No Escape from Vardanyan's Theorem

Albert Visser

Department of Philosophy, Utrecht University

Heidelberglaan 8, 3584 CS Utrecht, The Netherlands

email: `Albert.Visser@phil.uu.nl`

April 3, 2003

Abstract

Vardanyan's Theorem states that the set of PA-valid principles of Quantified Modal Logic, QML, is complete Π^0_2. We generalize this result to a wide class of theories. The crucial step in the generalization is avoiding the use of Tennenbaum's Theorem.

Key words: Predicate Provability Logic

MSC2000 codes: 03B25, 03F45

1 Introduction

The idea of assigning structures of kind Y to structures of kind X other in order to obtain information is a useful methodology in mathematics. The structures of kind Y are often simpler and better understood. Provability logic is a case in point: we assign to arithmetical theories certain propositional modal logics that are in many respects simpler than the original theories. These modal logics give us full information about arithmetical reasoning of a certain restricted sort. A disadvantage of this approach is that provability logic does not yield *differential* information: the modal logic assigned is the same for a wide class of reasonable theories. The situation becomes a bit better when we expand the modal language. One possible extension is to a propositional modal language with a binary predicate for relative interpretability. In this case salient classes of theories receive different but simple logics. See e.g. [JdJ98] and [Vis98]. Another possible expansion is to a predicate logical modal language. Here we certainly do have differential information about the original theories. Vardanyan's theorem concerns this situation. Vardanyan's Theorem tells us that the set of PA-valid principles of Quantified Modal Logic, QML, is complete Π^0_2. Thus, the PA-valid modal principles are more complicated than the theory PA itself. Clearly, the result can be viewed as a negative result for the case of PA.

Vardanyan's Theorem was discovered by V.A. Vardanyan in 1985. See [Var86]. Vardanyan's work used earlier ideas of V.E. Plisko (see [Pli77], [Pli78], [Pli83], [Pli93]) and S. Artemov (see [Art85]). Vardanyan's Theorem was discovered independently but slightly later by Vann McGee. The first published version of McGee's result is in his 1985 PhD Thesis. We will generalize Vardanyan's result to a wide class of arithmetical theories, containing all theories in the language of arithmetic that extend Elementary Arithmetic EA, ie. $I\Delta_0 + \mathsf{Exp}$.[1]

Let me point out that if we just consider the valid schemes for (non-modal) predicate logic, then, for a large class of theories, we get no specific information: e.g. for Σ^0_1-sound arithmetical theories you get precisely predicate logic. However, if you look at non-arithmetical theories there are many possibilities. See [Yav97]. It has been shown that there is a Σ^0_1-*unsound* arithmetical theory such that the set of its predicate logical valid principles is complete Π^0_2. See [Vis02].

the main theorem of the present paper extends Vardanyan's Theorem to a wide range of theories. Here is a first crude statement of our theorem.

Theorem 1.1 *Consider a theory T in the language of arithmetic. We assume that the axiom-set of T is simple, e.g. p-time decidable. If T is a Σ^0_1-sound extension of* EA*, then the set of T-valid* QML*-principles is complete Π^0_2.*

Closer inspection of the argument gives us a more refined, but less memorable statement.

Theorem 1.2 *Consider a theory T in the language of arithmetic. We assume that the axiom-set of T is simple, e.g. p-time decidable. Suppose (i) that T is an extension of* EA *such that $T + \mathsf{Con}^2(T)$ is consistent or (ii) that T is an finite extension of S^1_2 or $I\Delta_0 + \Omega_1$ such that T does not prove an iterated inconsistency statement $\mathsf{Bew}^n_T[\bot]$. Then, the set of T-valid* QML*-principles is complete Π^0_2.*

Acknowledgements

I thank Maartje de Jonge and Lev Beklemishev for enlightening conversations.

2 Prerequisites, Notations and Conventions

We will assume that the reader is familiar with provability logic and, in particular, with Boolos' beautiful presentation of Vardanyan's Theorem in the textbook [Boo93]. We will mostly follow Boolos' notational conventions. We will diverge at one point. Boolos writes $\{A\}$ for the translation of the arithmetical language $\mathcal{A}$ to a corresponding relational language $\mathcal{L}$ with predicates Z, S, A, M and E (for *identity*), which are resp. 1,2,3,3,2-ary. We will write $[\![A]\!]$ instead.

Section 6 requires some knowledge of weak theories. See e.g. [Bus86] or [HP91].

[1] We always assume that theories are axiomatized by an axiom-set that is simple, say p-time decidable.

3 The Basic Form of the Argument

In this section, we present the basic scheme of the argument. As given here the argument, and especially the soundness part of it, can still be improved. We will do this in section 5.

Let an arithmetical theory T be given. We demand that T is sufficiently strong and sufficiently sound. The amount of strongness and soundness needed is dependent on the further details of the proof. We will specify what is needed later. What we are looking for is a reduction of the truth of Π^0_2-sentences $P := \forall x \exists y\, P_0(x,y)$, for $P_0 \in \Delta_0$, to the T-valid principles of QML. We will, in an effective way, associate a suitable modal sentence $\tilde{P}$ to P such that P is true iff $\tilde{P}$ is T-valid.

