Friedman-reflexivity

Albert Visser [1]

Philosophy, Faculty of Humanities, Utrecht University, Janskerkhof 13, 3512BL Utrecht, the Netherlands

A R T I C L E   I N F O

A B S T R A C T

In the present paper, we explore an idea of Harvey Friedman to obtain a coordinate-free presentation of consistency. For some range of theories, Friedman's idea delivers actual consistency statements (modulo provable equivalence). For a wider range, it delivers consistency-like statements.

We say that a sentence C is an *interpreter* of a finitely axiomatised A over U iff it is the weakest statement C over U, with respect to U-provability, such that $U + C$ interprets A. A theory U is *Friedman-reflexive* iff every finitely axiomatised A has an interpreter over U. Friedman shows that Peano Arithmetic, PA, is Friedman-reflexive.

We study the question which theories are Friedman-reflexive. We show that a very weak theory, Peano Corto, is Friedman-reflexive. We do not get the usual consistency statements here, but bounded, cut-free, or Herbrand consistency statements. We illustrate that Peano Corto as a base theory has additional desirable properties.

We prove a characterisation theorem for the Friedman-reflexivity of sequential theories. We provide an example of a Friedman-reflexive sequential theory that substantially differs from the paradigm cases of Peano Arithmetic and Peano Corto. Interpreters over a Friedman-reflexive U can be used to define a provability-like notion for any finitely axiomatised A that interprets U. We explore what modal logics this idea gives rise to. We call such logics *interpreter logics*. We show that, generally, these logics satisfy the Löb Conditions, aka K4. We provide conditions for when interpreter logics extend S4, K45, and Löb's Logic. We show that, if either U or A is sequential, then the condition for extending Löb's Logic is fulfilled. Moreover, if our base theory U is sequential and if, in addition, its interpreters can be effectively found, we prove Solovay's Theorem. This holds even if the provability-like operator is not necessarily representable by a predicate of Gödel numbers.

At the end of the paper, we briefly discuss how successful the coordinate-free approach is.

© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

We study analogues of consistency statements that are characterised in a 'coordinate-free' way in the sense that their characterisation does not depend on arithmetisation. There are two reasons why such analogues are interesting. First, if we consider standard applications of the Second Incompleteness Theorem, aka G2, we would like to eliminate arithmetisation as a hidden parameter in the statement. The arithmetisation-free version is closer to an 'honest' mathematical theorem. Of course, the idea is that, for a reasonable range of standard applications, the analogues coincide, modulo provable equivalence, with ordinary consistency statements.

Secondly, on a more adventurous note, the analogues are able to live also in contexts where no (full) arithmetisation is possible. In other words, by considering coordinate-free versions, we can take consistency statements out of their comfort zone.[2] We think this is interesting in itself, but, in addition, the wider view allows us to look at the standard cases from a more general vantage point.

We develop a strategy proposed by Harvey Friedman. See [5]. The basic idea is that a consistency statement for a finitely axiomatised A over a base theory B is the weakest statement C such that $B + C$ interprets A. We will call such weakest statements: *interpreters*. We explain interpreters in more detail in Subsection 2.2.

We study Friedman's idea in a general setting. The comfort zone for his idea is the class of sequential theories. In this familiar context, interpreters are bounded (or cut-free or Herbrand) consistency statements that appear in *varying* interpretations of the natural numbers in the base theory. We show that essentially number-system-hopping interpreters may really occur.

Given a Friedman-reflexive base theory B and an interpretation K of B in a finitely axiomatised theory A, we can define the *interpreter logic* of A (w.r.t. K). This is an analogue of the provability logic of A. We will have a first look at what principles of interpreter logic we may obtain. We show that, generally, we have at least K4, i.e., the Löb Conditions. However, this need not yield Löb's Logic, GL, since we do not necessarily have a Fixed Point Lemma.[3] We provide general conditions for obtaining S4, K45, and GL. In case either B or A is sequential, the sufficient condition for obtaining Löb's Logic turns out to be fulfilled. Our Friedman-style version of the Second Incompleteness Theorem can be viewed as a direct consequence of this last insight, but also follows directly from the sufficient condition. See Corollary 8.15. In case B is sequential and the association of an interpreter to A is effective for the given Friedman-reflexive base U, we show that the proof of Solovay's Completeness Theorem can be given with few modifications. This is possible even if we do not have a definition of the provability-like modality corresponding to interpreters as a predicate of B.

We discuss the ins and outs of what is achieved and what is not achieved in the present paper in Section 9.

Remark 1.1. *Caveat emptor.* What we do not try to do in this paper is solve the philosophical problem of when a sentence expresses a consistency statement. The dialectic rather has this form. Show me your favourite arithmetisation of a consistency statement and I will characterise it, modulo provable equivalence, in a coordinate-free manner. Thus, a mathematically acceptable notion is provided, even if, in the motivation phase, arithmetisation with all its arbitrary choices may still play a role.   ○

1.1. On reading this paper

Appendix A gives basic definitions and basic facts and some references to further literature.

---

[2] A different example of a study of a G2-analogue that works in a wider context can be found in [11].

[3] We may lack the resources to formulate and prove a Gödel Fixed Point Lemma and, even if we have those, the provability-like notion need not be representable by a definable predicate.

We point out, at places, how our work links to elementary ideas from category theory (universal arrows and the like). The reader who wishes can skip this without losing the main thread of the paper.

Our main application to sequential theories demands some familiarity with *sequentiality*. See, e.g., [8] and [25]. However, the main development of Friedman-reflexivity and interpreter logics only asks for understanding predicate logic and translations/interpretations. The reader could elect to read the results concerning sequentiality, but ignore the proofs, and still get a good feeling for the main line of argument.

2. The main ideas of the paper

We explain in more detail two central ideas of the paper.

2.1. On the very idea of a base theory

To set the stage for our Friedman-style treatment of the Second Incompleteness Theorem, G2, and of a variant of provability logic, we discuss the base theory/main theory distinction.

A formulation of the no-interpretation version of G2 looks like this:

$$U \not\rhd (\mathsf{B} + \mathsf{Con}(U)),$$

in other words, U does not interpret the *base theory* B plus its own consistency.[4] Here $\mathsf{Con}(U)$ is the consistency statement, where we allow various further specifications of what it could be. Also, for a concrete formulation, there may be further conditions on U.

A first reason to set things up in this format is simply the nice general form of the statement. E.g., we have Pudlák-style G2: U does not interpret $\mathsf{Q} + \diamondsuit_\alpha \top$, or, using our preferred notation, $U \not\rhd (\mathsf{Q} + \diamondsuit_\alpha \top)$.[5] Here $\diamondsuit_\alpha \top$ is a specific form of the consistency statement for U, where we formalised consistency in arithmetic in a sufficiently good way and where α is a Σ_1^0-formula representing the axiom set of U.[6] The statement is general since there is no conditionalisation to 'theories that are sufficiently strong' and the like. Also, it is strong since Q is very weak, so the contribution of the base to the non-interpretability is minimal. We note that Pudlák's proof of this version of G2 does contain Gödel's original argument but extends it with new ideas like Solovay's method of shortening cuts. So, perhaps, we can also say that the Pudlák-style formulation counts as a *strengthening* of G2.

A second reason is more proof-oriented. Consider, for example, G2 for ZF. We can easily formalise G2 in ZF using the set-theoretical resources for coding sequences and the like. In fact, this is easier than formalisation in PA with only zero, successor, plus and times in the signature. On the other hand, ZF has a standard interpretation of PA in the finite von Neumann ordinals. We can simply import the arithmetisation of syntax as usually done in PA in ZF via the von Neumann interpretation. The advantage of doing it like this is that we see that Gödel's *proof* works uniformly across theories as soon as we have an interpretation of a suitable base. In this context, the best version is $U \not\rhd (\mathsf{S}_2^1 + \diamondsuit_\alpha \top)$. Here S_2^1 is Buss's weak arithmetical theory for the study of p-time computability.[7] Without any additional effort, as compared to, e.g., PA, Gödel's reasoning can be repeated in S_2^1. In fact, S_2^1 is better, since it prevents all kinds of silly and inefficient choices for doing the arithmetisation. Moreover, Pudlák's argument for his strengthened version can be framed as first

---

[4] A first form of the no-interpretation version is due to Feferman. See [4]. Feferman's form was stated for extensions of PA and for the case that main theory and base theory coincide.

[5] Pavel Pudlák contributed the main ingredient of the result in the present formulation. See [10].

[6] The various modal-style notations we use are explained in Appendix A or at the place where they are introduced as new concepts in the paper. Briefly: we use $\square$ for provability and $\diamondsuit$ for consistency. There are related notations $\boxdot$ and $\diamondsuit\!\!\!\diamond$, $\blacksquare$ and $\blacklozenge$ that represent provability-like and consistency-like notions based on interpretation power. We use $\triangle$ and $\triangledown$ for cut-free provability and cut-free consistency. If K is an interpretation, we will write $\square_\alpha^K B$ rather than $(\square_\alpha B)^K$, and, similarly, for the other modal notions.

[7] See [2] or [8] for the basic development of S_2^1.

showing that $\mathsf{Q} + \Diamond_\alpha \top$ is mutually interpretable with $\mathsf{S}_2^1 + \Diamond_\alpha \top$, or: $(\mathsf{Q} + \Diamond_\alpha \top) \bowtie (\mathsf{S}_2^1 + \Diamond_\alpha \top)$, and, then, concluding the Q-version from the S_2^1-version. We think that the S_2^1-version should be viewed as G2 proper, since it is proved simply by Gödel's original reasoning.

A third reason is mathematico-philosophical. The usual versions of G2 depend on certain design choices. Can we choose the base theory in such a way that either design choices become so natural that they are, as it were, intrinsically given, or, in such a way that they are fully eliminated?

- The first idea is explored by Volker Halbach and Graham Leigh in a forthcoming book, [7]. Let us point out how natural this idea is. Arithmetisation can be viewed as the development of a chain of interpretations ending in an appropriate syntax theory. Why not take this syntax theory as base? However, many syntax theories are possible. Which is the right one? Also, how do we get a syntax theory that really determines which choices to make, e.g., to represent a proof?
- The second idea is explored in this paper. We develop an idea of Harvey Friedman to eliminate design choices entirely in our base theory. Friedman's idea does not always deliver the usual consistency statements but also other things that are analogues of consistency statements. We elaborate on the idea in Subsection 2.2.

  In Subsection 6.2, we will explain how the idea of main/base works out when we consider provability-like logics based on the consistency analogues that we developed.

Finally, there is a fourth reason. There has been philosophical discussion on the question: when does a predicate logical sentence really express a consistency statement? One line of argument could be that we choose as base theory a *meaningful* theory and that it is the semantics of the base theory that carries the main burden of meaning giving. See also [29] for some further discussion.

2.2. Interpretation power

What makes a sentence an analogue of a consistency statement? We zoom in on an answer proposed by Harvey Friedman, in [5], to wit, that the hallmark of consistency statements is the kind *interpretation power* typical for consistency statements in virtue of the Interpretation Existence Lemma (see [31] for a detailed exposition of this lemma). The lemma says, roughly, that we can interpret a theory V in a suitable base theory plus a consistency statement for V.

In case B extends S_2^1 and α numerates the axioms of V over B, we have the following formulation of Interpretation Existence: $(\mathsf{B} + \Diamond_\alpha \top) \rhd V$. We will say that any B-sentence B such that $(\mathsf{B} + B) \rhd V$ is a *pro-interpreter* of V (over B). So, Interpretation Existence tells us that $\Diamond_\alpha \top$ is a pro-interpreter. A study of pro-interpreters was undertaken in [23].

A disadvantage of the idea of pro-interpreters is that they are not uniquely determined over the base theory. To make them unique, we have to impose an extra demand. The obvious one is that we consider the *weakest* pro-interpreter with respect to B-provability. Let us call such a weakest pro-interpreter simply *an interpreter*.

In this paper, we will restrict ourselves to the case of consistency analogues for *finitely axiomatised* theories A. Needless to say that this assumption simplifies a lot. We see that an interpreter of A over B is a B-sentence C such that $(\mathsf{B} + C) \rhd A$ and, for all B-sentences B, we have, if $(\mathsf{B} + B) \rhd A$, then $\mathsf{B} + B \vdash C$. Alternatively, we can say that C is an interpreter of A over B iff, for all B-sentences B, we have $(\mathsf{B} + B) \rhd A$ iff $\mathsf{B} + B \vdash C$.

Unfortunately, over the base theory S_2^1, this idea will not work. Consider any consistent A such that $\mathsf{S}_2^1 \not\rhd A$. E.g., A could be Elementary Arithmetic EA. Suppose we had an interpreter C of A. We have

$(\mathsf{S}_2^1 + \Diamond_A^I \top) \rhd A$, for any definable S_2^1-cut I.[8] So, for all I, we find $\mathsf{S}_2^1 + \Diamond_A^I \top \vdash C$. In other words, for all I, we find $\mathsf{S}_2^1 + \neg\, C \vdash \Box_A^I \bot$. Since $\mathsf{S}_2^1 \not\rhd A$, we have $\mathsf{S}_2^1 \nvdash C$ and, hence, $\mathsf{S}_2^1 + \neg\, C$ is consistent. By Theorem A.2 of the Appendix, we find that $\Box_A \bot$ is true, and, thus, that A is inconsistent, which contradicts our assumption.[9] Thus, Friedman's idea forces us towards other bases.

We call a theory that does have the desirable property that each finitely axiomatised A has an interpreter over the theory *Friedman-reflexive*.[10] Friedman's example of such a theory is Peano Arithmetic PA. He shows that, indeed, PA is Friedman-reflexive. See Section 4 for an exposition of Friedman's argument. A disadvantage of the choice of PA is that it is very strong. Our proposal for the ideal Friedman-reflexive theory is Peano Corto or $\mathsf{PA}^{\downarrow\downarrow}$. We will show, in Subsection 7.3, that $\mathsf{PA}^{\downarrow\downarrow}$ has some good further properties as a base.

3. Basics

In this section we give the basic definitions and state and prove some basic facts.

3.1. Definitions

The variables $T, U, V, \ldots$ range over theories of finite signature. These theories, generally, need not be RE. We allow inconsistent theories as values.

The variables $A, B, \ldots$ range ambiguously over sentences of predicate logic and over finitely axiomatised theories. We confuse the finitely axiomatised theory A with a single sentence A which is (the finite conjunction of) its axiom(s).

There is a subtle point here. Let U be any theory and let τ be an interpretation from the A-language to the U-language. Let E_A be the finite conjunction of the identity axioms for the signature of A.[11] We assume that $\exists x\, x = x$ is among these axioms. Then, $U + (E_A \wedge A)^\tau$ interprets A with an interpretation based on τ. The reason that we need E_A for this is that, in first order logic, identity is treated as a logical constant but that in the definition of translation this is ignored and identity may be translated to some U-formula. The inclusion of $\exists x\, x = x$ in the identity axioms is needed since we, somewhat unnaturally, assume non-empty domains.

We say that C is *an interpreter of A over U* iff, for all B in the language of U, we have: $(U + B) \rhd A$ iff $U + B \vdash C$.

A theory U is *Friedman-reflexive* iff all finitely axiomatised A have an interpreter C over U.

Suppose U is Friedman-reflexive. We will use $\Diamondblack(\cdot)$ for a function that selects, for each A, an interpreter C of A over U. So, $\Diamondblack$ is only uniquely determined modulo U-provability. We write $\Diamondblack_A B$ for $\Diamondblack(A \wedge B)$. If we want to make the dependence on U explicit, we write $\Diamondblack_{(U)}$, $\Diamondblack_{(U),A}$, etcetera. However, we prefer to treat the dependence on U as contextually given as much as possible. We write $\boxdot$ for $\neg\Diamondblack\neg$ and $\boxdot_A$ for $\neg\Diamondblack_A\neg$.

It is easy to see that $\Diamondblack A$ and $\Diamondblack_A \top$ are equivalent over U. We will use the second in contexts where we are interested in extensions of A.

The theory U is *effectively Friedman-reflexive* iff we can choose $\Diamondblack$ to be recursive. Note the existential quantifier here: there may very well be both recursive and non-recursive choices of $\Diamondblack$. In fact, we will see a salient example of that possibility.

---

[8] Our cuts are downward closed w.r.t. $<$ and are closed under zero, successor, addition, multiplication and ω_1, i.e., the function $x \mapsto 2^{(\log(x))^2}$, where $\log$ is the (entier of) the base 2 logarithm.

[9] A stronger version of this insight is given in Theorem 8.9.

[10] The ratio behind the second name will become clear in the paper.

[11] Here it is essential that we consider the signature of A *qua theory*. Not all symbols of this signature need to actually occur in the conjunction of the axioms of A.

3.2. The categorical viewpoint

We can view what is going on here in category theoretic terms as follows. Let $\mathbb{B}_U$ be the partial order category of all finite extensions of the base theory U (in the same language) with order $\subseteq$, i.e. extension in the same language. Let $\mathbb{D}$ be the partial pre-order category of theories ordered by interpretability $\lhd$.[12] Let proj_U be the projection functor from $\mathbb{B}_U$ into $\mathbb{D}$. Then, C is an interpreter of A iff $A \lhd (U + C)$ is a universal arrow from A to proj_U.

Remark 3.1. We note that if we restrict $\mathbb{D}$ to a subcategory that contains both the finitely axiomatised theories and the finite extensions of U, but that preserves the arrows between the objects, then universality is preserved in both directions.

So, e.g., in case U is RE, we can restrict $\mathbb{D}$ to the category of RE theories with the same effect. ◯

We can view Friedman-reflexivity as follows. Let $\mathbb{D}_{\mathsf{fin}}$ be the partial pre-order category of finitely axiomatised theories ordered by interpretability. Let emb be the embedding functor from $\mathbb{D}_{\mathsf{fin}}$ to $\mathbb{D}$. Friedman-reflexivity tells us that, for each A, there is a universal arrow from $\mathsf{emb}(A)$ to proj_U.

3.3. Semi normal form

Interpreters have a kind of non-unique normal forms. Let ν be any translation of the A language in the U-language. We note that $(E_A \wedge A)^\nu$ is always a pro-interpreter of A over U, but not necessarily an interpreter.

Theorem 3.2. *Suppose C is an interpreter of A over U. Then, there is a translation τ of the A-language in the U-language, such that $U \vdash C \leftrightarrow (E_A \wedge A)^\tau$.*

We note that τ need not be unique.

Proof. Suppose C is an interpreter of A over U. We have $(U + C) \rhd A$. Let τ be the translation underlying a witnessing interpretation. Then $U + C \vdash (E_A \wedge A)^\tau$.

On the other hand, $(U + (E_A \wedge A)^\tau) \rhd A$. So, by the defining property of interpreters, $U + (E_A \wedge A)^\tau \vdash C$. ❑

3.4. Some basic facts

We show that interpreters are unique.

Theorem 3.3. *Suppose both C and C' are interpreters of A over U. Then, we have $U \vdash C \leftrightarrow C'$.*

Proof. Since $(U + C) \vdash C$, it follows that $(U + C) \rhd A$, and, hence $(U + C') \vdash C$. The converse direction is similar. ❑

Remark 3.4. The above uniqueness argument is a special case of the usual argument for uniqueness of universal arrows. ◯

We show that $\lozenge\!\!\!\lozenge$ has a functorial property with respect to a Friedman-reflexive theory.

Theorem 3.5. *Let U be Friedman-reflexive and suppose $A \rhd B$. Then, we have $U \vdash \lozenge\!\!\!\lozenge A \rightarrow \lozenge\!\!\!\lozenge B$.*

---

[12] In the light of our specific application, we can replace $\lhd$ by $\lhd_{\mathsf{loc}}$.

