be any theory with designated natural numbers satisfying $I\Delta_0+\Omega_1+\mathrm{conV}$, where V is a theory whose language L is given by a $\Delta_0(\omega_1)$-formula. We assume that $V\vdash\phi \Rightarrow U\vdash \Box_V\phi$. Define an extension of L, L^+ as follows: L^+ is the smallest extension such that if ϕ is in L^+ then there are constants $c[\exists x\phi]$ and $c[\forall x\phi]$ in L^+. L^+ is again $\Delta_0(\Omega_1)$. We choose an efficient coding of 0,1-sequences, where 0 is the empty sequence. Sequences are written: $\mathbb{0}\mathbb{1}\mathbb{1}\mathbb{0}$, etc. . $|x|$ is the length of the sequence coded by x. $\prec$ is the 'initial sequence' ordering. 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 $\exists z\phi(z)$ such that $(x)_w=0$ and u codes $\phi(c[\exists z\phi])$) or (there is a w of the form $\forall z\phi(z)$ such that $(x)_w=1$ and u codes $\neg\phi(c[\forall z\phi])$).

Note that $u\in T[x]$ is $\Delta_0(\omega_1)$. Moreover: $U\vdash x\prec y \rightarrow T[x]\subseteq T[y]$ and $U\vdash T[0]=\emptyset$.

Define further: $x\in \mathrm{TREE} :\Leftrightarrow \mathrm{Con}(V+T[x])$. Clearly $U\vdash (x\prec y \wedge y\in\mathrm{TREE}) \rightarrow x\in\mathrm{TREE}$. Moreover: $U\vdash 0\in\mathrm{TREE}$. We show that $U\vdash x\in\mathrm{TREE} \rightarrow (x\mathbb{0}\in\mathrm{TREE} \vee x\mathbb{1}\in\mathrm{TREE})$. Reason in U:

> Suppose $x\in$TREE, i.e. Con(V+T[x]). Let $u:=|x|+1$. In case u does not code an L^+-sentence we have: $T[x\mathbb{0}]=T[x\mathbb{1}]=T[x]$, so we are done. We treat the case that u codes a sentence of the form $\exists z\phi(z)$, the other cases are analogous or easier. So

suppose u codes $\exists z\phi(z)$. Then $T[x\mathbb{0}] = T[x]+\{ \exists z\phi(z), \phi(c[\exists z\phi(z)]) \}$ (note that the existence of $\phi(c[\exists z\phi(z)])$ requires Ω_1) and $T[x\mathbb{1}] = T[x]+\{\neg \exists z\phi(z)\}$. The constant $c[\exists z\phi(z)]$ does not occur in $V+T[x]$ (because we used the natural Gödelnumbering), hence it is easy to convert a proof of falsity in $V+T[x]+\{\exists z\phi(z), \phi(c[\exists z\phi(z)])\}$ in a proof of falsity in $V+T[x]+\{\exists z\phi(z)\}$. Thus if both $V+T[x\mathbb{0}]$ and $V+T[x\mathbb{1}]$ 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 $I\Delta_0+\Omega_1$.) □

Define PATH:={x∈TREE|"there is no y in TREE to the left of x"}. As is easily seen: $U\vdash x\in PATH \rightarrow (x\mathbb{0}\in PATH \vee x\mathbb{1}\in PATH)$ and $U\vdash 0\in PATH$. Also $U\vdash (x\in PATH \wedge y\in PATH) \rightarrow (x\prec y \vee y\prec x \vee x=y)$.

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∈I and x is a code of an L-sentence.

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

$x\in F^1(y)$:⇔ x,y∈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.

$x\in D$:⇔ $x\in L^1$ and x codes a sentence of the form $\exists u\phi(u)$ or $\forall u\phi(u)$.

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

$R^K(x)$:⇔ x is in D, x codes ψ and $K(\ulcorner R(c[\psi])\urcorner)$.

We have:

(i) $U\vdash \forall x\in L^0\ Prov_V(x) \rightarrow K(x)$.

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 ψ. V+T[y] is consistent, and either ψ or ¬ψ is in T[y]. Clearly ¬ψ cannot be in T[y], so ψ is. □

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

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

□

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

(c) $U\vdash \forall x\in L^1\ K(NEG(x))\leftrightarrow\neg K(x)$
(d) $U\vdash \forall x,y\in L^1\ K(CONJ(x,y))\leftrightarrow(K(x)\wedge K(y))$
(e) $U\vdash \forall y\in I\ VAR(y)\rightarrow \forall x\in F^1(y)\ (\ K(UQ(y,x))\leftrightarrow \forall z\in D\ K(SUB(z,y,x))\ )$

(Here if z codes ψ, x codes $\phi(u)$ and y codes u: $SUB(z,y,x)=\ulcorner\phi(c[\psi])\urcorner$. Note that by Ω_1 both UQ(y,x) and SUB(z,y,x) 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. $\neg K(x)$ we have K(NEG(x)) and reasoning as before we find K(NEG(CONJ(x,y))), so $\neg K(CONJ(x,y))$.
e) Consider $y\in I$ with VAR(y) and $x\in F^1(y)$. First suppose K(UQ(y,x)). Clearly UQ(y,x) is in L^1. Consider z in D. As is easily seen SUB(z,y,x) is in L^1. Let v be the maximum of UQ(y,x) and SUB(z,y,x). There is a w in PATH with $|w|=v$. We have UQ(y,x) in T[w] and either SUB(z,y,x) or NEG(SUB(z,y,x)). By the consistency of T[w] we must have SUB(z,y,x) in T[w] and hence K(SUB(z,y,x)). Suppose for the converse that $\neg K(UQ(y,x))$. 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 $NEG(SUB(v,y,x))\in T[v]$. Clearly v is in D and we have $\neg K(SUB(v,y,x))$. □

We write ϕ^K for the interpretation in the language of U of sentences ϕ of L using D and the translation of the relation symbols described above. As is easily seen: $U\vdash \phi^K\leftrightarrow K(\phi)$. So we have by (i): $U\vdash \Box_V\phi\rightarrow\phi^K$. Conclude: $V\vdash\phi\Rightarrow U\vdash\phi^K$.

NOTE:

I) Clearly the above reasoning can be verified in any theory T extending $I\Delta_0+\Omega_1$ such that $T\vdash \Box_V\phi\rightarrow\Box_U\Box_V\phi$.

II) We didn't make any assumption on the complexity of the formula defining the axiom set of V. So we can use the result for Feferman style predicates.

III) If provability in V is representable by a Σ_1-predicate then by a result of Wilkie $I\Delta_0+\Omega_1+$ con(V) is interpretable on a cut in Q+con(V). So in this case we can reduce our assumption that $I\Delta_0+\Omega_1$+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 (simply take as the natural numbers of U the elements of this cut).

6.2.2.2 The principle W

Let U and V be theories axiomatized by R_1^+-formulas extending $I\Delta_0+\Omega_1$. Suppose V is interpretable in U. We show that $V+\Box_U\bot$ is interpretable in U. This result is called the principle W for 'Weak' because it is the strongest principle that we know to hold for all 'reasonable' arithmetics. The argument below is designed to be verifiable in $I\Delta_0+\Omega_1$.