The modal language $\mathcal{L}_\Box$ that we will employ, extends the relational variant $\mathcal{L}$ of the language of arithmetic which contains the predicates Z, S, A, M and E (for identity), which are resp. 1,2,3,3,2-ary. There will, of course, be the modal operator $\Box$, but also an additional predicate X which in the usual approaches is unary and in our approach is 0-ary. The general proof strategy is as follows: we have to find $\mathcal{L}_\Box$-formulas A and $B(x,X)$ with the following properties:

† A is in $\mathcal{L}$. We demand that A implies (in predicate logic) $[\![\mathsf{Q}]\!]$. Here Q is the single sentence axiomatization of Robinson's Arithmetic. Let e be the interpretation of $\mathcal{L}$ that sends $\mathsf{S}(x,y)$ to $\mathsf{S}x = y$, etc. We demand that $T + A^{\mathsf{e}}$ is sufficiently sound. We will specify later how much soundness is needed.

‡ Define $\nu_0(z) := \mathsf{Z}(z)$ and $\nu_{n+1}(z) := \exists u\, (\nu_n(u) \wedge \mathsf{S}(u,z))$.[2] We demand that, for every $(\cdot)^\star$, there is an n such that:

$$T + A^\star \vdash (\exists z\, (\nu_n(z) \wedge B(z,X)))^\star.$$

£ A certain class of formulas Γ is specified, such that, for every i, there is a T-formula C (of the arity of X) in Γ such that if $(\exists z(\nu_n(z) \wedge B(z,X))^{\mathsf{e}[X:=C]}$ is true, then $n \geq i$. Here $\mathsf{e}[X := C]$ has the obvious meaning: it is e extended to an interpretation that translates X to C.

We take as $\tilde{P}$ the formula: $(A \to \exists z\, (B(z,X) \wedge [\![\forall x{<}z\, \exists y\, P_0(x,y)]\!]))$. We reason as follows.

Completeness: Suppose P is true. Consider any $(\cdot)^\star$. By (‡), there is an n such that (a): $T \vdash (A \to \exists z\, (\nu_n(z) \wedge B(z)))^\star$. Clearly, $\forall x{<}n\, \exists y\, P_0(x,y)$. Hence, $\mathsf{Q} \vdash \forall x{<}\mathbf{n}\, \exists y\, P_0(x,y)$. Using (†), we can transform the Q-proof witnessing this last fact to a T-proof witnessing (b):

$$T \vdash (A \to \exists z\, (\nu_n(z) \wedge [\![\forall x{<}z\, \exists y\, P_0(x,y)]\!]))^\star.$$

Combining (a) and (b), we find: $T \vdash \tilde{P}^\star$.

[2]Note that $\nu_n(z)$ is equivalent over $[\![\mathsf{Q}]\!]$ to $[\![\mathbf{n} = z]\!]$. The reason that we explicitely use ν_n is to avoid scope confusions.

Soundness: Conversely, suppose $\tilde{P}$ is T-valid. Let i be an arbitrary number. Let C be the formula in Γ promised for i by (£). Consider the interpretation $\mathsf{e}_C := \mathsf{e}[X := C]$. We find:

$$T \vdash (A \rightarrow \exists z\,(B(z,X) \wedge [\![\forall x{<}z\,\exists y\,P_0(x,y)]\!]))^{\mathsf{e}_C}.$$

Hence, $T + A^{\mathsf{e}} \vdash \exists z\,(B^{\mathsf{e}_C}(z,X) \wedge \forall x{<}z\,\exists y\,P_0(x,y))$. We assume that $T + A^{\mathsf{e}}$ is sound for sentences of the form $\exists z\,(B^{\mathsf{e}_C}(z,X) \wedge \forall x{<}z\,\exists y\,P_0(x,y))$ for $C \in \Gamma$. It follows that $\forall x{<}i\,\exists y\,P_0(x,y)$. The number i was arbitrary, hence we find that P is true.

4 The Magic Formulas A and B

In this section we discuss various ways of constructing A and B. In subsection 4.2, we prove theorem 1.1.

4.1 Plisko and Vardanyan

The formula A associated to Vardanyan's work is the formula $[\![\mathsf{B}]\!] \wedge D \wedge E$. Here '$\mathsf{B}$' is the formula Boolos calls 'T' and D and E are as in Boolos' presentation. We have: $T + \mathsf{Con}(T) \vdash ([\![\mathsf{B}]\!] \wedge D \wedge E)^{\mathsf{e}}$.

Plisko used for B a formula expressing 'X is not a truthpredicate for sentences of complexity x'. Here we may take as complexity measure the one corresponding to the arithmetical hierarchy. One nice aspect of this approach is that B is entirely boxfree.[3] The unary formulas C provided by (£) will be truthpredicates for complexity i. To make this strategy work we need make A so strong that modulo provable equivalence $A^\star$ implies that $(\cdot)^\star$ restricted to the arithmetical part of the language is equivalent to e. To do this, we need to apply an arithmetized version of Tennenbaum's Theorem. The application of Tennenbaum's Theorem will need at least that T contains $I\Sigma_1$. A detailed verification of this fact will be given in Maartje de Jonge's forthcoming paper.

Note that, with Plisko's B, we need C of arbitrary complexity, so a reasonable choice of Γ will be the set of all formulas of the arithmetical language. Thus, if we follow the proof outlined in section 3, we will need true extensions of $I\Sigma_1$ in the arithmetical language. Closer inspection of the argument will reveal that a slightly adapted version works for extensions T of $I\Sigma_1$ such $T + \mathsf{Con}(T)$ is consistent. See also Maartje de Jonge's forthcoming paper.

