In the present paper, we explore an idea of Harvey Friedman to obtain a coordinate-free presentation of consistency. Friedman shows that, over Peano Arithmetic, the consistency statement for a finitely axiomatised theory $A$ can be characterised as the weakest statement $C$ over Peano Arithmetic such that $\text{PA} + C$ interprets $A$.

We study the question which base theories $U$ have the property that, for any finitely axiomatised $A$, there is a weakest $C$ such that $U + C$ interprets $A$. We call such theories Friedman-reflexive. We explore various implications of Friedman-reflexiveness.

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 Friedman-reflexive 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.

The consistency-like statements provided by a Friedman-reflexive base $U$ can be used to define a provability-like notion for a finitely axiomatised $A$ that interprets $U$ via an interpretation $K$ of $U$ in $A$. We explore what modal logics this idea gives rise to. We call such logics interpreter logics. We show that, generally, these logics satisfy the L"ob Conditions, aka K4. We provide conditions for when these logics extend $\text{S4}$, $\text{K45}$, and L"ob's Logic. We show that, if either $U$ or $A$ is sequential, then the condition for extending L"ob'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 G"odel numbers.

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

1. Introduction

'Consistoids' are 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. Secondly, on a more adventurous note, consistoids 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. Not only, do we think this is interesting in itself, but the wider view allows us also to look at the standard cases from a more general vantage point.

We develop a strategy proposed by Harvey Friedman. See [Fri21]. 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, consistoids 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 consistoids 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 get. 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. 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. In case $B$ sequential and the association of an interpreter to $A$ is effective for $U$, we show that the proof of Solovay’s 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 11.

**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 still plays a role.

1.1. **On Reading this Paper.** Appendix A gives basic definitions and basic facts and some references to further literature.

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., [HP93] and [Vis13]. However, the main development of

---

1 A different example of a study of a G2-analogue that works in a wider context can be found in [PV21].

2 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.
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 the main 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:

\[ U \not \vDash (B + \text{Con}(U)), \]

in other words, \( U \) does not interpret the base theory plus its own consistency. Here \( \text{Con}(U) \) 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: \( U \not \vDash (Q + \Diamond_\alpha \top) \)\footnote{A first form of the no-interpretation version is due to Feferman. See \cite{Fe60}. Feferman’s form was stated for extensions of \( \text{PA} \) and for the case that main theory and base coincide.} Here \( \Diamond_\alpha \top \) is a specific form of the consistency statement for \( U \), where we formalised consistency in arithmetic is a sufficiently good way and where \( \alpha \) is a \( \Sigma^0_1 \)-formula representing the axiom set of \( U \). 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 \( \text{PA} \) with only zero, successor, plus and times in the signature. On the other hand, \( ZF \) has a standard interpretation of \( \text{PA} \) in the finite von Neumann ordinals. We can simply import the arithmetisation of syntax as usually done in \( \text{PA} \) in \( ZF \) via the von Neumann interpretation. The advantage of doing it like this 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 \( U \not \vDash (S^1_2 + \Diamond_\alpha \top) \). Here \( S^1_2 \) is Buss’s weak arithmetical theory for the study of p-time computability.\footnote{Pavel Pudlák contributed the main ingredient of the result in the present formulation. See \cite{Pud85}.} Without any additional effort, as compared to, e.g., \( \text{PA} \), Gödel’s reasoning can be repeated in \( S^1_2 \). In fact, \( S^1_2 \) is better, since it prevents all kinds of silly and inefficient choices of doing the arithmetisation. Moreover, Pudlák’s argument for his strengthened version can be framed as first showing that \( (Q + \Diamond_\alpha \top) \not \vDash (S^1_2 + \Diamond_\alpha \top) \) and, then, concluding the Q-version from the \( S^1_2 \)-version. We think that the \( S^1_2 \)-version should be viewed as G2 proper, since it is proved simply by Gödel’s original reasoning.

\[3\text{Pavel Pudlák contributed the main ingredient of the result in the present formulation. See } \cite{Pud85}.\]

\[4\text{See } \cite{Bus86} \text{ or } \cite{HP93} \text{ for the basic development of } S^1_2.\]
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, [HL20]. 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: ‘consistoids’. 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 consistoids.

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 [Vis16] 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 [Fri21], 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 [Vis18] for a detailed exposition of this lemma). The lemma says, roughly, that we can interpret a theory $U$ in a suitable base theory plus a consistency statement for $U$.

In case $B$ extends $S_2^1$, $V$ is RE and $\alpha$ is $\Sigma^0_1$, we have the following formulation of Interpretation Existence: $(B + \phi_\alpha \top) \vdash V$. We will say that any $B$-sentence $B$ such that $(B + B) \vdash V$ is a pro-interpreter of $V$ (over $B$). So, Interpretation Existence tells us that $\phi_\alpha \top$ is a pro-interpreter. A study of pro-interpreters was undertaken in [Vis12a].

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 $(B + C) \vdash A$ and, for all $B$-sentences $B$, we have, if $(B + B) \vdash A$, then $B + B \vdash C$. Alternatively, we can say that $C$ is an interpreter of $A$ over $B$ iff, for all $B$-sentences $B$, $(B + B) \vdash A$ iff $B + B \vdash C$.

Unfortunately, over the base theory $S_2^1$, this idea will not work. Consider any consistent $A$ such that $S_2^1 \not\vdash A$. E.g., $A$ could be Elementary Arithmetic $EA$. 
Suppose we had an interpreter $C$ of $A$. Then, we have, for any $S^1_2$-cut $I$ that $(S^1_2 + \Diamond^A_1 \top) \models A$. So, for all $I$, we find $S^1_2 + \Diamond^A_1 \top \models C$. In other words, for all $I$, we find $S^1_2 + \neg C \models \Box^A_1 \bot$. Since $S^1_2 \not \models A$, we have $S^1_2 \not \models C$ and, hence, $S^1_2 + \neg C$ is consistent. By Theorem A.2, we find that $\Box^A_1 \bot$ is true, and, thus, that $A$ is inconsistent, which contradicts our assumption. 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 $\text{Friedman-reflexive}$. Friedman’s example of such a theory is Peano Arithmetic $\text{PA}$. He shows that, indeed, $\text{PA}$ is Friedman-reflexive. See Section 3. A disadvantage of the choice of $\text{PA}$ is that it is very strong. Our proposal for the ideal Friedman-reflexive theory is Peano Corto or $\text{PA}^\dagger$. We will show, in Subsection 3.3, that $\text{PA}^\dagger$ 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 in finite signature. These theories, generally, need not be RE. We allow inconsistent theories as values.

The variables $A, B, \ldots$ range ambiguously over sentences and over finitely axiomatised theories. We confuse the finite conjunction of the finitely axiomatised theory $A$ with a single sentence $A$.