The argument uses a trick that is due to Feferman. Let M be the interpretation of V in U. M is given by a finite amount of information and the associated translation of formulas is R_1^+-definable in $I\Delta_0+\Omega_1$. Define: $Prov_{V*}(x) :\Leftrightarrow Prov_V(x)\wedge Prov_U(x^M)$. Trivially V* is extensionally equal to V. So Id: $(V*+\Box_U\bot)\rhd(V+\Box_U\bot)$. Also: $\Box_{V*}(\Box_{V*}\bot\rightarrow\Box_U\bot)$. Clearly the principles of IL can be verified for $\Box_{V*}$ and $\rhd_{V*}$ (using the fact that $Prov_{V*}(x)$ can be written as an R_1^+-formula preceded by existential quantifiers). By K3: $V*\rhd(V*+\Box_{V*}\bot)$ and hence $V*\rhd(V*+\Box_U\bot)$. Con- clude: $U\rhd V\rhd V*\rhd(V*+\Box_U\bot)\rhd(V+\Box_U\bot)$. □

6.3 The Logic ILP

ILP is IL+P, where P is the Persistence Principle:

P $\vdash \phi\rhd\psi\rightarrow\Box(\phi\rhd\psi)$

ILP is arithmetically valid for interpretations in finitely axiomatized theories with designated natural numbers satisfying $I\Delta_0+\Omega_1$. The verification of the arithmetical validity of P is trivial. We will show that ILP is complete for interpretations in finitely axiomatized sequential theories with designated natural numbers satisfying

$I\Delta_0$+SUPEXP that do not prove their iterated inconsistency for any finite number of iterations. These include ACA_0 and GB.

ILP is also arithmetically sound and complete for a different interpretation, namely when we interpret $\Box$ as provability in PA and $\phi \rhd \psi$ as: for some primitive recursive term tx with just x free $PA \vdash \forall x\, Proof_{PA+\psi}(x, \ulcorner \bot \urcorner) \rightarrow Proof_{PA+\phi}(tx, \ulcorner \bot \urcorner)$. This strong notion of relative consistency is studied in Christian Bennet's Thesis (Bennet[1986a]). More on the alternative interpretation in section 8.3.

P characterizes IL structures with the following property: $yRzS_xu \Rightarrow yRu$. De Jongh & Veltman show the completeness of ILP w.r.t. (finite) IL structures satisfying this property (see de Jongh & Veltman[?], this volume).

We show that ILP extends ILW:

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

The desired result is immediate. $\Box$

6.4 The logic ILM

ILM is IL plus Montagna's Principle M:

M $\quad \vdash \phi \rhd \psi \rightarrow (\phi \wedge \Box\chi) \rhd (\psi \wedge \Box\chi)$

Fact: Montagna's Principle is arithmetically valid in verifiably essentially reflexive Δ_1-axiomatized theories with designated natural numbers satisfying $I\Delta_0+\Omega_1$.

Before we prove this fact first a few useful observations:

Observation 1: Suppose U has designated natural numbers satisfying $I\Delta_0+\Omega_1$. Let Q* be (Q+the axioms for linear ordering for the usual ordering on the natural numbers) extended to the language of U. Suppose U proves the Uniform Reflection Principle for Q*. Then U proves full Induction.

Proof of observation 1: Consider any formula $\phi(x)$ of the language of U. Let $X:=\{x|((\phi(\underline{0})\wedge \forall y(\phi(y)\rightarrow\phi(Sy))) \rightarrow \phi(x))\}$. We shorten X to a Q*-cut I and find: $U\vdash \forall x\Box_{Q^*}x\in I$. Ergo by URP for Q*: $U\vdash \forall x\ \ x\in I$. □

Observation 2: Let U be sequential, satisfying full Induction. Then U is essentially reflexive.

Proof of observation 2: This is just by the usual argument for the essential reflexiveness of PA, using the existence of partial truth-predicates in sequential theories. □

Proof of claim: Let T be an essentially reflexive Δ_1-axiomatized theory with designated natural numbers satisfying $I\Delta_0+\Omega_1$. We prove the slightly stronger principle: for S a Σ_1-sentence:

$\wedge\Sigma$ $\quad T\vdash \phi\rhd\psi\rightarrow(\phi\wedge S)\rhd(\psi\wedge S)$

By observation 1 T satsfies full induction. So the Orey-Hájek theorem is verifiable: $T\vdash \chi\rhd_T\rho\leftrightarrow\forall x\Box_T(\chi\rightarrow\Diamond_T\ulcorner x\rho)$. Reason in T:

Let S be a Σ_1-sentence. Suppose $\chi\rhd_T\rho$ so $\forall x\Box_T(\chi\rightarrow\Diamond_T\ulcorner x\rho)$. Let *q* be so big that for all $x>q$ $\Box_T(S\rightarrow\Box_T\ulcorner x(S\leftrightarrow\top))$. It follows that: $\Box_T(S\rightarrow(\Diamond_T\ulcorner x\rho\leftrightarrow\Diamond_T\ulcorner x(\rho\wedge S)))$, ergo: $\forall x\Box_T((\chi\wedge S)\rightarrow\Diamond_T\ulcorner x(\rho\wedge S))$ and thus $(\chi\wedge S)\rhd_T(\rho\wedge S)$. □

If T is sequential the following proof can be used: Reason in T:

Let S be a Σ_1-sentence. Suppose $M:\chi\rhd\rho$. The natural numbers of $T+\chi$ are on a $T+\chi$-cut isomorphic to the natural numbers of the interpretation on a suitable 'external' cut. $T+\chi$ satisfies full induction, so this means that the natural numbers of $T+\chi$ are isomorphic to the natural numbers of the interpretation on a suitable 'external' cut, say I*. We have $\Box_T(\chi\rightarrow(S\rightarrow(S^{I^*})^M))$, hence by upwards persistence of Σ_1-sentences: $\Box_T(\chi\rightarrow(S\rightarrow S^M))$. □

Question: Is some strengthened version of $\wedge\Sigma$ equivalent to essential reflexiveness?

I conjecture that ILM is complete for arithmetical interpretations in verifiably essentially reflexive Δ_1-axiomatized theories with designated natural numbers satisfying $I\Delta_0+\Omega_1$.

M characterizes IL-frames with the following property: $yS_xzRu \Rightarrow yRu$. De Jongh & Veltman show that ILM is complete w.r.t. (finite) IL-models satisfying this property.

6.4.1 Consequences of M

We leave the simple verification that W is derivable in ILM to the reader. Two important consequences of M are:

KM1 $\vdash \phi \rhd \Diamond \psi \rightarrow \Box(\phi \rightarrow \Diamond \psi)$

KM2 $\vdash \phi \rhd \psi \rightarrow (\Box(\psi \rightarrow \Diamond \chi) \rightarrow \Box(\phi \rightarrow \Diamond \chi))$

Clearly these principles show us what is 'visible' of the Π_1-conservativety of essentially reflexive theories over theories interpreted in them. First we prove KM1:

Proof: $\vdash \phi \rhd \Diamond \psi \rightarrow (\phi \wedge \Box \neg \psi) \rhd (\Diamond \psi \wedge \Box \neg \psi)$

$\rightarrow (\phi \wedge \Box \neg \psi) \rhd \bot$

$\rightarrow \Box \neg (\phi \wedge \Box \neg \psi)$

$\rightarrow \Box(\phi \rightarrow \Diamond \psi)$ $\Box$

Next we show derive KM2 from KM1:

Proof: $\vdash \phi \rhd \psi \rightarrow (\Box(\psi \rightarrow \Diamond \chi) \rightarrow (\psi \rhd \Diamond \chi)$

$\rightarrow (\phi \rhd \Diamond \chi)$

$\rightarrow \Box(\phi \rightarrow \Diamond \chi))$ $\Box$

Both KM1 and KM2 characterize IL-frames satisfying $yS_xzRu \Rightarrow yRu$. But it is unknown whether any of them implies M over IL.