The B used in Boolos' presentation of Vardanyan's argument is roughly 'x is a recursive index of the set $\{y \mid \Box X(y)\}$'. Using the argument of section 3 this argument would work for Σ^0_3-sound extensions of $I\Sigma_1$. However, one can refine the argument in such a way that the weaker demand that $T + \mathsf{Con}(T)$ is consistent, is sufficient. See also Maartje de Jonge's forthcoming paper.

[3] Plisko developed his beautiful idea for the study of interpretations of intuitionistic predicate logic. So he did have no boxes to play with.

4.2 Our Choice for A and B

We want to find A and B that work for a wide range of theories. It turns out that the wonderful idea of using an arithmetized version of Tennenbaum's theorem is the main obstacle to progress. In the construction below, we simply avoid the need to use of Tennenbaum's Theorem.

We demand that T is a Σ^0_1-sound extension of elementary arithmetic EA, also known as $I\Delta_0 + \mathsf{Exp}$. Here Exp is the axiom stating that exponentiation is total. Later we will show how to improve this condition: the result holds also for certain weak arithmetics and we need less soundness. Here is a brief description of EA.

1. EA is finitely axiomatizable. We will employ a finite axiomatization of EA.
2. The provably recursive functions of EA are precisely the elementary functions.
3. EA implies Σ^0_1-completeness for theories extending the theory R of Tarski, Mostowski and Robinson. In fact EA can formalize the standard argument for Σ^0_1-completeness.
4. The usual arithmetization of syntax is possible in EA. In addition, all Rosser style arguments involving witness-comparisons of Σ^0_1-sentences can be formalized in EA.
5. EA has a Σ^0_1 truthpredicate $\mathsf{True}_{\Sigma^0_1}$ for Σ^0_1-sentences.
6. EA does *not* prove cutelimination for predicate logic, which requires superexponentiation.
7. EA does *not* prove Σ^0_1-collection:

$$\vdash \forall x{<}a\, \exists y\, D_0(x,y) \to \exists b\, \forall x{<}a\, \exists y{<}b\, D_0(x,y),$$

   where $D_0 \in \Delta^0_1$.

A suitable finite axiomatization of the axioms of identity is supposed to be incorporated in EA. We take X 0-ary. To improve readability, we write $\boxdot G$ for $G \wedge \Box G$. Here are our A and B:

- $A := \boxdot([\![\mathsf{EA}]\!] \wedge \forall y\, (\mathsf{Z}(y) \to \Box \mathsf{Z}(y)) \wedge \forall y \forall z\, (\mathsf{S}(y,z) \to \Box \mathsf{S}(y,z)))$,
- $B(x,X) := \Box(\Box X \leftrightarrow \Box [\![\mathsf{True}_{\Sigma^0_1}(x)]\!])$.

To verify the satisfaction of (†), note that $T \vdash A^{\mathsf{e}}$. So, all we need is that T is sufficiently sound. Note that, for any C, $\exists z\, (B^{\mathsf{e}_C}(z,X) \wedge \forall x{<}z\, \exists y\, P_0(x,y))$ is Σ^0_1. Hence, we will only need that our theory is Σ^0_1-sound. We will improve this estimate in section 5.

We verify (‡). Let $G := \exists x\, G_0(x)$ and $H := \exists y\, H_0(y)$, for arbitrary G_0 and H_0. We write:

- $G \leq H :\leftrightarrow \exists x\, (G_0(x) \wedge \forall y{<}x\, \neg H_0(y))$.
- $G < H :\leftrightarrow \exists x\, (G_0(x) \wedge \forall y{\leq}x\, \neg H_0(y))$.
- $(G \leq H)^\perp :\leftrightarrow H < G$ and $(G < H)^\perp :\leftrightarrow H \leq G$.

Consider any $(\cdot)^\star$. By the Gödel Fixed Point Lemma, we may find a Σ^0_1-sentence R such that

$$T \vdash R \leftrightarrow \mathsf{Bew}_T[X^\star] \leq \mathsf{Bew}_T[[\![R]\!]^\star]. \tag{1}$$

We interpolate two lemmas.

Lemma 4.1 For any Σ^0_1-sentence S, we have: $T + A^\star \vdash S \rightarrow \mathsf{Bew}_T[[\![S]\!]^\star]$. ❑

Proof

Reason in $T + A^\star$. Given S we can find a Q-proof π of S. The proof π can be transformed into a T-proof π' of $[\![\mathsf{Q} \rightarrow S]\!]^\star$. (In fact this transformation is p-time.) By $A^\star$, there is a T-proof τ of $[\![\mathsf{Q}]\!]^\star$. Combining π' with τ, we find a T-proof τ' of $[\![S]\!]^\star$. ❑

The next lemma is a minor variation on a consequence of the Friedman-Goldfarb-Harrington Theorem. See [Vis02] for a detailed discussion of this theorem.