There is a subtle point here. Let $U$ be any theory and let $\tau$ 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 $A$. We assume that $\exists x \ x = x$ is among these axioms. Then, $U + (E_A \land A)^\tau$ interprets $A$ with an interpretation based on $\tau$. 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: $(U + B) \models A$ 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 $\diamond (\cdot)$ for a function that selects, for each $A$, an interpreter $C$ of $A$ over $U$. So, $\diamond$ is only uniquely determined modulo $U$-provability. We write $\diamond_A B$ for $\diamond(A \land B)$. If we want to make the dependence on $U$ explicit, we write $\diamond_{(U)}, \diamond_{(U), A}$, etcetera. However, we prefer to treat the dependence on $U$ as contextually given as much as possible. We write $\blacksquare$ for $\neg \diamond \neg$ and $\blacksquare_A$ for $\neg \diamond_A \neg$.

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

---

6Our cuts are downward closed w.r.t. $\prec$ and are closed under zero, successor, addition, multiplication and $\omega_1$, i.e., the function $x \mapsto 2^{(2 \log(x))^2}$.

7A stronger version of this insight is given in Theorem 10.9.

8The ratio behind the second name will become clear in the paper.

9Here 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 $A$. 
The theory $U$ is effectively Friedman-reflexive iff we can choose $\diamond$ to be recursive. Note the existential quantifier here: there may very well be both recursive and non-recursive choices of $\diamond$. In fact, we will see a salient example of that possibility.

3.2. The Categorical Viewpoint. We can view what is going on here in category theoretic terms as follows. Let $B_U$ 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 $D$ be the partial pre-order category of theories ordered by interpretability $\models$. Let $\text{proj}_U$ be the projection functor from $B_U$ into $D$. Then, $C$ is an interpreter of $A$ iff $A \models (U + C)$ is a universal arrow from $A$ to $\text{proj}_U$.

Remark 3.1. We note that if we restrict $D$ 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 $D$ to the category of RE theories with the same effect.

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

3.3. Semi Normal Form. Interpreters have a kind of non-unique normal forms.

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

We note that $\tau$ need not be unique.

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

On the other hand, $(U + (E_A \land A)^\tau) \models A$. So, by the defining property of interpreters, $U + (E_A \land A)^\tau \models C$.

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 $(U + C) \models C$, if follows that $(U + C) \models A$, and, hence $(U + C') \models C$. 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 $\diamond$ is has a functorial property w.r.t. a Friedman-reflexive theory.

Theorem 3.5. Let $U$ be Friedman-reflexive and suppose $A \models B$. Then, we have $U \vdash \diamond A \rightarrow \diamond B$.

Proof. Let $U$ be Friedman-reflexive. Suppose $A \models B$. Then, $(U + \diamond A) \models A \models B$. So, $U \vdash \diamond A \rightarrow \diamond B$. 

\footnote{In the light of our specific application, we can replace $\models$ by $\models_{\text{loc}}$.}
Remark 3.6. Theorem [5.5] is a special case of an elementary category theoretic insight. Suppose we have categories $A$, $B$, $C$ and functors $F : A \to C$, $G : B \to C$. Suppose further that, for every $a$ in $A$ there is a universal arrow from $F(a)$ to $G$. Then, there is a functor $H : A \to B$, such that our promised universal arrows are a natural transformation from $F$ to $G \circ H$.

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 $\Diamond$ commutes, modulo $U$-provable equivalence, with finite disjunctions of sentences in the same signature, including the empty one. In other words, $U \vdash \neg \Diamond \bot$ and, if $A$ and $B$ have the same signature, then $U \vdash (A \lor B) \leftrightarrow (\Diamond A \lor \Diamond B)$.

Proof. We have $(U + \Diamond \bot) \not\supset \bot$. So, $U \vdash \neg \Diamond \bot$.

We have $\vdash (A \lor B)$ and, hence, $A \not\supset (A \lor B)$. It follows, by Theorem [3.5] that $U \vdash \Diamond A \to \Diamond (A \lor B)$. Similarly, $U \vdash \Diamond B \to \Diamond (A \lor B)$. Ergo, $U \vdash (\Diamond A \lor \Diamond B) \to \Diamond (A \lor B)$.

We have $(U + \Diamond (A \lor B)) \not\supset (A \lor B)$. Let $\tau$ be the underlying translation of some witnessing interpretation. We have:

$$
U + \Diamond (A \lor B) \vdash (E_A \land (A \lor B))^\tau \\
\vdash (E_A \land A)^\tau \lor (E_A \land B)^\tau \\
\vdash \Diamond A \lor \Diamond B
$$

We define $W := U \otimes V$ 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 $\otimes$ is a supremum operator in $\mathcal{D}_{\text{fin}}$.

Theorem 3.8. Suppose $U$ is Friedman-reflexive and $A$ and $B$ are finitely axiomatised. Then, $U \vdash \Diamond (A \otimes B) \leftrightarrow (\Diamond A \land \Diamond B)$.

Proof.

$$
U + D \vdash \Diamond (A \otimes B) \leftrightarrow (U + D) \not\supset (A \otimes B) \\
\leftrightarrow (U + D) \not\supset A \text{ and } (U + D) \not\supset B \\
\leftrightarrow U + D \vdash \Diamond A \text{ and } U + D \vdash \Diamond B \\
\leftrightarrow U + D \vdash \Diamond A \land \Diamond B
$$

We will meet $\otimes$ again in Corollary [6.13].

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 $(\dag)$ for every theory $W$, there is a theory $\mathcal{C}(W) \supseteq U$, such that, for all $V \supseteq U$, we have $V \not\supset_{\text{loc}} W$ iff $V \supseteq \mathcal{C}(W)$.
Suppose $U$ is Friedman-reflexive. We prove $(†)$. Consider any $W$. We define $C(W) := U + \{ \Diamond A \mid W \vdash A \}$. Let $V \supseteq U$. We have:

$$V \vdash_W W \iff \forall A (W \vdash A \Rightarrow V \vdash A)$$

$$\iff \forall A (W \vdash A \Rightarrow \exists D (V \vdash D \text{ and } U + D \vdash \Diamond A))$$

$$\iff \forall A (W \vdash A \Rightarrow V \vdash A)$$

$$\iff V \supseteq C(W)$$

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

$$(U + B) \vdash A \iff (U + B) \vdash_{loc} A$$

$$\iff (U + B) \supseteq C(A)$$

$$\iff (U + B) \vdash C$$

So, we can take $\Diamond A := C$. 

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

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 $\tau$ such that $T$ is consistent with $(E_A \land A)^\tau$.

**Theorem 3.10.** $U$ is polyglot iff $U$ locally tolerates the theory $Q$ plus the true $\Pi^0_1$-sentences.

**Proof.** Suppose $U$ is polyglot. Let $A$ be a finite sub-theory of $Q$ plus the true $\Pi^0_1$-sentences. Then, for some $B$, we have $U + B$ is consistent and $(U + B) \vdash A$. Let $\tau$ be the translation on which the interpretation of $A$ in $U + B$ is based. Then, $U$ is consistent with $(E_A \land A)^\tau$.