7 $I\Delta_0+\Omega_1$, $I\Delta_0$+EXP & Friedman's Characterization

7.1 Tableaux provability in $I\Delta_0$+EXP

Consider any R_1^+-axiomatized theory T. Transforming an ordinary T-proof into a T-tableaux-proof is a superexponential process. To be precise it is of order: itexp(|x|,ρ(x)), where x is our original proof and where ρ(x) is the cut rank of x, i.e. the supremum of the lengths of the cut formulas in x. So in general $I\Delta_0$+EXP will not prove: $\Box_T\phi \rightarrow \Delta_T\phi$. On the other hand using the above estimate as a bound one can show for *sentences* ϕ and ψ:

$I\Delta_0$+EXP $\vdash \Delta_T(\phi \rightarrow \psi) \rightarrow (\Delta_T\phi \rightarrow \Delta_T\psi)$.

The point is of course that the cut-rank involved is standard and thus the rate of growth is just multi-exponential. It would be very pleasant if we had this fact also for formulas

(under our convention for free varables within the scope of Δ). It seems to me that the more general fact should hold, but I do not really know it. Another familiar principle is:

$$I\Delta_0+EXP \vdash \Delta_T\phi \rightarrow \Delta_T\Delta_T\phi .$$

In fact I conjecture that this principle is already verifiable in $I\Delta_0+\Omega_1$. (To prove this one would have to inspect how much cuts are involved in Paris & Wilkie's procedure to produce a proof of a R_1^+-formula ϕ given ϕ.)

The above observations imply that we have the usual provability logic for Δ_T with T extending $I\Delta_0+EXP$. One can verify Solovay's completeness proof in $I\Delta_0+EXP$, so it follows that Löb's Logic L is precisely the logic of such Δ_T. (The fact that we talk about tableaux proofs does not matter at all.)

So surprisingly $\Box_{EXP}$ and Δ_{EXP} satisfy the same provability principles without being provably the same over $I\Delta_0+EXP$. The next section's result will imply that one extra principle characterizes the logic of $\Box_{EXP}$ and Δ_{EXP} together.

9.2 Formalizing a result of Paris & Wilkie

We want to formalize 5.7.2: i.e.: for every $\psi(x,y)\in\Delta_0$ with only x,y free, there is an $I\Delta_0+\Omega_1$-cut I such that: $I\Delta_0+\Omega_1\vdash \forall x\in I\exists y\psi(x,y) \Leftrightarrow I\Delta_0+EXP\vdash \forall x\exists y\psi(x,y)$. An obvious first guess at the correct formulation of the formalization is e.g.: for every $\psi(x,y)\in\Delta_0$ with only x,y free:

$$I\Delta_0+EXP\vdash \Box_{EXP}\forall x\exists y\ \psi(x,y) \leftrightarrow \exists\ I\Delta_0+\Omega_1\text{-cut } I\ \Box_{\Omega}\forall x\in I\ \exists y\ \psi(x,y).$$

But this cannot be right. Taking $\psi:=\bot$ we would get: $I\Delta_0+EXP\vdash \neg\Box_{\Omega}\bot \rightarrow \neg\Box_{EXP}\bot$, con- tradicting theorem 8.19 of Paris & Wilkie[1987].

(A somewhat simplified proof of this theorem is as follows: suppose $I\Delta_0+EXP\vdash \neg\Box_{\Omega}\bot \rightarrow \neg\Box_{EXP}\bot$, then by 5.7.2 for some $I\Delta_0+\Omega_1$-cut I: $I\Delta_0+\Omega_1\vdash \neg\Box_{\Omega}\bot \rightarrow \neg\Box^{I}_{EXP}\bot$. Let J be an $I\Delta_0+EXP$-cut such that $I\Delta_0+EXP\vdash \forall x\in J$ supexp(x) exists. $I\Delta_0+EXP\vdash \neg\Delta_{\Omega}\bot$ by 5.7.1 so by cut-elimination: $I\Delta_0+EXP\vdash \neg\Box^{J}_{\Omega}\bot$. Because $I\Delta_0+\Omega_1$ is verifiable on J we find by composing cuts that for some $I\Delta_0+EXP$-cut J*: $I\Delta_0+EXP\vdash \neg\Box^{J^*}_{EXP}\bot$. This contradicts Pudlák's sharpening of the second incompleteness theorem (see Pudlák[1985]) (or alternatively: it contradicts Feferman's Principle F (see section 6.2.1)).)

The correct form for the formalization turns out to be this: for every $\psi(x,y)\in\Delta_0$ with only x,y free:

$I\Delta_0+EXP \vdash \Delta_{EXP}\forall x\exists y\ \psi(x,y) \leftrightarrow \exists\ I\Delta_0+\Omega_1\text{-cut } I \Box\forall x\in I\ \exists y\ \psi(x,y)$.

Proof: For the "$\rightarrow$"-direction I briefly sketch how this can be shown by transforming proofs and then give a more elaborate simulation of the model-theoretical argument of Paris & Wilkie. Reason in $I\Delta_0+EXP$:

Let z be an EXP-tableaux-proof of $\ulcorner\forall x\exists y\ \psi(x,y)\urcorner$. The tableaux will move once from $\ulcorner\neg\forall x\exists y\ \psi(x,y)\urcorner$ to $\ulcorner\neg\exists y\ \psi(c,y)\urcorner$ to obtain a contradiction from this last formula plus the axioms of $I\Delta_0+EXP$. The only principles used in the rest of the proof that are not Π_1 or negated Σ_1 are the axioms for S, +, . and EXP. So the only "growing constants" introduced are due to these axioms and their maximal rate of growth is due to EXP. Our tableaux system is assumed to be relational, so in every step the growth is only caused by one application of EXP. So the biggest constant in the proof will be something like: exp(exp(...c)...), where the exp is iterated |z| times. Using an estimate of Pudlák we can find for any u an $I\Delta_0+\Omega_1$-cut $I_u \leq \exp(u)$ such that $\Box(\forall v\in I_u\ \exp(v)\in I_{u-1})$. Choose $I:=I_{|z|}$.

Now we transform z into an $I\Delta_0+\Omega_1$-proof z* of $\ulcorner\forall x\in I\ \exists y\ \psi(x,y)\urcorner$ as follows. We may start from the axioms of $I\Delta_0+\Omega_1$ plus $\ulcorner c\in I\urcorner$ and $\ulcorner\neg\exists y\ \psi(c,y)\urcorner$. We follow z, but add on proofs for any constant e introduced that $e\leq\exp(\exp(...c)...)$ say for $u\leq|z|$ iterations of exp, plus proofs that: $\exp(\exp(...c)...)\in I_{|z|-u}$. Application of EXP to e can then be replaced by a use of the fact that e is in $I_{|z|-u}$.