Lemma 4.2 We have: $T + A^\star \vdash \mathsf{Bew}_T[X^\star] \leftrightarrow \mathsf{Bew}_T[[\![R]\!]^\star]$. ❑

Proof

Reason in $T + A^\star$. Suppose $\mathsf{Bew}_T[X^\star]$. It follows, by Δ_0-induction, that:

$$\mathsf{Bew}_T[X^\star] \leq \mathsf{Bew}_T[[\![R]\!]^\star] \text{ or } \mathsf{Bew}_T[[\![R]\!]^\star] < \mathsf{Bew}_T[X^\star].$$

In the first case, we have R. Hence, by lemma 4.1, $\mathsf{Bew}_T[[\![R]\!]^\star]$. In the second case, we have $\mathsf{Bew}_T[[\![R]\!]^\star]$ immediately.

For the converse, suppose $\mathsf{Bew}_T[[\![R]\!]^\star]$. It follows that:

$$\mathsf{Bew}_T[X^\star] \leq \mathsf{Bew}_T[[\![R]\!]^\star] \text{ or } \mathsf{Bew}_T[[\![R]\!]^\star] < \mathsf{Bew}_T[X^\star].$$

In the first case, we have $\mathsf{Bew}_T[X^\star]$. In the second case, we have, by lemma 4.1,

$$\mathsf{Bew}_T[\,[\![\mathsf{Bew}_T[\,[\![R]\!]^\star] < \mathsf{Bew}_T[X^\star]]\!]^\star\,].$$

In other words, we have $\mathsf{Bew}_T[[\![R^\perp]\!]^\star]$. In combination with our assumption that $\mathsf{Bew}_T[[\![R]\!]^\star]$, this gives us $\mathsf{Bew}_T[\perp]$. Hence, a fortiori, $\mathsf{Bew}_T[X^\star]$. ❑

Remark 4.3 One might be tempted to reason, in $T + A^\star$, as follows. Let U be the theory axiomatized by $\{E \in \mathsf{sent}_\mathcal{L} \mid \mathsf{Bew}_T[E^\star]\}$. This is a theory in predicate logic extending Q, hence we may apply the Friedman-Goldfarb-Harrington Theorem to U. The fallacy in this argument is a silent application of Σ^0_1-collection. The provability predicate of U will not be verifiably Σ^0_1. Of course, there are many ways to repair the argument, however, given the fact that the proofs of our two lemmas are quite short, why bother? ❏

Lemma 4.2 gives us:

$$T + A^\star \vdash \mathsf{Bew}_T[X^\star] \leftrightarrow \mathsf{Bew}_T[[\![R]\!]^\star]. \tag{2}$$

Suppose r is the Gödelnumber of R. We have: $\mathsf{EA} \vdash R \leftrightarrow \mathsf{True}_{\Sigma^0_1}(\mathbf{r})$, and, hence, $\mathsf{EA} \vdash R \leftrightarrow \exists z\,(z = \mathbf{r} \wedge \mathsf{True}_{\Sigma^0_1}(z))$. We can transform the witnessing EA-proof of this last fact into a $(T + A^\star)$-proof, thus obtaining:

$$T + A^\star \vdash (\ [\![R]\!] \leftrightarrow \exists z\,(\nu_r(z) \wedge [\![\mathsf{True}_{\Sigma^0_1}(z)]\!])\ )^\star. \tag{3}$$

Using the fact that $T + A^\star \vdash \mathsf{Bew}_T[A^\star]$, we now obtain:

$$T + A^\star \vdash \mathsf{Bew}_T[\ \mathsf{Bew}_T[[\![R]\!]^\star] \leftrightarrow \mathsf{Bew}_T[(\exists z\,(\nu_r(z) \wedge [\![\mathsf{True}_{\Sigma^0_1}(z)]\!]))^\star]\ ]. \tag{4}$$

We now want to 'export' the $\exists z\,(\nu^\star(z) \wedge \ldots$ to the outside of the $\mathsf{Bew}[\ldots]$'s. We indicate the obvious steps. Evidently, we have:

$$T + A^\star \vdash (\forall z \forall u\,((\nu_r(z) \wedge \nu_r(u)) \to \mathsf{E}(z,u)))^\star. \tag{5}$$

Moreover,

$$T + A^\star \vdash \forall z\,(\nu_r^\star(z) \to \mathsf{Bew}_T[\nu_r^\star(z)]). \tag{6}$$

Clearly, $T + A^\star \vdash \exists z\,\nu^\star(z)$. By equation (6) and the fact that $T + A^\star \vdash \mathsf{Bew}_T[A^\star]$, we obtain: $T + A^\star \vdash \exists z\,(\nu^\star(z) \wedge \mathsf{Bew}_T[\ \mathsf{Bew}_T[\nu_r^\star(z)]\ ])$. Hence, using equation (5), we find: $T + A^\star \vdash \exists z\,(\nu^\star(z) \wedge \mathsf{Bew}_T[\ \mathsf{Bew}_T[\forall v\,(\nu_r^\star(v) \to \mathsf{E}^\star(v,z))]\ ])$. Combining this with equation (4), using the fact that $A^\star$ implies that $\mathsf{E}^\star$ is a congruence w.r.t. the relations of $\mathcal{L}$, we find:

$$T + A^\star \vdash \exists z\,(\ \nu_r^\star(z) \wedge \mathsf{Bew}_T[\ \mathsf{Bew}_T[X^\star] \leftrightarrow \mathsf{Bew}_T[[\![\mathsf{True}_{\Sigma^0_1}(z)]\!]^\star)]\ ]\ ) \tag{7}$$

Clearly equation (7) gives us the promised $T \vdash (A \to \exists z\,(\nu_r(z) \wedge B(z,X)))^\star$.