Conversely, suppose $U$ locally tolerates the theory $Q$ plus the true $\Pi^0_1$-sentences. Consider any finitely axiomatised consistent theory $D$. Then, $\Diamond_D \top$ is true. So, for some $\tau$, we have $U + (E_Q \land Q \land \Diamond_D \top)^\tau$ is consistent. By the Interpretation Existence Lemma, we find that $(U + (E_Q \land Q \land \Diamond_D \top)^\tau) \vdash (Q + \Diamond_D \top) \vdash D$. 

Since $Q$ interprets $S^1_2$ on a cut, we have the same result with $S^1_2$ 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 $\phi$.

Let $[S]$ be a theory of a witness of $\Sigma^0_1$-sentence $S$. See [Vis17] for details. If $S$ is true, then $[S]$ has a finite model and, thus, any theory interprets $[S]$. It follows that $V \vdash \phi[S]$. On the other hand, if $[S]$ is false, then it extends $\mathcal{R}$ and, thus, since $V$ is decidable, $V \not\models [S]$. Ergo, $V \not\models \phi[S]$. It follows that using the decidability of $V$, we can solve the halting problem. Quod non. 

We remark the reader that $T$ tolerates $U$ iff, for some translation $\tau$, the theory $T + E^\tau_U + U^\tau$ 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 is is also strongly essentially undecidable. The same does not have to hold for RE theories. See, e.g., [Sho58] for examples. Cobham showed that the Tarski-Mostowski-Robinson theory $\mathcal{R}$ is strongly essentially undecidable. See [Vau62] or [Vis17]. 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 $V := T + E^\tau_U + U^\tau$ is consistent. In case $S$ is true, we have $V \vdash \phi^\tau[S]$. Now suppose $[S]$ is false. In case $V + \phi^\tau[S]$ were consistent, it would follow that $U$ tolerates $[S]$, contradicting the fact that $[S]$ is finitely axiomatised and essentially undecidable. So, $V \vdash \neg \phi^\tau[S]$. 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 4.4 is Theorem 2.7 of [Fri21].

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

**Proof.** Suppose $(\text{PA} + B) \vdash A$. Then, for some finite sub-theory $D$ of $\text{PA}$, we have $(D + B) \vdash A$. If follows that $\text{PA} \vdash ((D + B) \vdash A)$ and, so, $\text{PA} \vdash \Diamond_D B \rightarrow \Diamond_A \top$. By the essential reflexiveness of $\text{PA}$, we may conclude $\text{PA} + B \vdash \Diamond_A \top$.

For the other direction, suppose $\text{PA} + B \vdash \Diamond_A \top$. Then, by the Interpretation Existence Lemma (see [Vis18]), we find $(\text{PA} + B) \vdash A$.

We note that we have characterised the consistency statement $\Diamond_A \top$ among arithmetical sentences modulo $\text{PA}$-provable equivalence. We will later discuss how to improve this to $\text{EA}$-provable equivalence—see Theorem 9.4.
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, \( \Diamond \) can be chosen the same for \( U \) and \( V \).

**Proof.** Suppose \( U \subseteq V \) and \( U \) is Friedman-reflexive. We have:

\[
(V + B) \models A \iff \text{for some } D, \text{ we have } V \vdash D \text{ and } (U + (D \wedge B)) \models \Diamond A
\]

Since \( \models \) is preserved, we, *ipso facto*, preserve effectivity. \( \square \)

We note that polyglotticity is definitely not preserved to consistent extensions.

**Theorem 5.2.** Suppose \( U \) is Friedman-reflexive. Suppose \( A \) is consistent and \( U \not\models A \). Then, \( U + \neg \Diamond A \) is consistent and not polyglot.

**Proof.** We assume the conditions of the theorem. Since \( U \not\models A \), we have \( U \not\models \Diamond A \) and, so, \( U + \neg \Diamond A \) is consistent.

By Theorem 5.1, we may choose \( \Diamond \) for \( U + \neg \Diamond A \) the same as for \( U \), but, \( \Diamond A \) is inconsistent with \( U + \neg \Diamond A \). \( \square \)

An example of our theorem is the fact that no consistent extension in the same language of \( \text{PA} + \Box \text{ACA}_0 \perp \) interprets \( \text{ACA}_0 \).

Our next closure property is categorical in nature. Let \( \mathcal{E} := \text{INT}_3^\perp \) be the category of theories in finite signature and interpretations, where two interpretations \( K, K' : T \rightarrow W \) 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, \( \mathcal{E} \) 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.

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

**Proof.** Let \( K : V \rightarrow U \) and \( M : U \rightarrow V \) witness the retraction. So, \( M \circ K \) is the same, in the sense of \( \mathcal{E} \), as \( \text{Id}_V \). We have:

\[
(V + B) \models A \Rightarrow (U + B^K) \models A
\]

\[
\Rightarrow U + B^K \models \Diamond A
\]

\[
\Rightarrow V + B^{K_M} \models M A
\]

\[
\Rightarrow V + B \models M A
\]

\[
\Rightarrow (V + B) \models (U + M A)
\]

\[
\Rightarrow (V + B) \models A
\]

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

\[^{11}\text{For the case of finitely axiomatised theories, the interaction between interpretation and extension was studied in [Vis21].}\]
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 $C(D)C'$ be $(D \land C) \lor (\neg D \land C')$. We have:

$$(U + B) \vdash A \iff (U + D + B) \vdash A \text{ and } (U + \neg D + B) \vdash A$$

$$\iff (U + D + B) \vdash C \text{ and } (U + \neg D + B) \vdash C'$$

$$\iff (U + D + B) \vdash C(D)C' \text{ and } (U + \neg D + B) \vdash C(D)C'$$

$$\iff (U + B) \vdash C(D)C'$$

In the next two theorems, we verify general insights for universal arrows in our specific case.

Given theories $U$ and $V$, we define $W := U \oplus V$ 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 $P \rightarrow A$, for axioms $A$ of $U$ and $\neg P \rightarrow B$ for axioms $B$ of $V$. We have:

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

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

$$(W + B) \vdash A \iff (U + B[P := \top]) \vdash A \text{ and } (V + B[P := \bot]) \vdash A$$

$$\iff U + B[P := \top] \vdash \Diamond(U) A \text{ and } V + B[P := \bot] \vdash \Diamond(V) A$$

$$\iff W + B \vdash \Diamond(U) A \land \Diamond(V) A$$

So, we can take $\Diamond(W) A := (\Diamond(U) A)(\Diamond(V) A)$.

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 $A \triangleright U$. The interpretation $K : A \triangleright U$, where $\langle A, U \rangle$ is an FM-frame, is an FM-interpretation. We consider $U$ and $A$ as part of the data for $K$.

Consider an FM-interpretation $K : U \triangleleft A$. We define $\Diamond_{K,B} C := \Diamond_{U,B} K C$. Our default case is where $A$ is identical to $B$. In this case we write $\Diamond_K C$. Also, in many cases, we treat $K$ as contextually given and simply write $\Diamond$. As usual, we set $\Box := \neg \Diamond \neg$.

