CAN WE MAKE THE SECOND INCOMPLETENESS THEOREM COORDINATE FREE?

ALBERT VISSER

ABSTRACT. Is it possible to give a coordinate free formulation of the Second Incompleteness Theorem? We pursue one possible approach to this question. We show that (i) cutfree consistency for finitely axiomatized theories can be uniquely characterized modulo EA-provable equivalence, (ii) consistency for finitely axiomatized sequential theories can be uniquely characterized modulo EA-provable equivalence. The case of infinitely axiomatized ce theories is more delicate. We carefully discuss this in the paper.

1. INTRODUCTION

What does the Second Incompleteness Theorem say? Well, it tells us that a sufficiently strong consistent theory does not prove its own consistency. Clearly, this statement leaves much to be desired, but we can at least take some steps to improve upon it. First, a theory is a first-order theory with a ce axiom set. Secondly, the theory should contain Robinson's Arithmetic Q, a very weak and finitely axiomatized theory. 'Contain' is still too vague here. A theory like ZF does not contain Q in the strict sense, simply because it has a signature that differs from the signature of arithmetic. However, by translating the signature of arithmetic, say, via the von Neumann interpretation, we can make sense of Q occurring in ZF. Thus, we can replace 'contains' by 'interprets'.[1] But what if there is more than one interpretation of Q? (There always is.) It does not matter: the Second Incompleteness Theorem works for them all. Thus, we may rephrase the Second Incompleteness Theorem as: no ce theory U (in predicate logic) interprets $Q+\mathsf{con}(U)$, where $\mathsf{con}(U)$ is a sentence expressing the consistency of U. In symbols:

$$U \not\rhd (Q + \mathsf{con}(U)).$$

Let us for a moment postpone the discussion of the crucial question *what is* $\mathsf{con}(U)$?, and pause to marvel at the beautiful statement of the theorem. In this strong form, the result is due to Pavel Pudlák. See [Pud85], [HP91]. Pudlák's own proof of the theorem uses some trickery, but this is really unnecessary. One shows that $Q+\mathsf{con}(U)$ interprets $S_2^1+\mathsf{con}(U)$ on a definable cut,[2] where S_2^1 is a weak arithmetic invented by Sam Buss to study polynomial time computability. See [Bus86]. Now,

[1]We may take interpretability in its most general sense: multidimensional, piecewise, relative interpretability with parameters.

[2]The important point is to take care that the cut is initial w.r.t. the original ordering relation provided by the theory.

Buss' theory gives us what we need to formalize the proof of the Second Incompleteness Theorem in a natural and effortless way.[3] In fact, it is easier to do it in Buss' theory than in full Peano Arithmetic PA, since the restrictions present in S_2^1 prevent one from making wrong turns and inefficient choices.

Remark 1.1. Is this version the best and most general version of the Second Incompleteness Theorem? I guess there is not really such a thing as the best and most general version. There are always improvements that one tends to view as better versions of the same theorem. E.g., we can replace Q in the statement of the theorem by a weaker theory Q^- in which addition and multiplication are partial. See, e.g. [Šve07]. I think this point was first seen by Robert Solovay.

   Along a different line: there are versions of the result where we replace provability in PA by provability in PA that can call on an oracle for Σ_n^0-truth, or by provability in second order logic with the ω-rule, or by truth in all transitive models in set theory. Etc.

   Along yet another line: the statement of the theorem still holds when we replace Q by its constructive counterpart $i\mathsf{Q}$ and allow U to be a constructive theory. Here one axiom of Q has to be rephrased as: every number is either zero or a successor.   ⊠

We return to our reflection on the statement of the theorem. We still have to explicate the meaning of $\mathsf{con}(U)$. It is the statement of the consistency of U, but which one? There are three sources of indeterminacy.

   I. We have to choose a proof system.
  II. We have to choose a coding scheme.
 III. We have to choose a specific formula representing the axiom set of U.

Note that (III) presupposes (II). We can escape (III) in special cases where the axiomatization has a special form, e.g. if is is finite. Feferman's solution ([Fef60]) to deal with the indeterminacy is to employ a fixed choice for (I) and (II) and to make (III) part of the individuation of the theory. Just choose your favorite proof system and coding scheme. Of course, this does give the theorem a definite meaning, but it does not very well represent the generality of the insight we feel we have. We believe that for all reasonable choices we have the theorem. However, the quantifier over reasonable choices, is 'unmathematical'. We must articulate just what a reasonable choice is, and this seems scarcely possible.

In the present paper, we describe one possible solution strategy to the problem for the finitely axiomatized case. The consistency statement can be pinned down as the unique solution of a certain equation modulo a suitable equivalence relation. The diverse ingredients of the equation are not dependent upon conventional choices. Our solution is in a sense not quite satisfactory, since the equivalence relation is EA-provable equivalence, where one would have liked something like S_2^1-provable equivalence. Also, in the case of consistency (as opposed to cutfree-consistency) we only have a solution for sequential theories. But, I guess, to adapt a quote from dr. Johnson, even if it is not quite as desired, it surprising that such a result is possible at all. In the infinitely axiomatized case, we can characterize the set of restricted consistency statements of all finite subtheories of the theory modulo EA-provable equivalence. Similarly, we can characterize the set of consistency statements of finite subtheories of a sequential theory modulo EA-provable equivalence. Finally,

---

[3]We can formalize the Second Incompleteness Theorem for a certain finitely axiomatized fragment of $I\Delta_0$ in that same fragment. However, to do this is far from effortless and requires the development of alternative methods.

we show how to characterize the full consistency statement of the theory modulo EA-provable equivalence in a way that is independent of (I), but does not eliminate the dependency on (II) and (III).

Remark 1.2. I feel that it is important to clearly separate the demands of *philosophical understanding* from the demands of *mathematical understanding*. From the philosophical point of view, one asks what it means to say that a *formal* theory cannot prove its own consistency. How do we know that the statements not proved are consistency statements, How do we know that there are not entirely different statements, that are consistency statements and provable? Etc. Philosophical reflection may lead one to deny that the Second Incompleteness Theorem for Q is about consistency at all.

The demands of mathematical understanding, on the other hand, are about finding the proper level of generality for the Second Incompleteness Theorem. We want to find a representation independent of arbitrary choices. The present paper is clearly concerned with the quest for mathematical understanding. ⬛

About the Paper. In Section 3, we study how cutfree-consistency can be given as unique solution modulo EA-provable equivalence. We also show how to characterize the set of restricted consistency statements of finite subtheories of a ce axiomatized theory. In Section 4, we do the same for consistency statements, restricting ourselves to sequential theories. Some technical preliminaries are in Section 2 and in Appendix A.