We proceed to derive (£). We define, for any arithmetical formula G,

- $\mathsf{Bew}^0_T[G] :\leftrightarrow G$,
- $\mathsf{Bew}^{n+1}_T[G] :\leftrightarrow \mathsf{Bew}_T[\mathsf{Bew}^n_T[G]]$.

Consider any natural number i. Suppose that for every C there would be an $n < i$ such that $(\exists z\,(\nu_n(z) \wedge B(z,X))^{\mathsf{e}[X:=C]}$ is true. It would follow, by the Pidgeon Hole Principle, that for some n, m and k with $m > k > 0$, we would have:

$$T \vdash \mathsf{Bew}^m_T[\bot] \leftrightarrow \mathsf{Bew}_T[\mathsf{True}_{\Sigma^0_1}(\mathbf{n})] \text{ and } T \vdash \mathsf{Bew}^k_T[\bot] \leftrightarrow \mathsf{Bew}_T[\mathsf{True}_{\Sigma^0_1}(\mathbf{n})].$$

Ergo, $T \vdash \mathsf{Bew}^m_T[\bot] \leftrightarrow \mathsf{Bew}^k_T[\bot]$. Hence, by Löb's Principle, $T \vdash \mathsf{Bew}^k_T[\bot]$. Quod non, by Σ^0_1-soundness.

5 Sharpening the Soundness Condition

How sound should our theories T be? We will look into the matter by refining the completeness part of the argument in section 3 a bit.

It seems fair to me to demand that the theories we consider satisfy the demand that $T + \mathsf{Con}(T)$ is consistent. If it is not, the modal part of the language trivializes and we are really looking at the non-modal predicate logics of theories.

Let's briefly look at the excluded case that $T + \mathsf{Con}(T)$ is inconsistent, i.o.w. that $T \vdash \mathsf{Bew}_T[\bot]$. Clearly, the question of complexity reduces to the question whether the ordinary, non-modal, predicate logic of T is complete Π^0_2. Well, anything is possible. For example, if T is finitely axiomatized, the logic of $T + \mathsf{Bew}_T[\bot]$ is precisely predicate logic plus the axioms $\Box A$ for all A. This is an immediate consequence of the results of [Vis02]. On the other hand, the results of [Vis02] show that *there is* a consistent arithmetical extension of, say, $I\Sigma_1 + \mathsf{Bew}_{I\Sigma_1}[\bot]$ such that the set of valid predicate logical principles is complete Π^0_1.

As will be shown in Maartje de Jonge's forthcoming paper, a refinement of the original Vardanyan argument works for extensions T of $I\Sigma_1$ such that $T + \mathsf{Con}(T)$ is consistent. Thus, w.r.t. soundness the original result is, in a sense, optimal. The result of this paper asks for less strength. However, we will need a bit more soundness: $T + \mathsf{Con}^2(T)$ has to be consistent.

Suppose T is a theory in the language of arithmetic, extending EA. Suppose further that $T^+ := T + \mathsf{Con}^2(T)$ is consistent. We will show that the set of T-valid QML-sentences is complete Π^0_2. This proves (i) of theorem 1.2.

Consider $P := \forall x \exists y\, P_0(x,y)$, where $P_0 \in \Delta^0_1$. Let $P_1(x) := \exists y\, P_0(x,y)$. By the Gödel Fixed Point Lemma we may find $Q_1(x) \in \Delta^0_1$ such that

$$\mathsf{Q} \vdash Q_1(x) \leftrightarrow P_1(x) \leq \mathsf{Bew}_{T^+}[Q_1(x)].$$

By the FGH theorem (see[Vis02]; see also lemma 4.2), we have, for any n, $P_1(n) \Leftrightarrow T^+ \vdash Q_1(\mathbf{n})$. Suppose $Q_1(x) := \exists y\, Q_0(x,y)$, where $Q_0(x,y) \in \Delta^0_1$. Let $Q := \forall x\, \exists y\, Q_0(x,y)$. Again, by the FGH theorem, we have that P is true iff Q is true.

Now we repeat the argument of section 3 for T with Q in the role of P. The *completeness* part remains the same. We sharpen the *soundness* part as follows. Suppose that $\tilde{Q}$ is T-valid. Specializing this to the e_C, we obtain, for any C:

$$T \vdash \exists z\, (\, \mathsf{Bew}_T[\, \mathsf{Bew}_T[C] \leftrightarrow \mathsf{Bew}_T[\mathsf{True}_{\Sigma^0_1}(z)]\,] \wedge \forall x{<}z\, \exists y\, Q_0(x,y)\,).$$

Consider some number N. We have:

$$T + \exists x{<}\mathbf{N}\, \forall y\, \neg Q_0(x,y) \vdash \exists z{<}\mathbf{N}\, \mathsf{Bew}_T[\, \mathsf{Bew}_T[C] \leftrightarrow \mathsf{Bew}_T[\mathsf{True}_{\Sigma^0_1}(z)]\,].$$

Hence:

$$T + \exists x{<}\mathbf{N}\, \forall y\, \neg Q_0(x,y) \vdash \bigvee_{k<N} \mathsf{Bew}_T[\, \mathsf{Bew}_T[C] \leftrightarrow \mathsf{Bew}_T[\mathsf{True}_{\Sigma^0_1}(\mathbf{k})]\,].$$