We turn to the alternative argument: by contraposition it is sufficient to show that for χ in Δ_0 with only x,y free:

$I\Delta_0+EXP \vdash\ \forall\ I\Delta_0+\Omega_1\text{-cuts } I \Diamond_\Omega\exists x\in I\ \forall y\ \chi(x,y) \rightarrow \nabla_{EXP}\exists x\forall y\ \chi(x,y)$,

hence by 5.7.1 it is sufficient to show that: for some $I\Delta_0+\Omega_1$-cut J:

$I\Delta_0+\Omega_1 \vdash\ \forall\ I\Delta_0+\Omega_1\text{-cuts } I \Diamond_\Omega\exists x\in I\ \forall y\ \chi(x,y) \rightarrow \nabla^J_{EXP}\exists x\forall y\ \chi(x,y)$.

The above in its turn is immediate by Π_1-cut-conservativity from:

$(\forall\ I\Delta_0+\Omega_1\text{-cuts } I \Diamond_\Omega\exists x\in I\ \forall y\ \chi(x,y)) \rhd_\Omega \nabla_{EXP}\exists x\forall y\ \chi(x,y)$,

Because $\nabla_{EXP}\exists x\forall y\ \chi(x,y) \equiv (EXP\wedge\exists x\forall y\ \chi(x,y))$ (see section 5.6) this last statement follows from:

$(\forall\ I\Delta_0+\Omega_1\text{-cuts } I \Diamond_\Omega\exists x\in I\ \forall y\ \chi(x,y)) \rhd_\Omega (EXP\wedge\exists x\forall y\ \chi(x,y))$.

Reason in $I\Delta_0+\Omega_1$:

Suppose that for every $I\Delta_0+\Omega_1$-cut I: $\Diamond_\Omega\exists x\in I\ \forall y\ \chi(x,y)$. By 5.6.4 we can find a (standard) $I\Delta_0+\Omega_1$-cut J such that $\forall u\in J\ \exists\ I\Delta_0+\Omega_1\text{-cut } I\ \Box_\Omega(\forall v\in I\ \text{itexp}(v,u)$ exists). It follows that: $\forall u\in J\ \Diamond_\Omega\exists x$ (itexp(x,u) exists $\wedge\ \forall y\ \chi(x,y)$). Let c be a new constant and let $V:=I\Delta_0+\Omega_1+\forall y\ \chi(c,y)+\{\text{itexp}(c,u)\text{ exists} \mid u\in J\}$. As is

easily seen V is consistent. Let I, K, D, f be as in 6.2.2.1 and let $c^*:=\ulcorner\exists x\ x=c\urcorner$. Define:

$$x\in D^* :\Leftrightarrow x\in D \wedge \exists y\in I\ x\leq^K \mathrm{itexp}^K(c^*,f(y)).$$

Let K* be the interpretation based on I, K and D*. As is easily seen D* is closed under $\exp^K$ and thus under $\exp^{K^*}$. Moreover $(\forall y\ \chi(c,y))^{K^*}$, and thus $(\exists x\forall y\ \chi(x,y))^{K^*}$. The (standard) instances ρ of Δ_0-induction have Π_1 form, moreover we have ρ^K, so: ρ^{K^*}.

"$\leftarrow$" Let $\Im$ be an $I\Delta_0$+EXP-cut such that $I\Delta_0$+EXP$\vdash \forall u\in J\forall v$ itexp(v,u) exists. We first show: $I\Delta_0$+EXP$\vdash \forall I\in\Im(\ \Box_\Omega^\Im$"I is a cut" $\rightarrow (\ \exists z\in\Im\ \Box_{\Omega,z}\forall x\in I\ \exists y\ \psi(x,y) \rightarrow \forall x\ \exists y\ \psi(x,y)\ )\ )$. Reason in $I\Delta_0$+EXP:

Suppose $I\in\Im$, $\Box_\Omega^\Im$"I is a cut", $z\in\Im$ and $\Box_{\Omega,z}\forall x\in I\ \exists y\ \psi(x,y)$. Inspecting the argument for 5.6.2 we find that for some $u\in\Im$ and for all v $\Box_{\Omega,u}v\in I$. It follows that for some $w\in\Im$: $\forall x\Box_{\Omega,w}\exists y\ \psi(x,y)$. Using the estimate on cut-elimination in Paris and Wilkie[1978], p293 we may conclude: $\forall x\Delta_\Omega\exists y\ \psi(x,y)$. By 5.7.1: $\forall x\exists y\ \psi(x,y)$.

To prove our theorem reason again in $I\Delta_0$+EXP:

Suppose that for some $I\Delta_0+\Omega_1$-cut I $\Box_\Omega\forall x\in I\ \exists y\ \psi(x,y)$. By the sharp version of R_1^+- completeness we find that for some standard $\underline{m}$ and for some u and z: $\Box_{EXP,\underline{m}}\mathrm{Proof}_\Omega$(u,"I is a cut") and $\Box_{EXP,\underline{m}}\Box_{\Omega,z}\forall x\in I\ \exists y\ \psi(x,y)$. By 5.7.1 for some standard $\underline{k}$: $\Box_{EXP,\underline{k}}I\in\Im$, $\Box_{EXP,\underline{k}}u\in\Im$, $\Box_{EXP,\underline{k}}z\in\Im$. By our auxiliary result for some standard $\underline{n}$:

$\Box_{EXP,\underline{n}}\forall I\in\Im(\ \Box_\Omega^\Im$"I is a cut" $\rightarrow (\ \exists z\in\Im\ \Box_{\Omega,z}\forall x\in I\ \exists y\ \psi(x,y) \rightarrow \forall x\ \exists y\ \psi(x,y)\ )\ )$.

Conclude that for some standard $\underline{p}$: $\Box_{EXP,\underline{p}}\forall x\ \exists y\ \psi(x,y)$. By applying cut elimination we find: $\Delta_{EXP}\forall x\ \exists y\ \psi(x,y)$. $\Box$

A variant of our theorem can be easily obtained as follows: by 5.7.3 we find: for every $\psi(x,y)\in\Delta_0$ with only x,y free:

$$I\Delta_0+\Omega_1\vdash \Box_\Omega\Delta_{EXP}\forall x\exists y\ \psi(x,y) \leftrightarrow \Box_\Omega\exists\ I\Delta_0+\Omega_1\text{-cut } I\Box_\Omega\forall x\in I\ \exists y\ \psi(x,y).$$

Combining this with 5.8.2 we get:

$$I\Delta_0+\Omega_1\vdash \Box_{EXP}\forall x\exists y\ \psi(x,y) \leftrightarrow \Box_\Omega\exists\ I\Delta_0+\Omega_1\text{-cut } I\Box_\Omega\forall x\in I\ \exists y\ \psi(x,y).$$

7.3 Some Consequences

1 $(\forall\ I\Delta_0+\Omega_1\text{-cut } I\Diamond_\Omega\exists x\in I\ \forall y\ \chi(x,y))\equiv_\Omega(EXP\wedge\exists x\forall y\ \chi(x,y))$.

2 $\Diamond_\Omega \top \equiv_\Omega \mathrm{EXP}$.