Many of the results employed in this paper have been known for a long time, even if, perhaps, not precisely in the form I gave them. What is new is the arrangement of the technical material around the theme of characterizing a consistency statement. This arrangement is inspired by some FOM postings in 2006 by Harvey Friedman, I think the idea of uniquely characterizing the consistency statements is original in this paper.

I wrote the paper as a presentation of ideas. This means that much of the technical background is only sketched with references to locations where more details are given. I always did provide the basic idea of a proof.

The book [HP91] is the locus classicus for most of the background material needed.

2. Preliminaries

In the body of the paper, we consider theories in *one-sorted* first order predicate logic. The results of the paper also work in the many-sorted case, however this would make our treatment more complicated. In Appendix A, we give some definitions for the many-sorted case. In Remark A.2, we point out that every many-sorted theory U is mutually interpretable with its one-sorted flattening $U^\flat$. Many of our results can be induced via the $(\cdot)^\flat$ functor from the one-sorted to the many-sorted world.

We assume that theories have *officially* a relational signature. Unofficially, we use function symbols, but these can be eliminated using a well-known unwinding procedure.

The axiom set of a theory is supposed to be given by a Σ_1^0-formula α. We consider α to be part of the data for the theory. Here is a list of the most important theories we will meet in the paper.

- Q, or Robinson's Arithmetic: this is a very weak arithmetical theory introduced in [TMR53]

- S_2^1: this is a theory introduced by Sam Buss to study polynomial time computability. See [Bus86].
- EA, or Elementary Arithmetic: this is $I\Delta_0 + \mathsf{exp}$. The theory is also called EFA, Elementary Function Arithmetic, by Harvey Friedman. It is also known as Kalmár Arithmetic or Kalmár Elementary Arithmetic.

Our notion of reduction of theories is *piecewise, more-dimensional, relative interpretation with parameters, where identity is not necessarily translated as identity.* This notion is explained in Appendix A. We write:

- $K : U \lhd V$ or $K : V \rhd U$, for: K is an interpretation of U in V.
- $V \rhd U :\Leftrightarrow U \lhd V :\Leftrightarrow \exists K\ K : U \lhd V$. We read $U \lhd V$ as: U is interpretable in V. We read $V \rhd U$ as: V interprets U.
- $U \equiv V :\Leftrightarrow U \rhd V$ and $V \rhd U$. We read $U \equiv V$ as: U and V are mutually interpretable.

We will sometimes view the structure of theories preordered by $\lhd$ as a preorder category.

A theory U is *locally interpretable* in V, if every finitely axiomatized subtheory U_0 of U is interpretable in V. We write $U \lhd_{\mathsf{loc}} V$, etc.

An important notion is *direct* interpretability. An interpretation is direct if it is one-dimensional, non-piecewise, unrelativized and sends identity to identity (i.o.w., identity is absolute for this interpretation).

We will use a fundamental theorem.

Theorem 2.1. *For any Π_1^0-sentences P, P', we have:*
$$(\mathsf{Q} + P) \rhd (\mathsf{Q} + P') \quad \Leftrightarrow \quad \mathsf{EA} \vdash P \to P'.$$

This theorem is due to Wilkie and Paris. See [WP87]. For a generalization, see: [Vis92].

3. Cutfree Consistency and Restricted Consistency

In this section, we study a characterization of cutfree consistency. For our characterization, we need the functor SEQ from theories to theories. We define $\mathsf{SEQ}(U)$ as follows. We first extend the language of U by a new predicate δ. Then, we relativize all the quantifiers of the axioms of U to δ, except the axioms for identity. We add the axiom that δ is non-empty. Say, the resulting theory is U^δ. Next we add a good (unrelativized) theory of sequences to U^δ, obtaining $\mathsf{SEQ}(U)$. These sequences are not supposed to be 'extensional': 'the same sequence' can have several carriers. The necessity of relativization can be seen by contemplating the example of a theory saying that there are precisely three objects.[4]

Since (i) writing down the appropriate theory of sequences is somewhat laborious and (ii) since there is a much simpler characterization of the thing we want (modulo mutual interpretability), we will not give the theory of sequences here. We will give

---

[4]Modulo mutual interpretability, we can also do it a little bit differently. We do not only add a new domain δ, but we also add a new binary relation symbol E. Now we do not only relativize but we also replace identity in the axioms of U by E. This time we do include the axioms of identity. We add a good theory of sequences including the ordinary theory of identity to the result. To see the equivalence of the notions, first we use of the presence of sequences to make our class of sequences and the U-objects disjoint. Then, we divide out the equivalence relation on the U-objects. Note that we do not need to divide out an equivalence relation on the sequences, since sequences need not be extensional.

the simpler characterization below. The resulting theory $\mathsf{SEQ}(U)$ is sequential in the sense of Pavel Pudlák. See [Pud83], [HP91]. This notion was independently invented by Harvey Friedman: see [Smo85]. Friedman calls the notion *adequate*. See [Vis07] for a brief history of the concept.

We can view SEQ as the left adjoint of the embedding functor of the degrees of interpretability of sequential theories into the degrees of interpretability tout court:

$$U \lhd \mathsf{emb}(V) \quad \Leftrightarrow \quad \mathsf{SEQ}(U) \lhd_{\mathsf{seq}} V.$$

Note that it follows that SEQ is a endo-functor of the preorder category of interpretability. Moreover, we have: $U \lhd \mathsf{SEQ}(U)$, and $\mathsf{SEQ}(\mathsf{SEQ}(U)) \equiv \mathsf{SEQ}(U)$.

A simpler but equivalent way to get the effect of adding sequences is to add 'adjunctive sets'. Adjunctive set theory AS is the following theory:

$\mathsf{AS}1. \vdash \exists x \, \forall y \, \neg \, y \in x$,
$\mathsf{AS}2. \vdash \forall u, v \, \exists x \, \forall y \, (y \in x \leftrightarrow (y \in u \lor y = v))$.

This theory has a long history. See [Vis08]. To 'add' AS, we first relativize U and then add AS.

We will use 'A' and 'B' for sentences, or if you wish finitely axiomatized theories. We have the following theorem.

Theorem 3.1. *We have:* $\mathsf{SEQ}(A) \equiv (\mathsf{Q} + \mathsf{cutfreecon}(A))$.

Proof. From left to right, the proof uses the fact that there is a standard bound on the complexity of the formulas occurring on any cutfree proof of $\neg A$. In $\mathsf{SEQ}(A)$ we can develop a truth predicate for formulas of restricted complexity. We do not have enough induction to prove the cutfree consistency of A, but we can get the same effect by going to a definable cut.[5] Thus, on some definable cut, we have $\mathsf{cutfreecon}(A)$. See, e.g., [Pud85], [HP91].