Consider $\omega \cup \{\infty\}$ with the obvious ordering. We construct the following Kripke frame $\mathcal{F}$ on $\omega \cup \{\infty\}$:

$$n \prec m \Leftrightarrow (m \leq N \text{ and } n = m + N + 1) \text{ or } (n > 2N + 1 \text{ and } m < n).$$

The frame specified is the frame of a tail-model (see [Vis84]). By the result of [Vis84], we can embed the algebra of propositions given by the finite and the co-finite sets of nodes into T using a Solovay function h.[4] The statements $S_n :\leftrightarrow \exists u\, hu = \mathbf{n}$ for $n \leq N$ represent the propositions $p_n := \{n\}$. Clearly, for $n, m \leq N$ with $n \neq m$, we have $\mathcal{F} \models \Box(\Box p_n \leftrightarrow \Box p_m) \to \Box\Box\bot$. Hence, by the embedding lemma,

$$T \vdash \mathsf{Bew}_T[\,\mathsf{Bew}_T[S_n] \leftrightarrow \mathsf{Bew}_T[S_m]\,] \to \mathsf{Bew}^2_T[\bot].$$

By elementary reasoning involving the Pigheon Hole Principle we find:

$$T + \exists x{<}\mathbf{N}\, \forall y\, \neg Q_0(x, y) \vdash \bigvee_{m,n \leq \mathbf{N},\ m \neq n} \mathsf{Bew}_T[\,\mathsf{Bew}_T[S_n] \leftrightarrow \mathsf{Bew}_T[S_m]\,].$$

Ergo: $T + \exists x{<}\mathbf{N}\, \forall y\, \neg Q_0(x, y) \vdash \mathsf{Bew}^2_T[\bot]$. Contraposing, we find:

$$T + \mathsf{Con}^2(T) \vdash \forall x{<}\mathbf{N}\, \exists y\, Q_0(x, y).$$

Since N was arbitrary, we may conclude that, for any n, $T^+ \vdash Q_1(\mathbf{n})$. By the FGH property, it follows that P is true. Hence, P is true iff Q is true iff $\tilde{Q}$ is T-valid.

Open Question 5.1 Is there a theory T in the language of arithmetic that extends EA such that $T + \mathsf{Con}(T)$ is consistent, for which Vardanyan's Theorem does not hold? ❑

A curious aspect of our argument is that I do not know how to produce the analogues of the S_i for S^1_2. See [BV93] for an extensive discussion of the problems involved. However, if we use the sentences $\mathsf{Bew}^i_T[\bot]$, we still get a minor improvement. By the above considerations and by the discussion in section 6, we find that Vardanyan's Theorem holds when T is a finite extension of either S^1_2 or $I\Delta_0 + \Omega_1$ such that T does not prove any iterated inconsistency statement $\mathsf{Bew}^n_T[\bot]$.

6 Adapting the Argument to Weak Theories

We show how to adapt our argument to certain extensions of Buss' theory S^1_2. Regretably the full description of the reasoning is rather long, so we will just sketch the main ingredients of the adaptation.

[4] Alternatively, we could have used the much stronger embedding theorem due to V. Shavrukov. See [Sha93]. For a simplified proof, see [Zam94]. Using this result we could even have produced infinitely many S_i with the desired property.

The results of this section in combination with the considerations of the previous section will give us (ii) of theorem 1.2. Here is a brief description of S^1_2. For a detailed discussion, see [Bus86] or [HP91].

1. S^1_2 is finitely axiomatizable. We employ a finite axiomatization of S^1_2 in our argument.
2. The provably recursive functions of S^1_2 are precisely the p-time computable functions.
3. Arithmetization of syntax can be executed without any problems in S^1_2.
4. We do not know how to formalize Rosser style arguments in S^1_2. However, some Rosser style theorems, such as Rosser's original theorem, can be formalized using a different argument. See subsection 6.3.

The theories for which we can also prove Vardanyan's Theorem are the theories in the arithmetical language that are either finite extensions of S^1_2 or finite extensions of the stronger theory $I\Delta_0 + \Omega_1$. We will describe below the points where we have to change the argument. We will aim at adapting our argument, replacing EA by S^1_2 everywhere.

6.1 Efficient Numerals

The definition of quantifying in involves the num-function which sends a number to the Gödelnumber of its numeral. This function will, when naively formalized, be of exponential growth. However, we do not have exponentiation in weak theories. The usual way to solve this problem is to employ dyadic numerals given by the mapping BNUM:

- $\mathsf{BNUM}(0) := 0$,
- $\mathsf{BNUM}(2n + 1) := \mathsf{S}(\mathsf{BNUM}(n) \times \mathsf{SS}0)$,
- $\mathsf{BNUM}(2n + 2) := \mathsf{SS}(\mathsf{BNUM}(n) \times \mathsf{SS}0)$.

Let bnum be the function sending n to the Gödel number of $\mathsf{BNUM}(n)$. Under a reasonable coding scheme the bnum function is polynomial. We replace the use of num, by the use of bnum.

We might wish also to replace the predicates ν_n with predicates β_n corresponding to efficient numerals. For our specific argument this is not necessary, since we are only using ν_n for standard n.