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

* $p^{(\sigma,K)} := \sigma(p)$.

12 FM stands for ‘Feferman and Montague’ who initiated the idea of looking at the combination of G2 and interpretability.
\begin{itemize}
  \item \((\cdot)^{\sigma,K}\) commutes with the truth-functional connectives.
  \item \((\Box \psi)^{\sigma,K} := \Diamond_{K,A} \psi^{\sigma,K}\).
\end{itemize}

We will call the logic of \(\Diamond_{K,A}\) over \(A\): \(\Lambda^K\). So,

\[\Lambda^K = \{\varphi \mid \text{for all } \sigma, \text{ we have } A \vdash \varphi^{\sigma,K}\}\]

We also define the logic of a frame \((A,U)\).

\[\Lambda^A_{A,U} = \{\varphi \mid \text{for all } M : A \triangleright U \text{ and } \sigma, \text{ we have } A \vdash \varphi^{\sigma,M}\}\]

In other words, \(\Lambda^A_{A,U} = \bigcap_{M : A \triangleright U} \Lambda^M_{M}\).

We consider one further notion in Appendix C.

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 \(\alpha^B\), where \(\alpha\) is a suitable presentation of the axioms of \(V\). The base theory is ‘in’ \(V\) via an interpretation \(K : V \triangleright B\). So, \(V\)-provability gets the form \(K^\alpha B\) in \(V\).

If we switch to interpreter logic, the idea is precisely the same: the necessity operator gets the form \(K A B\) for main theory \(A\). We note that here we have \(K : A \triangleright B\), on the one hand, and that, on the other hand, the \(\Diamond A B\) are defined using interpretations of \((A \land B)\) 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.

Both in the case of ordinary provability logic and 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 \((A,U)\).

6.3. The L"ob Conditions. We will show that every \(\Lambda^A_K\) satisfies the L"ob 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 L"ob Conditions and then prove the equivalence.

**Lemma 6.1.** Let \(K : U \triangleright A\) be an FM-interpretation. We write \(\Diamond := \Diamond_{K,A}\). We have:

\begin{enumerate}
  \item \(A \vdash B \rightarrow C\) implies \(A \vdash \Diamond B \rightarrow \Diamond C\) and \(A \vdash \Box B \rightarrow \Box C\).
  \item \(A \vdash \neg \bot\) and \(A \vdash \Box \top\).
  \item \(A \vdash \Diamond (B \lor C) \leftrightarrow (\Box B \lor \Diamond C)\) and \(A \vdash \Box (B \land C) \leftrightarrow (\Diamond B \land \Box C)\).
  \item \(A \vdash \Box \Box B \rightarrow \Diamond B\) and \(A \vdash \Box B \rightarrow \Box \Box B\).
\end{enumerate}

**Proof.** Ad (a). Suppose \(A \vdash B \rightarrow C\). Then,

\[(U + \Diamond_A B) \triangleright (A \land B) \triangleright (A \land C)\]

So, \(U + \Diamond_A B \vdash \Diamond_A C\), and, thus, \(A \vdash \Diamond B \rightarrow \Diamond C\).

We note that \(\neg \Box \neg \Box B\) is \(\neg \neg \neg \neg \neg \neg B\). So, by the first conjunct of (a), we have \(U \vdash \Diamond B \leftrightarrow \neg \Box \neg B\). As a consequence, we may switch between \(\Diamond\)-versions of principles and their dual \(\Box\)-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 6.7.

Ad (d). We have

\[(U + \Diamond_A \Diamond B) \triangleright (A + \Diamond B) \triangleright (U + \Diamond_A B) \triangleright (A \land B)\]
So, \( U + \Diamond A B \vdash \Diamond A B \). If follows that \( A \vdash \Box \Box B \rightarrow \Box B \). \(\square\)

**Theorem 6.2.** Let \( K : U \triangleleft A \) be an FM-interpretation. Then, \( \Lambda^f_K \) is a normal modal logic extending \( K4 \), in other words, \( \Lambda^f_K \) satisfies the Löb Conditions.

**Proof.** It is sufficient to show that (a-d) of Lemma 6.1 imply the Löb Conditions. Clearly, \( \Box1 \) or necessitation follows from (a,b).

We verify \( L2 \). We have \( A \vdash (B \land (B \rightarrow C)) \rightarrow C \). Ergo, by (a), we find \( A \vdash \Box(B \land (B \rightarrow C)) \rightarrow \Box C \). Applying the second conjunct of (c), we obtain \( A \vdash (\Box B \land \Box(B \rightarrow C)) \rightarrow \Box C \).

Finally, \( L3 \) is identical to (d). \(\square\)

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 and \( K45 \) 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 \( U' \models V' \). 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 6.2 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. The next theorem directly connects irreconcilability to Löb’s Principle.

**Theorem 6.3.** Consider an FM-frame \( \langle A, U \rangle \). Then, \( A \) and \( U \) are irreconcilable iff \( \Lambda^f_{A,U} \) 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 \( K : A \triangleright U \). It is sufficient to show that \( \Lambda^f_K \) extends Löb’s Logic. It is well-known that, over \( K4 \), Löb’s Principle and Löb’s Rule are equivalent. So, it suffices to prove closure of \( \Lambda^f_K \) under Löb’s Rule. Suppose \( A \vdash B \rightarrow \Box B \). Then,

\[
\vdash (A \rightarrow B) \triangleright (U + \Diamond A B) \triangleright (A + B).
\]

So, \( (A + B) \triangleright (U + \Diamond A B) \) 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 \( (A + B) \triangleright (U + C) \), where \( A + B \) is consistent.

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

\[
A \not\vdash \Box K_A (B \rightarrow \Box K_A B) \rightarrow \Box K_A \neg B,
\]

\[\text{(6.3)}\]

See [Sm83], Chapter 1, Section 1, Exercise 5(ii), p75. Smoryński attributes the result to Macintyre and Simmons. Alternatively, see [Boo93], Chapter 3, p59.
by the fact that Löb’s Rule follows from Löb’s Principle. So, we do not have Löb’s Principle in $\Lambda^f_K$.  

Remark 6.4. We note that Theorem 6.3 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 $K : A \triangleright U$ in combination with a mapping $\sigma$ from the propositional atoms to the sentences of the language of $A$ is the analogue of a model on the frame.

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

Theorem 6.5. Suppose $\langle A, U \rangle$ is an FM-frame and $A$ and $U$ are irreconcilable. We have:

a. Suppose $K : A \triangleright U$. Then $A \triangleright (A + \Box_{K,A} \bot)$.

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

Proof. Ad (a). We have:

$$\quad (A + \Box_{K,A} \top) \triangleright (A + \Box_{K,A} \Box_{K,A} \bot) \triangleright (A + \Box_{K,A} \bot)$$

Trivially, $(A + \Box_{K,A} \bot) \triangleright (A + \Box_{K,A} \bot)$. So, by a disjunctive interpretation, we find $A \triangleright (A + \Box_{K,A} \bot)$.