From right to left, we use the fact that we can use a consistency statement for cutfree provability to produce a model. Since the model is entirely syntactic, on closer inspection, we have constructed an interpretation. A careful verification can be found in [Kal89]. This gives us an interpretation of A. The objects of A are coded on some definable cut of $\mathsf{Q} + \mathsf{cutfreecon}(A)$. We can choose this cut so deep that, on it, we interpret a sequential theory like Buss' S_2^1. The sequences of numbers on the cut can be used to induce sequences of the objects of the interpretation. □

Suppose that for Π_1^0-sentences P and P', we have:

$$\mathsf{SEQ}(A) \equiv (\mathsf{Q} + P) \text{ and } \mathsf{SEQ}(A) \equiv (\mathsf{Q} + P').$$

It follows that $(\mathsf{Q} + P) \equiv (\mathsf{Q} + P')$. Hence, according to Theorem 2.1, we have $\mathsf{EA} \vdash P \leftrightarrow P'$. Combining this insight with Theorem 3.1, we find that, modulo EA-provable equivalence, there is a unique Π_1^0-sentence P, such that $\mathsf{SEQ}(A)$ is mutally interpretable with $(\mathsf{Q} + P)$. Thus, we have characterized the cutfree consistency statement as the unique solution of an equation. We formulate this as a theorem.

Theorem 3.2. *The sentence* $\mathsf{cutfreecon}(A)$ *is, modulo* EA*-provable equivalence, the unique* Π_1^0*-sentence* P, *such that* $\mathsf{SEQ}(A) \equiv (\mathsf{Q} + P)$.

Here is the well known Friedman characterization of interpretability.

---

[5]The awkwardness of the presence of two different meanings of 'cut' seems to be unavoidable here. Caveat lector!

Theorem 3.3. *Suppose A is sequential. We have:*
$A \rhd B \quad \Leftrightarrow \quad \mathsf{EA} \vdash \mathsf{cutfreecon}(A) \to \mathsf{cutfreecon}(B).$

Proof. Since SEQ is a functor, $A \rhd B$ implies $\mathsf{SEQ}(A) \rhd \mathsf{SEQ}(B)$. We also have the converse of this fact: suppose (i) $\mathsf{SEQ}(A) \rhd \mathsf{SEQ}(B)$. Since A is sequential, we have (ii) $A \rhd \mathsf{SEQ}(A)$. On general grounds, we have (iii) $\mathsf{SEQ}(B) \rhd B$. Combining (ii), (i) and (iii), we find: $A \rhd B$. We have:

$$
\begin{aligned}
A \rhd B \quad &\Leftrightarrow \quad \mathsf{SEQ}(A) \rhd \mathsf{SEQ}(B) \\
&\Leftrightarrow \quad (\mathsf{Q} + \mathsf{cutfreecon}(A)) \rhd (\mathsf{Q} + \mathsf{cutfreecon}(B)) \\
&\Leftrightarrow \quad \mathsf{EA} \vdash \mathsf{cutfreecon}(A) \to \mathsf{cutfreecon}(B). \qquad \square
\end{aligned}
$$

The following theorem is in Harvey Friedman's Tarski lectures.

Theorem 3.4. *The degrees of interpretability of sequential sentences are isomorphic to the degrees of provability of Π_1^0-sentences over EA.*

Proof. We consider the mapping $A \mapsto \mathsf{cutfreecon}(A)$. By Theorem 3.3, this mapping is an embedding. It is surjective since:

$$
(\mathsf{Q} + P) \equiv \mathsf{SEQ}(\mathsf{Q} + P) \equiv (\mathsf{Q} + \mathsf{cutfreecon}(\mathsf{Q} + P)),
$$

and, hence $\mathsf{EA} \vdash P \leftrightarrow \mathsf{cutfreecon}(\mathsf{Q} + P)$. $\qquad \square$

We can reproduce the above results employing various other kinds of consistency: Herbrand consistency, tableaux consistency and restricted consistency. For restricted consistency we employ ordinary provability but restrict the complexity of the formulas occurring in the proof. Let's take as measure of complexity ρ: depth of quantifier alternations. We say that a proof is an n-proof if it contains only formulas of complexity less than or equal to n. (We will employ a slightly modified notion of n-proof, when dealing with infinitely axiomatized theories.) We write '$\mathsf{con}_n(A)$' for: there is no n-proof of absurdum from A. We can now repeat the above development for $\mathsf{con}_{\rho(A)}(A)$. E.g., we have:

$$
A \rhd B \quad \Leftrightarrow \quad \mathsf{EA} \vdash \mathsf{con}_{\rho(A)}(A) \to \mathsf{con}_{\rho(B)}(B).
$$

We can formulate an alternative Friedman style characterization for restricted consistency. This characterization avoids exponentiation at the cost of raising one of the bounds.

Theorem 3.5. *Suppose A is sequential. Then:*
$A \rhd B \quad \Leftrightarrow \quad \exists m \; \mathsf{S}_2^1 \vdash \mathsf{con}_m(A) \to \mathsf{con}_{\rho(B)}(B).$

Proof. Suppose A is sequential. Suppose $K : A \rhd B$. In S_2^1, we can use K to transform a $\rho(B)$-proof of $\perp$ from B into an m-proof of $\perp$ from A, for sufficiently large m. This uses essentially the fact that we choose ρ to be: depth of quantifier alternations. Note that this direction does not use that A is sequential.

Conversely, suppose $\mathsf{S}_2^1 \vdash \mathsf{con}_m(A) \to \mathsf{con}_{\rho(B)}(B)$, for some m. Since A is sequential, we have $A \rhd (\mathsf{S}_2^1 + \mathsf{con}_m(A))$. Hence, $A \rhd (\mathsf{S}_2^1 + \mathsf{con}_{\rho(B)}(B))$. It follows that $A \rhd B$. $\qquad \square$

What can we say in the infinite case? To answer this question we need the functor $\mho$. We also introduce the functor $\mho^+$ that will play a role in the next section. Consider a theory U where the set of Gödelnumbers of axioms is represented by a Σ_1^0-formula αx. We may assume that αx is of the form $\exists y \, \alpha_0 y x$, where α_0 is Δ_0.

Wlog, we may assume that $\alpha_0 xy$ implies $x \leq y$. Let $U \upharpoonright n$ be the theory with axioms witnessed below n. We write $\mathsf{con}_n^+(U)$ for $\mathsf{con}(U \upharpoonright n)$.