6.2 Truth Predicate

A somewhat more serious business is the fact that S^1_2 cannot verify the properties of the Σ^0_1-truth predicate. There are two ways to get around thus restriction. The first is by noting that our sentences R_C have the specific form:

$$\mathsf{Bew}_T[C] \leq \mathsf{Bew}_T[R_C].$$

Now take as measure of complexity ρ which counts just depth of quantifier changes. Note that all sentences of the form R_C have complexity below K, for some standard K. We can build a truth predicate True such that, for all sentences H of complexity $\leq K$, we have: $\mathsf{S}^1_2 \vdash H \leftrightarrow \mathsf{True}(\mathbf{h})$, where h is the Gödel number of H. (See fact 2.4.4 of [Vis93].) Now we can use True instead of $\mathsf{True}_{\Sigma^0_1}$ in our argument.

The alternative strategy is as follows. We can find an S^1_2-definable cut I and a predicate $\mathsf{TRUE}_{\Sigma^0_1}$ such that, for any Σ^0_1-sentence S, we have $\mathsf{S}^1_2 \vdash S^I \leftrightarrow \mathsf{True}^*_{\Sigma^0_1}(\mathbf{s})$, where s is the Gödel number of S. Now we use the alternative fixed point Z_C such that $\mathsf{Q} \vdash Z_C \leftrightarrow \mathsf{Bew}_T[C] \leq \mathsf{Bew}_T[[\![Z^I_C]\!]^\star]$.

6.3 Adapting the FGH Theorem

We have to find an alternative proof for: $T + A^\star \vdash \mathsf{Bew}_T[C] \leftrightarrow \mathsf{Bew}_T[[\![R]\!]^\star]$. This is one of the corollaries of the FGH theorem. Our earlier proof used the principles

$$T + A^\star \vdash R \rightarrow \mathsf{Bew}_T[[\![R]\!]^\star] \text{ and } T \vdash R^\perp \rightarrow \mathsf{Bew}_T[[\![R^\perp]\!]^\star].$$

However, we only know how to prove these facts in the presence of Exp as part of T. Also we used:

$$T + A^\star \vdash (\mathsf{Bew}_T[C] \vee \mathsf{Bew}_T[[\![R]\!]^\star]) \rightarrow (R \vee R^\perp).$$

This fact can only be verified if T extends $I\Delta_0 + \Omega_1$ (as far as we know). Fortunately, for a number of theories, there is another road. We use Švejdar's Principle.[5] Inspecting the main proof of [VV94], one can show, for any finitely axiomatized theory U extending S^1_2 (or even a weaker theory like Q), or for any extension U by finitely many axioms of $I\Delta_0 + \Omega_1$, that:

Švejdar's Principle: $\mathsf{S}^1_2 \vdash \mathsf{Bew}_U[G] \rightarrow \mathsf{Bew}_U[\,\mathsf{Bew}_U[H] \leq \mathsf{Bew}_U[G] \rightarrow H\,]$.

The argument uses the fact that inside Bew_U in the consequent, we will have a U-proof of G in any definable U-cut and, hence, a U-proof of H in any definable U-cut. We can adapt the proof of the Švejdar's Principle to our context to obtain:

- $T + A^\star \vdash \mathsf{Bew}_T[G] \rightarrow \mathsf{Bew}_T[\,[\![\mathsf{Bew}_T[H] \leq \mathsf{Bew}_T[G]]\!]^\star \rightarrow H\,]$

The reason that this adaptation works is that for the original argument it is irrelevant what the interpretation of number theory employed for the formalization of syntax inside U is. Also we have:

- $T + A^\star \vdash \mathsf{Bew}_T[C] \rightarrow \mathsf{Bew}_T[[\![R \vee R^\perp]\!]^\star]$.

The argument for this fact is similar to the proof of fact 2.7 of [Vis02].

[5] First formulated in [Šve83].

Using Švejdar's Principle, we cannot recover the full FGH principle, but we can get the principle we need. We reproduce the argument for the case we are interested in. We have:

$$
\begin{aligned}
T + A^\star \vdash \ \mathsf{Bew}_T[\llbracket R \rrbracket^\star] \ &\rightarrow \ \mathsf{Bew}_T[\llbracket R \rrbracket^\star] \wedge \\
& \ \mathsf{Bew}_T[\,\llbracket \mathsf{Bew}_T[C] \leq \mathsf{Bew}_T[\llbracket R \rrbracket^\star] \rrbracket^\star \rightarrow C\,] \\
&\rightarrow \ \mathsf{Bew}_T[\llbracket R \rrbracket^\star] \wedge \mathsf{Bew}_T[\llbracket R \rrbracket^\star \rightarrow C] \\
&\rightarrow \ \mathsf{Bew}_T[C]
\end{aligned}
$$

$$
\begin{aligned}
T + A^\star \vdash \ \mathsf{Bew}_T[C] \ &\rightarrow \ \mathsf{Bew}_T[C] \wedge \\
& \ \mathsf{Bew}_T[\,\llbracket \mathsf{Bew}_T[\llbracket R \rrbracket^\star] < \mathsf{Bew}_T[C] \rrbracket^\star \rightarrow \llbracket R \rrbracket^\star\,] \\
&\rightarrow \ \mathsf{Bew}_T[\llbracket R \vee R^\perp \rrbracket^\star] \wedge \mathsf{Bew}_T[\llbracket R^\perp \rightarrow R \rrbracket^\star] \\
&\rightarrow \ \mathsf{Bew}_T[\llbracket R \rrbracket^\star]
\end{aligned}
$$

The argument for the adapted version of Švejdar's Principle also yields, for any S^1_2-cut I,

$$T + A^\star \vdash \mathsf{Bew}_T[G] \rightarrow \mathsf{Bew}_T[\,\llbracket (\mathsf{Bew}_T[H] \leq \mathsf{Bew}_T[G])^I \rrbracket^\star \rightarrow H\,].$$

Using this variant we can easily get the desired facts for Z of subsection 6.2.