3 For S in Σ_1 we have: $I\Delta_0+\mathrm{EXP} \vdash \Delta_{\mathrm{EXP}} S \leftrightarrow \Box_\Omega S$.

4 For S in Σ_1 we have: $I\Delta_0+\Omega_1 \vdash \Box_{\mathrm{EXP}} S \leftrightarrow \Box_\Omega \Box_\Omega S$.

5 For S and S' in Σ_1 we have:
$I\Delta_0+\Omega_1 \vdash \Box_{\mathrm{EXP}}(S \to S') \to \Box_\Omega(\Box_\Omega S \to \Box_\Omega S')$.

Proof: $I\Delta_0+ \Omega_1 \vdash \Box_{\mathrm{EXP}}(S \to S') \to \Box_\Omega(\exists\ I\Delta_0+\Omega_1\text{-cut } I\ \Box_\Omega(S^I \to S')$
$\to \Box_\Omega(\exists\ I\Delta_0+\Omega_1\text{-cut } I\ (\Box_\Omega S^I \to \Box_\Omega S')$
$\to \Box_\Omega(\Box_\Omega S \to \Box_\Omega S')$ $\Box$

6 For ψ in Π_2 we have: $I\Delta_0+\Omega_1 \vdash \Box_\Omega \Box_\Omega \psi \to \Box_{\mathrm{EXP}} \psi$.

7 $I\Delta_0+\mathrm{SUPEXP}$ proves Π_2-reflection for $I\Delta_0+\mathrm{EXP}$. Let $\phi(x,y)$ be Δ_0, having only x,y free:

Proof: $I\Delta_0+\mathrm{SUPEXP} \vdash \Box_{\mathrm{EXP}} \forall x \exists y\ \phi(x,y) \to \Delta_{\mathrm{EXP}} \forall x \exists y\ \phi(x,y)$
$\to \exists\ I\Delta_0+\Omega_1\text{-cut } I \Box_\Omega \forall x \in I\ \exists y\ \phi(x,y))$
$\to \forall x \Box_\Omega \exists y\ \phi(x,y)$
$\to \forall x \Delta_\Omega \exists y\ \phi(x,y)$
$\to \forall x \exists y\ \phi(x,y)$ $\Box$

Combining 5.8.2 and 7.3.3 we get the desired missing principle for the combined provability logic of $\Box_{\mathrm{EXP}}$ and Δ_{EXP}:

8 $I\Delta_0+\mathrm{EXP} \vdash \Box_{\mathrm{EXP}} \psi \leftrightarrow \Delta_{\mathrm{EXP}} \Delta_{\mathrm{EXP}} \psi$

It is immediately clear that this last principle together with the complete set of principles for Δ_{EXP} fully describes the mixed logic. Note that we have this variation of Löb's Principle:

9 $I\Delta_0+\Omega_1 \vdash \Box_{\mathrm{EXP}}(\Box_{\mathrm{EXP}} \psi \to \Delta_{\mathrm{EXP}} \psi) \to \Box_{\mathrm{EXP}} \Delta_{\mathrm{EXP}} \psi$.

This principle "says" in some sense that $I\Delta_0+\mathrm{EXP}$ does not prove cut-elimination.

7.4 Friedman's Characterization

Let U and V be finitely axiomatized sequential theories. Combining 5.8.1 with 7.2 we find:

$I\Delta_0+EXP \vdash U \rhd V \leftrightarrow \Delta_{EXP}(Tcon(U) \rightarrow Tcon(V))$.

A variant is:

$I\Delta_0+\Omega_1 \vdash \Box_\Omega U \rhd V \leftrightarrow \Box_{EXP}(Tcon(U) \rightarrow Tcon(V))$.

Note that for interpretability over $I\Delta_0+EXP$ this implies:

$I\Delta_0+EXP \vdash \phi \rhd_{EXP} \psi \leftrightarrow \Delta_{EXP}(\nabla_{EXP}\phi \rightarrow \nabla_{EXP}\psi)$.

Combining this with the fact that we know the complete provability logic of Δ_{EXP} we get a *Kripke model characterization* of the interpretability logic of $I\Delta_0+EXP$. It is unknown what modal theory corresponds precisely with this Kripke semantics.

8 Arithmetical Completeness for Interpretations in Finitely Axiomatized, Sequential Theories extending $I\Delta_0+SUPEXP$

In this section we prove: ILP is complete for arithmetical interpretations in any finitely axiomatized sequential theory with designated natural numbers that satisfy $I\Delta_0+SUPEXP$ (plus a minimal extra condition). It is convenient to use a slightly different Kripke semantics in the proof of the arithmetical completeness theorem. Because this semantics is strongly suggested by Friedman's characterization of interpretability, I propose to call it Friedman Semantics.

8.1 Friedman semantics

A *Friedman structure* is a tuple <K,b,P,Q>, where:

i) K is a non-empty set.
ii) $b \in K$ and for every $x \in K$ bWQx
iii) R is a binary relation on K.
iv) Q is a transitive, irreflexive, upwards wellfounded, binary relation
v) $P \subseteq Q$
vi) $xQyPz \Rightarrow xPz$

Note that (v) and (vi) imply that P is transitive. Let $R:=Q \circ P$, i.e. $xRy \Leftrightarrow$ for some z xQzPy.

A relation $\Vdash$ between K and *L* is a *forcing relation* if it satisfies the usual clauses, with R as the accessibility relation for the $\Box$, plus:

$$x \Vdash \phi \rhd \psi \Leftrightarrow \forall u\,(xQu \Rightarrow (\exists y\,(uPy \wedge y \Vdash \phi) \Rightarrow \exists z\,(uPz \wedge z \Vdash \psi)))$$

If F is a Friedman structure and $\Vdash$ is a forcing relation on F, then $<F,\Vdash>$ is a *Friedman model*. It is easy to check that ILP is valid in any Friedman model.

Consider an IL-model $W=<K,R,S,\Vdash>$ and a Friedman model G'. β is a *bisimulation* between W and G' if:

i) β is a relation between K and K'.

ii) $b\beta b'$.

iii) Let x,y,... range over K, let x',y',... range over K':
we have for any x,x' with $x\beta x'$:
$\forall y(xRy \Rightarrow \exists u',y'(y\beta y' \wedge x'Q'u'P'y' \wedge \forall z'(u'P'z' \Rightarrow \exists z(z\beta z' \wedge yS_x z))))$.

iv) We have for any x,x' with $x\beta x'$:
$\forall u',y'(x'Q'u'P'y' \Rightarrow \exists y(y\beta y' \wedge xRy \wedge \forall z(yS_x z \Rightarrow \exists z'(z\beta z' \wedge u'P'z'))))$.

v) $x\beta x' \Rightarrow (x \Vdash p \Leftrightarrow x' \Vdash' p)$, for all atoms p.

Note that for every $x \in K$ there is an $x' \in K'$ with $x\beta x'$, but that possibly there are $u' \in K'$ such that for no $u \in K$ $u\beta u'$. We do have: for all $x' \in \{b'\} \cup$ range R' there is an $x \in K$ such that $x\beta x'$.

We have: $x\beta x' \Rightarrow$ for all ϕ $x \Vdash \phi \Leftrightarrow x' \Vdash' \phi$. The proof is by a trivial induction on ϕ.

To prove completeness for ILP w.r.t finite Friedman models it is clearly sufficient to show that every finite IL-model W satisfying: $xRyRzS_xu \Rightarrow zS_yu$, can be bisimulated by a finite Friedman model G'.