Proof. Let U be Friedman-reflexive. Suppose $A \triangleright B$. Then, $(U + \Diamondblack A) \triangleright A \triangleright B$. So, $U \vdash \Diamondblack A \to \Diamondblack B$. $\quad \square$

Remark 3.6. Theorem 3.5 is a special case of an elementary category theoretic insight. Suppose we have categories $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ and functors $F : \mathcal{A} \to \mathcal{C}$, $G : \mathcal{B} \to \mathcal{C}$. Suppose further that, for every a in A there is a universal arrow from $F(a)$ to G. Then, there is a functor $H : \mathcal{A} \to \mathcal{B}$, such that our promised universal arrows are a natural transformation from F to $G \circ H$. $\quad \bigcirc$

Here is a fact that does not follow from the general categorical ideas but is specific to our setting.

Theorem 3.7. *Let U be Friedman-reflexive. Then $\Diamondblack$ commutes, modulo U-provable equivalence, with finite disjunctions of sentences in the same signature, including the empty one. In other words, $U \vdash \neg \Diamondblack \bot$ and, if A and B have the same signature, then $U \vdash \Diamondblack(A \vee B) \leftrightarrow (\Diamondblack A \vee \Diamondblack B)$.*

Proof. We have $(U + \Diamondblack \bot) \triangleright \bot$. So, $U \vdash \neg \Diamondblack \bot$.

We have $A \vdash (A \vee B)$ and, hence, $A \triangleright (A \vee B)$. It follows, by Theorem 3.5, that $U \vdash \Diamondblack A \to \Diamondblack(A \vee B)$. Similarly, $U \vdash \Diamondblack B \to \Diamondblack(A \vee B)$. Ergo, $U \vdash (\Diamondblack A \vee \Diamondblack B) \to \Diamondblack(A \vee B)$.

We have $(U + \Diamondblack(A \vee B)) \triangleright (A \vee B)$. Let τ be the underlying translation of some witnessing interpretation. We have:

$$\begin{aligned} U + \Diamondblack(A \vee B) &\vdash (E_A \wedge (A \vee B))^\tau \\ &\vdash (E_A \wedge A)^\tau \vee (E_A \wedge B)^\tau \\ &\vdash \Diamondblack A \vee \Diamondblack B \end{aligned}$$

The last step uses that the pro-interpreter $(E_A \wedge A)^\tau$ implies the interpreter $\Diamondblack A$ and, similarly, for B. $\quad \square$

We define $W := U \oslash V$ as follows. The signature of W is the disjoint union of the signatures of U and V plus two unary domain predicates $\triangle_0$ and $\triangle_1$. We have the axioms of U relativised to $\triangle_0$, the axioms of V relativised to $\triangle_1$ plus axioms that say that the $\triangle_i$ form a partition of the domain.

The following fact again follows from the categorical framework alone in combination with the fact that $\oslash$ is a supremum operator in $\mathbb{D}_{\mathsf{fin}}$.

Theorem 3.8. *Suppose U is Friedman-reflexive and A and B are finitely axiomatised. Then, we have $U \vdash \Diamondblack(A \oslash B) \leftrightarrow (\Diamondblack A \wedge \Diamondblack B)$.*

Proof.
$$\begin{aligned} U + D \vdash \Diamondblack(A \oslash B) &\Leftrightarrow (U + D) \triangleright (A \oslash B) \\ &\Leftrightarrow (U + D) \triangleright A \text{ and } (U + D) \triangleright B \\ &\Leftrightarrow U + D \vdash \Diamondblack A \text{ and } U + D \vdash \Diamondblack B \\ &\Leftrightarrow U + D \vdash \Diamondblack A \wedge \Diamondblack B \quad\quad \square \end{aligned}$$

We will meet $\oslash$ again in Corollary C.5.

3.5. Alternative characterisation

We have an alternative characterisation of Friedman-reflexivity that gives us a full adjunction. We write $U \subseteq V$ for V is an extension of U in the same language.

Theorem 3.9. *A theory U is Friedman-reflexive iff* (†) *for every theory W, there is a theory $\mathfrak{C}(W) \supseteq U$, such that, for all $V \supseteq U$, we have $V \rhd_{\mathsf{loc}} W$ iff $V \supseteq \mathfrak{C}(W)$.*

Proof. Suppose U is Friedman-reflexive. We prove (†). Consider any W. We define the theory $\mathfrak{C}(W) := U + \{\lozenge\!\!\!\lozenge\, A \mid W \vdash A\}$. Let $V \supseteq U$. We have:

$$
\begin{aligned}
V \rhd_{\mathsf{loc}} W &\Leftrightarrow \forall A\,(W \vdash A \;\Rightarrow\; V \rhd A) \\
&\Leftrightarrow \forall A\,(W \vdash A \;\Rightarrow\; \exists D\,(V \vdash D \text{ and } (U+D) \rhd A)) \\
&\Leftrightarrow \forall A\,(W \vdash A \;\Rightarrow\; \exists D\,(V \vdash D \text{ and } U+D \vdash \lozenge\!\!\!\lozenge\, A)) \\
&\Leftrightarrow \forall A\,(W \vdash A \;\Rightarrow\; V \vdash \lozenge\!\!\!\lozenge\, A) \\
&\Leftrightarrow V \supseteq \mathfrak{C}(W)
\end{aligned}
$$

Suppose (†). Consider any finitely axiomatised A. We have $\mathfrak{C}(A) \rhd A$. It follows that for some C, we have $\mathfrak{C}(A) \vdash C$ and $C \rhd A$. Since $(U+C) \rhd A$, we have $(U+C) \supseteq \mathfrak{C}(A)$. Thus, C axiomatises $\mathfrak{C}(A)$ over U. It follows that:

$$
\begin{aligned}
(U+B) \rhd A &\Leftrightarrow (U+B) \rhd_{\mathsf{loc}} A \\
&\Leftrightarrow (U+B) \supseteq \mathfrak{C}(A) \\
&\Leftrightarrow (U+B) \vdash C
\end{aligned}
$$

So, we can take $\lozenge\!\!\!\lozenge\, A := C$. □

We translate our alternative characterisation in categorical terms. Let $\mathbb{B}_U^+$ be the category of all extensions of U in the same language with as arrows $\subseteq$. Let $\mathbb{D}_{\mathsf{loc}}$ be the category of all theories with the local interpretability relation as arrows. Let proj_U^+ be the projection functor of $\mathbb{B}_U^+$ into $\mathbb{D}_{\mathsf{loc}}$. Then U is Friedman-reflexive iff proj_U^+ has a left adjoint $\mathfrak{C}$.

3.6. Polyglotticity

A theory U is *polyglot* or *polyglottic* if, for every consistent finitely axiomatised A, there is a pro-interpreter B of A such that $U + B$ is consistent.

We remind the reader that T locally tolerates V if, for every finite sub-theory A of V, there is a translation τ such that T is consistent with $(E_A \wedge A)^\tau$.

Theorem 3.10. *U is polyglot iff U locally tolerates the theory Q plus the true Π_1^0-sentences.*

Proof. Suppose U is polyglot. Let A be a finite sub-theory of Q plus the true Π_1^0-sentences. Then, for some B, we have $U + B$ is consistent and $(U + B) \rhd A$. Let τ be the translation on which the interpretation of A in $U + B$ is based. Then, U is consistent with $(E_A + A)^\tau$.

Conversely, suppose U locally tolerates the theory Q plus the true Π_1^0-sentences. Consider any finitely axiomatised consistent theory D. Then, $\lozenge_D \top$ is true. So, for some τ, we have $U + (E_\mathsf{Q} \wedge \mathsf{Q} \wedge \lozenge_D \top)^\tau$ is consistent. By the Interpretation Existence Lemma, we find that $(U + (E_\mathsf{Q} \wedge \mathsf{Q} \wedge \lozenge_D \top)^\tau) \rhd (\mathsf{Q} + \lozenge_D \top) \rhd D$. □

Since Q interprets S_2^1 on a cut, we have the same result with S_2^1 substituted for Q.

3.7. A computational insight

We end this section with two computational results about effectively Friedman-reflexive theories.

Theorem 3.11. *If U is consistent and effectively Friedman-reflexive, then U is essentially undecidable.*

Proof. Suppose U is consistent and effectively Friedman reflexive. Suppose U had a consistent and decidable extension V. By Theorem 5.1, the theory V is effectively Friedman reflexive with recursive $\diamondsuit\!\!\!\!\diamondsuit$.

Let $[S]$ be a theory of a witness of the Σ_1^0-sentence S. See [30] for details. If S is true, then $[S]$ has a finite model and, thus, any theory interprets $[S]$. It follows that $V \vdash \diamondsuit\!\!\!\!\diamondsuit[S]$. On the other hand, if $[S]$ is false, then it extends R and, thus, since V is decidable, $V \ntriangleright [S]$. Ergo, $V \nvdash \diamondsuit\!\!\!\!\diamondsuit[S]$. It follows that, using the decidability of V, we can solve the halting problem. *Quod non.* ❑

We remind the reader that T tolerates U iff, for some translation τ, the theory $T + E_U^\tau + U^\tau$ is consistent. In other words, T tolerates U, if some consistent extension of T interprets U. We define:

- A theory U is *strongly essentially undecidable* iff every theory T that tolerates U is undecidable.

If a finitely axiomatised theory is essentially undecidable it is easily seen that it is also strongly essentially undecidable. The same does not have to hold for RE theories. See, e.g., [13] for examples. Cobham showed that the Tarski-Mostowski-Robinson theory R is strongly essentially undecidable. See [16] or [30]. We have:

Theorem 3.12. *Suppose U is consistent, RE, and effectively Friedman-reflexive. Then, U is strongly essentially undecidable.*

Proof. Let U be consistent, RE, and effectively Friedman-reflexive. Suppose T is decidable and suppose $V := T + E_U^\tau + U^\tau$ is consistent. In case S is true, we have $V \vdash \diamondsuit\!\!\!\!\diamondsuit^\tau[S]$. Now suppose $[S]$ is false. In case $V + \diamondsuit\!\!\!\!\diamondsuit^\tau[S]$ were consistent, it would follow that T tolerates $[S]$, contradicting the fact that $[S]$ is finitely axiomatised and essentially undecidable. So, $V \vdash \neg \diamondsuit\!\!\!\!\diamondsuit^\tau[S]$. Since V is RE, this gives us a procedure to decide the halting problem. ❑

Open Question 3.13. Is there a theory U that is consistent, effectively Friedman-reflexive and not strongly essentially undecidable? ◯

4. The paradigm case of Peano arithmetic

Peano Arithmetic is a paradigmatic theory that is Friedman-reflexive. This is the theory for which Harvey's original observation was made. Our Theorem 4.1 is Theorem 2.7 of [5].

Theorem 4.1 (Friedman). *The sentence $\diamondsuit_A \top$ is an interpreter of A over PA. So, PA is effectively Friedman-reflexive.*

Proof. Suppose $(\mathsf{PA} + B) \triangleright A$. Then, for some finite sub-theory D of PA, we have $(D + B) \triangleright A$. It follows that $\mathsf{PA} \vdash ((D+B) \triangleright A)$ and, so, $\mathsf{PA} \vdash \diamondsuit_D B \to \diamondsuit_A \top$. By the essential reflexiveness of PA, we may conclude $\mathsf{PA} + B \vdash \diamondsuit_A \top$.

For the other direction, suppose $\mathsf{PA} + B \vdash \diamondsuit_A \top$. Then, by the Interpretation Existence Lemma (see [31]), we find $(\mathsf{PA} + B) \triangleright A$. ❑

We note that we have characterised the consistency statement $\Diamond_A \top$ among arithmetical sentences modulo PA-provable equivalence. We will later discuss how to improve this to EA-provable equivalence—see Theorem 7.4.

5. Closure properties

In this section, we study various closure properties of Friedman reflexive theories.

Theorem 5.1. *Suppose $U \subseteq V$, where V is in the same language as U, and U is (effectively) Friedman-reflexive. Then, V is (effectively) Friedman-reflexive with the same interpreters. In other words, $\Diamond\!\!\!\Diamond$ can be chosen the same for U and V.*

Proof. Suppose $U \subseteq V$ and U is Friedman-reflexive. We have:

$$(V + B) \rhd A \Leftrightarrow \text{for some } D, \text{ we have } V \vdash D \text{ and } (U + (D \wedge B)) \rhd A$$

$$\Leftrightarrow \text{for some } D, \text{ we have } V \vdash D \text{ and } (U + (D \wedge B)) \vdash \Diamond\!\!\!\Diamond A$$

$$\Leftrightarrow (V + B) \vdash \Diamond\!\!\!\Diamond A$$

Since $\Diamond\!\!\!\Diamond$ is preserved, we, *ipso facto*, preserve effectivity. ❏

The next theorem illustrates that polyglotticity is definitely *not* preserved to consistent extensions.

Theorem 5.2. *Suppose U is Friedman-reflexive. Suppose A is consistent and $U \ntriangleright A$. Then, $U + \neg \Diamond\!\!\!\Diamond A$ is consistent and not polyglot.*

Proof. We assume the conditions of the theorem. Since $U \ntriangleright A$, we have $U \nvdash \Diamond\!\!\!\Diamond A$ and, so, $U + \neg \Diamond\!\!\!\Diamond A$ is consistent.

By Theorem 5.1, we may choose $\Diamond\!\!\!\Diamond$ for $U + \neg \Diamond\!\!\!\Diamond A$ the same as for U, but, $\Diamond\!\!\!\Diamond A$ is inconsistent with $U + \neg \Diamond\!\!\!\Diamond A$. ❏

An example of our theorem is the fact that no consistent extension in the same language of $\mathsf{PA} + \Box_{\mathsf{ACA}_0} \bot$ interprets ACA_0. We note that we can always find an A such that $U \ntriangleright A$, if U is RE and consistent. In contrast, EA plus all true Π_1^0-sentences is consistent and Friedman-reflexive, but this theory does interpret every consistent A. Thus, it is polyglottic and so is every extension.

Our next closure property is categorical in nature. Let $\mathbb{E} := \mathsf{INT}_3^+$ be the category of theories in finite signature and interpretations, where two interpretations $K, K' : T \to W$ are the same iff, for all T-sentences A, we have $W \vdash A^K \leftrightarrow A^{K'}$. We note that we do not demand that the theories are RE. In a sense, $\mathbb{E}$ in combination with its sub-category of all theories with as arrows theory-extensions-in-the-same-language is the natural home for the study of Friedman-reflexivity.[13]

We remind the reader that, in a category, a morphism $a \xrightarrow{f} b$ is a *retraction* or *split epimorphism* iff there is a $b \xrightarrow{g} a$, such that $f \circ g = \mathsf{id}_b$. Here g is called *section*, *co-retraction*, or *split monomorphism*.

Theorem 5.3. *Suppose V is an $\mathbb{E}$-retract of U and suppose U is (effectively) Friedman-reflexive. Then, V is (effectively) Friedman-reflexive.*

---

[13] For the case of finitely axiomatised theories, the interaction between interpretation and extension was studied in [33].

Proof. Let $K : V \to U$ and $M : U \to V$ witness the retraction. So, $M \circ K$ is the same, in the sense of $\mathbb{E}$, as Id_V. We have:

$$(V + B) \rhd A \Rightarrow (U + B^K) \rhd A$$
$$\Rightarrow U + B^K \vdash \lozenge\!\!\!\lozenge A$$
$$\Rightarrow V + B^{KM} \vdash \lozenge\!\!\!\lozenge^M A$$
$$\Rightarrow V + B \vdash \lozenge\!\!\!\lozenge^M A$$
$$\Rightarrow (V + B) \rhd (U + \lozenge\!\!\!\lozenge A)$$
$$\Rightarrow (V + B) \rhd A$$

It follows that $\lozenge\!\!\!\lozenge^M A$ is the desired interpreter of A over V. Clearly, composition of $\lozenge\!\!\!\lozenge$ with $(\cdot)^M$ preserves effectiveness. ❑

Theorem 5.4. *Suppose $U + D$ and $U + \neg D$ are both (effectively) Friedman-reflexive. Then, U is also (effectively) Friedman-reflexive.*

Proof. Consider any A and let C and C' be the interpreters of A over $U + D$, resp. $U + \neg D$. Let $C \langle D \rangle C'$ be $(D \wedge C) \vee (\neg D \wedge C')$. We have:

$$(U + B) \rhd A \Leftrightarrow (U + D + B) \rhd A \text{ and } (U + \neg D + B) \rhd A$$
$$\Leftrightarrow (U + D + B) \vdash C \text{ and } (U + \neg D + B) \vdash C'$$
$$\Leftrightarrow (U + D + B) \vdash C \langle D \rangle C' \text{ and } (U + \neg D + B) \vdash C \langle D \rangle C'$$
$$\Leftrightarrow (U + B) \vdash C \langle D \rangle C' \qquad\qquad ❑$$

In the next theorem, we verify a general insight for universal arrows in our specific case.

Given theories U and V, we define $W := U \oslash V$ as follows. The signature of W is the disjoint union of the signatures of U and V plus a new 0-ary predicate symbol P. The axioms of W are $P \to A$, for axioms A of U and $\neg P \to B$ for axioms B of V. We have:

Theorem 5.5. *Suppose U and V are (effectively) Friedman-reflexive. Then, $U \oslash V$ is (effectively) Friedman-reflexive.*

Proof. Let $W := U \oslash V$. We have:

$$(W + B) \rhd A \Leftrightarrow (U + B[P := \top]) \rhd A \text{ and } (V + B[P := \bot]) \rhd A$$
$$\Leftrightarrow U + B[P := \top] \vdash \lozenge\!\!\!\lozenge_{(U)} A \text{ and } V + B[P := \bot] \vdash \lozenge\!\!\!\lozenge_{(V)} A$$
$$\Leftrightarrow W + B \vdash (\lozenge\!\!\!\lozenge_{(U)} A) \langle P \rangle (\lozenge\!\!\!\lozenge_{(V)} A)$$

So, we can take $\lozenge\!\!\!\lozenge_{(W)} A := (\lozenge\!\!\!\lozenge_{(U)} A) \langle P \rangle (\lozenge\!\!\!\lozenge_{(V)} A)$. ❑

6. Interpreter logics

Can something like a modal logic be based on interpreters as an analog of provability logic? Since we only consider interpreters for finitely axiomatised theories, this should be a modal logic interpreted in a finitely axiomatised theory. We first give the definitions and then some motivating remarks.

6.1. Definitions

An *FM-frame* is a pair $\langle A, U \rangle$, where A is finitely axiomatised, U is Friedman-reflexive and $A \rhd U$.[14] An interpretation $K : A \rhd U$, where $\langle A, U \rangle$ is an FM-frame, is an FM-interpretation. We consider U and A as part of the data for K. The interpretation on the frame in our context is the analogue of the Kripke model on the Kripke frame in ordinary modal logic.

Consider an FM-interpretation $K : U \lhd A$. We define $\blacklozenge_{K,B} C := \lozenge\!\!\!\!\lozenge^K_{(U),B} C$. Our default case is where A is identical to B. In this case we write $\blacklozenge_K C$. Also, in many cases, we treat K as contextually given and simply write $\blacklozenge$. As usual, we set $\blacksquare := \neg \blacklozenge \neg$.

We consider the usual modal language with possibility operator $\lozenge$ and with necessity defined by $\neg \lozenge \neg$. Let σ be a function from the propositional atoms to the A-language. We define $\varphi^{(\sigma,K)}$ as follows.

- $p^{(\sigma,K)} := \sigma(p)$.
- $(\cdot)^{(\sigma,K)}$ commutes with the truth-functional connectives.
- $(\lozenge\psi)^{(\sigma,K)} := \blacklozenge_{K,A}\, \psi^{(\sigma,K)}$.

We will call the logic of $\blacklozenge_{K,A}$ over A: Λ^{fr}_K. So,

- $\Lambda^{\mathsf{fr}}_K = \{\varphi \mid \text{ for all } \sigma, \text{ we have } A \vdash \varphi^{(\sigma,K)}\}$.