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

$$\quad A \triangleright (A + \Box_{M \otimes P,A} \bot) \triangleright (B + \Box_{M,A} \bot).$$

6.5. $S_4$. 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 $\mathcal{E}$ enriched by designated arrows for finite extensions as indicated in the diagram below.

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

$$\quad U + C \xrightarrow{K} A + B$$

Here we require no commutation for $M$.

Theorem 6.6. Consider an FM-interpretation $K : A \triangleright U$. Then, $K$ is companionable iff $\Lambda^f_K$ extends $S_4$.

\footnote{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 to show that a principle has a counter-instance.}
Proof. Suppose $K$ is companionable. Consider any $B$ and suppose $K’ : (A + B) \vdash (U + C)$, where $K’$ is based on the same translation as $K$ and that $(U + C) \vdash (A + B)$. It follows that $U + C \vdash \lozenge_A B$. Hence, $A + B \vdash \lozenge_{K,A} B$.

Conversely, suppose $\Lambda_{K}^L$ contains the reflection principle, aka $M$. We have that $A + B \vdash \lozenge_{K,A} B$, so, there is an interpretation $K’$ based on the same translation as $K$, such that $K’ : (A + B) \vdash (U + \lozenge_{K,A} B)$. Conversely, $(U + \lozenge_{K,A} B) \vdash (A + B)$. $\blacksquare$

Remark 6.7. We note that the characterisation provided by Theorem 6.6 is rather different in nature from the one given of L"ob’s Principle in Theorem 6.3. First, in Theorem 6.6, we consider a property of interpretations rather than a property of frames as in Theorem 6.3. Secondly, we use more notions to formulate companion-ship than for irreconcilability. In Appendix C, we prove a result for the reflection principle that is more in the spirit of Theorem 6.6. 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. $\blacksquare$

Example 6.8. Let $A$ be e.g. the theory of the ordering of the natural numbers. The theory $A$ is finitely axiomatisable. Theorem 7.1 will tell us that $A$ is Friedman-reflexive. The identical interpretation $\text{Id}_A$ of $A$ in itself is clearly companionable. So, $\Lambda_{\text{Id}_A}^L$ extends $S_4$.

In fact, we can show that the modality trivialises for this example. Consider any $B$ in the $A$-language. The sentence $\lozenge_{\text{Id}_A,A} B$ is either provable or refutable in $A$. If it is refutable, we have $A \vdash \lozenge_{\text{Id}_A,A} B \rightarrow B$. Suppose it is provable. So, $A \vdash \lozenge_{\text{Id}_A,A} B$.

It follows that $A \vdash (A + B)$. So, $A + B$ is consistent, and, hence, $B$ is provable in $A$. Thus, $A \vdash \lozenge_{\text{Id}_A,A} B \rightarrow B$. So, in both cases, we have $A \vdash \lozenge_{\text{Id}_A,A} B \rightarrow B$. $\blacksquare$

Open Question 6.9. Can we find a more inspiring example of a theory with logic $S_4$ than Example 6.8? Is it perhaps possible to find an FM-interpretation with interpreter logic precisely $S_4$?

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 $f \circ g = \text{id}_b$. Here $g$ is called section, co-retraction, or split monomorphism.

Corollary 6.10. Suppose the FM-interpretation $U \xrightarrow{K} A$ is a retraction in $\mathbb{E}$. Then, $K$ is companionable and, hence, $\Lambda_{K}^L$ extends $S_4$.

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., $K \circ M = \text{Id}_A$. Consider any $B$ in the $A$-language. We have: $(U + B^M) \vdash (A + B)$. Moreover, writing $\equiv$ for having the same theorems, we have:

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

Here $K’ : (A + B) \vdash (U + B^M)$ has the same underlying translation as $K$. $\blacksquare$
6.6. Relations between Logics. The notion of sameness of theories that is relevant in the present paper is sentential congruence or $\mathcal{E}$-isomorphism (see Appendix [A] for more on these notions). 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 $V_0 \xrightarrow{M} V_1, V_1 \xrightarrow{\bar{M}} V_0, W_0 \xrightarrow{P} W_1, W_1 \xrightarrow{\bar{P}} W_0$, such that $M, \bar{M}$ and $P, \bar{P}$ are pairs of inverses in $\mathcal{E}$ and $K_1 \circ M = P \circ K_0$ in $\mathcal{E}$.

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

We have the following theorem.

**Theorem 6.11.** 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 $\Lambda_{K_0}^{fr} = \Lambda_{K_1}^{fr}$.

The theorem is one of these examples where the truth is immediately clear but it still requires some work to really prove it. We give the proof in Appendix [B].

We have a second preservation theorem.

**Theorem 6.12.** Let $F$ be an endofunctor $F$ of $\mathcal{D}$. 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 $\Gamma$ is a set of sentences in the $V$-language. Then, $F(V + \Gamma) = F(V) + \Gamma^{\eta_V}$.

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: $\Lambda_{\eta_K \circ K}^{fr} \subseteq \Lambda_K^{fr}$.

We give the proof in Appendix [B].

Let $\text{in}_{V_0,V_1}$ be the obvious interpretation of $V_i$ in $V_0 \otimes V_1$.

**Corollary 6.13.** Suppose $K : A \triangleright U$ is an FM-interpretation. Then, we have $\Lambda_{\eta_K \circ K}^{fr} \subseteq \Lambda_K^{fr}$.

We will see a further application of Theorem 6.12 in Subsection 10.4.

We end with two simple observations.

**Theorem 6.14.** Suppose $U$ is Friedman-reflexive and $K : U \triangleleft A$. Let $B$ be a sentence in the $A$-language. Then, $A \vdash K(A \triangleleft B \wedge C) \leftrightarrow K(A \vdash B \wedge C)$.

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

**Theorem 6.15.** Suppose $U$ is Friedman-reflexive and $K : U \triangleleft A$. Let $K^* : \text{Th}(K) \triangleleft A$ be the interpretation based on $\tau_K$ the translation given with $K$. Then, $\Lambda_K^{fr} = \Lambda_K^{fr}$.  

\[\text{This condition does not look very ‘categorical’, since it goes into the hardware. In Appendix [B] we discuss a more categorical formulation.}\]
We leave the proofs to the reader.

**Open Question 6.16.** Suppose $K, M : U \triangleright A$ are FM-interpretations and $\text{Th}(K) = \text{Th}(M)$. Do we have $\Lambda^r_K = \Lambda^r_M$, or is there a counter-example? $\square$

In the rest of this paper, we will have a closer look at interpreter logics over certain special classes of base theories.

## 7. Complete Theories

In this section, we discuss complete theories.

**Theorem 7.1.** 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 $\top$. $\square$

**Example 7.2.** Presburger Arithmetic is complete and decidable. So, it is Friedman-reflexive but not effectively so.

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

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 $\diamond$ must be non-computable.

We show that the interpreter logics for complete bases extend $K45$.