Construction: Let W be a finite IL-model for ILP. We construct a bisimulating Friedman model G':

$K' := \{<x_1,...,x_n> \mid n \geq 1,\ x_1 = b,\ x_{2i-1}Rx_{2i}$ (for $1 \leq i$ and $2i \leq n$) and
$x_{2i}S[x_{2i-1}]x_{2i+1}$ for $1 \leq i$ and $2i < n\}$

$b' = <b>$

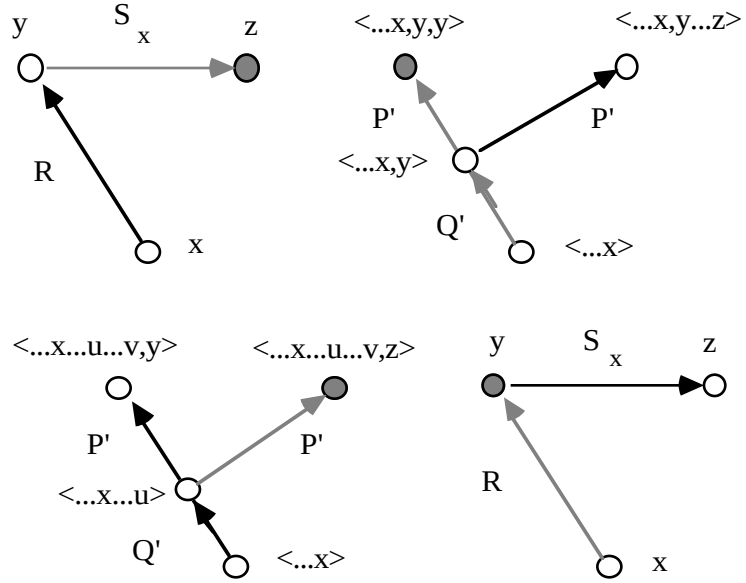
$<x_1,...,x_n>\ Q'\ <y_1,...,y_m> :\Leftrightarrow\ m > n$; for all $i \leq n$ $x_i = y_i$

$<x_1,...,x_n>\ P'\ <y_1,...,y_m> :\Leftrightarrow <x_1,...,x_n>\ Q'\ <y_1,...,y_m>$ and m is odd

$<x_1,...,x_n> \Vdash p :\Leftrightarrow x_n \Vdash p$

It is easy to see that the model constructed is a finite Friedman model. Note that Q' is a tree and that G' is really a Carlson model, i.e. there is a set $X' \subseteq K'$ such that $x'R'y' \Leftrightarrow x'Q'y'$ and $y' \in X'$. We always take $X' := \{\langle y_1,...,y_m\rangle \mid m \text{ is odd}\}$, so that $b' \in X'$. The nodes x' of X' also satisfy the additional property: if x'P'y' then x'R'y'. This property will be needed in our arithmetical completeness proof for reasons to be explained later.

Define: $x\beta x' :\Leftrightarrow x' = \langle x_1,...,x_{2m-1}\rangle$ and $x_{2m-1}=x$. We show that β is a bisimulation between W and G'.



The conventions for interpreting these pictures are as follows: the black arrows and open nodes are 'given', 'univer- sal'; the grey arrows and grey nodes are 'produced', 'existential'. If a 'given' arrow (node) in the left (right) half of the picture corresponds to a 'produced' arrow (node) in the left (right) half of the picture, then the 'produced' arrow (node) is produced, given the corresponding 'given' arrow (node). Corresponding nodes bisimulate under the given bisimula- tion β.

Let's first discuss the upmost picture. Here it is to be shown that indeed yS_xz. We have: either $\langle...x,y...z\rangle = \langle...x,y,z\rangle$ or $\langle...x,y...z\rangle = \langle...x,y,u,v...z\rangle$. In the first case we have trivially yS_xz. In the second case one easily shows uRz. So we have: yS_xuRz, hence yS_xz. Secondly we look at the second picture. We only have to show that for the unique w with $\langle...x...u...v,y\rangle = \langle...x...w,v,y\rangle$ we have: vS_wz. It is possible that x=w: this case is easy. So suppose $x \neq w$. As is easily seen: xRw, so $xRwRvS_wyS_xz$, and thus $xRwRyS_xz$. Hence yS_wz and thus vS_wz. □

8.2 A Solovay-style Completeness Proof

Let U be a finite sequential extension of $I\Delta_0$+SUPEXP. We only need to assume that U in fact extends $I\Delta_0$+SUPEXP (as an infinitely axiomatized theory), we do not need to stipulate that $I\Delta_0+EXP \vdash \Delta_{SUPEXP}\psi \rightarrow \Delta_U\psi$. This because we will only need actual theorems of $I\Delta_0$+SUP- EXP. We will need however the following lemma:

Lemma: For sentences $\psi \in \Pi_2$: $I\Delta_0+EXP \vdash \Delta_{EXP}\psi \rightarrow \Delta_U\psi$.

Proof: We have: $I\Delta_0+EXP \vdash \Delta_{EXP}\psi \rightarrow \Delta_U\Delta_{EXP}\psi$ and
$I\Delta_0+EXP \vdash \Delta_U(\Delta_{EXP}\psi \rightarrow \psi)$. Hence: $I\Delta_0+EXP \vdash \Delta_{EXP}\psi \rightarrow \Delta_U\psi$. □

(Note that even a stronger version of the lemma is provable with $\Box_{EXP}$ instead of Δ_{EXP}.)

We assume that for no k $U \vdash \Box^k_U \bot$. (If $U \vdash \Box^k_U \bot$ for some k, let k* be the smallest such k. The corresponding logic will then be ILP+$\Box^{k*}\bot$, as is seen by an easy adaptation of the argument below.)

Arithmetical Completeness Theorem : ILP$\vdash \phi \Leftrightarrow$ for all * $U \vdash \phi^*$.

Proof: the proof is a refinement of an earlier proof by Smorynski and the author for the case that U=GB or U=ACA_0. Its basic idea is close to that behind the completeness proof for PRL_{ZF} due to Carlson (see Smorynski[1985], p205-214).

The "$\Rightarrow$" side is clear. For the "$\Leftarrow$" part suppose ILP$\nvdash \phi$. Let G_0 be a finite Friedman countermodel to ϕ with Q_0 upwards wellfounded. We may assume that K_0 is {1,...,N}, that 1 is the bottom element and $1 \nVdash_0 \phi$, and that our model satisfies:

- (i) Q is a tree,
- (ii) P is given "Carlson-style" by a set X: so $xPy \Leftrightarrow xQy$ and $y \in X$; $1 \in X$.
- (iii) if $x \in X$ and xPy, then xRy.

From the arithmetical point of view the nodes in X correspond with the 'point of view' of U. P considered as an accessibility relation corresponds to Δ_U, Q similarly corresponds to Δ_{EXP}. Property (iii) corresponds to the fact that U proves cut-elimination and thus proves the equivalence of Δ_U and $\Box_U$.

G is the result of hanging a new node 0 under G_0. Formally: $K := K_0 \cup \{0\}$, xQy :$\Leftrightarrow$ (x=0 and y≠0) or xQ_0y, $X := X_0 \cup \{0\}$, xPy :$\Leftrightarrow$ xQy and $x \in X$,

$x \Vdash p :\Leftrightarrow x\neq 0$ and $x \Vdash_0 p$. Clearly $1 \nVdash \phi$.