We also define the logic of a frame $\langle A, U \rangle$.

- $\Lambda^{\mathsf{fr}}_{A,U} = \{\varphi \mid \text{ for all } M : A \rhd U \text{ and } \sigma, \text{ we have } A \vdash \varphi^{(\sigma,M)}\}$.
  In other words, $\Lambda^{\mathsf{fr}}_{A,U} = \bigcap_{M:A \rhd U} \Lambda^{\mathsf{fr}}_M$.

We consider one further notion in Appendix D.

6.2. Motivating remarks

Let us first think about ordinary provability logic. What is the provability logic of a given theory V? The arithmetisation of provability is provided by some base-theory B. Let us say this is $\square_\alpha B$, where α is a suitable presentation of the axioms of V in B. The base theory is 'in' V via an interpretation $K : V \rhd \mathsf{B}$. So, V-provability gets the form $\square^K_\alpha B$ in V, where we write $\square^K_\alpha B$ for $(\square_\alpha B)^K$.

If we switch to interpreter logic, the idea is precisely the same: the necessity operator gets the form $\boxdot^K_A B$, for main theory A. We note that here we have $K : A \rhd \mathsf{B}$, on the one hand, and that, on the other hand, the $\lozenge\!\!\!\!\lozenge_A B$ are defined using interpretations of $(A \wedge B)$ in finite extensions of B. So, both an interpretation *of* B *in* A and interpretations of extensions of A *in* extensions of B play a role.

In the case of ordinary provability logic *and* in the case of interpreter logic, we can quantify out the interpretations of the base theory B in the main theory V, resp. A. This leads to the frame provability/interpreter logic. A frame is the pair $\langle A, U \rangle$ with $A \rhd U$. So, we abstract away from the specific interpretation. The ordinary provability logic of a frame was studied in [28]. In the present paper, we will consider the interpreter logic of a frame.

---

[14] 'FM' stands for 'Feferman and Montague' who initiated the idea of looking at the combination of G2 and interpretability.

6.3. The Löb Conditions

We will show that every Λ_K^{fr} satisfies the Löb Conditions, in other words, it is a normal modal logic extending K4.

We will first verify a set of conditions that are equivalent to the Löb Conditions and then prove the equivalence.

Lemma 6.1. *Let $K : U \lhd A$ be an FM-interpretation. We write $\blacklozenge := \blacklozenge_{K,A}$. We have:*

a. $A \vdash B \to C$ *implies* $A \vdash \blacklozenge B \to \blacklozenge C$ *and* $A \vdash \blacksquare B \to \blacksquare C$.
b. $A \vdash \neg \blacklozenge \bot$ *and* $A \vdash \blacksquare \top$.
c. $A \vdash \blacklozenge (B \vee C) \leftrightarrow (\blacklozenge B \vee \blacklozenge C)$ *and* $A \vdash \blacksquare (B \wedge C) \leftrightarrow (\blacksquare B \wedge \blacksquare C)$
d. $A \vdash \blacklozenge \blacklozenge B \to \blacklozenge B$ *and* $A \vdash \blacksquare B \to \blacksquare \blacksquare B$.

Proof. Ad (a). Suppose $A \vdash B \to C$. Then,

$$(U + \lozenge\!\!\!\lozenge_A B) \rhd (A \wedge B) \rhd (A \wedge C).$$

So, $U + \lozenge\!\!\!\lozenge_A B \vdash \lozenge\!\!\!\lozenge_A C$, and, thus, $A \vdash \blacklozenge B \to \blacklozenge C$.

We note that $\neg \blacksquare \neg B$ is $\neg\neg \blacklozenge \neg\neg B$. So, by the first conjunct of (a), we have $U \vdash \blacklozenge B \leftrightarrow \neg \blacksquare \neg B$. As a consequence, we may switch between $\blacklozenge$-versions of principles and their dual $\blacksquare$-formulations in a confident way. Thus, we omit the verification of the second conjuncts of (a-d).

Principles (b) and (c) are immediate from Theorem 3.7.

Ad (d). We have

$$(U + \lozenge\!\!\!\lozenge_A \blacklozenge B) \rhd (A + \blacklozenge B) \rhd (U + \lozenge\!\!\!\lozenge_A B) \rhd (A \wedge B).$$

So, $U + \lozenge\!\!\!\lozenge_A \blacklozenge B \vdash \lozenge\!\!\!\lozenge_A B$. It follows that $A \vdash \blacklozenge \blacklozenge B \to \blacklozenge B$. ❑

Theorem 6.2. *Let $K : U \lhd A$ be an FM-interpretation. Then, Λ_K^{fr} is a normal modal logic extending K4, in other words, Λ_K^{fr} satisfies the Löb Conditions.*

Proof. It is sufficient to show that (a-d) of Lemma 6.1 imply the Löb Conditions. Clearly, L1, also known as Necessitation, follows from (a,b).

We verify L2. We have $A \vdash (B \wedge (B \to C)) \to C$. Ergo, by (a), we find $A \vdash \blacksquare (B \wedge (B \to C)) \to \blacksquare C$. Applying the second conjunct of (c), we obtain $A \vdash (\blacksquare B \wedge \blacksquare (B \to C)) \to \blacksquare C$.

Finally, L3 is identical to (d). ❑

We note that, conversely, the Löb Conditions imply (a-d) of Lemma 6.1.

We will see that it is possible to get as extensions Löb's Logic GL and the logics K45 and S4.

6.4. Löb's Logic

In what circumstances do we have interpreter logics that extend Löb's Logic GL? We provide a basic result concerning that question.

We say that theories U and V are *reconcilable* iff there are consistent finite extensions-in-the-same-language $U' \supseteq U$ and $V' \supseteq V$, such that $U' \bowtie V'$. The theories U and V are *irreconcilable* iff they are not reconcilable.

The simplest case of irreconcilability is when one of U, V is inconsistent. The second simplest case is when one of U, V is RE and essentially undecidable and the other is decidable and complete. Theorem 8.13 will tell us that, if one of A, U of an FR-frame $\langle A, U \rangle$ is sequential, then A and U are irreconcilable.

Before we turn to the proof of Löb's Theorem, let us briefly look at a trivial theorem that is close in spirit to the Second Incompleteness Theorem.

Theorem 6.3. *Suppose U and A are irreconcilable, U is Friedman-reflexive, and A is finitely axiomatised and consistent. Then $A \not\rhd (U + \lozenge_A \top)$.*

Proof. We assume the conditions of the theorem. Suppose $A \rhd (U + \lozenge_A \top)$. Then, $A \bowtie (U + \lozenge_A \top)$. This contradicts irreconcilability. ❑

Why is the above theorem *not* a version of the Second Incompleteness Theorem? It does not seem reasonable to say that it is, since none of Gödel's hard work (nor any new thing replacing it) occurs in the trivial proof. My proposal is to demand from a version of the Second Incompleteness Theorem in this style that it also supply *a concrete range* of irreconcilable U's and A's. Such a range will be provided by Theorem 8.13. Then, Corollary 8.15 will be our version of the Second Incompleteness Theorem.

The next theorem directly connects irreconcilability to Löb's Principle.

Theorem 6.4. *Consider an FM-frame $\langle A, U \rangle$. Then, A and U are irreconcilable iff $\Lambda_{A,U}^{\mathsf{fr}}$ extends Löb's Logic.*

Proof. Let $\langle A, U \rangle$ be an FM-frame.

We prove the left-to-right direction. Suppose A and U are irreconcilable. Consider any K such that $K : A \rhd U$. It is sufficient to show that Λ_K^{fr} extends Löb's Logic. It is well-known that, over K4, Löb's Principle and Löb's Rule are equivalent.[15] So, it suffices to prove closure of Λ_K^{fr} under Löb's Rule. Suppose $A \vdash B \to \blacklozenge_{K,A} B$. Then,

$$(A + B) \rhd (U + \lozenge_A B) \rhd (A + B).$$

So, $(A + B) \bowtie (U + \lozenge_A B)$ and, thus, by irreconcilability, $A + B$ is inconsistent, i.e., $A \vdash \neg B$.

We prove the right-to left direction. Suppose A and U are reconcilable. Suppose $(A + B) \bowtie (U + C)$, where $A + B$ is consistent.

Suppose that $K_0 : A \rhd U$ and $K_1 : (A + B) \rhd (U + C)$. We define $K := K_1 \langle B \rangle K_0$, i.e., the interpretation that is K_1 if B and K_0 otherwise. Clearly, $K : A \rhd U$ and $K' : (A + B) \rhd (U + C)$, where K' has the same underlying translation as K. Since $(U + C) \rhd (A + B)$, we have $U + C \vdash \lozenge_A B$ and, so, $A + B \vdash \blacklozenge_{K,A} B$. We also have $A \nvdash \neg B$. So, A is not closed under Löb's Rule. We have:

$$A \nvdash \blacksquare_{K,A}(B \to \blacklozenge_{K,A} B) \to \blacksquare_{K,A} \neg B,$$

by the fact that Löb's Rule follows from Löb's Principle. So, we do not have Löb's Principle in Λ_K^{fr}.[16] ❑

Remark 6.5. We note that Theorem 6.4 is a correspondence result for Löb's Principle. Here the FM-frame $\langle A, U \rangle$, is the analogue of a Kripke frame. The interpretation $K : A \rhd U$ in combination with a mapping σ from the propositional atoms to the sentences of the language of A is the analogue of a model on the frame.

○

---

[15] See [15, Chapter 1, Section 1, Exercise 5(iii), p75]. Smoryński attributes the result to Macintyre and Simmons. Alternatively, see [1, Chapter 3, p59].

[16] We note that the detour over Löb's Principle is necessary here. Closure of A under a rule implies closure of the associated logic under the same rule, but not *vice versa*. So, we need a principle to provide a counter-instance.

When we have Löb's Principle, we also have an analogue of Feferman's Theorem of the interpretability of inconsistency.

Theorem 6.6. *Suppose $\langle A, U \rangle$ is an FM-frame and A and U are irreconcilable. We have:*

a. *Suppose $K : A \triangleright U$. Then $A \triangleright (A + \blacksquare_{K,A} \bot)$.*
b. *Suppose $M : B \triangleright U$ and $A \triangleright B$. Then $A \triangleright (B + \blacksquare_{M,A} \bot)$.*

Proof. Ad (a). We have:

$$(A + \blacklozenge_{K,A} \top) \vdash (A + \blacklozenge_{K,A} \blacksquare_{K,A} \bot)$$
$$\triangleright (A + \blacksquare_{K,A} \bot)$$

Trivially, $(A + \neg \blacklozenge_{K,A} \top) \triangleright (A + \blacksquare_{K,A} \bot)$. By a disjunctive interpretation, we find $A \triangleright (A + \blacksquare_{K,A} \bot)$.

Ad (b). Suppose $P : A \triangleright B$. By (a), we have:

$$A \triangleright (A + \blacksquare_{M \circ P, A} \bot) \triangleright (B + \blacksquare_{M,A} \bot). \qquad \square$$

6.5. S4

An FM-interpretation $K : A \triangleright U$ is *companionable* iff, for every B of the A-language, there is a C of the U-language such that $(A + B) \bowtie (U + C)$, where the interpretation of $U + C$ in $A + B$ has the same underlying translation as K.

We can define companionship in terms of the category $\mathbb{E}$ enriched by designated arrows for finite extensions as indicated in the diagram below.

$$U + C \xleftarrow{\quad M \quad} A + B$$

$$\begin{array}{ccc} U + C & \xrightarrow{\quad K' \quad} & A + B \\ \subseteq \uparrow & & \uparrow \subseteq \\ U & \xrightarrow{\quad K \quad} & A \end{array}$$

Here we require no commutation for M.

Theorem 6.7. *Consider an FM-interpretation $K : A \triangleright U$. Then, K is companionable iff Λ_K^{fr} extends $\mathsf{S4}$.*

Proof. Suppose K is companionable. Consider any B and suppose $K' : (A + B) \triangleright (U + C)$, where K' is based on the same translation as K and that $(U + C) \triangleright (A + B)$. It follows that $U + C \vdash \lozenge_A B$. Hence, $A + B \vdash \blacklozenge_{K,A} B$.

Conversely, suppose Λ_K^{fr} contains the reflection principle, aka M. We have that $A + B \vdash \blacklozenge_{K,A} B$, so, there is an interpretation K' based on the same translation as K, such that $K' : (A + B) \triangleright (U + \lozenge_A B)$. Conversely, $(U + \lozenge_A B) \triangleright (A + B)$. $\qquad \square$

Remark 6.8. We note that the characterisation provided by Theorem 6.7 is rather different in nature from the one given of Löb's Principle in Theorem 6.4. First, in Theorem 6.7, we consider a property of interpretations rather than a property of frames as in Theorem 6.4. Secondly, we use more notions to formulate companionship than for irreconcilability. In Appendix D, we prove a result for the reflection principle that

is more in the spirit of Theorem 6.7. However, to do that we need to consider local interpreter logics of an FM-frame rather than the unique (global) interpreter logic of the frame.   ❍

Example 6.9. Let A be e.g. the theory of the ordering of the natural numbers. The theory A is finitely axiomatisable. Theorem B.1 will tell us that A is Friedman-reflexive. The identical interpretation Id_A of A in itself is clearly companionable. So, $\Lambda^{\mathsf{fr}}_{\mathsf{Id}_A}$ extends S_4.

In fact, we can show that the modality trivialises for this example. Consider any B in the A-language. The sentence $\blacklozenge_{\mathsf{Id}_A, A} B$ is either provable or refutable in A. If it is refutable, we have $A \vdash \blacklozenge_{\mathsf{Id}_A, A} B \to B$. Suppose it is provable. So, $A \vdash \blacklozenge_{\mathsf{Id}_A, A} B$. It follows that $A \rhd (A + B)$. So, $A + B$ is consistent, and, hence, B is provable in A. Thus, $A \vdash \blacklozenge_{\mathsf{Id}_A, A} B \to B$. So, in both cases, we have $A \vdash \blacklozenge_{\mathsf{Id}_A, A} B \to B$.   ❍

Open Question 6.10. Can we find a more inspiring example of a theory with logic $\mathsf{S}4$ than Example 6.9? Is it perhaps possible to find an FM-interpretation with interpreter logic precisely $\mathsf{S}4$?   ❍

Corollary 6.11. *Suppose the FM-interpretation $U \xrightarrow{K} A$ is a retraction in $\mathbb{E}$. Then, K is companionable and, hence, Λ^{fr}_K extends $\mathsf{S}4$.*

Proof. Suppose $U \xrightarrow{K} A$ is an FM-interpretation which is a retraction in $\mathbb{E}$. Let M be the corresponding section, i.e., $K \circ M = \mathsf{Id}_A$. Consider any B in the A-language. We have: $(U + B^M) \rhd (A + B)$. Moreover, writing $\equiv$ for having the same theorems, we have:

$$(U + B^M) \xrightarrow{K'} (A + B^{MK})$$
$$\equiv (A + B)$$

Here $K' : (A + B) \rhd (U + B^M)$ has the same underlying translation as K.   ❏

In Appendix C, we will treat a notion of sameness between interpretations that leads to a notion of sameness of interpreter logics.

7. Essentially Sententially Reflexive Theories

In this section we study essential sentential reflexiveness. A theory U is *essentially sententially reflexive* if, for some N, we have $N : U \rhd \mathsf{S}^1_2$ and, for all U-sentences A and for all n, $U \vdash \square_n^N A \to A$. Here A ranges over U-sentences and $\square_n$ means provability in predicate logic using proofs which only involve formulas with depth of quantifier alternations $\leq n$. As a default, we assume in our notation that n exceeds $\rho(A)$, the depth of quantifier alternations of A. The definition of $\square_{n,A} B$ is similar, where now we consider provability from A and not just from predicate logic. (See e.g. [32] for details about the notion of depth of quantifier alternations.)

We will write $\triangle_A B$ for $\square_{\rho(A \to B), A} B$ and $\nabla_A B$ for $\lozenge_{\rho(A \wedge B), A} B$.

7.1. A basic fact

We have the following theorems. The second result provides a coordinate-free characterisation of Essentially Sententially Reflexive Theories in the sequential case.

Theorem 7.1. *Suppose U is essentially sententially reflexive. Then, there is an $N : U \rhd \mathsf{S}^1_2$, such that, for all Σ^0_1-sentences S and for all $M : U \rhd \mathsf{S}^1_2$, we have $U \vdash S^N \to S^M$.*

Proof. Suppose that U is essentially sententially reflexive with witness $N_0 : U \rhd \mathsf{S}_2^1$. We start with the observation that, for all U-sentences A and for all $m \geq \rho(A)$, we have $U \vdash \Box_m^{N_0} A \to A$. This follows immediately by replacing A by $(A \wedge B)$, where B is a tautology with $\rho(B) = m$. Let N be a logarithmic cut of N_0. Consider any $M : U \rhd \mathsf{S}_2^1$. We have, for a sufficiently large m and for some U-theorem D, that $U \vdash S^N \to \Box_{m,D}^{N_0} S^M$ (see Theorem A.1). Here we can take $D := (E_{\mathsf{S}_2^1} \wedge \bigwedge \mathsf{S}_2^1)$. It follows that $U \vdash S^N \to S^M$. ❑

Theorem 7.2. *Suppose U is sequential. Then the following conditions are equivalent.*

a. *U is essentially sententially reflexive.*
b. *There is an $N : U \rhd \mathsf{S}_2^1$, such that for all Σ_1^0-sentences S and all $M : U \rhd \mathsf{S}_2^1$, we have $U \vdash S^N \to S^M$.*
c. *Consider any $N^* : U \rhd \mathsf{S}_2^1$. There is an N^*-cut I, such that, for all Σ_1^0-sentences S and all N^*-cuts J, we have $U \vdash S^I \to S^J$.*

Proof. Suppose U is sequential.

Theorem 7.1 tells us that (a) implies (b). We prove the other direction. Suppose N witnesses (b). Consider any U-sentence A. Since U is sequential, there is an N-cut I, such that $U \vdash \triangle^I A \to A$. Since we have $U \vdash \triangle^N A \to \triangle^I A$, we find $U \vdash \triangle^N A \to A$. So N witnesses the essential sentential reflexivity of U.

We prove the implication from (b) to (c). Let N witness (b). Consider any $N^* : U \rhd \mathsf{S}_2^1$. By a result of Pudlák, there is an N-cut I and an N^*-cut I^*, such that I and I^* are U-definably, U-provably isomorphic. We take I^* as our witness for (c). Reason in U. Let J be any N^*-cut. Suppose S^{I^*}. Then, S^I, and, hence S^N. It follows that S^J.

We prove the implication from (c) to (b). We take as witness for (a), the N^*-cut I promised by (c). Consider any $M : U \rhd \mathsf{S}_2^1$. Let J and J^* be U-definably, U-provably isomorphic cuts of M and I. Reason in U. Suppose S^I. Then, S^{J^*}. So, S^J, and, hence, S^M. ❑

7.2. Essential sentential reflexiveness implies Friedman-reflexiveness

We have the following theorem.

Theorem 7.3. *Suppose U is essentially sententially reflexive with witnessing interpretation N. Then, $\Diamond A := \nabla_A^N \top$ witnesses that U is effectively Friedman-reflexive.*

The argument is, of course, just Harvey Friedman's argument for the case of PA.

Proof. Suppose N witnesses the sentential essential reflexiveness of U.

Suppose $(U + B) \rhd A$. Then, for some finite sub-theory D of U, we have that $(D + B) \rhd A$. It follows that $U \vdash ((D + B) \rhd A)^N$ and, so, $U \vdash \Diamond_{m,D}^N B \to \nabla_A^N \top$, for sufficiently large m. By the sentential essential reflexiveness of U, it follows that $U + B \vdash \nabla_A^N \top$.