We redefine restricted provability as follows. π is an n-proof from U if π is an n-proof in the original sense from $U \upharpoonright n$. So we both restrict the axioms used and the complexity of the formulas in the proof. We write $\mathsf{con}_n(U)$ for: there is no n-proof of $\bot$ from U. We define:

- $\mho(U) := \mathsf{S}_2^1 + \{\mathsf{con}_n(U) \mid n \in \omega\}$.
- $\mho^+(U) := \mathsf{S}_2^1 + \{\mathsf{con}_n^+(U) \mid n \in \omega\}$.
- U is *reflexive* if $U \rhd \mho(U)$.
- U is *strongly reflexive* if $U \rhd \mho^+(U)$.

We pronounce '$\mho(U)$' as: *mho U*, where 'mho' rhymes with 'Joe'. It is easy to see that $\mho(U)$ is, modulo provable equivalence, independent of the choice of α. Instead of S_2^1, we could have used e.g., Q or $I\Delta_0 + \Omega_1$. The theory S_2^1 is convenient since unlike Q we can formalize many intuitive arguments. Specifically, unlike Q, the theory S_2^1 is sequential. Unlike $I\Delta_0 + \Omega_1$, the theory S_2^1 is known to be finitely axiomatizable. In fact we will assume that it is finitely axiomatized.

Remark 3.6. Since, originally reflexivity was studied in the context of strong theories containing superexponentiation, without the relativization to some choice of the natural numbers, the notions of *reflexive* and *strongly reflexive* were not clearly distinguished. My choice of terminology reflects my conviction that *reflexivity* is the more fundamental notion.

An example of a reflexive theory is $I\Delta_0 + \Omega_1 + \Omega_2 + \dots$ Examples of strongly reflexive theories are PRA, PA, ZF.

Pavel Pudlák has shown, using an adaptation of the proof of the First Incompleteness Theorem, that reflexive theories are not finitely axiomatizable. One can easily see, that no theory that is mutually locally interpretable with a strongly reflexive theory can be finitely axiomatized. I.o.w., the degree of local interpretability of a strongly reflexive theory does not contain a finitely axiomatizable theory. As we will see this is not true for reflexive theories. ❏

Here is the central insight concerning the relationship between U and $\mho(U)$.

Theorem 3.7. *We have: (a)* $\mathsf{SEQ}(U) \rhd_{\mathsf{loc}} \mho(U)$ *and (b)* $\mho(U) \rhd \mathsf{SEQ}(U)$. *It follows that* $\mathsf{SEQ}(U) \equiv_{\mathsf{loc}} \mho(U)$.

Proof. The proof is a strengthening of the proof of Theorem 3.1. For (a), we build partial truth predicates and prove the restricted consistency statements by compensating the lack of induction by going to a definable cut. Note that as n grows, we need deeper and deeper cuts to prove $\mathsf{con}_n(U)$. Thus, we only construct local interpretations.

For (b), we work in $\mathsf{SEQ}(U)$. We consider the set X of all x, such that $\mathsf{con}_x(U)$. We note that $\mathsf{con}_X(U)$, i.e., U is consistent when we restrict ourselves to proofs that (i) employ only axioms witnessed in X and (ii) only contain formulas of complexity in X. We use the Henkin-Feferman construction on $\mathsf{con}_X(U)$ to build an interpretation H_U of U. We add sequences to H_U as in the proof of Theorem 3.1. Note that, externally, we can prove: $\alpha(A) \Rightarrow \mho(U) \vdash X(\ulcorner \underline{A} \urcorner)$, and hence $\mho(U) \vdash A^{\mathsf{H}_U}$.

The basic idea of the Feferman style proof comes from Sol Feferman's classical paper [Fef60]. The extension to restricted provability was added later. For more details see [Vis92]. ❏

We can now formulate and prove the analogue of Theorem 3.2.

Theorem 3.8. *Let $\mathcal{P}$ be a set of Π_1^0-sentences. We have:*
$$\mathsf{SEQ}(U) \equiv (\mathsf{S}_2^1 + \mathcal{P}) \; \textit{iff} \; (\mathsf{EA} + \mho(U)) = (\mathsf{EA} + \mathcal{P}).$$

In the rest of this section, we will collect some further facts about $\mho$. These facts are only in part relevant for the rest of the paper and could be skipped by the reader mainly interested in Gödel's Second Incompleteness Theorem.

Theorem 3.9. *Suppose U is sequential. The following are equivalent:*

a. $U \rhd_{\mathsf{loc}} V$,
b. $\mho(U) \supseteq \mho(V)$,
c. $\mho(U) \rhd \mho(V)$,
d. $\mho(U) \rhd V$,
e. $(\mathsf{EA} + \mho(U)) \supseteq (\mathsf{EA} + \mho(V))$,

Proof. $(a) \Rightarrow (b)$: Suppose $U \rhd_{\mathsf{loc}} V$, Then, we can use, inside S_2^1, the local interpretations to transform an n-proof (for standard n) of $\bot$ from V, into an m-proof of $\bot$ from U, for sufficiently large standard m.

$(b) \Rightarrow (c)$: Trivial.

$(c) \Rightarrow (d)$: This follows from the fact that, by the Henkin-Feferman argument, we have $\mho(V) \rhd V$.

$(d) \Rightarrow (a)$: Since U is sequential, we have: $U \rhd_{\mathsf{loc}} \mho(U)$.

$(c) \Leftrightarrow (e)$: Immediate by Theorem 2.1.   ❑

We have a brief look at the meaning of two of the above equivalences. The equivalence between (a) and (b) is a sort of double Orey-Hájek characterization of local interpretability. The equivalence between (a) and (c) can be viewed as follows. Consider the preorder category of global interpretablity restricted to the the sequential theories. We have the projection functor π, which is given by the identical mapping, from this category to the preorder category of local interpretability, We have:
$$\pi(V) \lhd_{\mathsf{loc}} U \quad \Leftrightarrow \quad V \lhd \mho(U).$$

Thus, $\mho$ is right adjoint of π and as such uniquely determined modulo mutual interpretability. If we consider the theories in the degree of local interpretability of U, ordered by interpretability, then $\mho(U)$ is the maximal element.

We end this section by providing another functor that is the same as $\mho$ modulo mutual interpretability. We define the first order comprehension functor FOC as follows. We start with a sequential theory U. We add two unary predicates ob and class and one binary predicate $\in^\star$ to the language. We relativize U to ob. We write, e.g., $\forall x{:}\mathsf{ob}\, B$ for $\forall x\,(\mathsf{ob}(x) \to B)$. Using the sequences of U we can add axioms for comprehension over ob to the language of the following form:
$$\forall \vec{z}{:}\mathsf{ob}\, \exists y{:}\mathsf{class}\, \forall x{:}\mathsf{ob}\, (x \in^\star y \leftrightarrow A^{\mathsf{ob}}x\vec{z}).$$

Here A is a formula of the language of U. The superscript ob means that we relativize all quantifiers in A to ob. Thus we obtain $\mathsf{FOC}(U)$.