7 Variations

The result above is remarkably robust against all kinds of variations. The first possible variation is to consider relative interpretations in the definition of schematic validity instead of unrelativized interpretations as we did. Everything simply goes through. We might wish to move in the other direction, e.g. restricting ourselves to interpretations where identity translates only to identity. Again there are no poblems.

A somewhat more interesting variation is to consider as our theories pairs $\langle \mathcal{T}, \mathcal{N} \rangle$, where T is a theory in some arbitrary language and $\mathcal{N}$ is a relative interpretation of EA in T. We redefine our notion of validity for the modal language in such a way that the quantifiers of the modal language are interpreted as ranging over $\delta_{\mathcal{N}}$, the domain associated with $\mathcal{N}$. Again Vardanyan's Theorem will work for this notion of validity. Under this interpretation, we can extend Vardanyan's Theorem to such theories as ZFC and GB.

It is unknown to me whether our result can be sharpened —as Vardanyan improved his original theorem— to a predicate logical language with just one unary predicate symbol. It could very well be that Vardanyan's argument can be copied here, but I did not try.

References

[Art85] S.N. Artëmov. Nonarithmeticity of truth predicate logics of provability. *Doklady Akad. Nauk SSSR*, 284(2):270–271, 1985. In Russian. English translation in *Soviet Mathematics Doklady* 33:403–405, 1985.

[Boo93] G. Boolos. *The logic of provability.* Cambridge University Press, Cambridge, 1993.

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

[BV93] A. Berarducci and R. Verbrugge. On the provability logic of bounded arithmetic. *Annals of Pure and Applied Logic*, 61:75–93, 1993.

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

[JdJ98] G. Japaridze and D. de Jongh. The logic of provability. In S. Buss, editor, *Handbook of proof theory*, pages 475–546. North-Holland Publishing Co., amsterdam edition, 1998.

[Pli77] V. E. Plisko. The nonarithmeticity of the class of realizable formulas. *Math. of USSR Izv.*, 11:453–471, 1977.

[Pli78] V. E. Plisko. Some variants of the notion of realizability for predicate formulas. *Math. of USSR Izv.*, 12:588–604, 1978.

[Pli83] V. E. Plisko. Absolute realizability of predicate formulas. *Math. of USSR Izv.*, 22:291–308, 1983.

[Pli93] V. E. Plisko. On arithmetic complexity of certain constructive logics. *Mathematical Notes*, pages 701–709, 1993. Translated from *Mat. Zametki*, vol. 52, pp. 94–104, 1992.

[Sha93] V.Yu. Shavrukov. Subalgebras of diagonalizable algebras of theories containing arithmetic. *Dissertationes mathematicae (Rozprawy matematycne)*, CCCXXIII, 1993.

[Šve83] V. Švejdar. Modal analysis of generalized rosser sentences. *The Journal of Symbolic Logic*, 48:986–999, 1983.

[Var86] V.A. Vardanyan. Arithmetic comlexity of predicate logics of provability and their fragments. *Doklady Akad. Nauk SSSR*, 288(1):11–14, 1986. In Russian. English translation in *Soviet Mathematics Doklady* 33:569–572, 1986.

[Vis84] A. Visser. The provability logics of recursively enumerable theories extending Peano arithmetic at arbitrary theories extending Peano arithmetic. *Journal of Philosophical Logic*, 13:97–113, 1984.

[Vis93] A. Visser. The unprovability of small inconsistency. *Archive for Mathematical Logic*, 32:275–298, 1993.

[Vis98] A. Visser. An Overview of Interpretability Logic. In M. Kracht, M. de Rijke, H. Wansing, and M. Zakharyaschev, editors, *Advances in Modal Logic, vol 1*, CSLI Lecture Notes, no. 87, pages 307–359. Center for the Study of Language and Information, Stanford, 1998.

[Vis02] A. Visser. Faith & Falsity: a study of faithful interpretations and false Σ^0_1-sentences. Logic Group Preprint Series 216, Department of Philosophy, Utrecht University, Heidelberglaan 8, 3584 CS Utrecht, October 2002.

[VV94] L.C. Verbrugge and A. Visser. A small reflection principle for bounded arithmetic. *The Journal of Symbolic Logic*, 59:785–812, 1994.

[Yav97] R.E. Yavorsky. Logical schemes for first order theories. In *Springer LNCS (Yaroslavl'97 volume)*, volume 1234, pages 410–418, 1997.

[Zam94] D. Zambella. Shavrukov's theorem on the subalgebras of diagonalizable algebras for theories containing $I\Delta_0 + \mathsf{EXP}$. *The Notre Dame Journal of Formal Logic*, 35:147–157, 1994.