For the opposite direction, suppose $U + B \vdash \nabla_A^N \top$. Then, by the Interpretation Existence Lemma, we find $(U + B) \rhd A$. ❑

7.3. Peano Corto

In this subsection, we discuss the sententially essentially reflexive theory $\mathsf{PA}^{\downarrow\downarrow}$. We introduce two other examples of sententially essentially reflexive theories, however, since we mainly want to illustrate the idea of a weak sententially essentially reflexive theory, it seemed good to zoom in om $\mathsf{PA}^{\downarrow\downarrow}$.

In our paper [26], we used PA^- as starting point. Here, we use S_2^1 as starting point since it fits the set-up of the present paper better.

We introduce $\mathsf{PA}^{\downarrow\downarrow}$, its little brother $\mathsf{PA}^{\downarrow\downarrow\downarrow}$, and its big brother $\mathsf{PA}^{\downarrow}$.

- Peanissimo or $\mathsf{PA}^{\downarrow\downarrow\downarrow}$ is the theory

$$\mathsf{S}_2^1 + \{(S \to S^I) \mid S \text{ is an } \exists\Sigma_1^{\mathsf{b}}\text{-sentence and } I \text{ is an } \mathsf{S}_2^1\text{-cut}\}.$$

  This theory is identical to $\mathsf{S}_2^1 + \{\triangle A \to A \mid A \text{ is an arithmetical sentence}\}$.
- Peano Corto or $\mathsf{PA}^{\downarrow\downarrow}$ is the theory

$$\mathsf{S}_2^1 + \{(S \to S^I) \mid S \text{ is a } \Sigma_1^0\text{-sentence and } I \text{ is an } \mathsf{S}_2^1\text{-cut}\}.$$

  Here Σ_1^0 means a block of existential quantifiers followed by a Δ_0-formula. In fact, Peano Corto is Peanissimo plus the scheme

$$S \to \exists x\, \exists z\, (2^x = z \wedge S_0(x)),$$

  where $S = \exists x\, S_0(x)$.
- Peano Basso or $\mathsf{PA}^{\downarrow}$ is the theory

$$\mathsf{S}_2^1 + \{(S \to S^I) \mid S \text{ is a } \Sigma_{1,\infty}^0\text{-sentence and } I \text{ is an } \mathsf{S}_2^1\text{-cut}\}.$$

  Here $\Sigma_{1,\infty}^0$ is the class of formulas given by a block of existential quantifiers and bounded universal quantifiers, where both sorts may occur in an alternating way, followed by a Δ_0-formula. In [26], it is shown that Peano Basso is Peano Corto plus Σ_1^0-collection.

In [26], it is shown that Peano Corto and Peano Basso are essentially sententially reflexive w.r.t. the identical cut. A similar argument shows the same for Peanissimo.

Consider arithmetical theories V and W. We confuse the translation and the interpretation based on relativisation of quantifiers to a W-cut I with I. We say that V is *cut-interpretable* in W or $W \rhd_{\mathsf{cut}} V$, if W interprets V using a translation based on a W-cut. Similarly, we say that V is *locally cut-interpretable* in W or $W \rhd_{\mathsf{cut,loc}} V$, if W locally interprets V using only interpretations based on W-cuts. See [26] for more information.

All three theories Peanissimo, Peano Corto and Peano Basso are locally cut-interpretable in S_2^1, i.o.w., $\mathsf{S}_2^1 \rhd_{\mathsf{cut,loc}} \mathsf{PA}^{\downarrow\downarrow\downarrow}$ and $\mathsf{S}_2^1 \rhd_{\mathsf{cut,loc}} \mathsf{PA}^{\downarrow\downarrow}$ and $\mathsf{S}_2^1 \rhd_{\mathsf{cut,loc}} \mathsf{PA}^{\downarrow}$. Also, all three theories are mutually cut-interpretable.

We remind the reader that each theory is recursively axiomatisable, since we can replace the cuts I in our formulation by $E\langle\mathsf{cut}_x(E)\rangle(x = x)$, where E ranges over all formulas with at most the free variable x. Here $\mathsf{cut}_x(E)$ is the S_2^1-sentence that expresses '$\{x \mid E(x)\}$ is a cut' and $F\langle G\rangle H$ is $((G \to F) \wedge (\neg G \to H))$.

Since, the identical cut is the designated cut, we can, by Theorem 7.3, take $\diamondsuit A := \nabla_A\top$ in each of our theories.

Theorem 7.4. *Suppose P is a Π_1^0 interpreter of A over Peano Corto. Then,* $\mathsf{EA} \vdash P \leftrightarrow \nabla_A\top$.

Proof. Suppose P is a Π_1^0 interpreter of A over $\mathsf{PA}^{\downarrow\downarrow}$. Then, by uniqueness, we have $\mathsf{PA}^{\downarrow\downarrow} \vdash P \leftrightarrow \nabla_A\top$. So, for some S_2^1-cut J, we have $\mathsf{S}_2^1 \vdash (P \leftrightarrow \nabla_A\top)^J$. Hence, by a meta-theorem of Paris and Wilkie, we have $\mathsf{EA} \vdash P \leftrightarrow \nabla_A\top$. Application of the meta-theorem uses that $(P \leftrightarrow \nabla_A\top)$ is equivalent to a Π_2^0-sentence. See [34] and [18]. □

So, we have characterised $\nabla_A\top$ as a Π_1^0-sentence up to EA-provable equivalence in a coordinate-free way. This improves the result of [22], where this was only done for finitely axiomatised *sequential* theories.

We have a version of the Friedman characterisation over Peano Corto.

Theorem 7.5. *Suppose A is sequential. Then,*

$$A \rhd B \quad \text{iff} \quad \mathsf{PA}^{\downarrow\downarrow} \vdash \triangledown_A\top \to \triangledown_B\top.$$

Proof. The left-to-right direction is just Theorem 3.5 in combination with the fact that we can take $\diamondsuit A :=$ $\triangledown_A\top$ over Peano Corto.

From right to left: suppose $\mathsf{PA}^{\downarrow\downarrow} \vdash \triangledown_A\top \to \triangledown_B\top$. Then, $\mathsf{S}_2^1 \vdash (\triangledown_A\top \to \triangledown_B\top)^I$, for some cut I. So, since $\mathsf{S}_2^1 \vdash \triangledown_A\top \to \triangledown_A^I\top$, we find $\mathsf{S}_2^1 \vdash \triangledown_A\top \to \triangledown_B^I\top$. We have:

$$A \rhd (\mathsf{S}_2^1 + \triangledown_A\top) \rhd (\mathsf{S}_2^1 + \triangledown_B\top) \rhd B.$$

The first step uses the sequentiality of A. □

Remark 7.6. The original Friedman characterisation had EA in place of Peano Corto. A result due to Paris and Wilkie (see [34] and [18]) shows that we have, for Π_1^0-sentences P and Q:

$$\mathsf{PA}^{\downarrow\downarrow} \vdash P \to Q \quad \Leftrightarrow \quad \mathsf{EA} \vdash P \to Q.$$

So, the connection between the two results is obvious. However, the *internal* version of the characterisation in EA needs cut-free EA-provability and ordinary $\mathsf{PA}^{\downarrow\downarrow}$-provability. ○

Remark 7.7. The theory $\mathsf{Seq}(V)$ is specified as follows. We add a unary predicate D and a binary predicate $\in$ to the signature of V, we relativise V to D and we add the (unrelativised) axioms for Adjunctive Set Theory AS plus an axiom that states that every element of D is an empty set. We can show that Seq supports a functor from $\mathbb{D}$ to $\mathbb{D}_{\mathsf{seq}}$ and from $\mathbb{D}_{\mathsf{fin}}$ to $\mathbb{D}_{\mathsf{fin,seq}}$. See Appendix A for details. It is easily seen that we have $\mathsf{Seq}(A) \bowtie (\mathsf{S}_2^1 + \triangledown_A\top)$.

We see that we can split the functor H of Remark 3.6 in two stages. First, we have a projection π of $\mathbb{D}_{\mathsf{fin}}$ to $\mathbb{D}_{\mathsf{fin,seq}}$. This can be either $A \mapsto \mathsf{Seq}(A)$ or $A \mapsto (\mathsf{S}_2^1 + \triangledown_A\top)$. Then, we have a (lax) embedding of $\mathbb{D}_{\mathsf{fin,seq}}$ into $\mathbb{B}_{\mathsf{PA}^{\downarrow\downarrow}}$. ○

Suppose $K : A \rhd \mathsf{PA}^{\downarrow\downarrow}$. Since, we can apply the Gödel Fixed Point Theorem in the usual way because $\diamondsuit$ can be represented by a predicate, we have Löb's Logic. This also follows from Corollary 8.16 in combination with the fact that $\mathsf{PA}^{\downarrow\downarrow}$ is sequential. That theorem, however, has a more involved proof.

If K is Σ_1^0-sound, Λ_K^{fr} is precisely Löb's Logic. In the case of $\mathsf{PA}^{\downarrow\downarrow}$ we can verify Solovay's Theorem simply using Solovay's proof. The reason is that $\mathsf{PA}^{\downarrow\downarrow}$ proves that $\exists x, y\, (2^{2^x} = y \wedge S_0(x))$ from $\exists x\, S_0(x)$, where S_0 is Δ_0 or $\Delta_0(\omega_1)$. This delivers Σ_1^0-completeness. This argument is not present for $\mathsf{PA}^{\downarrow\downarrow\downarrow}$. However, we still have Solovay's Theorem for $\mathsf{PA}^{\downarrow\downarrow\downarrow}$ as a special case of Theorem 8.19.

We can see that Peano Corto has some definite advantages over S_2^1 in the role of base theory. We have a coordinate-free representation of the interpreter variant of provability. Moreover, we have the insights contained in Theorems 7.4 and 7.5 and the good properties of the interpreter logics over Peano Corto. However, there is a down-side too.

  I. Peano Corto is not finitely axiomatisable.
 II. Peano Corto is not interpretable in S_2^1. If it were, it would be mutually interpretable with S_2^1 and this contradicts Theorem 8.13. In fact, no sequential Friedman-reflexive theory is interpretable in S_2^1 by the same argument. As a consequence, there are no interpreter logics for S_2^1 with Peano Corto as base (or, with any sequential Friedman-reflexive theory as base).
III. Even if Peano Corto is interpretable in some reasonably weak concrete A, like EA, it is not always clear that we can find an interpretation that does not involve arithmetisation. We discuss this kind of problem in Section 9.

8. Friedman-reflexivity meets sequentiality

We already met some specific sequential Friedman-reflexive theories that are essentially sententially reflexive. In this section, we look at sequential theories that are Friedman-reflexive in general. Moreover, we look at interpreter logics for sequential A, also in cases where the base is not itself sequential.

8.1. Characterisation

In this subsection, we provide characterisations both of the interpreters provided by sequential Friedman-reflexive bases and of such bases themselves.

We show that $\Diamond A$ always has the form of a restricted consistency statement of A on some cut.

Theorem 8.1. *Suppose U is sequential and Friedman-reflexive. Let $N : \mathsf{S}^1_2 \lhd U$. Then, for some N-cut I, we have $U \vdash \Diamond A \leftrightarrow \triangledown^I_A \top$.*

In case U is RE and effectively Friedman-reflexive, we can find I effectively from A.

Proof. We have $(U + \Diamond A) \rhd A$. Let K be a witnessing interpretation. It follows, for some N-cut I, that $(U + \Diamond A) \vdash \triangledown^I_A \top$. We can see that by choosing I so short that we can verify reflection-inside-K for proofs involving only formulas of ρ-complexity $\leq m := \rho(A) + \rho(K)$ w.r.t. a truth-predicate for formulas of ρ-complexity $\leq m$. This truth-predicate works on an appropriate N-cut I^*. We choose I smaller than I^*. So, we have $U + \Diamond A \vdash \triangledown^I_A \top$.

Conversely, since, $(U + \triangledown^I_A \top) \rhd A$, it follows that $U + \triangledown^I_A \top \vdash \Diamond A$.

Trivially, I can be effectively found when $\Diamond A$ is given and U is RE.   □

We provide a characterisation of Friedman-reflexivity in the sequential case.

Theorem 8.2. *Suppose U is sequential.*

A. The following are equivalent.

> *a. U is Friedman-reflexive.*
> *b. For all Σ^0_1-sentences S, there is an $N : U \rhd \mathsf{S}^1_2$, such that, for all $M : U \rhd \mathsf{S}^1_2$, we have $U \vdash S^N \to S^M$.*
> *c. Consider any $N^* : \mathsf{S}^1_2 \lhd U$. Then, for all Σ^0_1-sentences S, there is an N^*-cut I such that for all N^*-cuts J, we have $U \vdash S^I \to S^J$.*

B. The following are equivalent.

> *a. U is effectively Friedman-reflexive.*
> *b. There is a recursive function F such that, for all Σ^0_1-sentences S, we have $F(S) = N : U \rhd \mathsf{S}^1_2$ and, for all $M : U \rhd \mathsf{S}^1_2$, we have $U \vdash S^N \to S^M$.*
> *c. Consider any $N^* : \mathsf{S}^1_2 \lhd U$. There is a recursive function G such that, for all Σ^0_1-sentences S, we have $G(S) = I$, where I is an N^*-cut such that for all N^*-cuts J, we have $U \vdash S^I \to S^J$.*

Proof. We will just prove the equivalence between (Aa) and (Ac). The equivalence between (Ab) and (Ac) is immediate using the fact that any two interpretations of S^1_2 in U have U-definably, U-verifiably isomorphic cuts. The proof of (B) is by inspection of the proof of (A).

Suppose that U is Friedman-reflexive and $N^* : U \rhd \mathsf{S}^1_2$. Consider any Σ^0_1-sentence S. We note that $A := \mathsf{S}^1_2 + \neg S$ is finitely axiomatised. Let I_0 be the N-cut such that $\triangledown^{I_0}_{\mathsf{S}^1_2} \neg S$ is U-provably equivalent to $\Diamond A$. We find $U \vdash \triangledown^J_{\mathsf{S}^1_2} \neg S \to \triangledown^{I_0}_{\mathsf{S}^1_2} \neg S$, for all N-cuts J. Thus, $U \vdash \triangle^{I_0}_{\mathsf{S}^1_2} S \to \triangle^J_{\mathsf{S}^1_2} S$, for all N-cuts J.

Let I be an N-cut so that $U \vdash S^I \to \triangle^{I_0}_{\mathsf{S}^1_2} S$: see Theorem A.1. Consider any N-cut J. By sequentiality, we can find a J-cut J_0 so that $U \vdash \triangle^{J_0}_{\mathsf{S}^1_2} S \to S^J$. This uses again a soundness proof involving a truth-predicate. Putting everything together we find:

$$U \vdash S^I \to \triangle^{I_0}_{\mathsf{S}^1_2} S$$
$$\to \triangle^{J_0}_{\mathsf{S}^1_2} S$$
$$\to S^J$$

Conversely, suppose U satisfies (c). Let I be the N^*-cut guaranteed by (c) for $S := \triangle_A \bot$. Suppose $(U + B) \rhd A$. Then, for some N-cut J, we have $U + B \vdash \triangledown^J_A \top$ and hence $U + B \vdash \triangledown^I_A \top$. The other direction is immediate by Interpretation Existence.   ❑

Remark 8.3. We note that a complete and consistent sequential theory will automatically have property (c) of Theorem 8.2. It should, by Theorems B.1 and 8.2.   ◯

Remark 8.4. Our result is rather robust for the precise notion of Σ^0_1 used. The result works both for smaller classes and for larger ones.

It works for Σ^b_1 and even for Diophantine sentences consisting of a block of existential quantifiers followed by an equation $t = u$.

In the other direction, the result also applies when we admit ω_1-terms in our definition of Σ^0_1. Finally, it works when we define our Σ^0_1-analogue X as follows:

- $X ::= \top \mid \bot \mid t = u \mid \neg Y \mid (X \wedge X) \mid (X \vee X) \mid (Y \to X) \mid \forall x < t\, X \mid \exists x\, X$
- $Y ::= \top \mid \bot \mid t = u \mid \neg X \mid (Y \wedge Y) \mid (Y \vee Y) \mid (X \to Y) \mid \exists x < t\, Y \mid \forall x\, Y$ ◯

We give a slightly modified version of our characterisation.

Theorem 8.5. *Suppose U is sequential and let $N : U \rhd \mathsf{S}^1_2$. Then, U is Friedman-reflexive iff (†) for all Σ^0_1-sentences S, there is a U-sentence A, such that, for all U-sentences B, we have $U + \{S^I \mid I \text{ is an } N\text{-cut}\} \vdash B$ iff $U + A \vdash B$.*

Proof. Suppose U is Friedman-reflexive. Let I^* be the cut guaranteed for S by Theorem 8.2(c). Then, it is easy to see that S^{I^*} can be chosen as our A to satisfy (†).

Conversely, suppose (†). It is clear that $U + A$ proves all S^I. On the other hand, taking $B := A$, we see that some finite conjunction of the S^I will imply A over U. We now take J the intersection of all cuts occurring in this finite conjunction. We find that $U \vdash A \leftrightarrow S^J$. We take J as witness for satisfaction of the characterisation of Theorem 8.2(c).   ❑

We give a final version of our characterisation that is both useful and enlightening. We need the notion of intersection of all cuts. Consider any sequential model $\mathcal{M}$. We define $\mathfrak{I}_{\mathcal{M}}$ as follows. First, we choose an internal model $\mathcal{N}$ of S^1_2 and then we take $\mathfrak{I}_{\mathcal{M}}$ to be the intersection of all $\mathcal{M}$-definable $\mathcal{N}$-cuts. Using elementary facts about sequentiality, one can easily show that $\mathfrak{I}_{\mathcal{M}}$ is independent of the choice of $\mathcal{N}$ in the sense that all versions are isomorphic by $\mathcal{M}$-definable isomorphism. Moreover, this isomorphism is unique when restricted to $\mathfrak{I}_{\mathcal{M}}$. See also [32, Section 5.1].

Consider a sequential theory U and let $N : \mathsf{S}^1_2 \lhd U$. We extend the language of U with a new unary predicate $\mathfrak{I}$ that is interpreted in each U-model $\mathcal{M}$ as $\mathfrak{I}_{\mathcal{M}}$. Here we think of $\mathfrak{I}_{\mathcal{M}}$ as given by $\mathcal{N} := N^{\mathcal{M}}$. Let U^{e} be the set of all sentences in the extended language true in all $\mathcal{M}, \mathfrak{I}_{\mathcal{M}}$, where $\mathcal{M}$ is an U-model. Let $\mathcal{I}_U$ be the set of arithmetical sentences A such that $U^{\mathsf{e}} \vdash A^{\mathfrak{I}}$.

An important insight is that $\mathcal{I}_U$ contains $\mathsf{EA} + \mathsf{B}\Sigma_1$. See [32, Section 5.1].

Theorem 8.6. *Suppose U is sequential and let $N : U \rhd \mathsf{S}^1_2$. Then, U is Friedman-reflexive iff, for all Σ^0_1-sentences S, there is a U-sentence B, such that we have $U^{\mathsf{e}} \vdash S^{\mathfrak{I}} \leftrightarrow B$. Moreover, B can always be taken to be of the form S^I for some N-cut I.*

U is effectively Friedman-reflexive iff we can find B (or, if you wish, I) effectively.

Proof. Suppose U is sequential. If U is Friedman-reflexive, then, the S^I provided by Theorem 8.1 gives us the B we are looking for.