**Theorem 7.3.** Suppose $K : A \triangleright U$ is an FM-interpretation and $U$ is complete. Then, $A \vdash_{K,A} \top \rightarrow \Box_{K,A} \top $.

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

We note that, if $U \not\models A$, then $U \not\models \diamond A$ and, so, $U \not\models \neg \diamond A$. It follows that $A \vdash \Box_{K,A} \bot$. So, the interpreter logic for $A$ trivialises. Suppose, on the other hand, that $U \triangleright A$. Then, $A \vdash \diamond_{K,A} \top$. It follows that $U$ 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 [PZ20]. (So, the interpreter frame logic of Presburger Arithmetic trivialises.)

There are, of course, 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 7.4.** Is there an example of an FM-interpretation $K : A \triangleright U$, where $U$ is complete, with an interesting interpreter logic? $\square$
8. 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 8.1. Is there a consistent finitely axiomatised theory that is effectively Friedman-reflexive?

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 $\mathcal{S}4$, since $\text{Id}_A$ is clearly companionable.

9. ESSENTIALLY SENTENTIALLY REFLEXIVE THEORIES

In this section we study essential sentential reflexivity. A theory $U$ is essentially sententially reflexive if, for some $N$, we have $N : U \vdash S^1_2$ and, for all $U$-sentences $A$, $U \vdash \square^N_n A \rightarrow A$. Here $A$ ranges over $U$-sentences and $\square^N_n$ means provability in predicate logic using only involving 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 $A$.

We will write $\Delta A B$ for $\rho(A \rightarrow B), A B$ and $\nabla A B$ for $\rho(A \land B), A B$.

9.1. 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 9.1.** Suppose $U$ is essentially sententially reflexive. Then, there is an $N : U \vdash S^1_2$, such that, for all $\Sigma^0_1$-sentences $S$ and for all $M : U \vdash S^1_2$, we have $U \vdash S^N \rightarrow S^M$.

**Proof.** Suppose that $U$ is essentially sententially reflexive with witness $N_0 : U \vdash S^1_2$. We start with the observation that, for all $U$-sentences $A$ and for all $m \geq \rho(A)$, we have $U \vdash \square^{N_0}_m A \rightarrow A$. This follows immediately by replacing $A$ by $(A \land B)$, where $B$ is a tautology with $\rho(B) = m$. Let $N$ be a logarithmic cut of $N_0$. Consider any $M : U \vdash S^1_2$. We have, for a sufficiently large $m$ and for some $U$-theorem $D$, that $U \vdash S^N \rightarrow \square^{N_0}_{m,D} S^M$ (see Theorem [A.1]). Here we can take $D := (E_S^1 \land S_2^1)$. It follows that $U \vdash S^N \rightarrow S^M$. □

**Theorem 9.2.** Suppose $U$ is sequential. Then the following conditions are equivalent.

a. $U$ is essentially sententially reflexive.

b. There is an $N : U \vdash S^1_2$, such that for all $\Sigma^0_1$-sentences $S$ and all $M : U \vdash S^1_2$, we have $U \vdash S^N \rightarrow S^M$.

c. Consider any $N^* : U \vdash S^1_2$. There is an $N^*$-cut $I$, such that, for all $\Sigma^0_1$-sentences $S$ and all $N^*$-cuts $J$, we have $U \vdash S^I \rightarrow S^J$. □
Proof. Suppose \( U \) is sequential.

Theorem 9.1 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 \( U \vdash \Delta^I A \rightarrow A \). Since \( U \vdash \Delta^N A \rightarrow \Delta^I A \), we find \( U \vdash \Delta^N A \rightarrow A \). So \( N \) witnesses the essential sentential reflexivity of \( U \).

We prove the implication from (b) to (c). Let \( N \) witness (b). Consider any \( N^* : U \models S^I_2 \). 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 \( M : U \models S^I_2 \). Let \( J \) and \( J^* \) be \( U \)-definably, \( U \)-provably isomorphic cuts of \( M \) and \( I \). Reason in \( U \). Suppose \( S^J \). Then, \( S^{J^*} \). So, \( S^J \), and, hence, \( S^M \). \( \square \)

9.2. Essential Sentential Reflexiveness implies Friedman-reflexiveness. We have the following theorem.

Theorem 9.3. Suppose \( U \) is essentially sententially reflexive with witnessing interpretation \( N \). Then, \( \Phi A := \forall^N_A \top \) witnesses that \( U \) is effectively Friedman-reflexive.

The argument is, of course, just Harvey Friedman’s argument for the case of \( \text{PA} \).

Proof. Suppose \( N \) witnesses the sentential essential reflexiveness of \( U \).

Suppose \( (U + B) \models A \). Then, for some finite sub-theory \( D \) of \( U \), we have that \( (D + B) \models A \). If follows that \( U \vdash (D + B) \models A \) and, so, \( U \vdash \forall^N_B \top \). Then, by the Interpretation Existence Lemma, we find \( (U + B) \models A \). \( \square \)

9.3. Peano Corto. In this subsection, we discuss the sententially essentially reflexive theory \( \text{PA}^{\perp \perp} \). However, since we mainly want to illustrate the idea of a weak sententially essentially reflexive theory, it seemed good to zoom in on one.

In our paper [Vis14a], we used \( \text{PA}^{\perp} \) as starting point. Here, we use \( S^I_2 \) as starting point since it fits the set-up of the present paper better.

We introduce \( \text{PA}^{\perp \perp} \) and its little brother \( \text{PA}^{\perp \perp \perp} \) and its big brother \( \text{PA}^{\perp} \).

- Peanissimo or \( \text{PA}^{\perp \perp} \) is the theory
  
  \( S^I_2 + \{(S \to S^I) | S \text{ is an } \Sigma^0_1 \text{-sentence and } I \text{ is an } S^I_2 \text{-cut} \}. \)

  This theory is identical to \( S^I_2 + \{\Delta A \to A | A \text{ is an arithmetical sentence} \} \).

- Peano Corto or \( \text{PA}^{\perp \perp} \) is the theory
  
  \( S^I_2 + \{(S \to S^I) | S \text{ is a } \Sigma^0_1 \text{-sentence and } I \text{ is an } S^I_2 \text{-cut} \}. \)

  Here \( \Sigma^0_1 \) means a block of existential quantifiers followed by a \( \Delta_0 \)-formula. In fact, Peano Corto is Peanissimo plus the scheme

  \[ S \to \exists x \exists z (2^x = z \land S_0(x)), \]

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

$$S_2^1 + \{(S \rightarrow S') \mid S \text{ is a } \Sigma_{1,\infty}^0 \text{-sentence and } I \text{ is an } S_2^1 \text{-cut}\}.$$  

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 $\Delta_0^0$-formula. In [Vis14a], it is shown that Peano Basso is Peano Corto plus $\Sigma_1^0$-collection.

In [Vis14a], 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.