Define by the recursion theorem:

$h(0):=0$,
$h(x+1):=y$ if $h(x)Py$ and $Tproof_U(x,L\neq y)$ or
if $h(x)Qy$ and $Tproof_{EXP}(x,L\neq y)$,
$h(x+1):=h(x)$ otherwise,
$L:=$ the unique x such that $\exists y \forall z>y\ h(z)=x$.

The definition of h can be given in such a way that 'h(x)=y'is a $\Delta_0(exp)$-formula and its elemen- tary properties are verifiable in $I\Delta_0+EXP$ (in fact we can do much better, but that is not relevant here). We have e.g. $I\Delta_0+EXP \vdash x<y \rightarrow h(x)WQh(y)$ and $I\Delta_0+EXP \vdash$ "L exists".

Note that: $I\Delta_0+EXP \vdash L=x \leftrightarrow \exists y\ h(y)=x \wedge \forall u,v((h(u)=x \wedge v>u)\rightarrow h(v)=x)$, so L=x is in fact the conjunction of a Σ_1- and a Π_1-formula.

Define for atoms p: $p^*:=\bigvee\{L=\underline{i}\,|\,i \Vdash p\}$.

Lemma: $U \vdash L\in X$.

Proof of lemma: Reason in U:

Suppose $L=\underline{i}\notin X$. By the definition of h: $\Delta_{EXP}L\neq\underline{i}$. By Π_2-reflection: $L\neq\underline{i}$. Contradiction, so $L\neq\underline{i}$. □

We show by induction on ψ for $1\leq i\leq N$:

$i \Vdash \psi \Rightarrow I\Delta_0+EXP \vdash L=\underline{i}\rightarrow\psi^*$,
$i \nVdash \psi \Rightarrow I\Delta_0+EXP \vdash L=\underline{i}\rightarrow\neg\psi^*$.

The cases of the atoms and the propositional connectives are trivial. The case of □ follows immediately from the case of $\rhd$. Assume the IH for χ. We show for $i\neq 0$:

for all j with iPj $j \Vdash \chi \Rightarrow I\Delta_0+EXP \vdash L=\underline{i}\rightarrow\Delta_U\chi^*$,
for some j iPj and $j \nVdash \chi \Rightarrow I\Delta_0+EXP \vdash L=\underline{i}\rightarrow\neg\Delta_U\chi^*$.

First suppose that $i\neq 0$ and for all j with iPj $j \Vdash \chi$. Reason in $I\Delta_0+EXP$:

Suppose $L=\underline{i}$. By the definition of h we have $\Delta_{EXP}L\neq\underline{i}$ or $\Delta_U L\neq\underline{i}$. In both cases $\Delta_U L\neq\underline{i}$. For some x $h(x)=\underline{i}$, so $\Delta_U\exists x\ h(x)=\underline{i}$ and thus by an easy argument:

$\Delta_U \underline{i}$QL. By the lemma: $\Delta_U \underline{i}$PL. Conclude $\Delta_U \bigvee\{L=\underline{j}|iPj\}$. By the Induction Hypothesis: $\Delta_U\chi^*$.

Assume that $i\neq 0$ and for some j iPj and $j \Vdash \chi$. Reason in $I\Delta_0$+EXP:

Suppose $L=\underline{i}$ and $\Delta_U\chi^*$. By the Induction Hypothesis: $\Delta_U(L=\underline{j}\rightarrow\neg\chi^*)$. Hence $\Delta_U L\neq j$. Suppose $Tproof_U(z,L\neq j)$. Because $L=\underline{i}$ clearly for every y h(y)Pj, hence $h(z+1)=\underline{j}$. Contra- diction. Conclude $\neg\Delta_U\chi^*$.

Suppose $\psi=\chi \triangleright \rho$. First we assume $i\neq 0$ and $i \Vdash \psi$. By the above: for every j with iQj: $I\Delta_0$+EXP$\vdash L=\underline{j}\rightarrow(\nabla_U\chi^*\rightarrow\nabla_U\rho^*)$. Moreover if $i\in X$ by the special property (iii) of our model: $I\Delta_0$+EXP$\vdash L=\underline{i}\rightarrow(\nabla_U\chi^*\rightarrow\nabla_U\rho^*)$. Reason in $I\Delta_0$+EXP:

Suppose $L=\underline{i}$. For some x $h(x)=\underline{i}$, so $\Delta_{EXP}\exists x\ h(x)=\underline{i}$. Conclude: $\Delta_{EXP}\bigvee\{L=\underline{j}|iWQj\}$. In case $i\in X$, we have immediately by the above: $\Delta_{EXP}(\nabla_U\chi^*\rightarrow\nabla_U\rho^*)$, i.e. $\chi^* \triangleright_U \rho^*$. If $i\notin X$, it follows that $\Delta_{EXP}L\neq\underline{i}$, and thus $\Delta_{EXP}\bigvee\{L=\underline{j}|iQj\}$. So again: $\Delta_{EXP}(\nabla_U\chi^*\rightarrow\nabla_U\rho^*)$, i.e. $\chi^* \triangleright_U \rho^*$.

Secondly assume $i\neq 0$ and $i \Vdash \psi$. So for some j with iQj for some k with jPk: $k \Vdash \chi$, and for all k' with jPk': $k' \Vdash \rho$. Ergo $I\Delta_0$+EXP$\vdash L=\underline{j}\rightarrow\nabla_U\chi^*$ and $I\Delta_0$+EXP$\vdash L=\underline{j}\rightarrow\Delta_U\neg\rho^*$. Reason in $I\Delta_0$+EXP:

Suppose $L=\underline{i}$ and $\Delta_{EXP}(\nabla_U\chi^*\rightarrow\nabla_U\rho^*)$. We have: $\Delta_{EXP}(L=\underline{j}\rightarrow\neg(\nabla_U\chi^*\rightarrow\nabla_U\rho^*))$, so $\Delta_{EXP}L\neq\underline{j}$. Suppose $Tproof_{EXP}(z,L\neq\underline{j})$. From $L=\underline{i}$, we have: for all y $h(y)Q\underline{j}$, so $h(z+1)=\underline{j}$. Contradiction. Conclude: $\neg\Delta_{EXP}(\nabla_U\chi^*\rightarrow\nabla_U\rho^*)$.

Finally: suppose $U\vdash\phi^*$. By the above we find: $U\vdash L\neq\underline{1}$. So by the definition of h and the fact that $1\in X$: $U\vdash L\neq\underline{0}$. Thus $U\vdash\bigvee\{L=\underline{i}|1<i\leq N\}$. Clearly for some k and for all i with $1<i\leq N$: $i \Vdash \Box^k\bot$. So: $U\vdash\Box^k{}_U\bot$. Quod non. □

8.3 Another Interpretation

Christian Bennet studies in his thesis (see Bennet[1986a]) the following notion of strong relative consistency over Peano Arithmetic: for A,B sentences of the language of Peano:

$\phi \triangleright^{SRC} \psi :\Leftrightarrow$ there is a primitive recursive term t such that
$PA\vdash\forall x(Proof_{PA+\psi}(x,\bot)\rightarrow Proof_{PA+\phi}(tx,\bot))$

(Actually Bennet defines his notion for arbitrary theories which are verifiably in PA RE extensions of PA; we scaled his notion down to fit our framework.)