Suppose $U^{\mathsf{e}} \vdash S^{\mathfrak{I}} \leftrightarrow B$. Then, $U + \{S^I \mid I \text{ is an } N\text{-cut}\} \vdash B$ and, conversely, $U + B \vdash S^I$, for all N-cuts I. By Theorem 8.5, it follows that U is Friedman-reflexive. Clearly, if we can find the B effectively, then U is effectively Friedman-reflexive. $\quad\square$

So if we view the $S^{\mathfrak{I}}$ as a second-order or as an infinitary statement, then Friedman-reflexiveness means a reduction of a certain second-order or infinitary statement to a first-order finitary statement.

8.2. An example: the theory DA

We provide an example of an effectively Friedman reflexive theory that is not essentially sententially reflexive. We call the theory of our example DA (Descending Arithmetic). Giving it a name does make it seem like a definite thing. So, it is good to point out that the theory does depend on two arbitrarily chosen enumerations.

Let $S_0, S_1, \ldots$ effectively enumerate the Σ^0_1-sentences and let $I_0, I_1, \ldots$ be an effective enumeration of S^1_2-cuts such that $\mathsf{S}^1_2 \vdash I_{n+1} \subseteq I_n$ and such that, for each S^1_2-cut J, we can find a k such that $\mathsf{S}^1_2 \vdash I_k \subseteq J$. Briefly said, $(I_k)_{k \in \omega}$ is effective, descending, and co-initial with all cuts.

We note that, because, in sequential theories, we have truth-predicates for formulas with ρ-complexity below a given number, we can take I_n to be the intersection of all definable cuts with ρ-complexity $\leq n$.

Let DA be $\mathsf{S}^1_2 + \{S^{I_i}_i \to S^J_i \mid i \in \omega \text{ and } J \text{ is a definable cut}\}$. Clearly, DA is effectively Friedman-reflexive. We note that DA is a sub-theory of $\mathsf{PA}^{\downarrow\downarrow}$ and, thus, locally cut-interpretable in S^1_2.

Let us say that a theory V is *restrictedly* Friedman-reflexive iff there is an n and a mapping $A \mapsto \lozenge\!\!\!\lozenge A$, where $\rho(\lozenge\!\!\!\lozenge A) \leq n$, for all A. It is easy to see that in case V is a *sequential* restrictedly Friedman-reflexive theory, then, for any $N : V \rhd \mathsf{S}^1_2$, there is a formula $C(x)$ such that, for all A, we have $V \vdash \lozenge\!\!\!\lozenge A \leftrightarrow C(\ulcorner A \urcorner)$, where the Gödel numbers are chosen w.r.t. N. Another immediate insight is that, if V is essentially sententially reflexive, then V is restrictedly Friedman-reflexive.

Theorem 8.7. *The theory DA is not restrictedly Friedman-reflexive and, hence, not essentially sententially reflexive.*

Proof. Suppose DA were restrictedly Friedman-reflexive with bound k_0. Let C be such that we have $\mathsf{DA} \vdash \lozenge\!\!\!\lozenge A \leftrightarrow C(\ulcorner A \urcorner)$ and let $k_0 := \rho(C) + 1$. Let $\rho(\mathsf{S}^1_2) = k_1$. Suppose I_p is the first logarithmic cut in the sequence. And let k_2 be the maximum of the ρ-complexities of the S_i for $i < p$. Finally, let k_3 be the complexity of a standard Σ^0_1-truth predicate true. Let k be the maximum of k_0, k_1, k_2, k_3. We pick n so large that I_n is Σ^0_1-sound for every consistent extension of S^1_2 with complexity $\leq k$. The existence of such a cut is guaranteed by Theorem A.2. We may assume that $n > p$. Let

$$A := \mathsf{S}^1_2 + \{\neg S_i \mid i < p \text{ and } S_i \text{ is false}\} + \{\neg\,\mathsf{true}(S_j) \mid p \leq j < n \text{ and } S_j \text{ is false}\}$$

and let $B := \mathsf{S}^1_2 + \lozenge_A \top$. We note that $\rho(A) \leq k$.

We claim that $A + \neg C(\ulcorner \underline{B} \urcorner)$ is consistent. Suppose it were not. Then, we would have $A \vdash C(\ulcorner \underline{B} \urcorner)$. It follows that $\mathsf{DA} + A \vdash \diamondsuit B$, and, hence, $(\mathsf{DA} + A) \rhd B$. Since, A locally cut-interprets $\mathsf{PA}^{\downarrow\downarrow} + A$ and, hence, $\mathsf{DA} + A$, we find that $A \rhd B$. *Quod non*, by the usual no-interpretation version of G2.

By the special property of I_n, it follows that

$$V := A + \neg C(\ulcorner \underline{B} \urcorner) + \{\neg S_i^{I_n} \mid i \geq n \text{ and } S_i \text{ is false}\}$$

is consistent. By Theorem A.1, we have $\mathsf{S}_2^1 \vdash \neg\,\mathsf{true}(S) \to \neg\,S^{I_p}$, for any Σ_1^0-sentence S. It follows that V extends DA. Hence, V is Friedman-reflexive with $\diamondsuit$ as selection function and $V \vdash \neg\diamondsuit B$. Since V proves every true Π_1^0-sentence, including $\diamondsuit_B \top$, on I_n, we find $V \rhd B$ and, hence, $V \vdash \diamondsuit B$. A contradiction. ❏

Open Question 8.8.

  i. Is DA reflexive? If, against expectation, it turns out to be reflexive, can we modify the construction to find a non-reflexive, Friedman-reflexive, sequential theory?
 ii. Is there a finitely axiomatised A and $K : A \rhd \mathsf{DA}$, such that, for no $D(x)$ in the A-language, we have, for all B in the A-language, $A \vdash D(\ulcorner \underline{B} \urcorner) \leftrightarrow \blacklozenge_{K,A} B$? Here the numerals are the K-numerals.
iii. Is there an RE sequential theory that is Friedman-reflexive but not effectively so?
 iv. Suppose U is sequential and restrictedly (effectively) Friedman-reflexive. Does it follow that U is essentially sententially reflexive?

We note that our proof of Theorem 8.7 uses special features of DA. So, the proof does not generalise, in an obvious way, to a proof of a positive answer to (iv). ○

8.3. Constraints

In this subsection, we prove two results that constrain the form of consistent, sequential, Friedman-reflexive theories.

Theorem 8.9. *Suppose U is consistent, Friedman-reflexive, sequential and RE. Then, any axiomatisation of U must have axioms of ρ-complexity $> n$, for any n.*

Proof. Suppose that U is consistent, Friedman-reflexive, sequential and RE and that U has a restricted axiomatisation. We fix $N : U \rhd \mathsf{S}_2^1$. Let A be any consistent finitely axiomatised theory such that $U \not\rhd A$. We find that $U \not\vdash \diamondsuit A$. So, $U + \neg \diamondsuit A$ is consistent.

We have, for all N-cuts I, that $(U + \diamondsuit_A^I \top) \rhd A$. So, $U + \diamondsuit_A^I \top \vdash \diamondsuit A$. It follows that $U + \neg \diamondsuit A \vdash \square_A^I \bot$. So, by Theorem A.2, we find that $\square_A \bot$ is true and, thus, that A is inconsistent. *Quod non.* ❏

Example 8.10. Neither PRA, nor $\mathsf{I}\Sigma_n$, nor ACA_0 is Friedman-reflexive. However, EA plus all true Π_1^0-sentences has a restricted axiomatisation and is consistent and Friedman-reflexive. So, the condition that U is RE in Theorem 8.9 is really needed.[17] ○

Remark 8.11. We note that we could, alternatively, have framed the proof of Theorem 8.9 as follows. Theorem A.2 tells us that each consistent finite extension of U tolerates S_2^1 plus all true Π_1-sentences and, hence, is polyglottic. On the other hand, clearly, there is an A such that $U \not\rhd A$ and, so, $U + \neg \diamondsuit A$ is consistent. By Theorem 5.2, this last theory is not polyglottic. A contradiction. ○

---

[17] I thank Leszek Kołodziejscyk for pointing out this example.

Example 8.12. Since, Peano Corto is both reflexive and mutually locally interpretable with S^1_2, we find that $\mathsf{PA}^{\downarrow\downarrow} \bowtie \mho(\mathsf{S}^1_2)$. Also, $\mho(\mathsf{S}^1_2)$ is restricted and, hence, not Friedman-reflexive. Ergo, Friedman-reflexivity is not preserved by mutual interpretability.   ⃝

We proceed with a central result of this paper.

Theorem 8.13. *Suppose A is finitely axiomatised and U is Friedman-reflexive. Suppose further that one of A, U is sequential. Then, A and U are irreconcilable.*

Proof. Suppose A is finitely axiomatised and U is Friedman-reflexive. Since, both finite axiomatisability and Friedman-reflexiveness are closed under finite extensions, it is sufficient to show that A and U, if consistent, are not mutually interpretable. Suppose A and U are consistent.

We first consider the case, where A is sequential. Suppose $K : U \rhd A$ and $M : A \rhd U$. Consider any finitely axiomatised B such that B is consistent and $A \not\rhd B$. For example, we could take $B := (\mathsf{S}^1_2 + \Diamond_A \top)$.

We have (a) $A + \neg \diamondsuit\!\!\!\!\diamondsuit^M B$ is consistent, since otherwise

$$A \vdash (A + \diamondsuit\!\!\!\!\diamondsuit^M B) \rhd (U + \diamondsuit\!\!\!\!\diamondsuit B) \rhd B.$$

Quod non. We have:

$$(U + \diamondsuit\!\!\!\!\diamondsuit^{MK} B) \rhd (A + \diamondsuit\!\!\!\!\diamondsuit^M B) \rhd (U + \diamondsuit\!\!\!\!\diamondsuit B) \rhd B.$$

So, $U + \diamondsuit\!\!\!\!\diamondsuit^{MK} B \vdash \diamondsuit\!\!\!\!\diamondsuit B$. Ergo, $A + \diamondsuit\!\!\!\!\diamondsuit^{MKM} B \vdash \diamondsuit\!\!\!\!\diamondsuit^M B$, and hence, $A + \neg \diamondsuit\!\!\!\!\diamondsuit^M B \vdash \neg \diamondsuit\!\!\!\!\diamondsuit^{MKM} B$. It follows that (b) $(M \circ K)' : (A + \neg \diamondsuit\!\!\!\!\diamondsuit^M B) \rhd (A + \neg \diamondsuit\!\!\!\!\diamondsuit^M B)$, where $(M \circ K)'$ is the interpretation based on the same translation as $M \circ K$.

Suppose $N : A \rhd \mathsf{S}^1_2$. Let I be any N-cut in A. We have

$$(U + \Diamond^{IK}_B \top) \rhd (A + \Diamond^I_B \top) \rhd (\mathsf{S}^1_2 + \Diamond_B \top) \rhd B.$$

So $U + \Diamond^{IK}_B \top \vdash \diamondsuit\!\!\!\!\diamondsuit B$. It follows that $U + \neg \diamondsuit\!\!\!\!\diamondsuit B \vdash \Box^{IK}_B \bot$, and, hence, we have $A + \neg \diamondsuit\!\!\!\!\diamondsuit^M B \vdash \Box^{IKM}_B \bot$.

We have shown: (c) for all N-cuts I, we have $A + \neg \diamondsuit\!\!\!\!\diamondsuit^M B \vdash \Box^{IKM}_B \bot$. Combining (a), (b), and (c), we find, by Theorem A.4, that $\Box_B \bot$ is true and, thus, that B is inconsistent. *Quod non.*

We now turn to the case that U is sequential. Suppose $A \bowtie U$. Then, we can find a finitely axiomatised sequential $U_0 \subseteq U$, such that $A \bowtie U_0$ and, hence $U_0 \bowtie U$, contradicting the first case.   ❑

An alternative argument, for the case that U is sequential, is to note that $\mathsf{Seq}(A) \bowtie U$, where Seq is the functor that adds sequentiality to A.

Remark 8.14. We note Theorem 8.13 implies that there is no Friedman-reflexive and sequential U such that $\mathsf{S}^1_2 \rhd U$. The reason is that any sequential theory interprets S^1_2. Of course, S^1_2 does interpret the theory of the ordering of the natural numbers. However, that only gives the trivial interpreter logic that proves $\Box \bot$. So, one might wonder whether there is a more interesting base for S^1_2. We will discuss some hopeful signs that there is in Section 9.   ⃝

If we combine Theorems 6.3 with 8.13, we get what I would count as a version of the Second Incompleteness Theorem.

Corollary 8.15. *Suppose that A is finitely axiomatised and consistent and that U is Friedman-reflexive. Suppose further that one of A, U is sequential. Then, $A \not\rhd (U + \diamondsuit\!\!\!\!\diamondsuit A)$.*

8.4. Interpreter logic

In this subsection, we will be concerned with interpreter logics over sequential Friedman-reflexive bases. These interpreter logics satisfy Löb's Logic. We will show that, if the base is, in addition, *effectively* Friedman-reflexive, then the proof of Solovay's Theorem works with minor adaptations.

8.4.1. Soundness

Combining Theorems 6.4 and 8.13, we find:

Theorem 8.16. *Let* $\langle A, U \rangle$ *be an FM-frame. Suppose that one of* A, U *is sequential. Then,* $\Lambda^{\mathsf{fr}}_{\langle A, U \rangle}$ *extends* GL.

Since we have Löb's Logic when either A or U is sequential, we also have explicit solutions for modal equations where the fixed point variable is guarded. E.g., $A \vdash \blacklozenge \top \leftrightarrow \neg \blacksquare \blacklozenge \top$. The Gödel Fixed Point Lemma plays a role in our proof of this fact via the proof of Theorem 8.13, but it is still not a *direct* application.

Can we prove Löb's Principle more directly? Remember that we do not even know whether our provability-like operator can be represented in A by a formula. E.g., S^1_2 and the theory of the ordering of the natural numbers would provide an example. Surprisingly, we can employ the usual argument in case U is *effectively* Friedman-reflexive. We give the argument below. Even if we, thus, prove a result that is weaker than what we already know, we think the alternative proof is of independent interest. E.g., it could have other generalisations. We first prove a Fixed Point Lemma.

Theorem 8.17. *Suppose* U *is sequential and effectively Friedman-reflexive. Let* $N : U \rhd \mathsf{S}^1_2$. *We define* $\mathfrak{I}$ *w.r.t.* N. *Suppose* $A(x)$ *is a boolean combination of* Σ^0_1*-formulas with just* x *free. Then, there is a* B *in the* U*-language, such that* $U^{\mathsf{e}} \vdash B \leftrightarrow A^{\mathfrak{I}}(\underline{\ulcorner B \urcorner})$.

Proof. Suppose U is sequential and effectively Friedman-reflexive. Let N and $\mathfrak{I}$ be as in the statement of the Theorem.

By effectivity, we can find a recursive F that sends any Σ^0_1-sentences S to S^I, where S^I is equivalent over U^{e} to $S^{\mathfrak{I}}$. We can lift this function to Boolean combinations of Σ^0_1-sentences. Let's say the result is G.

Suppose $A(x)$ is a boolean combination of Σ^0_1-formulas with just x free. We write $A(G(x))$ for the result of replacing each Σ^0_1-component Sx of the Boolean combination by $\exists y, z, u\, (G_0 xyz \wedge S_0 yu)$, where $Gxyz$ is a Δ^0_1-formula such that $\exists z\, Gxyz$ represents the graph of G and $S_0 yu$ is a Δ_0-formula such that Sx is (equivalent to) $\exists u\, S_0 xu$.

We can find a C such that $\mathsf{EA} \vdash C \leftrightarrow A(G(\ulcorner C \urcorner))$, by the Gödel Fixed Point Lemma. We note that the Fixed Point Lemma yields a sentence of the form $A(G(t))$, where t is a substitution term. Since, this term is not really in the language, we have to eliminate it. We do this in the same way as we did for the function G, so that C is again a boolean combination of Σ^0_1-sentences. Let $B := G(C)$. Then,

$$
\begin{aligned}
U^{\mathsf{e}} \vdash\ & (C \leftrightarrow A(G(\underline{\ulcorner C \urcorner})))^{\mathfrak{I}} \\
\vdash\ & C^{\mathfrak{I}} \leftrightarrow (A(G(\underline{\ulcorner C \urcorner})))^{\mathfrak{I}} \\
\vdash\ & B \leftrightarrow (A(\underline{\ulcorner B \urcorner}))^{\mathfrak{I}} \qquad \square
\end{aligned}
$$

Theorem 8.18 (*Alternative Proof for Löb's Principle in the effective Case*). *Suppose* U *is sequential and effectively Friedman-reflexive. Let* $K : A \rhd U$ *be an FM-interpretation. Then,* Λ^{fr}_K *proves Löb's Principle.*

Proof. Suppose U is sequential and effectively Friedman-reflexive. Let $K : A \rhd U$ be an FM-interpretation. Consider any B in the A-language. By Theorem 8.17, we can find a D such that $U^{\mathsf{e}} \vdash D \leftrightarrow \triangle^{\mathfrak{I}}_A(D^K \to B)$.

Thus, $U \vdash D \leftrightarrow \boxdot_A(D^K \to B)$. Setting $E := D^K$, we find $A \vdash E \leftrightarrow \blacksquare_A(E \to B)$. So, we have one variant of the Löb Fixed Point in A. Since we have K4, Löb's Principle follows.   ❏

8.4.2. Solovay's Theorem

We can prove Solovay's Theorem in case the base theory U is both sequential and effectively Friedman-reflexive. We first formulate the theorem.

Let α range over $0, 1, \ldots, \infty$. We define $\blacksquare^0 \perp := \perp$, $\blacksquare^{k+1} \perp := \blacksquare\blacksquare^k \perp$, and $\blacksquare^\infty \perp := \top$ and, similarly, for $\Box$. Suppose $K : A \rhd U$ is an FM-interpretation. Let $\mathsf{d}(K)$ be the smallest α such that $A \vdash \blacksquare_{K,A}^\alpha \perp$.

Theorem 8.19 (*Solovay's Theorem for sequential effectively Friedman-reflexive bases*). *Suppose that $K : A \rhd U$ is an FM-interpretation and that U is effectively Friedman-reflexive and sequential. Then, $\Lambda_K^{\mathsf{fr}} = \mathsf{GL} + \Box^{\mathsf{d}(K)} \perp$.*

We will prove Solovay's Theorem by verifying the conditions for it given in [3]. See also [28].

The idea of de Jongh, Jumelet and Montagna is that Solovay's embedding result can be verified in an extension of the modal logic R⁻ enriched with certain fixed points. We introduce the logic R⁻ of Guaspari and Solovay. See [6]. The language of R⁻ is given by:

- $\alpha ::= p_0 \mid p_1 \mid \ldots$
- $\varphi ::= \alpha \mid \perp \mid \top \mid \neg\varphi \mid \Box\varphi \mid (\varphi \wedge \varphi) \mid (\varphi \vee \varphi) \mid (\varphi \to \varphi) \mid (\Box\varphi < \Box\varphi) \mid$ $(\Box\varphi \leq \Box\varphi)$

The logic R⁻ is axiomatised by the axioms and rules of GL (for the extended language) plus the following axioms.

R⁻1. $\vdash (\Box\varphi \leq \Box\psi) \to \Box\varphi$
R⁻2. $\vdash ((\Box\varphi \leq \Box\psi) \wedge (\Box\psi \leq \Box\chi)) \to (\Box\varphi \leq \Box\chi)$
R⁻3. $\vdash (\Box\varphi < \Box\psi) \leftrightarrow ((\Box\varphi \leq \Box\psi) \wedge \neg(\Box\psi \leq \Box\varphi))$
R⁻4. $\vdash \Box\varphi \to ((\Box\varphi \leq \Box\psi) \vee (\Box\psi \leq \Box\varphi))$
R⁻5. $\vdash (\Box\varphi \leq \Box\psi) \to \Box(\Box\varphi \leq \Box\psi)$
R⁻6. $\vdash (\Box\varphi < \Box\psi) \to \Box(\Box\varphi < \Box\psi)$

We will usually omit the brackets around $(\Box\varphi < \Box\psi)$ taking $<$ to bind stronger than all other connectives. Similarly, for $\leq$.