All three theories are locally cut-interpretable in $S_1^{2\infty}$, i.o.w., $S_2^1 \nvdash_{\text{cut,loc}} \text{PA}^\downarrow$ and $S_2^1 \nvdash_{\text{cut,loc}} \text{PA}^\uparrow$ and $S_2^1 \nvdash_{\text{cut,loc}} \text{PA}^\uparrow$. 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 $E\langle \text{cut}_x(E) \rangle (x = x)$, where $E$ ranges over formulas with at most the free variable $x$. Here $\text{cut}_x(E)$ is the $S_1^{2\infty}$-sentence that expresses ‘$\{x \mid E(x)\}$ is a cut’ and $F\langle G \rangle H$ is $((G \rightarrow F) \land (\neg G \rightarrow H))$.

Since, the identical cut is the designated cut, we can, by Theorem 9.3, take $A := \nabla_A \top$ in each of our theories.

**Theorem 9.4.** Suppose $P$ is a $\Pi_1^0$ interpreter of $A$ over Peano Corto. Then, $\text{EA} \vdash P \Leftrightarrow \nabla_A \top$.

**Proof.** Suppose $P$ is a $\Pi_1^0$ interpreter of $A$ over $\text{PA}^\uparrow$. Then, by uniqueness, we have $\text{PA}^\uparrow \vdash P \leftrightarrow \nabla_A \top$. So, for some $S_2^1$-cut $J$, we have $S_2^1 \vdash (P \leftrightarrow \nabla_A \top)^J$. Hence, by a meta-theorem of Paris and Wilkie, we have $\text{EA} \vdash P \leftrightarrow \nabla_A \top$. See [WPS7] and [Vis92].  

So, we have characterised $\nabla_A \top$ as a $\Pi_1^0$-sentence up to $\text{EA}$-provable equivalence in a coordinate-free way. This improves the result of [Vis11], where this was only done for finitely axiomatised sequential theories.

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

**Theorem 9.5.** Suppose $A$ is sequential. Then,

$$A \nvdash B \iff \text{PA}^\downarrow \vdash \nabla_A \top \rightarrow \nabla_B \top.$$  

**Proof.** The left-to-right direction is just Theorem 9.4 in combination with the fact that we can take $A := \nabla_A \top$ over Peano Corto.

From right to left: suppose $\text{PA}^\downarrow \vdash \nabla_A \top \rightarrow \nabla_B \top$. Then, $S_2^1 \vdash (\nabla_A \top \rightarrow \nabla_B \top)^J$, for some cut $I$. So, $S_2^1 \vdash \nabla_A \top \rightarrow \nabla_B ^I \top$. We have:

$$A \nvdash (S_2^1 + \nabla_A \top) \nvdash (S_2^1 + \nabla_B \top) \nvdash B.$$  

The first step uses the sequentiality of $A$.  

**Remark 9.6.** The original Friedman characterisation had $\text{EA}$ in place of Peano Corto. A result due to Paris and Wilkie (see [WPS7] and [Vis92]) shows that we have, for $\Pi_1^0$-sentences $P$ and $Q$:

$$\text{PA}^\downarrow \vdash P \rightarrow Q \iff \text{EA} \vdash P \rightarrow Q.$$
So, the connection between the two results is obvious. However, the internal version of the characterisation in \( \text{EA} \) needs cut-free \( \text{EA} \)-provability and ordinary \( \text{PA}^{\uparrow \downarrow} \)-provability.

**Remark 9.7.** The theory \( \text{Seq}(V) \) is specified as follows. We add a unary predicate \( \mathcal{D} \) and a binary predicate \( \in \) to the signature of \( V \), we relativise \( V \) to \( \mathcal{D} \) and we add the (unrelativised) axioms for Adjunctive Set Theory \( \text{AS} \) plus an axiom that states that every element of \( \mathcal{D} \) is an empty set. We can show that \( \text{Seq} \) supports a functor from \( \mathcal{D} \) to \( \mathcal{D}_{\text{seq}} \) and from \( \mathcal{D}_{\text{fin}} \) to \( \mathcal{D}_{\text{fin}, \text{seq}} \). See Appendix A for details. It is easily seen that we have \( \text{Seq}(A) \models (S^2 + \nabla A \uparrow) \).

We see that we can split the functor \( H \) of Remark 3.6 in two stages. First, we have a projection \( \pi \) of \( \mathcal{D}_{\text{fin}} \) to \( \mathcal{D}_{\text{fin}, \text{seq}} \). This can be either \( A \mapsto \text{Seq}(A) \) or \( A \mapsto (S^2 + \nabla A \uparrow) \). Then, we have a (lax) embedding of \( \mathcal{D}_{\text{fin}, \text{seq}} \) into \( \mathcal{B}_{\text{PA}^{\uparrow \downarrow}} \).

Suppose \( K : A \vdash \text{PA}^{\uparrow \downarrow} \). Since, we can apply the Gödel Fixed Point Theorem in the usual way because \( \phi \) can be represented by a predicate, we have Löb’s Logic. This also follows from Corollary 10.12 in combination with the fact that \( \text{PA}^{\uparrow \downarrow} \) is sequential. That theorem, however, has a much more involved proof.

If \( K \) is \( \Sigma_1 \)-sound, \( \mathcal{A}^L_K \) is precisely Löb’s Logic. In the case of \( \text{PA}^{\uparrow \downarrow} \) we can verify Solovay’s Theorem simply using Solovay’s proof. The reason is that \( \text{PA}^{\uparrow \downarrow} \) proves that \( \exists x, y (2^x = y \land S_0(x)) \) from \( \exists x S_0(x) \), where \( S_0 \) is \( \Delta_0 \) or \( \Delta_0(\omega_1) \). This delivers \( \Sigma_1 \)-completeness. This argument is not present for \( \text{PA}^{\uparrow \downarrow} \). However, we still have Solovay’s Theorem for \( \text{PA}^{\uparrow \downarrow} \) as a special case of Theorem 10.16.

We can see that Peano Corto has some definite advantages over \( S^2 \) 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 9.4 and 9.5 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 \( S^2 \). If it were it would be mutually interpretable with \( S^2 \) and this contradicts Theorem 10.12 that we will prove later. In fact, no sequential Friedman-reflexive theory is interpretable in \( S^2 \) by the same argument. As a consequence, there are no interpreter logics for \( S^2 \) with Peano Corto as base (or, with any sequential Friedman-reflexive base).
III. Even if Peano Corto is interpretable in some reasonably weak concrete \( A \), like \( \text{EA} \), it is not always clear that we can find an interpretation that does not involve arithmetisation. We discuss this kind of problem in Section 11.

10. **Friedman-reflexivity meets Sequentiality**

We already met some specific sequential 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.

10.1. **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 \( \phi A \) always has the form of a restricted consistency statement of \( A \) on some cut.
Theorem 10.1. Suppose $U$ is sequential and Friedman-reflexive. Let $N : S^1_2 ∼ U$. Then, for some $N$-cut $I$, we have $U ⊩ ° A ↔ \nabla^I_A \top$.

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

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

Conversely, since, $(U + \nabla^I_A \top) ∼ A$, it follows that $U + \nabla^I_A \top ⊢ ° A$.