Theorem 3.10. *Suppose U is sequential. Then, $\mathsf{FOC}(U) \equiv \mho(U)$.*

Proof. We will use capitals for variables relativized to class and small letters for variables relativized to ob.

We start with the left to right direction. Since U is sequential, it contains an interpretation N of S_2^1. We can find, for every n, a definable cut I of N, such that $U \vdash \mathsf{con}_n^I(U)$. Here we demand that cuts are closed under $0, +, \times, \#$ and downwards closed w.r.t. $\leq$. Thus, they preserve S_2^1. We consider, in $\mathsf{FOC}(U)$, the cut $\mathcal{J}$ with $Jx :\leftrightarrow \forall X \, (\mathsf{cut}_N(X) \to x \in^\star X)$. It is easily seen that:

$$\mathcal{J} : \mathsf{FOC}(U) \rhd \mho(U).$$

We proceed with the right to left direction. It is clearly sufficient to show that $U \rhd_{\mathsf{loc}} \mathsf{FOC}(U)$. We enumerate the formulas of U as $A_0, A_1, \dots$. Let ℓ_i be the length of the parameter sequence $\vec{z}_i$ in $A_i x \vec{z}_i$. We show that the comprehension scheme for the first n formulas is interpretable. We employ the version of sequentiality defined using Adjunctive Set Theory. We set:

- $\mathsf{ob}(x) :\leftrightarrow (x = x)$,
- $\mathsf{class}(x) :\leftrightarrow (x = x)$.
- $n(x) :\leftrightarrow \exists y_0 \dots y_{n-1} \, (\bigwedge_{i<j<n} y_i \neq y_j \wedge \forall z \, (z \in x \to \bigvee_{i<n} y_i = z))$.
- We define pairs and n-sequences in the usual way (not forgetting the fact that we do not have extensionality).
- $x \in^\star y :\leftrightarrow \bigvee_{i<n} ((\ell_i + 1)\text{-}\mathsf{seq}(y) \wedge \exists u, v \, (0(u) \wedge i(v) \wedge \mathsf{proj}(w, u, v)) \wedge$
  $\bigwedge_{j<\ell_i} \exists w \, ((j+1)(w) \wedge \mathsf{proj}(w, y, z_{ij})) \wedge A_i x \vec{z}_i)$. ❑

Open Question 3.11. We have seen that, for sequential theories, FOC and $\mho$ coincide modulo mutual interpretability. Outside of this range they do not coincide. Could it be that FOC has some of the good properties of $\mho$ outside of the range where $\mho$ has these properties?

We consider the case of *pair-theories*. Here is their definition. The theory TP is axiomatized by:

tp1. $\vdash \exists x \, \forall y \, y \notin x$,
tp2. $\vdash \forall u, v \, \exists x \, \forall y \, (y \in x \leftrightarrow (x = u \vee x = v))$.

A pair-theory is a theory that directly interprets TP. It is easy to see that any pair-theory U locally interprets $\mathsf{FOC}(U)$. We would conjecture that, for a pair-theory U, the theory $\mathsf{FOC}(U)$ is maximal in the interpretability ordering among theories mutually locally interpretable with U. Hence, the following question: *Do we have, for every pair-theory U, that, for every V, $U \rhd_{\mathsf{loc}} V \Leftrightarrow \mathsf{FOC}(U) \rhd V$?* ❑

4. Consistency

In this section we study consistency statements. We look for an equation with a consistency statement as its unique solution. Over sequential theories, we can define a functor $U \mapsto \mathsf{PC}(U)$ that adds predicative comprehension to U. First we relativize U to a new predicate ob. We also add a new predicate class of classes and $\in^\star$ for an object being an element of a class. We add an axiom that $x \in^\star y$ implies $\mathsf{ob}(x)$ and $\mathsf{class}(y)$. We say that A is sorted iff there is a function σ that assigns to each variable of A either ob or class in such a way that (i) if $x = y$ is a subformula of A, then $\sigma(x) = \sigma(y)$, (ii) if $P\vec{x}$ is a subformula of A, then $\sigma(x_i) = \mathsf{ob}$, (iii) if

$x \in^\star y$ occurs in A, then $\sigma(x) = \mathsf{ob}$ and $\sigma(y) = \mathsf{class}$, and (iv) all quantifiers with variable x are relativized to $\sigma(x)$.[6]

Predicative comprehension is the principle:

- $\vdash \exists x{:}\mathsf{class}\, \forall y{:}\mathsf{ob}\ (y \in^\star x \leftrightarrow Ay\vec{z})$, where, A is sorted, y is assigned ob in the sorting of A, and all quantifiers in A are relativized to ob.

Since U is sequential, we can find a finite axiomatization for $\mathsf{PC}(U)$. See e.g. [Vis06], for details. It is easy to show that PC is an endofunctor on the preorder category of interpretability between sequential theories.

Theorem 4.1. *Suppose A is sequential. We have:* $\mathsf{PC}(A) \equiv (\mathsf{Q} + \mathsf{con}(A))$.

Proof. Using our classes, we can build a satisfaction predicate for (the relativization of) the language of A. See e.g. [Vis06], for details. We can use this predicate to prove the consistency of A on a $\mathsf{PC}(A)$-definable cut.

In the other direction, we make a Henkin-Feferman interpretation in $\mathsf{Q}+\mathsf{con}(A)$. This interpretation comes naturally equipped with a truth predicate for the interpretation. We use this truth predicate to define the classes as formulas in the (internally defined) language of U. See [Vis06], for details. ❑

Open Question 4.2. Can we lift the restriction to sequential A in the above theorem. I.o.w., can we find an appropriate 'coordinate free' functor F such that $F(A) \equiv (\mathsf{Q} + \mathsf{con}(A))$, for all A? ❑

Open Question 4.3. Suppose A is sequential. Can we prove $A \not\rhd \mathsf{PC}(A)$ directly, without using the Second Incompleteness Theorem? (Jan Krajíček asked a similar question in conversation.) ❑

Using the same considerations as in the case of cutfree-consistency, we find:

Theorem 4.4. *Suppose A is sequential. The statement $\mathsf{con}(A)$ is, modulo EA-provable equivalence, the unique Π^0_1-sentence P, such that $\mathsf{PC}(A) \equiv (\mathsf{Q} + P)$.*

There is an alternative characterization of $\mathsf{Q} + \mathsf{con}(A)$.

Theorem 4.5. *Suppose A is sequential. Then,*
$(\mathsf{Q} + \mathsf{con}(A)) \equiv (\mathsf{EA} + \mathsf{cutfreecon}(A))$.