Now consider a finite Kripke model $\mathcal{K}$ of GL with nodes $0, \ldots, n-1$. Here 0 is the bottom node. Let $\prec$ be its accessibility relation. We want to 'embed' this model in our modal logic. To realise this purpose, we add constants ℓ_i, for $i < n$, to the language of R⁻ and we extend the schemes to the extended language. We demand that the constants satisfy the following equations. We write $j \parallel i$ for j is incompatible with i w.r.t. $\prec$.

fp1. $\vdash \ell_i \leftrightarrow (\Box\neg\ell_i \wedge \bigwedge_{j \succ i} \Diamond \ell_j \wedge \bigwedge_{j \parallel i} \bigvee_{k \preceq i,\, k \parallel j} \Box\neg\ell_k < \Box\neg\ell_j)$.
fp2. For $i \neq j$, we have $\vdash \Box\neg\ell_i \leq \Box\neg\ell_j \to \Box\neg\ell_i < \Box\neg\ell_j$.

We proceed to introduce our intended interpretation of $\leq$ and $<$. We remind the reader of the witness comparison notation. We define, for any $C = \exists x\, C_0(x)$ and $D = \exists y\, D_0(y)$:

- $C \leq D := \exists x\, (C_0(x) \wedge \forall y < x \, \neg D_0(y))$,
- $C < D := \exists x\, (C_0(x) \wedge \forall y \leq x \, \neg D_0(y))$,

Suppose $K : A \rhd U$ is an FM-interpretation, where U is sequential. As usual we work with a fixed $N : U \rhd \mathsf{S}_2^1$.

- We define $\boxdot_A B < \boxdot_A C$ by $(\triangle_A B < \triangle_A C)^I$, where I is a cut such that

$$U^{\mathsf{e}} \vdash (\triangle_A B < \triangle_A C)^I \leftrightarrow (\triangle_A B < \triangle_A C)^{\mathfrak{J}}.$$

- We define $\blacksquare_{K,A} B < \blacksquare_{K,A} C$ by $(\boxdot_A B < \boxdot_A C)^K$.
- We define $\boxdot_A B \leq \boxdot_A C$ and $\blacksquare_{K,A} B \leq \blacksquare_{K,A} C$ similarly.

We extend the notion of translation of the modal language to the richer vocabulary in the obvious way.

Theorem 8.20. *Suppose $K : A \rhd U$ is an FM-interpretation, where U is sequential. Let $N : U \rhd \mathsf{S}_2^1$. Then, we have R^- for all K, A-translations of the modal language.*

Proof. Ad R^-1. We have $\mathsf{EA} \vdash \triangle_A B \leq \triangle_A C \rightarrow \triangle_A B$. It follows that

$$\mathsf{DA}^{\mathsf{e}} \vdash (\triangle_A B \leq \triangle_A C \rightarrow \triangle_A B)^{\mathfrak{J}}.$$

Hence, $U \vdash \boxdot_A B \leq \boxdot_A C \rightarrow \boxdot_A B$. We may conclude that

$$A \vdash \blacksquare_{K,A} B \leq \blacksquare_{K,A} C \rightarrow \blacksquare_{K,A} B.$$

The proofs of R^-2, R^-3, and R^-4 are similar.

We treat R^-5. Let I be a cut such that

$$U^{\mathsf{e}} \vdash (\triangle_A B < \triangle_A C)^I \leftrightarrow (\triangle_A B < \triangle_A C)^{\mathfrak{J}}.$$

We have: $\mathsf{EA} \vdash \triangle_A B \leq \triangle_A C \rightarrow \triangle_A (\triangle_A B \leq \triangle_A C)^{IK}$. So,

$$U^{\mathsf{e}} \vdash (\triangle_A B \leq \triangle_A C \rightarrow \triangle_A (\triangle_A B \leq \triangle_A C)^{IK})^{\mathfrak{J}}.$$

And, thus,

$$U \vdash \boxdot_A B \leq \boxdot_A C \rightarrow \boxdot_A (\blacksquare_{K,A} B \leq \blacksquare_{K,A} C).$$

We may conclude that $A \vdash \blacksquare_{K,A} B \leq \blacksquare_{K,A} C \rightarrow \blacksquare_{K,A} (\blacksquare_{K,A} B \leq \blacksquare_{K,A} C)$.

The proof of R^-6 is similar to that of R^-5. ❑

To provide the ℓ_i we need an extension of Theorem 8.17.

Theorem 8.21. *Suppose U is sequential and effectively Friedman-reflexive. Let $N : U \rhd \mathsf{S}_2^1$. We define $\mathfrak{J}$ w.r.t. N. Suppose we have formulas $A_i(x_0, \ldots, x_{k-1})$, for $i < k$, where each A_i is a boolean combination of Σ_1^0-formulas with just $x_0, \ldots, x_{k-1}$ free. Then, for $i < k$, there are B_i in the U-language, such that*

$$U^{\mathsf{e}} \vdash B_i \leftrightarrow A_i^{\mathfrak{J}}(\underline{\ulcorner B_0 \urcorner}, \ldots, \underline{\ulcorner B_{k-1} \urcorner}),$$

for each $i < k$.

The proof of the Theorem is similar to the proof of Theorem 8.17 using the Multiple Fixed Point Lemma as it is available in EA.

In the proof of Theorem 8.22, we need the assumptions that we work with a single conclusion system and that the fixed point construction is a usual construction.

Theorem 8.22. *Suppose $K : A \rhd U$ is an FM-interpretation, where U is sequential. Suppose U is effectively Friedman-reflexive. Let $N : U \rhd \mathsf{S}_2^1$. Let $\mathcal{K}$ be a Kripke model for GL with nodes $0, \ldots, n-1$ and accessibility relation $\prec$, where 0 is the bottom node. Then, we can find sentences L_i, for $i < n$, such that:*

a. $A \vdash L_i \leftrightarrow (\blacksquare_{K,A} \neg L_i \wedge \bigwedge_{j \succ i} \blacklozenge_{K,A} L_j \wedge \bigwedge_{j \| i} \bigvee_{k \preceq i,\, k \| j} \blacksquare_{K,A} \neg L_k < \blacksquare \neg L_j).$
b. For $i \neq j$, we have $A \vdash \blacksquare_{K,A} \neg L_i \leq \blacksquare_{K,A} \neg L_j \rightarrow \blacksquare_{K,A} \neg L_i < \blacksquare_{K,A} \neg L_j.$

Proof. Using Theorem 8.21, we can find sentences λ_i such that:

$$U^{\mathsf{e}} \vdash \lambda_i \leftrightarrow (\underline{i} = \underline{i} \wedge \triangle_A \neg \lambda_i^K \wedge \bigwedge_{j \succ i} \triangledown_A \lambda_j^K \wedge \bigwedge_{j \| i} \bigvee_{k \preceq i,\, k \| j} \triangle_A \neg \lambda_k^K < \triangle_A \neg \lambda_j^K)^{\mathfrak{I}}.$$

Hence,

$$U \vdash \lambda_i \leftrightarrow (\boxdot_A \neg \lambda_i^K \wedge \bigwedge_{j \succ i} \lozenge_A \lambda_j^K \wedge \bigwedge_{j \| i} \bigvee_{k \preceq i,\, k \| j} \boxdot_A \neg \lambda_k^K < \boxdot_A \neg \lambda_j^K).$$

Taking $L_i := \lambda_i^K$, we find (a).

Claim (b) follows from the fact that, for $i \neq j$:

$$\mathsf{EA} \vdash \triangle_A \neg L_i \leq \triangle_A \neg L_j \rightarrow \triangle_A \neg L_i < \triangle_A \neg L_j.$$

We note that this last fact uses that our proof system is single conclusion and that the L_i are pairwise distinct. This last insight follows from the fact that the λ_i are pairwise distinct. This is because the first conjunct of Λ_i has the form $(\underline{i} = \underline{i})^I$, for some N-cut I. ❑

Solovay's Theorem now follows from the results of [3]. We have the following corollary.

Corollary 8.23. *Suppose $K : A \rhd U$ is a faithful FM-interpretation and A is consistent and U is effectively Friedman-reflexive, sequential, and polyglottic. Then, $\Lambda_K^{\mathsf{fr}} = \mathsf{GL}$.*

Proof. We assume the conditions of the corollary. Clearly $A + \blacklozenge_{K,A}^0 \top$ is consistent. Suppose $A + \blacklozenge_{K,A}^n \top$ is consistent. So, by polyglotticity, $U + \lozenge_A \blacklozenge_{K,A}^n \top$ is consistent. But, then, by faithfulness, $A + \blacklozenge_{K,A}^{n+1} \top$ is consistent. It follows that $\mathsf{d}(K) = \infty$. ❑

Corollary 8.24. *Suppose $\langle A, U \rangle$ is an FM-frame, A is consistent and sequential, and U is RE, effectively Friedman-reflexive, sequential, and polyglottic. Then, $\Lambda_{A,U}^{\mathsf{fr}} = \mathsf{GL}$.*

Proof. By a result of Friedman, if there is an interpretation of an RE theory U in a finitely axiomatised A, then there is also a faithful one. See [14] or [20]. ❑

9. Concluding remarks

In this section, we look briefly backward at what is and what is not achieved and we look forward at a possible next step in the program.

9.1. Coordinate-free?

Did we succeed in giving a treatment of the Second Incompleteness Theorem and of provability logic that is indeed coordinate-free?

The results of our general framework as, for example, provided in Sections 3, 6 and B.1 clearly do not involve arithmetisation, neither in their statement nor in their proof (with the exception of the proof of Theorem 3.10). In the remaining sections we have general results that do not depend on arithmetisation in their statement but that do ask for a proof involving arithmetisation.

Sometimes the concrete statement of an application does involve arithmetisation. This is not unlike the situation for the Cartesian product in Category Theory. The general treatment of the product is clearly implementation-free, but if we want to apply it, e.g., to the hereditarily finite sets, we have to define the category from the sets and, for this, we need to code pairing . . .

A first point of discussion is that the usual finite axiomatisations of many salient finitely axiomatisable theories employ a truth predicate. Examples are S_2^1, EA, and $\mathsf{I}\Sigma_1$.[18] So, statements involving these theories would not be coordinate-free to begin with. Now, of course, we could also axiomatise these with the first so-and-so-many instances of their coordinate-free schematic axiomatisations. That would be a bit unnatural perhaps, but it would be at least a coordinate-free specification. Maybe a better way of looking at the matter is as follows. All finite axiomatisations of a theory are provably equivalent. Since, our framework works modulo provable equivalence, we are not dependent on a specific finite axiomatisation. Moreover, the class of all finite axiomatisations is uniquely determined by the infinite axiomatisation, which in the cases above, is coordinate-free.

Here are some further examples as food for thought.

- Peano Corto is interpretable in $\mathsf{S}_2^1 + \Diamond_{\mathsf{S}_2^1} \top$. However, neither $\mathsf{S}_2^1 + \Diamond_{\mathsf{S}_2^1} \top$ itself nor the Henkin interpretation that we employ is coordinate-free.
- Peano Corto is interpretable in EA, since EA interprets $\mathsf{S}_2^1 + \Diamond_{\mathsf{S}_2^1} \top$ on a super-logarithmic cut. Here EA is coordinate-free (ignoring the worry articulated above) but the only interpretation that we know of is not.
- By Vaught's Theorem (see [17,24]), there is a finite axiom scheme that axiomatises Peano Corto. We replace the schematic variables in the scheme by class variables and take the universal closure. Also we add Predicative Comprehension. This results in a finitely axiomatised theory, say W, that conservatively extends Peano Corto in an extended language. One can show that W is mutually interpretable with EA. The interpretation of EA in W can be taken to be coordinate-free: we have EA on the intersection of all cuts that are classes. The interpretation of Peano Corto in W is also coordinate-free, however the specification of W is not, since the Vaught construction uses truth-predicates.

We note that, if we only know arithmetisation-involving interpretations, but both theories are coordinate-free, then the frame properties are unproblematic. For example, our version of G2 for EA and Peano Corto is $\mathsf{EA} \not\rhd (\mathsf{PA}^{\downarrow\downarrow} + \Diamondblack\, \mathsf{EA})$, which is perfectly fine, and, similarly, for the fact that the frame-logic of $\langle \mathsf{EA}, \mathsf{PA}^{\downarrow\downarrow} \rangle$ is GL.

Remark 9.1. We consider the following theory IIA (Initial Isomorphism Arithmetic). We start with S_2^1. We expand the language with second-order variables and we expand the theory with predicative comprehension obtaining $\mathsf{PC}(\mathsf{S}_2^1)$. Now we expand the language with a new unary predicate J and a binary predicate F. We add an axiom $\forall I\,(\mathsf{cut}(I) \to \mathsf{J} \subseteq I)$ and an axiom stating that F is an isomorphism between ID and J.

---

[18] Similarly, the finite axiomatisation of predicative comprehension over a pair theory does involve various implementation details. See [21, Appendix A].

By Theorem 5.9 of [26], it follows that IIA and Peano Basso have the same arithmetical consequences. So, the pair IIA and Peano Basso form a nice coordinate free pair.[19]

Since $\mathsf{PC}(\mathsf{S}_2^1)$ is mutually interpretable with $\mathsf{S}_2^1 + \Diamond_{\mathsf{S}_2^1} \top$ and $\mathsf{S}_2^1 + \Diamond_{\mathsf{S}_2^1} \top$ is mutually interpretable with EA, we find that IIA interprets EA.   ⭕

Open Question 9.2.

   i. Can we find a more canonical axiom scheme for Peano Corto in a coordinate-free way? We note that the cuts are already schematic. The whole problem is in replacing the schematic variable that ranges over Σ_1^0-*sentences* by an unrestricted one that ranges over arbitrary formulas.
  ii. Is there a coordinate-free specification of an interpretation of Peano Corto in EA or in some $\mathrm{I}\Sigma_n$?
 iii. Do we have $\mathsf{EA} \rhd \mathsf{IIA}$?   ⭕

9.2. New insights

Finding coordinate-free representations is a worthy aim, but it should not stand alone. We also want new insights. The present paper does indeed produce some new insights.

For example, the usual form of G2, for the case of EA, is (a) $\mathsf{EA} \not\rhd (\mathsf{S}_2^1 + \Diamond_{\mathsf{EA}} \top)$. However, we do have $\mathsf{EA} \rhd (\mathsf{S}_2^1 + \triangledown_{\mathsf{EA}} \top)$. Our version of G2 with base $\mathsf{PA}^{\downarrow\downarrow}$ is (b) $\mathsf{EA} \not\rhd (\mathsf{PA}^{\downarrow\downarrow} + \diamondsuit_{\mathsf{EA}} \top) = (\mathsf{PA}^{\downarrow\downarrow} + \triangledown_{\mathsf{EA}} \top)$. We see that in (a) the base theory is weaker and the consistency statement stronger. In (b) it is the other way around. Both (a) and (b) follow from a version of G2 due to Pudlák: (c) $\mathsf{EA} \not\rhd \mho(\mathsf{EA}) := (\mathsf{S}_2^1 + \{\Diamond_{n,\mathsf{EA}} \top \mid n \in \omega\})$. We suspect that the version of (b) with $\mathsf{PA}^{\downarrow\downarrow}$ replaced by DA does not directly follow from (c). However, this depends on a negative answer to Question 8.8(i).

The most convincing example of a new phenomenon is Solovay's Theorem for interpreter logics over a sequential, effectively Friedman-reflexive base.

9.3. Perspectives

We think the obvious next step in the project should be the study of Friedman-reflexivity for pair theories. There is hope for progress, since Fedor Pakhomov suggested a very natural construction of an effectively Friedman-reflexive pair theory. This theory is interpretable in S_2^1.

Of course, the present paper also left a list of open questions. We collected them in Appendix E.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Notations, notions and imported results

Theories will be theories in predicate logic of finite signature. We allow theories with sets of axioms of arbitrary complexity. The variables $T, U, V, \ldots$ will range over theories. The variables $A, B, C, \ldots$ will ambiguously range over sentences and finitely axiomatised theories.

Interpretations will be multi-dimensional piece-wise interpretations. If we assume that a theory proves that there are at least two objects the piece-wiseness can be eliminated. A property—typical for piece-wise

---

[19]   We could even take PA^- as starting point rather than S_2^1, so removing all doubts regarding coordinate freedom. In fact we need only add closure under ω_1 to the PA^--based version of IIA to obtain precisely IIA. Moreover, we could add ω_1 using a new symbol and adding appropriate recursion axioms, thus even avoiding sequence coding and the like.

interpretations—we use in the paper is that predicate logic can interpret the theory of any finite model. The reader is referred to our paper [32] for a quick introduction to the details. The idea of a piece-wise interpretation is explained in [30] and in [27].

We will use depth-of-quantifier-alternations as our measure ρ of complexity of formulas. The notion is worked out in detail in [32].

Here are our main notions and notations.

- $U \subseteq V$ means that the theory V considered as a set of theorems extends the theory U considered as a set of theorems, *where V has the same language as U.*
- We write $\Box_\alpha B$ for the formalised statement that B is provable from the theory with axiom-set represented by α. In case A is a finitely axiomatised theory, we write $\Box_A B$ for $\Box_{\alpha_0} B$, where $\alpha_0 = \bigvee_{i<n} x = \ulcorner A_i \urcorner$, where $A_0, \ldots, A_{n-1}$ are the axioms of A. We write $\diamond$ for $\neg \Box \neg$.
  We use $\Box_\alpha^K B$ for $(\Box_\alpha B)^K$, where the superscript K means relativisation to interpretation K. Note that here α is interpreted inside K.
- We write $\Box_{n,A} B$ for the formalised statement that B is provable from A using only statements of ρ-complexity $\leq n$, where ρ measures the depth of quantifier alternations. We will usually implicitly assume that $n \geq \rho(A)$.
- We write $\triangle_A B$ for $\Box_{\rho(A \to B),A} B$ and $\triangledown_A B$ for $\neg \triangle_A \neg B$.
- We write $K : U \triangleright V$ or $K : V \triangleleft U$ for K is an interpretation of V in U. We write $U \triangleright V$ for $\exists K \, K : U \triangleright V$, etcetera, We write $U \bowtie V$ for U and V are mutually interpretable. In other words, $U \bowtie V$ iff $U \triangleright V$ and $U \triangleleft V$.[20] In the context of a category of interpretations, we will use $V \xrightarrow{K} U$, for $K : V \triangleleft U$.
- $\mathbb{E} := \mathsf{INT}_3^+$ is the category of theories and interpretations, where two interpretations $K, K' : V \to U$ are the same iff, for all V-sentences A, we have $U \vdash A^K \leftrightarrow A^{K'}$.
- E_{VU} is the interpretation based on the identical translation that witnesses $V \subseteq U$.
- U and V are *sententially congruent* or *elementarily congruent* iff they are isomorphic in $\mathbb{E}$.
- A theory U is *restricted* if, for some m all its axioms are of ρ-complexity $\leq m$.
- Let an interpretation $N : \mathsf{S}_2^1 \triangleleft U$ be given. A definable N-cut in U is given by a formula I such that U proves that I is a subclass of N and is closed under 0_N, S_N, $+_N$, $\times$ and ω_{1N} and downwards closed w.r.t. $\leq_N$. The formula defining I need not be of the form J^N. If I is definable inside N, the cut is N-internal. In case N is a multi-dimensional interpretation, the cut I is also multi-dimensional. Similarly, if N is piecewise, then I need not be given by a single formula but by a number of pieces. Since, we need our notion mostly in the context of sequential theories these fine points can be safely ignored.