Let PRA be Primitive Recursive Arithmetic, a theory which is for our purposes equal to $I\Sigma_1$. By a remark of Kreisel we have: $\phi \rhd^{SRC} \psi \Leftrightarrow PRA \vdash \Diamond_{PA}\phi \rightarrow \Diamond_{PA}\psi$. Formalizing this in PA we get:

$$PA \vdash \phi \rhd^{SRC} \psi \leftrightarrow \Box_{PRA}(\Diamond_{PA}\phi \rightarrow \Diamond_{PA}\psi)$$

Comparing this characterization with Friedman's characterization of $\rhd_U$, for finitely axiomatized sequential U, it is easy to adapt the above proof to show completeness of ILP for arithmetical interpretations where $\Box$ is interpreted as $\Box_{PA}$ and $\rhd$ is interpreted as $\rhd^{SRC}$.

References:

Bennet, C., 1986a, *On some orderings of extensions of arithmetic,* Thesis, Department of Philosophy, University of Göteborg, Göteborg.

Bennet, C., 1986b, *On a problem by D. Guaspari*, in: Furberg, M. &al eds., *Logic and Abstraction,* Acta Philosophica Gothoburgensia 1, Göteborg, 61-69.

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

Bezboruah, A., Shepherdson, J.C., 1976, *Gödel's second incompleteness theorem for Q*, JSL 41, 503-512.

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

Diaconescu, R., Kirby, L.A.S., 1987, *Models of arithmetic and categories with finiteness conditions,* Annals of pure and applied logic 35, 123-148.

Dimitricopoulos, C., 1980, *Matijasevic's theorem and fragments of arithmetic,* Thesis, Univ. of Manchester, Manchester.

Ehrenfeucht, A., Mycielski, J., ?, *Theorems and problems of the lattice of local interpretability.*

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

Feferman, S., Kreisel G., Orey, S., 1960, *1-consistency and faithful interpretations,* Archiv für Mathematische Logik und Grundlagen der Mathematik 6, 52-63.

Feferman, S., 1988, *Hilbert's program relativized: proof-theoretical and foundational reductions*, JSL 53, 364-384.

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

Gaifman, H., Dimitracopoulos, C., 1982, *Fragments of Peano's arithmetic and the MRDP theorem,* in: 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., 1984, *On a new notion of partial conservativity,* in: Richter, M.M. &al eds, *Computation and Proof Theory,* Logic Colloquium '83, Lecture Notes in Mathematics 1104, Springer Verlag, Berlin, 217-232.

Hájek, P., Kucera, A., ?, *On recursion theory in* IS_1.

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

Hájková, M., 1971a, *The lattice of binumerations of arithmetic I,* Commentationes Ma- thematicae Universitatis Carolinae 12, 81-104.

Hájková, M., 1971b, *The lattice of binumerations of arithmetic II,* Commentationes Ma- thematicae Universitatis Carolinae 12, 281-306.

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

Harrow, K., 1987, *The bounded arithmetic hierarchy,* Information and Control 36.

Jongh, D.H.J. de & Veltman F., ?, *Provability logics for relative interpretability,* this volume.

Krajícek, J., 1985, *Some theorems on the lattice of local interpretability types,* Zeitschrift für Mathematische Logik und Grundlagen der Mathematik 31, 449-460.

Krajícek, J., ?, *A note on proofs of falsehood.*

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 snd 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.

Macintyre, A., 1987, *On the strength of weak systems*, from: *Logic, Philosophy of Science and Epistemology*, Hölder-Pichler-Tempsky, 43-59.

Minc, G., 1972, *Quantifier-free and one-quantifier induction,* Journal of Soviet Mathematics, volume 1, 71-84.

Misercque, D., 1983, *Answer to a problem by D. Guaspari*, in: Guzicki, W. &al eds., *Open days in Model Theory and Set Theory,* Proceedings of a conference held in September 1981 in Jadwisin, Poland, Leeds University, 181-183.

Misercque, D., 1985, *Sur le treillis des formules fermées universelles de l' arithmétique de Peano,* Thesis, Université Libre de Bruxelles.

Montagna, F., 1987, *Provability in finite subtheories of PA and relative interpretability: a modal investigation,* JSL 52, 494-511.

Montague, R., 1958, *The continuum of relative interpretability types,* JSL 23, 460.

Montague, R., 1962, *Theories incomparable with respect to relative interpretability,* JSL 27, 195-211

Mycielski, J., 1962, *A lattice connected with relative interpretability of theories,* Notices of the AMS 9, 407-408. [*errata,* 1971,ibidem, 18, 984].

Mycielski, J., 1977, *A lattice of interpretability types of theories,* JSL 42, 297-305.

Mycielski, J., ?, *Finististic consistency proofs*

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.

Palúch, S., 1973, *The lattices of numerations of theories containing Peano's Arithmetic,* Commentationes Mathematicae Universitatis Carolinae 14, 339-359.

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

Paris, J.B., Dimitracopoulos, C., 1982, *Truth definitions for Δ_0-formulae*, in: *Logic and Algorithmic*, Monography 30 de l'Enseignement Mathematique Genève, 319-329.

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

Paris, J., Kirby, L.A.S., 1978, *S_n-collection schema's in arithmetic*, in: Logic Colloquium '77, North Holland, Amsterdam.

Paris, J., Wilkie, A., 1981, *Models of arithmetic and the rudimentary sets*, Bulletin Soc. Math. Belg. 33, 157-169.

Paris, J., Wilkie A., 1983, *Δ_0-sets and induction*, in: Guzicki, W. &al eds., *Open days in Model Theory and Set Theory*, Proceedings of a conference held in September 1981 in Jadwisin, Poland, Leeds University, 237-248.

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*, Commen- tationes Mathematicae Universitatis Carolinae 24, 667-671.

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

Pudlák, P., 1986, *On the length of proofs of finitistic consistency statements in finitistic theories*, in: Paris, J.B. &al eds., *Logic Colloquium '84*, North Holland, 165-196.

Pudlák, P., ?, *A note on bounded arithmetic.*

Quincy, J., 1981, *Sets of S_k-conservative sentences are P_2 complete*, JSL 46, [abstract], 442.

Schwichtenberg, H., *Proof Theory: Some Applications of Cut-Elimination*, in Barwise, J. ed., *Handbook of Mathematical Logic*, North Holland, 867-895.

Simpson, S.G., 1988, *Partial realizations of Hilbert's Program*, JSL 53, 349-363.

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.

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., 1986, *Peano's Smart Children, a provability logical study of systems with built-in consistency,* Logic Group Preprint Series nr 14, Dept. of Philosophy, University of Utrecht, Heidelberglaan 2, 3584CS Utrecht. To appear in the Notre Dame Journal of Formal Logic.

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

Wilkie, A., 1980a, *Some results and problems on weak systems of arithmetic,* in: MacIntyre, A. &al eds., *Logic Colloquium '77,* North Holland, Amsterdam, 285-296.

Wilkie, A., 1980b, *Applications of complexity theory to S_0-definability problems in arithmetic,* in: *Model theory of algebra and arithmetic,* Lecture Notes in Mathematics 834, Springer, Berlin, 363-369.

Wilkie, A., 1986, *On sentences interpretable in systems of arithmetic,* in: Paris, J.B. &al eds., *Logic Colloquium '84,* North Holland, 329-342.

Woods, A.R., 1981, *Some problems in logic and number theory, and their connections,* Thesis, Manchester University.