Proof. Suppose A is sequential.
From left to right. As a first step, we interpret $\mathsf{S}^1_2 + \mathsf{con}(A)$ in $\mathsf{Q} + \mathsf{con}(A)$. We build a Henkin-Feferman interpretation $\mathsf{H}^+ := \mathsf{H}^+_A$ of A in $\mathsf{S}^1_2 + \mathsf{con}(A)$. Since A is sequential, we have in A an interpretation N of $\mathsf{Q} + \mathsf{cutfreecon}(A)$.

By an argument of Pavel Pudlák in [Pud85], we have:

$$(1) \qquad\qquad \mathsf{S}^1_2 \vdash \forall x\ (2^x \downarrow\, \rightarrow \Box_\mathsf{Q}(\mathsf{itexp}(x)\downarrow)),$$

where $\mathsf{itexp}(0) = 0$ and $\mathsf{itexp}(x+1) = 2^{\mathsf{itexp}(x)}$.

There is a $(\mathsf{S}^1_2 + \mathsf{con}(A))$-definable cut J and a standard embedding e of the J numbers into an external cut of the H^+-numbers. By the construction of H^+, we have:

$$(2) \qquad\qquad \mathsf{S}^1_2 + \mathsf{con}(A) \vdash ((\vec{x} : J \wedge \Box_\mathsf{Q} B\vec{x}) \rightarrow B^{\mathsf{H}^+}\mathsf{e}\vec{x}).$$

---

[6]Of course, the notion of sortedness just says that A is the translation of the more natural corresponding sorted formula. We are punished for our choice for one-sorted theories. Our $\mathsf{PC}(U)$ is the flattening of the more natural two sorted predicative extension of U. Note, however, that if we start with a many sorted U, it becomes problematic to define sequentiality.

(We are not fully precise here: e is really a relation between equivalence classes of different equivalence relations.) Since we can always shorten J, we may assume that, for every x in J, S_2^1 proves that 2^x exists. Combining equations (1) and (2), we find:

$$(3) \qquad\qquad \mathsf{S}_2^1 \vdash \forall x{\in}J \; (\mathsf{itexp}(\mathsf{e}(x)) \downarrow)^{\mathsf{H}^+}.$$

Now we restrict the domain of H^+ to:

- $\eta := \{x{\in}\delta_{\mathsf{H}^+} \mid \exists y{\in}J \; (x \leq \mathsf{itexp}(\mathsf{e}(y)))^{\mathsf{H}^+}\}.$

This restriction evidently gives us an interpretation of $\mathsf{EA} + \mathsf{cutfreecon}(A)$.[7] For more details of this argument, see [Vis92].

From right to left. The theory EA has a cut $\mathcal{J}$, such that EA proves that for all x in $\mathcal{J}$, the superexponent of x exists. This cut is called *the superexponential cut.* We find, by the formalization of cutelimination:

$$\mathsf{EA} \vdash \mathsf{cutfreecon}(A) \to \mathsf{con}^{\mathcal{J}}(A).$$

The desired result is now immediate.   ❑

It immediately follows that:

Theorem 4.6. *Let A be sequential. The sentences $\mathsf{cutfreecon}(A)$, $\mathsf{con}(A)$ are the unique pair, modulo EA-provable equivalence, of Π_1^0-sentences P, Q, such that:*
$A \equiv (\mathsf{Q} + P)$ *and* $(\mathsf{EA} + P) \equiv (\mathsf{Q} + Q)$.

What happens in the infinite case? I can see two approaches. The first, extensional one, employs $\mho^+$. The other one employs the notion of scheme. Here is the $\mho^+$-approach.

Theorem 4.7. *Suppose U is sequential. Then, $\mathsf{PC}(U) \equiv \mho^+(U)$.*

Proof. Suppose U is sequential.

From left to right, we use the truth predicate for the language of U, that we can build using our classes, to prove the consistency statements on a definable cut. Note that we have one cut for all consistencies since we have truth for the full language of U. Note also that we do not get $\mathsf{con}(U)$, since we can prove of each axiom that it is true, but we need not be able to prove that all axioms are true.

From right to left: we use the Henkin-Feferman construction as in the proof of Theorem 4.1.   ❑

We may conclude:

Theorem 4.8. *Let $\mathcal{P}$ be a set of Π_1^0-sentences. We have:*
$(\mathsf{Q} + \mathcal{P}) \equiv \mathsf{PC}(U) \quad \Leftrightarrow \quad (\mathsf{EA} + \mathcal{P}) = (\mathsf{EA} + \mho^+(U))$.

Here is the analogue of Theorem 4.5.

Theorem 4.9. *Suppose U is sequential. Then, $\mho^+(U) \equiv (\mathsf{EA} + \mho(U))$.*

Proof. The proof is an adaptation of the proof of Theorem 4.5.   ❑

Here is the analogue of Theorem 4.6.

---

[7]Our argument would not have worked with $\mathsf{Q} + \mathsf{cutfreecon}(A)$ in the place of $\mathsf{Q} + \mathsf{con}(A)$ as the interpreting theory, since the efficient S_2^1-proofs of $\mathsf{itexp}(x) \downarrow$, may have non-standard cutrank.

Theorem 4.10. *Let U be a sequential theory. The theories $\mho(U)$, $\mho^+(U)$ are the unique pair, modulo EA-provable equivalence, of classes of Π_1^0-sentences $\mathcal{P}, \mathcal{Q}$, such that $U \equiv (\mathsf{S}_2^1 + \mathcal{P})$ and $(\mathsf{EA} + \mathcal{P}) \equiv (\mathsf{S}_2^1 + \mathcal{Q})$.*

Proof. The proof is an adaptation of the proof of Theorem 4.6. ❑

We end with the schematic approach to the infinitely axiomatized case. Consider any sequential theory U. By a theorem of Vaught (see [Vau67]), U is axiomatizable by a scheme.[8] We can think of a scheme as a sentence containing certain schematic variables $\vec{X}$. Vaught's theorem works even when we allow just one schematic variable and even if we do not allow parameters.

We consider a slightly more detailed statement, Consider any sequential ce theory U. Remember that a Σ_1^0-sentence α representing the axiom set of U is part of the data for U. Vaught provides an effective construction transforming α into a scheme Σ_α, such that the theory $\widetilde{U}$ axiomatized by Σ_α is extensionally equal to U. We take Σ_α to be part of the data for $\widetilde{U}$. Inspection of Vaught's argument shows that the extensional equality of U and $\widetilde{U}$ is EA-verifiable.