We turn to the statement of some central results that we import from outside in this paper. Let true be a standard Σ_1^0-truth predicate. The following theorem is a direct consequence of the estimate of the transformation of witnesses, when we move from S to $\mathsf{true}(S)$, for Σ_1^0-sentences S. See [8, V.5(b)] for details. We can give a similar estimate for the case of provability.

Theorem A.1. *Let J be a logarithmic cut in S_2^1. Then,*

a. $\mathsf{S}_2^1 \vdash S^J \to \mathsf{true}(\ulcorner S \urcorner)$.

b. $\mathsf{S}_2^1 \vdash S^J \to \Box_{m,\mathsf{S}_2^1} S$, *for sufficiently large m.*
   (*The number m will be* $\mathsf{max}(\rho(\mathsf{S}_2^1), \rho(S))$ *plus some constant for the overhead.*)

---

[20] We used to employ $\equiv$ for mutual interpretability. However, in the writings of Lev Beklemishev and his school, $\equiv$ is used for extensional sameness of theories.

The following theorem is a watered-down and inessentially modified version of Theorem 6.2. of [32]. This theorem is an extension of earlier results independently obtained by Harvey Friedman (see [14]) and Jan Krajíček (see [9]). See also [19,20].

Theorem A.2. *Suppose U is a consistent restricted sequential RE theory with bound m. Let K be a witness of sequentiality for U. Let $N : \mathsf{S}_2^1 \lhd U$. Then, from the data m, K, and N, we can effectively find an N-cut I such that, for all Σ_1^0-sentences S, if $U \vdash S^I$, then S is true.*

The following theorem is Theorem 4.15 of [20]

Theorem A.3. *Suppose A is a consistent, finitely axiomatised sequential theory. Suppose $A \bowtie U$. Then, there is an $N : \mathsf{S}_2^1 \lhd U$, such that N is Σ_1^0-sound. In other words, U tolerates S_2^1 plus all true Π_1^0-sentences.*

In our paper, we use the following immediate consequence of Theorem A.3.

Theorem A.4. *Suppose A is consistent, finitely axiomatised and sequential. Suppose further that $K : A \rhd A$ and $N : A \rhd \mathsf{S}_2^1$. Then there is an N-cut I such that $K \circ I$ is Σ_1^0-sound.*

Proof. Let U be axiomatised by the B such that $A \vdash B^K$. Clearly, $A \bowtie U$. Let N' be the interpretation promised by Theorem A.3. Take I the common cut of N and N'. ❑

Remark A.5. It is easy to see that Theorem A.4 immediately implies Theorem A.3. In fact, the natural order is to prove Theorem A.4 first. ◯

Finally, we sketch some of the details of the treatment of the functor Seq. We remind the reader that the mapping Seq is defined as follows. We start with a theory V. We extend the signature of V with a unary predicate D and a binary predicate $\in$. The axioms of $\mathsf{Seq}(V)$ are the axioms of V relativised to D plus the unrelativised axioms of Adjunctive Set Theory AS plus the axiom that says that all elements of D are empty sets. (For an introduction to Adjunctive Set Theory, further references, and some history, see [25].) Let η_V be the interpretation of V in $\mathsf{Seq}(V)$ given by relativisation to D.

Theorem A.6. *Suppose $V \xrightarrow{K} W$. Then there is an interpretation $\mathsf{Seq}(K)$ such that $\mathsf{Seq}(V) \xrightarrow{\mathsf{Seq}(K)} \mathsf{Seq}(W)$.*

Proof. Suppose $V \xrightarrow{K} W$. Consider the interpretation $K' := \theta_V \circ K$ of V in $\mathsf{Seq}(W)$. We note that K may be piece-wise and multidimensional (with possibly different dimensions for the pieces). Since in $\mathsf{Seq}(W)$ we have the full machinery of sequences available, we can rebuild K' to a one-dimensional, one-piece interpretation K^* that is $\mathsf{Seq}(W)$-provably definably isomorphic to K'. We extend K^* to the interpretation $M := \mathsf{Seq}(K)$ as follows.

- M has two pieces $\mathfrak{s}$ and $\mathfrak{o}$ both of dimension 1.
- $\delta_M^{\mathfrak{o}}$ is δ_{K^*} and $\delta_M^{\mathfrak{s}}$ is the complement of D.
- Identity between elements of different pieces is always $\bot$. We have: $x =_M^{\mathfrak{oo}} y$ iff x and y are in $\delta_M^{\mathfrak{o}}$ and $x =_{K^*} y$, and $x =_M^{\mathfrak{ss}} y$ iff x and y are in $\delta_M^{\mathfrak{s}}$ and $x = y$.
- $\mathsf{D}_M^{\mathfrak{o}}$ is $\delta_M^{\mathfrak{o}}$ and $\mathsf{D}_M^{\mathfrak{s}}$ is empty.
- Let P be an n-ary predicate of the V-language. Then $P_M^{\mathfrak{o};\dots;\mathfrak{o}}(\vec{x})$ tells us that each x_i is in $\delta_M^{\mathfrak{o}}$ and that $P_{K^*}(\vec{x})$. If any variable in $P(\vec{x})$ is assigned a non-$\mathfrak{o}$ piece, then $P_M(\vec{x})$ is false.
- We have:
  ○ $x \in_M^{\mathfrak{oo}} y$ and $x \in_M^{\mathfrak{so}} y$ are false.

○ $x \in_M^{\mathfrak{os}} y$ iff $x \in \delta_M^{\mathfrak{o}}$, $y \in \delta_M^{\mathfrak{s}}$, and there is an x' with $x =_{K^*} x'$ and "$\langle 0, x' \rangle \in y$". Here the scare quotes are there to remind us that we do not have extensionality. So we really mean here: there is an empty set z and there is a pair of the form $\langle z, x' \rangle$ such that . . .

○ $x \in_M^{\mathfrak{ss}} y$ iff $x \in \delta_M^{\mathfrak{s}}$, $y \in \delta_M^{\mathfrak{s}}$, and "$\langle 1, x \rangle \in y$".

It is easy to verify that the interpretation sketched here indeed is an interpretation of $\mathsf{Seq}(V)$ in $\mathsf{Seq}(W)$. ❏

It would be nice if we could prove that Seq lifts to a functor in $\mathbb{E}$. However, we do not quite see how that can work for the $\in$-part.

Open Question A.7. Does Seq or, some appropriate variant of it, lift to a functor on $\mathbb{E}$? ○

Theorem A.8. *Suppose V is sequential. Then, $\mathsf{Seq}(V) \bowtie V$.*

Proof. We have $V \xrightarrow{\eta_V} \mathsf{Seq}(V)$. Let $\in^\star$ be a V-formula that witnesses that V is sequential. Our interpretation M of $\mathsf{Seq}(V)$ in V looks as follows. We duplicate the domain by having two pieces $\mathfrak{s}$ and $\mathfrak{o}$ with domain $x = x$. We let D correspond to the elements in the $\mathfrak{o}$-piece and have the predicates of V on the $\mathfrak{o}$-piece. We take $x \in_M^{\mathfrak{os}} y$ iff x and y are in the correct domains and "$\langle 0, x \rangle \in^\star y$". Similarly, $x \in_M^{\mathfrak{ss}} y$ iff x and y are in the correct domains and "$\langle 1, x \rangle \in^\star y$". We set $\in_M$ to $\bot$ in the other cases. ❏

Appendix B. Two special classes of theories

In this appendix we have a brief look at Friedman-reflexive theories that are complete and at Friedman-reflexive theories that are finitely axiomatised.

B.1. Complete theories

We discuss the case of complete theories.

Theorem B.1. *Suppose U is a complete theory. Then, U is Friedman-reflexive.*

Proof. Suppose U is complete. So, modulo U-provable equivalence, we only have propositions $\top$ and $\bot$. So, A has as pro-interpreters, modulo provable equivalence, either just $\bot$ or both $\bot$ and $\top$. In the first case, the desired interpreter is $\bot$, in the second it is $\top$. ❏

Example B.2. Presburger Arithmetic is complete and decidable. So, it is Friedman-reflexive but not effectively so.

True arithmetic $\mathsf{Th}(\mathbb{N})$ extends PA and is, hence, effectively Friedman-reflexive. Thus, we have an example of a complete theory that is effectively Friedman-reflexive. We note that the $\lozenge$ that takes $\top$ and $\bot$ as values cannot be recursive, providing an example of a salient non-recursive choice of $\lozenge$. ○

We consider the interpreter logic for a complete base. We note that if $\langle U, A \rangle$ is an FM-frame and U is complete, then U must be decidable and, hence any choice for $\lozenge$ must be non-computable.

We show that the interpreter logics for complete bases extend $\mathsf{K45}$.

Theorem B.3. *Suppose $K : A \triangleright U$ is an FM-interpretation and U is complete. Then, we have $A \vdash \blacklozenge_{K,A} B \to \blacksquare_{K,A} \blacklozenge_{K,A} B$.*

Proof. Suppose $K : A \rhd U$ is an FM-interpretation and U is complete. Consider any A-sentence B. If $A + \blacklozenge_{K,A} B$ is inconsistent, we are done. If not, it follows that $U + \lozenge_A B$ is consistent. Hence, by completeness, $U \vdash \lozenge_A B$. So, $A \vdash \blacklozenge_{K,A} B$ and, hence, $A \vdash \blacksquare_{K,A} \blacklozenge_{K,A} B$. We may conclude that in both cases, we have $A \vdash \blacklozenge_{K,A} B \rightarrow \blacksquare_{K,A} \blacklozenge_{K,A} B$. $\qquad \square$

We note that, if $U \not\rhd A$, then $U \nvdash \lozenge A$ and, so, $U \vdash \neg \lozenge A$. It follows that $A \vdash \blacksquare_{K,A} \bot$. So, the interpreter logic for A trivialises. Suppose, on the other hand, that $U \rhd A$. Then, $A \vdash \blacklozenge_{K,A} \top$. We note that, in this case, U is mutually interpretable with a finite sub-theory.

We suspect that most non-finitely axiomatisable complete and decidable theories in the literature have the property that they are *not* interpretable in a finite sub-theory. This has been verified for Presburger Arithmetic. See [12]. (So, the interpreter frame logic of Presburger Arithmetic trivialises.)

There are, of course, also examples of consistent, finitely axiomatised, complete and decidable theories like the theory of dense linear orderings without end-points and the theory of the ordering of the natural numbers.

Open Question B.4. Is there an example of an FM-interpretation $K : A \rhd U$, where U is complete, with an interesting interpreter logic? $\quad \bigcirc$

B.2. Finitely axiomatised theories

If the Friedman-reflexive base is finitely axiomatised, we can view the embedding functor as an embedding of the finite extensions of the base A in the finitely axiomatised theories. So, we can view the Friedman-reflexivity of the base as the existence of an adjoint of this functor.

There are plenty of consistent complete finitely axiomatised theories, so we do not lack examples of the phenomenon of a consistent finitely axiomatised Friedman-reflexive theory. However, these examples cannot be effectively Friedman-reflexive, since they, clearly, cannot be essentially undecidable.

Open Question B.5. Is there a consistent finitely axiomatised theory that is effectively Friedman-reflexive?
$\bigcirc$

In the case of a finitely axiomatised base, there is, of course, the salient interpreter logic of A over A via the identical interpretation. This logic satisfies $\mathsf{S4}$, since Id_A is clearly companiable.

Appendix C. A notion of sameness of interpretations

The notion of sameness of theories that is relevant in the present paper is sentential congruence or $\mathbb{E}$-isomorphism (see Appendix A, for the definition of $\mathbb{E}$). In the case of interpreter logics, the relevant notion of sameness of interpretations is as follows. Suppose $V_0 \xrightarrow{K_0} W_0$, $V_1 \xrightarrow{K_1} W_1$.

- $K_0 \approx K_1$ iff, there are $V_0 \xrightarrow{M} V_1$, $V_1 \xrightarrow{\breve{M}} V_0$, $W_0 \xrightarrow{P} W_1$, $W_1 \xrightarrow{\breve{P}} W_0$, such that $M, \breve{M}$ and $P, \breve{P}$ are pairs of inverses in $\mathbb{E}$ and $K_1 \circ M = P \circ K_0$ in $\mathbb{E}$.

$$
\begin{array}{ccc}
V_0 & \xrightarrow{\ K_0\ } & W_0 \\
{\scriptstyle M,\breve{M}}\big\uparrow & & \big\uparrow{\scriptstyle P,\breve{P}} \\
V_1 & \xrightarrow{\ K_1\ } & W_1
\end{array}
$$

We note that $\approx$ is simply isomorphism in the arrow category $\mathsf{Arr}(\mathbb{E})$.

We have the following theorem.

Theorem C.1. *Suppose U_0 is Friedman-reflexive and A_0, A_1 are finitely axiomatised. Suppose further that $U_0 \xrightarrow{K_0} A_0$, $U_1 \xrightarrow{K_1} A_1$ and $K_0 \approx K_1$. Then U_1 is Friedman-reflexive and $\Lambda_{K_0}^{\mathsf{fr}} = \Lambda_{K_1}^{\mathsf{fr}}$.*

The theorem is one of these somewhat annoying examples where the truth is immediately clear but it still requires some work to really prove it.

To prove the theorem, we need a few lemmas. We assume the conditions of the theorem. Since, $K_0 \approx K_1$, we can find $M, \check{M}, P, \check{P}$ such that the following diagram commutes.

$$
\begin{array}{ccc}
U_0 & \xrightarrow{K_0} & A_0 \\
{\scriptstyle M,\check{M}}\big\uparrow\big\downarrow & & \big\uparrow\big\downarrow{\scriptstyle P,\check{P}} \\
U_1 & \xrightarrow{K_1} & A_1
\end{array}
$$

Here M and P correspond to the down-direction of the isomorphism and $\check{M}$ and $\check{P}$ go up.

The fact that U_1 is Friedman reflexive is immediate from Theorem 5.3.

Lemma C.2. $A_0 \vdash \blacklozenge_{K_0,A_0} B \leftrightarrow \blacklozenge_{K_1,A_1}^{\check{P}} B^P$ and $A_1 \vdash \blacklozenge_{K_1,A_1} D \leftrightarrow \blacklozenge_{K_0,A_0}^{P} D^{\check{P}}$.

Proof. By symmetry, we only have to prove the first conjunct of the lemma. We first prove:

$$(\dagger) \quad U_0 \vdash \lozenge_{(U_0),A_0} B \leftrightarrow \lozenge_{(U_1),A_1}^{\check{M}} B^P.$$

We have:

$$
\begin{aligned}
U_0 + \lozenge_{(U_1),A_1}^{\check{M}} B^P &\rhd U_1 + \lozenge_{(U_1),A_1} B^P \\
&\rhd (A_1 + B^P) \\
&\rhd (A_0 + B)
\end{aligned}
$$

It follows that

$$(\ddagger) \quad U_0 + \lozenge_{(U_1),A_1}^{\check{M}} B^P \vdash \lozenge_{(U_0),A_0} B.$$

By symmetry, we find $U_1 + \lozenge_{(U_0),A_0}^{M} D^Q \vdash \lozenge_{(U_1),A_1} D$. Substituting B^P for D gives $U_1 + \lozenge_{U_0,A_0}^{M} B^{PQ} \vdash \lozenge_{(U_1),A_1} B^P$. By Theorem 3.5, we may replace provable equivalents under $\lozenge_{(U_0)}$, so: $U_1 + \lozenge_{(U_0),A_0}^{M} B \vdash \lozenge_{(U_1),A_1} B^P$. It follows that

$$U_0 + \lozenge_{(U_0),A_0}^{M\check{M}} B \vdash \lozenge_{(U_1).A_1}^{\check{M}} B^P$$

and, hence,

$$(\$) \quad U_0 + \lozenge_{(U_0),A_0} B \vdash \lozenge_{(U_1),A_1}^{\check{M}} B^P.$$

Combining ($\ddagger$) and ($\$$), we find ($\dagger$). In its turn ($\dagger$) gives us:

$$
\begin{aligned}
A_0 \vdash \blacklozenge_{K_0,A_0} B &\leftrightarrow \lozenge_{(U_0),A_0}^{K_0} B \\
&\leftrightarrow \lozenge_{(U_1),A_1}^{\check{M}K_0} B^P \\
&\leftrightarrow \lozenge_{(U_1),A_1}^{K_1\check{P}} B^P \\
&\leftrightarrow \blacklozenge_{K_1,A_1}^{\check{P}} B^P \qquad \Box
\end{aligned}
$$

Suppose τ_i is a function from the propositional atoms to the A_i-language. To simplify notations a bit we will confuse, e.g., P and the mapping $D \mapsto D^P$.

Lemma C.3. *We have:*

$$A_0 \vdash \varphi^{(\tau_0, K_0)} \leftrightarrow \varphi^{(P \circ \tau_0, K_1)\breve{P}} \ and \ A_1 \vdash \varphi^{(\tau_1, K_1)} \leftrightarrow \varphi^{(\breve{P} \circ \tau_1, K_0)P}.$$

Proof. We prove the first conjunct. Our proof is by induction on φ. In the atomic case we have:

$$A_0 \vdash p^{(\tau_0, K_0)} \leftrightarrow \tau_0(p)$$
$$\leftrightarrow (\tau_0(p))^{P\breve{P}}$$
$$\leftrightarrow p^{(P \circ \tau_0, K_1)\breve{P}}$$

The cases of the truth-functional connectives are simple. Finally, for the case of $\diamond$, we employ Lemma C.2.

$$A_0 \vdash (\diamond \psi)^{(\tau_0, K_0)} \leftrightarrow \blacklozenge_{K_0, A_0} \psi^{(\tau_0, K_0)}$$
$$\leftrightarrow \blacklozenge_{K_0, A_0} \psi^{(P \circ \tau_0, K_1)\breve{P}}$$
$$\leftrightarrow (\blacklozenge_{K_1, A_1} \psi^{(P \circ \tau_0, K_1)\breve{P}P})^{\breve{P}}$$
$$\leftrightarrow (\blacklozenge_{K_1, A_1} \psi^{(P \circ \tau_0, K_1)})^{\breve{P}}$$
$$\leftrightarrow (\diamond \psi)^{(P \circ \tau_0, K_1)\breve{P}} \qquad \square$$

Finally, we prove the theorem.

Proof of Theorem C.1. We have:

$$A_1 \vdash \varphi^{(P \circ \tau, K_1)} \Rightarrow A_0 \vdash \varphi^{(P \circ \tau, K_1)\breve{P}}$$
$$\Rightarrow A_0 \vdash \varphi^{(\tau, K_0)}$$

It follows that, if $\Lambda^{\mathsf{fr}}_{K_1} \vdash \varphi$, then $\Lambda^{\mathsf{fr}}_{K_0} \vdash \varphi$. By symmetry, we also have the other direction. $\square$

We have a second preservation theorem.

Theorem C.4. *Let F be an endofunctor of $\mathbb{D}$. Here we suppose that F is specified on concrete theories and interpretations. We assume that:*

- *F preserves finite axiomatisability.*
- *For each theory V, there is a faithful interpretation $V \xrightarrow{\eta_V} F(V)$.*
- *Suppose Γ is a set of sentences in the V-language. Then, $F(V + \Gamma) = F(V) + \Gamma^{\eta_V}$.[21]*

We note that we can drop the 'faithful' in Condition (B), if we demand that F preserves consistency.

Let U be Friedman-reflexive and suppose $F(V) \xrightarrow{M_V} V$, for all extensions V of U in the same language. Here there is no further constraint on the M_V. Let A be finitely axiomatised and suppose $U \xrightarrow{K} A$.

We have: $\Lambda^{\mathsf{fr}}_{\eta_A \circ K} \subseteq \Lambda^{\mathsf{fr}}_K$.

---

[21] This condition does not look very 'categorical', since it goes into the hardware. In Appendix C, we discuss a more categorical formulation.

Proof. For some P, we have $(A + B) \xrightarrow{P} (U + \Diamond_A B)$. Let $M' := M_{U + \Diamond_A B}$. We have:

$$(F(A) + B^{\eta_A}) = F(A + B) \xrightarrow{F(P)} F(U + \Diamond_A B) \xrightarrow{M'} (U + \Diamond_A B)$$

So, (a) $U \vdash \Diamond_A B \to \Diamond_{F(A)} B^{\eta_A}$.

For some Q, we have $Q : (F(A) + B^{\eta_A}) \xrightarrow{Q} (U + \Diamond_{F(A)} B^{\eta_A})$. So, we have:

$$(A + B) \xrightarrow{\eta_{A+B}} F(A + B) = (F(A) + B^{\eta_A}) \xrightarrow{Q} (U + \Diamond_{F(A)} B^{\eta_A})$$

So, (b) $U \vdash \Diamond_{F(A)} B^{\eta_A} \to \Diamond_A B$. Combining (a) and (b), we find:

$$U \vdash \Diamond_{F(A)} B^{\eta_A} \leftrightarrow \Diamond_A B.$$

We may conclude that:

$$(\dagger) \quad F(A) \vdash \blacklozenge_{\eta_A \circ K, F(A)} B^{\eta_A} \leftrightarrow \blacklozenge^{\eta_A}_{K,A} B.$$

Let σ be any function from propositional variables to sentences of the A-language. To lighten our notational burdens a bit, we will confuse η_A and the mapping: $D \mapsto D^{\eta_A}$. By induction, we prove that $F(A) \vdash \varphi^{(\eta_A \circ \sigma, \eta_A \circ K)} \leftrightarrow (\varphi^{(\sigma,K)})^{\eta_A}$. We treat the case of $\Diamond$. Let $\varphi := \Diamond \psi$. Using $(\dagger)$, we find:

$$
\begin{aligned}
F(A) \vdash \varphi^{(\eta_A \circ \sigma, \eta_A \circ K)} &\leftrightarrow \blacklozenge_{\eta_A \circ K, F(A)} \psi^{(\eta_A \circ \sigma, \eta_A \circ K)} \\
&\leftrightarrow \blacklozenge_{\eta_A \circ K, F(A)} (\psi^{(\sigma,K)})^{\eta_A} \\
&\leftrightarrow \blacklozenge^{\eta_A}_{K,A} \psi^{(\sigma,K)} \\
&\leftrightarrow (\varphi^{(\sigma,K)})^{\eta_A}
\end{aligned}
$$

From the fact that η_A is faithful, we now have:

$$
\begin{aligned}
(\ddagger) \quad F(U) \vdash \varphi^{(\eta_A \circ \sigma, \eta_A \circ K)} &\Leftrightarrow F(U) \vdash (\varphi^{(\sigma,K)})^{\eta_A} \\
&\Leftrightarrow U \vdash \varphi^{(\sigma,K)}
\end{aligned}
$$

Our theorem is immediate from $(\ddagger)$. ❏

Let $\mathsf{in}_i^{V_0,V_1}$ be the obvious interpretation of V_i in $V_0 \oslash V_1$.

Corollary C.5. *Suppose $K : A \rhd U$ is an FM-interpretation. Then, we have $\Lambda^{\mathsf{fr}}_{\mathsf{in}_0^{A,B} \circ K} \subseteq \Lambda^{\mathsf{fr}}_K$.*

Here is a direct application of Theorem C.4. We consider the functor Seq of Remark 7.7. Let $\jmath_V$ be the interpretation based on relativisation to D.

Theorem C.6. *Suppose U is sequential and Friedman-reflexive and $K : A \rhd U$. Then, $\Lambda^{\mathsf{fr}}_{\jmath_A \circ K} \subseteq \Lambda^{\mathsf{fr}}_K$.*

Proof. We apply Theorem C.4 with $\jmath_A$ in the role of η_A. ❏

Here are two simple observations.

Theorem C.7. *Suppose U is Friedman-reflexive and $K : U \lhd A$. Let B be a sentence in the A-language. Then, $A \vdash \blacklozenge_{K,A}(B \wedge C) \leftrightarrow \blacklozenge_{K,A+B} C$.*

Suppose $K : A \triangleright U$. Let $\mathsf{Th}(K) := \{B \mid A \vdash B^K\}$.

Theorem C.8. *Suppose U is Friedman-reflexive and $K : U \triangleleft A$. Let $K^* : \mathsf{Th}(K) \triangleleft A$ be the interpretation based on τ_K the translation given with K. Then, $\Lambda_K^{\mathsf{fr}} = \Lambda_{K^*}^{\mathsf{fr}}$.*

We leave the proofs to the reader.

Open Question C.9. Suppose $K, M : U \triangleleft A$ are FM-interpretations and $\mathsf{Th}(K) = \mathsf{Th}(M)$. Do we have $\Lambda_K^{\mathsf{fr}} = \Lambda_M^{\mathsf{fr}}$, or is there a counter-example? ◯

Remark C.10. The conditions for Theorem C.4 look somewhat *ad hoc*. Especially, Condition (C) goes into the hardware in an unelegant way. The following (more specific) conditions look a little bit better. However, they have the disadvantage that they do not (yet) apply to our main application: we did not supply a version of Seq that is an endofunctor of $\mathbb{E}$.

We work in category $\mathbb{E}$ enriched with designated embedding arrows $V \xrightarrow{\mathsf{E}_{VW}} W$ between theories of the same signature that are based on the identity translation. We demand that F is a functor on the enriched category that preserves (modulo sameness) the embedding arrows. Let $\bot$ be the theory in the language of identity axiomatised by $\bot$.

Our new conditions are as follows.

A*. F preserves finite axiomatisability.
B*. If $\bot \longrightarrow F(V)$, then $\bot \longrightarrow V$.
C*. For each V, there is an interpretation $V \xrightarrow{\eta_V} F(V)$ such that:
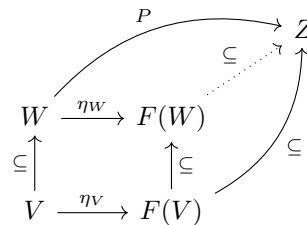
  If $V \subseteq W$, then, $\mathsf{E}_{F(V)F(W)} \circ \eta_V = \eta_W \circ \mathsf{E}_{VW}$. Moreover, whenever $\mathsf{E}_{F(V)Z} \circ \eta_V = P \circ \mathsf{E}_{VW}$, then $F(W) \subseteq Z$ and $P = \mathsf{E}_{F(W)Z} \circ \eta_W$.

It is easy to see that the new conditions imply the old ones. ◯

Appendix D. Local logics for a frame

There are other notions that can be considered than just the logic associated with an FM-frame or an FM-interpretation. Let $\langle A, U \rangle$ be an FM-frame. Let X range over finite sets of assignments from the propositional variables to A-sentences and let K range over interpretations $K : A \triangleright U$. We define:

- $\Lambda_{A,U}^\star := \{\varphi \mid \forall X \, \exists K \, \forall \sigma \in X \; A \vdash \varphi^{(\sigma, K)}\}$.
- Λ *is a local logic for* $\langle A, U \rangle$ if Λ is a subset of $\Lambda_{A,U}^\star$ that contains $\mathsf{K4}$ and is closed under modus ponens, necessitation, and substitution.

We have the following obvious facts.

Theorem D.1. $\Lambda_{A,U}^{\mathsf{fr}} \subseteq \Lambda_{A,U}^\star$.

Thus, the class of local logics is non-empty.

Theorem D.2. *Consider an FM-frame $\langle A, U \rangle$. Then, $\Lambda^{\star}_{A,U}$ is closed under substitution and necessitation. Moreover, if φ is in $\Lambda^{\star}_{A,U}$ and ψ is in $\Lambda^{\mathsf{fr}}_{A,U}$, then $(\varphi \wedge \psi)$ is in $\Lambda^{\star}_{A,U}$.*

Theorem D.3. *Consider an FM-frame $\langle A, U \rangle$. A local logic for $\langle A, U \rangle$ is a logic. Moreover, for every $\varphi \in \Lambda^{\star}_{A,U}$, there is a minimal local logic Λ such that $\varphi \in \Lambda$.*

Theorem D.4. *Consider an FM-frame $\langle A, U \rangle$. Let Λ be a local logic for $\langle A, U \rangle$. Then, the logic generated by Λ and $\Lambda^{\mathsf{fr}}_{A,U}$ is a local logic for $\langle A, U \rangle$.*

We proceed with a characterisation of reflection. Consider theories V and W. We say that $\bowtie$ has the *forward property* for V, W iff, for all B in the V-language, there is a C in the W-language, such that $(V + B) \bowtie (W + C)$.

Theorem D.5. *Consider an FM-frame $\langle A, U \rangle$. Then, $\bowtie$ has the forward property for A, U iff $\mathsf{S4}$ is a local logic for $\langle A, U \rangle$.*

Proof. Suppose $\bowtie$ has the forward property for A, U. Consider any sentences $B_0, \ldots, B_{n-1}$ in the language of A. Let $B_0^{\star}, \ldots, B_{k-1}^{\star}$ enumerate all conjunctions of the form $\bigwedge_{i<n} \pm B_i$, where $\pm B_i$ is either B_i or $\neg B_i$.

Consider any $B_j^{\star}$. We have, for some $C_j^{\star}$ that $(A + B_j^{\star}) \bowtie (U + C_j^{\star})$. It follows that $U \vdash C_j^{\star} \to \lozenge\!\!\lozenge_A B_j^{\star}$. Ergo, $(A + B_j^{\star}) \rhd (U + \lozenge\!\!\lozenge_A B_j^{\star})$. Let K_j be the witnessing interpretation. It follows that $A + B_j^{\star} \vdash \blacklozenge_{K_j,A} B_j^{\star}$. If B_i is unnegated in $B_j^{\star}$, then $A \vdash \blacklozenge_{K_j,A} B_j^{\star} \to \blacklozenge_{K_j,A} B_i$. If B_i is negated in $B_j^{\star}$, then $A + B_j^{\star} \vdash B_i \to \blacklozenge_{K_j,A} B_i$. So, $A + B_j^{\star} \vdash \bigwedge_{i<n}(B_i \to \blacklozenge_{K_j,A} B_i)$

Now let $K := K_0 \langle B_0^{\star} \rangle (K_1 \langle B_1^{\star} \rangle \ldots)$. Because the $B_j^{\star}$ are mutually exclusive, we find that, for each $j < k$, we have $A + B_j^{\star} \vdash \bigwedge_{i<n}(B_i \to \blacklozenge_{K,A} B_i)$ and, thus $A \vdash \bigwedge_{i<n}(B_i \to \blacklozenge_{K,A} B_i)$. From this, it is immediate that $p \to \lozenge p$ is in $\Lambda^{\star}_{A,U}$. Moreover, $\mathsf{S4}$ is the logic generated by $p \to \lozenge p$ over $\mathsf{K4}$.

In the other direction, suppose $\mathsf{S4}$ is a local logic for $\langle A, U \rangle$. Consider any B in the A-language. For some $K : A \rhd U$, we have $A + B \vdash \blacklozenge_{K,A} B$. So, $(A + B) \rhd (U + \lozenge\!\!\lozenge_A B)$. Moreover, $(U + \lozenge\!\!\lozenge_A B) \rhd (A + B)$. □

Open Question D.6. The second part of the proof of Theorem D.5 only uses a singleton set X. We wonder whether that means that our approach may be simplified. ◯

We note that it follows that the logic generated by $\mathsf{S4}$ and $\Lambda^{\star}_{A,U}$ is a local logic for $\langle A, U \rangle$.

Appendix E. List of questions

Here are all questions asked in the paper.

Q1. Is there a theory U that is consistent, effectively Friedman-reflexive and not strongly essentially reflexive? This is Question 3.13.

Q2. Can we find a more inspiring example of a theory with logic $\mathsf{S4}$ than Example 6.9? Is it, perhaps, possible to find an FM-interpretation with interpreter logic precisely $\mathsf{S4}$? This is Question 6.10.

Q3. Is DA reflexive? If, against expectation, it turns out to be reflexive, can we modify the construction to find a non-reflexive, Friedman-reflexive, sequential theory? This is Question 8.8(i).

Q4. Is there a finitely axiomatised A and $K : A \rhd \mathsf{DA}$, such that, for no $D(x)$ in the A language, we have, for all B in the A-language, $A \vdash D(\ulcorner B \urcorner) \leftrightarrow \blacklozenge_{K,A} B$? Here the numerals are the K-numerals. This is Question 8.8(ii).

Q5. Is there an RE sequential theory that is Friedman-reflexive but not effectively so? This is Question 8.8(iii).

Q6. Suppose U is sequential and restrictedly (effectively) Friedman-reflexive. Does it follow that U is essentially sententially reflexive? This is Question 8.8(iv).

Q7. Can we find a more canonical axiom scheme for Peano Corto in a coordinate-free way? We note that the cuts are already schematic. The whole problem is in replacing the schematic variable that ranges over Σ_1^0-*sentences* by an unrestricted one that ranges over arbitrary formulas. This is Question 9.2(i).

Q8. Is there a coordinate-free specification of an interpretation of Peano Corto in EA or in some $I\Sigma_n$? This is Question 9.2(ii).

Q9. Do we have $EA \rhd IIA$? This is Question 9.2(iii).

Q10. Does Seq or, some appropriate variant of it, lift to a functor on $\mathbb{E}$? This is Question A.7.

Q11. Is there an example of an FM-interpretation $K : A \rhd U$, where U is complete, with an interesting interpreter logic? This is Question B.4.

Q12. Is there a consistent finitely axiomatised theory that is effectively Friedman-reflexive? This is Question B.5.

Q13. Suppose $K, M : U \lhd A$ are FM-interpretations and $\mathsf{Th}(K) = \mathsf{Th}(M)$. Do we have $\Lambda_K^{\mathsf{fr}} = \Lambda_M^{\mathsf{fr}}$, or is there a counter-example? This is Question C.9.

Q14. The second part of the proof of Theorem D.5 only uses a singleton set X. We wonder whether that means that our approach may be simplified. This is Question D.6.

References

[1] G. Boolos, The Logic of Provability, Cambridge University Press, Cambridge, 1993.
[2] S.R. Buss, Bounded Arithmetic, Bibliopolis, Napoli, 1986.
[3] D.H.J. de Jongh, M. Jumelet, F. Montagna, On the proof of Solovay's theorem, Stud. Log. 50 (1991) 51–70.
[4] S. Feferman, Arithmetization of metamathematics in a general setting, Fundam. Math. 49 (1960) 35–92.
[5] H. Friedman, Aspects of Gödel incompleteness. This is the third revised lecture notes for Harvey Friedman's talk delivered on 8/16/21 at the Online International Workshop on Gödel's Incompleteness Theorems, in Wuhan, The lecture is available at https://u.osu.edu/friedman.8/foundational-adventures/downloadable-lecture-notes-2/, 2021.
[6] D. Guaspari, R.M. Solovay, Rosser sentences, Ann. Math. Log. 16 (1979) 81–99.
[7] V. Halbach, G. Leigh, The Road to Paradox: A Guide to Syntax, Truth, and Modality, Cambridge University Press, 2022, in press.
[8] P. Hájek, P. Pudlák, Metamathematics of First-Order Arithmetic, Perspectives in Mathematical Logic, Springer, Berlin, 1993.
[9] J. Krajíček, A note on proofs of falsehood, Arch. Math. Log. Grundl.forsch. 26 (1) (1987) 169–176.
[10] P. Pudlák, Cuts, consistency statements and interpretations, J. Symb. Log. 50 (2) (1985) 423–441.
[11] P. Pakhomov, A. Visser, Finitely axiomatized theories lack self-comprehension, arXiv preprint, arXiv:2109.02548, 2021, Bull. Lond. Math. Soc. (2022), in press.
[12] F. Pakhomov, A. Zapryagaev, Multi-dimensional interpretations of Presburger Arithmetic in itself, J. Log. Comput. 30 (8) (2020) 1681–1693.
[13] J.R. Schoenfield, Degrees of formal systems 1, J. Symb. Log. 23 (4) (1958) 389–392.
[14] C. Smoryński, Nonstandard models and related developments, in: L.A. Harrington, M.D. Morley, A. Scedrov, S.G. Simpson (Eds.), Harvey Friedman's Research on the Foundations of Mathematics, North Holland, Amsterdam, 1985, pp. 179–229.
[15] C. Smoryński, Self-Reference and Modal Logic, Universitext, Springer, New York, 1985.
[16] R.L. Vaught, On a theorem of Cobham concerning undecidable theories, in: E. Nagel, P. Suppes, A. Tarski (Eds.), Logic, Methodology and Philosophy of Science, in: Proceedings of the 1960 International Congress, Stanford University Press, Stanford, 1962, pp. 14–25.
[17] R.A. Vaught, Axiomatizability by a schema, J. Symb. Log. 32 (4) (1967) 473–479.
[18] A. Visser, An inside view of EXP, J. Symb. Log. 57 (1) (1992) 131–165.
[19] A. Visser, The unprovability of small inconsistency, Arch. Math. Log. 32 (4) (1993) 275–298.
[20] A. Visser, Faith & falsity: a study of faithful interpretations and false Σ_1^0-sentences, Ann. Pure Appl. Log. 131 (1–3) (2005) 103–131.
[21] A. Visser, The predicative Frege hierarchy, Ann. Pure Appl. Log. 160 (2) (2009) 129–153, https://doi.org/10.1016/j.apal.2009.02.001.
[22] A. Visser, Can we make the Second Incompleteness Theorem coordinate free? J. Log. Comput. 21 (4) (2011) 543–560.
[23] A. Visser, The Second Incompleteness Theorem and bounded interpretations, Stud. Log. 100 (1–2) (2012) 399–418, https://doi.org/10.1007/s11225-012-9385-z.
[24] A. Visser, Vaught's theorem on axiomatizability by a scheme, Bull. Symb. Log. 18 (3) (2012) 382–402.

[25] A. Visser, What is sequentiality? in: P. Cégielski, Ch. Cornaros, C. Dimitracopoulos (Eds.), New Studies in Weak Arithmetics, in: CSLI Lecture Notes, vol. 211, CSLI Publications and Presses Universitaires du Pôle de Recherche et d'Enseingement Supérieur Paris-est, Stanford, 2013, pp. 229–269.

[26] A. Visser, Peano Corto, Peano Basso, A study of local induction in the context of weak theories, Math. Log. Q. 60 (1–2) (2014) 92–117.

[27] A. Visser, Why the theory R is special, in: Neil Tennant (Ed.), Foundational Adventures. Essays in Honour of Harvey Friedman, College Publications, UK, 2014, pp. 7–23, Originally published online by Templeton Press in 2012, http://foundationaladventures.com/.

[28] A. Visser, The arithmetics of a theory, Notre Dame J. Form. Log. 56 (1) (2015) 81–119.

[29] A. Visser, The second incompleteness theorem: reflections and ruminations, in: Gödel's Disjunction: The Scope and Limits of Mathematical Knowledge, Oxford University Press, 2016, pp. 67–90.

[30] A. Visser, On Q, Soft Comput. 21 (1) (2017) 39–56.

[31] A. Visser, The interpretation existence lemma, in: Feferman on Foundations, in: Outstanding Contributions to Logic, vol. 13, Springer, New York, 2018, pp. 101–144.

[32] A. Visser, The small-is-very-small principle, Math. Log. Q. 65 (4) (2019) 453–478.

[33] A. Visser, Extension and interpretability, in: Mathematics, Logic and Their Philosophies, in: Logic, Epistemology, and the Unity of Science, vol. 49, Springer Nature, 2021, pp. 53–92.

[34] A.J. Wilkie, J.B. Paris, On the scheme of induction for bounded arithmetic formulas, Ann. Pure Appl. Log. 35 (1987) 261–302.