Consider any theory V given by a single scheme, which is part of the data of V. We define $\mathsf{PC}^+(V)$ as $\mathsf{PC}(V)$ plus the universal closure of the scheme treating the schematic variables as class variables. This is precisely the transformation of PA to ACA_0 and of ZF to GB. Note that $\mathsf{PC}^+(V)$ is finitely axiomatizable. We have:

Theorem 4.11. *Suppose V is sequential and given by a finite scheme. Then, $\mathsf{PC}^+(V) \equiv (\mathsf{Q} + \mathsf{con}(V))$.*

For a proof, see [Vis06]. So we immediately get:

Theorem 4.12. *Let U be a sequential theory. Then, $\mathsf{con}(U)$ is the unique Π_1^0-sentence P, modulo EA-provable equivalence, such that $\mathsf{PC}^+(\widetilde{U}) \equiv (\mathsf{Q} + P)$.*

Note that that this construction is intensional —in contrast to the $\mho^+$-approach— in the sense that $\mathsf{PC}^+(\widetilde{U})$ is crucially dependent on α. E.g., consider $\mathsf{PA}^{\mathsf{zf}}$, a theory that is axiomatized as follows. We proceed in stages. At stage n we add the nth PA-axiom and if n codes an inconsistency-proof of ZF we add $\perp$. Clearly, $\mathsf{PA}^{\mathsf{zf}}$ is extensionally equal to PA. However it is equally clear that $\mathsf{con}(\mathsf{PA}^{\mathsf{zf}})$ is EA-provably equivalent to $\mathsf{con}(\mathsf{ZF})$. Hence: $\mathsf{PC}^+(\widetilde{\mathsf{PA}^{\mathsf{zf}}}) \equiv (\mathsf{Q}+\mathsf{con}(\mathsf{ZF}))$. Thus, the same theory is axiomatizable by many schemes that are not 'predicatively second order equivalent'.

Thus, the present construction is dependent on the coding of syntax and on α, but it is independent of the choice of the proof system.

References

[Bus86] S. Buss. *Bounded Arithmetic.* Bibliopolis, Napoli, 1986.

[Fef60] S. Feferman. Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, 49:35–92, 1960.

[HP91] P. Hájek and P. Pudlák. *Metamathematics of First-Order Arithmetic.* Perspectives in Mathematical Logic. Springer, Berlin, 1991.

[Kal89] M.B. Kalsbeek. An orey sentence for predicative arithmetic. Technical Report Report X-89-01, ITLI, University of Amsterdam, 1989.

---

[8]Actually Vaught's Theorem is stronger, since it uses a somewhat weaker assumption on the theory. In fact, one can prove an even sharper result than Vaught's: the theorem works for all pair-theories,

[Pud83] P. Pudlák. Some prime elements in the lattice of interpretability types. *Transactions of the American Mathematical Society*, 280:255–275, 1983.

[Pud85] P. Pudlák. Cuts, consistency statements and interpretations. *The Journal of Symbolic Logic*, 50:423–441, 1985.

[Smo85] C. Smoryński. Nonstandard models and related developments. In L.A. Harrington, M.D. Morley, A. Scedrov, and S.G. Simpson, editors, *Harvey Friedman's Research on the Foundations of Mathematics*, pages 179–229. North Holland, Amsterdam, 1985.

[Šve07] V. Švejdar. An interpretation of Robinson's Arithmetic in Grzegorczyk's weaker variant. *Fundamenta Informaticae*, 81:347–354, 2007.

[TMR53] A. Tarski, A. Mostowski, and R.M. Robinson. *Undecidable theories*. North–Holland, Amsterdam, 1953.

[Vau67] R.A. Vaught. Axiomatizability by a schema. *The Journal of Symbolic Logic*, 32(4):473–479, 1967.

[Vis92] A. Visser. An inside view of exp. *The Journal of Symbolic Logic*, 57:131–165, 1992.

[Vis06] A. Visser. The predicative Frege hierarchy. Logic Group Preprint Series 246, Department of Philosophy, Utrecht University, Heidelberglaan 8, 3584 CS Utrecht, ⟨http://www.phil.uu.nl/preprints/lgps/⟩, 2006.

[Vis07] A. Visser. Pairs, sets and sequences in first order theories. Logic Group Preprint Series 251, Department of Philosophy, Utrecht University, Heidelberglaan 8, 3584 CS Utrecht, ⟨http://www.phil.uu.nl/preprints/lgps/⟩, 2007.

[Vis08] A. Visser. Cardinal arithmetic in weak theories. Logic Group Preprint Series 266, Department of Philosophy, Utrecht University, Heidelberglaan 8, 3584 CS Utrecht, ⟨ http://www.phil.uu.nl/preprints/lgps/⟨, 2008.

[WP87] A. Wilkie and J.B. Paris. On the scheme of of induction for bounded arithmetic formulas. *Annals of Pure and Applied Logic*, 35:261–302, 1987.

Appendix A. Interpretations

We will give a careful definition of *one-dimensional, many-sorted, relative interpretations without parameters, where identity is not necessarily translated as identity*. We will briefly indicate how to extend the framework to being piecewise, more-dimensionality and parameters.

A.1. **Translations.** To define an interpretation, we first need the notion of translation. Let Σ and Ξ be finite signatures for many-sorted predicate logic with finitely many sorts. We assume that the sorts are specified with the signature. We also assume that the designated sort is also given by the signature. A *relative translation* $\tau : \Sigma \to \Xi$ is given by a triple $\langle \sigma, \delta, F \rangle$. Here σ is a mapping of the Σ-sorts to the Ξ-sorts. The mapping δ assigns to every Σ-sort $\mathfrak{a}$ a Ξ-formula $\delta^{\mathfrak{a}}$ representing the *domain* for sort $\mathfrak{a}$ of the translation. We demand that $\delta^{\mathfrak{a}}$ contains at most a designated variable $v_0^{\sigma\mathfrak{a}}$ of sort $\sigma\mathfrak{a}$ free. The mapping F associates to each relation symbol R of Σ a Ξ-formula $F(R)$. The relation symbol R comes equipped a sequence $\vec{\mathfrak{a}}$ of sorts. We demand that $F(R)$ has at most the variables $v_i^{\sigma\mathfrak{a}_i}$ free. We translate Σ-formulas to Ξ-formulas as follows:

- $(R(y_0^{\mathfrak{a}_0}, \cdots, y_{n-1}^{\mathfrak{a}_{n-1}}))^{\tau} := F(R)(y_0^{\sigma\mathfrak{a}_0}, \cdots, y_{n-1}^{\sigma\mathfrak{a}_{n-1}})$.
  (We assume that some mechanism for α-conversion is built into our definition of substitution to avoid variable-clashes.)
- $(\cdot)^{\tau}$ commutes with the propositional connectives;
- $(\forall y^{\mathfrak{a}} A)^{\tau} := \forall y^{\sigma\mathfrak{a}} (\delta^{\mathfrak{a}}(y) \to A^{\tau})$;
- $(\exists y^{\mathfrak{a}} A)^{\tau} := \exists y^{\sigma\mathfrak{a}} (\delta^{\mathfrak{a}}(y) \wedge A^{\tau})$.

Suppose τ is $\langle \sigma, \delta, F \rangle$. Here are some convenient conventions and notations.

- We write δ_{τ} for δ and F_{τ} for F.
- We write R_{τ} for $F_{\tau}(R)$.

- We write $\vec{x} : \delta^{\vec{\mathfrak{a}}}$ for: $\delta^{\mathfrak{a}_0}(x_0^{\sigma \mathfrak{a}_0}) \wedge \ldots \wedge \delta^{\mathfrak{a}_{n-1}}(x_{n-1}^{\sigma \mathfrak{a}_{n-1}})$.
- We write $\forall \vec{x} : \delta^{\vec{\mathfrak{a}}} \ A$ for: $\forall x_0^{\sigma \mathfrak{a}_0} \ldots \forall x_{n-1}^{\sigma \mathfrak{a}_{n-1}} \ (\vec{x}:\delta^{\vec{\mathfrak{a}}} \to A)$. Similarly for the existential case.

A.2. Interpretations and Interpretability. A translation τ supports a *relative interpretation* of a theory U in a theory V, if, for all axioms A of U, we have $U \vdash A \Rightarrow V \vdash A^{\tau}$. (Note that this automatically takes care of the theory of identity. Moreover, it follows that $V \vdash \exists v_0 \ \delta_{\tau}^{\mathfrak{a}} v_0$, for each Σ-sort $\mathfrak{a}$.) Thus, an interpretation has the form: $K = \langle U, \tau, V \rangle$.

Par abus de langage, we write 'δ_K' for: δ_{τ_K}; 'P_K' for: P_{τ_K}; 'A^K' for: A^{τ_K}, etc. We define: $K : U \lhd V$ or $K : V \rhd U$, for: K is an interpretation of the form $\langle U, \tau, V \rangle$.

A.3. Local Interpretability. We say that a theory V locally interprets a theory U if, for any finite subtheory U_0 of U, we have $V \rhd_{(\mathsf{dir})} U_0$. We write $V \rhd_{\mathsf{loc}} U$ for: V locally interprets U.

A.4. Multidimensionality. We can extend the notion of interpretation to the case of multidimensional interpretations by sending a sort $\mathfrak{a}$ of the interpreted theory via σ to *a sequence of sorts* of the interpreting theory. A domain formula $\delta^{\mathfrak{a}}$ will have a sequence of variables as arguments. These variables have the sorts given by $\sigma(\mathfrak{a})$.

A.5. Parameters. We can extend the notion of interpretation to the case of interpretation with parameters by allowing extra parameters from a given finite sequence $\vec{w}$ into the δ_{τ} and P_{τ}. As extra data we need a fixed formula $A\vec{w}$ of the language of the interpreting theory representing the intended range of the parameters. E.g., in the Poincaré disk interpretation of the hyperbolic plane in the Euclidean plane, the parameters could be w_0, w_1, where w_0 is the centre of the Poincaré disk and where w_0 is a point on the circumference of the disk. The formula A would be $w_0 \neq w_1$. We define: $\langle U, A\vec{w}, \tau\vec{w}, V \rangle$ is an interpretation iff $V \vdash \exists \vec{w} \ A\vec{w}$ and, for any axiom B of U, we have $V \vdash \forall \vec{w} \ (A\vec{w} \to B^{\tau\vec{w}})$.

A.6. Piecewise Interpretations. The idea of piecewise interpretations[9] is that we can develop the domains for each sort as built up from a finite number of pieces. These pieces may overlap. The same original object may pose as two different objects depending on which piece we are considering. The pieces also may be of different sorts and of different dimensionality.

For example, let us define a one dimensional, parameter free translation from a one sorted language to a one sorted language with two pieces, for a theory with just one binary predicate P. Our translation τ provides two domain pieces δ_0 and δ_1. The δ_i are formulas of the target language containing just v_0 free. We have a function F that assigns formulas $A_{ij}(v_0, v_1)$ to P and 0,1-pairs ij. Similarly for the identity. Our translation function has as inputs formulas B of source language and assignments α of 0 or 1 to the free variables of B. We give the clauses for Pxy, conjunction and universal quantification.

- $(Pxy)^{\tau,\alpha} := A_{\alpha(x),\alpha(y)}(x,y)$, where α has domain $\{x,y\}$;
- $(B \wedge C)^{\tau,\alpha} := B^{\tau,\alpha \restriction \mathsf{FV}(B)} \wedge C^{\tau,\alpha \restriction \mathsf{FV}(C)}$, where α has domain $\mathsf{FV}(B \wedge C)$;

---

[9]I learned the idea of piecewise interpretability in a slightly less general form from Harvey Friedman.

- $(\forall x\ B)^{\tau,\alpha} := \forall x{:}\delta_0\ (B)^{\tau,\alpha\cup\{\langle x,0\rangle\}} \wedge \forall x{:}\delta_1\ (B)^{\tau,\alpha\cup\{\langle x,1\rangle\}}$, where α has domain $\mathsf{FV}(\forall x\ B)$.

Note that τ will be the empty assignment for sentences. The rest of the development is as expected. Note that any theory with a finite model has a piecewise, one-dimensional, parameter free interpretation in EQ, the one-sorted theory of equality.

A natural example of a piecewise interpretation is 'adding the unit to the theory of semigroups'.

Remark A.1. Prima facie, piecewise translations are rather costly. Translations that are not piecewise yield p-time transformations of formulas (if we handle the needed alpha-conversions in a sufficiently smart way). Piecewise translations, on the other hand may be exponential.

Fortunately, in rather general circumstances piecewise translations can be eliminated. E.g., if the target theory T, i.e., the interpreting theory T, does not have a one-element model, then a piecewise translation can always be replaced by a multi-dimensional interpretation with parameters that is 'the same' in the sense that there is a T-definable, T-verifiable isomorphism between these translations (considered as inner model constructions).   ❏

Remark A.2. Consider the well-known flattening construction, in which we transform a many-sorted theory U into a one-sorted theory $U^\flat$. We have that $U \rhd U^\flat$, by a simple one dimensional, non-piecewise, interpretation. We have $U \rhd U^\flat$, via a piecewise interpretation. (If U proves that every domain has at least two elements, we can replace the use of pieces by more dimensionality.)   ❏

DEPARTMENT OF PHILOSOPHY, UTRECHT UNIVERSITY, HEIDELBERGLAAN 8, 3584 CS UTRECHT, THE NETHERLANDS

E-mail address: `Albert.Visser@phil.uu.nl`