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

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

Theorem 10.2. Suppose $U$ is sequential.

A. The following are equivalent.

a. $U$ is Friedman-reflexive.

b. For all $\Sigma^0_1$-sentences $S$, there is an $N : U ∼ S^1_2$, such that, for all $M : U ∼ S^1_2$, we have $U ⊢ △^N S \rightarrow △^M S$.

c. Consider any $N^* : S^1_2 ∼ U$. Then, for all $\Sigma^0_1$-sentences $S$, there is an $N^*$-cut $I$ such that for all $N^*$-cuts $J$, we have $U ⊢ S^I \rightarrow S^J$.

B. The following are equivalent.

a. $U$ is effectively Friedman-reflexive.

b. There is a recursive function $F$ such that, for all $\Sigma^0_1$-sentences $S$, we have $F(S) = N : U ∼ S^1_2$ and, for all $M : U ∼ S^1_2$, we have $U ⊢ △^N S \rightarrow △^M S$.

c. Consider any $N^* : S^1_2 ∼ U$. There is a recursive function $G$ such that, for all $\Sigma^0_1$-sentences $S$, we have $G(S) = I$, where $I$ is an $N^*$-cut such that for all $N^*$-cuts $J$, we have $U ⊢ S^I \rightarrow S^J$.

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 $N^* : U ∼ S^1_2$. Consider any $\Sigma^0_1$-sentence $S$. We note that $A := S^1_2 + \neg S$ is finitely axiomatised. Let $I_0$ be the $N$-cut such that $\nabla^{I_0}_{S^1_2} \neg S$ is $U$-provably equivalent to $° A$. We find $U ⊢ \nabla^{I_0}_{S^1_2} \neg S \rightarrow \nabla^{I_0}_{S^1_2} \neg S$, for all $N$-cuts $J$. Thus, $U ⊢ \nabla^{I_0}_{S^1_2} S \rightarrow \nabla^J_{S^1_2} S$, for all $N$-cuts $J$.

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

\[
U ⊢ S^I \quad \rightarrow \quad \nabla^{I_0}_{S^1_2} S \\
\qquad \rightarrow \quad \nabla^{I_0}_{S^1_2} S \\
\qquad \rightarrow \quad S^I
\]

Conversely, suppose $U$ satisfies (c). Let $I$ be the $N^*$-cut guaranteed by (c) for $S := \nabla_A \bot$. Suppose $(U + B) ∼ A$. Then, for some $N$-cut $J$, we have $U + B ⊢ \nabla^I_A \bot$.
and hence $U + B \vdash \forall^1_A \top$. The other direction is immediate by Interpretation Existence. 

**Remark 10.3.** We note that a complete and consistent sequential theory will automatically have property (c) of Theorem 10.2. Of course, it should, by Theorems 7.1 and 10.2. 

**Remark 10.4.** Our result is rather robust for the precise notion of $\Sigma^0_1$ used. The result works both for smaller classes and for larger ones. 

It works for $\Sigma^0_1$ 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 $\omega_1$-terms in our definition of $\Sigma^0_1$. Finally, it works when we define our class $X$ as follows:

- $X := \top \iff t = u \iff \neg X \iff (X \land X) \iff (X \lor X) \iff (Y \to X) \iff \forall X < t X \iff \exists x X $

- $Y := \top \iff t = u \iff \neg X \iff (Y \land Y) \iff (Y \lor Y) \iff (X \to Y) \iff \exists x < t Y \iff \forall x Y$

We give a slightly modified version of our characterisation. 

**Theorem 10.5.** Suppose $U$ is sequential and let $N : U \models S^I_1$. Then, $U$ is Friedman-reflexive if and only if for all $S^I_1$-sentences $S$, there is an $U$-sentence $A$, such that, for all $U$-sentences $B$, we have $U + \{S^I \mid I \text{ is an } N \text{-cut}\} \vdash B$ if and only if $U + A \vdash B$.

**Proof.** Suppose $U$ is Friedman-reflexive. Let $I^*$ be the cut guaranteed for $S$ by Theorem 10.2(c). Then, it is easy to see that $S^{I^*}$ can be chosen as our $A$ to satisfy (1).

Conversely, suppose (1). 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 $U \vdash A \iff S^I$. We take $J$ as witness for satisfaction of the characterisation of Theorem 10.2(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 a sequential model $\mathcal{M}$. We define $\mathfrak{M}$ as follows. First, we choose an internal model $\mathcal{N}$ of $S^I_2$ and then we take $\mathfrak{M}$ to be the intersection of all $\mathcal{M}$-definable $N$-cuts. Using elementary facts about sequentiality, one can easily show that $\mathfrak{M}$ 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 the intersection. See also [Vis19, Section 5.1].

Consider a sequential theory $U$ and let $N : S^I_2 \models U$. We extend the language of $U$ with a new unary predicate $\exists$ that is interpreted in each $U$-model $\mathcal{M}$ as $\mathfrak{M}$. Here we think of $\mathfrak{M}$ as given by $\mathcal{N} := N^{\mathcal{M}}$. Let $U^*$ be the set of all sentences in the extended language true in all $\mathcal{M}, \mathfrak{M}$, where $\mathcal{M}$ is an $U$-model. Let $\mathcal{U}$ be the set of arithmetical sentences $A$ such that $U^* \vdash A^3$.

An important insight is that $\mathcal{U}$ contains $\exists A + B\Sigma_1$. See [Vis19, Section 5.1].

**Theorem 10.6.** Suppose $U$ is sequential and let $N : U \models S^I_2$. Then, $U$ is Friedman-reflexive iff, for all $S^I_1$-sentences $S$, there is an $U$-sentence $B$, such that we have $U^* \vdash S^I \iff B$. 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 10.1 gives us the $B$ we are looking for.

Suppose $U^+ \vdash S^3 \leftrightarrow B$. Then, $U + \{S^I \mid I$ is an $N$-cut$\} \vdash B$ and, conversely, $U + B \vdash S^I$, for all $N$-cuts $I$. By Theorem 10.5, it follows that $U$ is Friedman-reflexive. Clearly, if we can find the $B$ effectively, then $U$ is effectively Friedman-reflexive. 

So if we view the $S^3$ as a second-order or as an infinitary statement, then Friedman-reflexiveness means a reduction to first-order or finitary statements.

### 10.2. 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 $S_0,S_1,\ldots$ enumerate the $\Sigma^0_1$-sentences and let $I_0, I_1, \ldots$ be an effective enumeration of $S^1_2$-cuts such that $S^1_2 \vdash I_{n+1} \subseteq I_n$ and such that, for each $S^1_2$-cut $J$, we can find a $k$ such that $S^1_2 \vdash I_k \subseteq J$. 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 $\rho$-complexity below a given number, we can take $I_n$ to be the intersection of all definable cuts with $\rho$-complexity $\leq n$.

Let DA be $S^1_2 + \{S^1_i \rightarrow S^1_j \mid i \in \omega$ and $J$ is a definable cut$\}$. Clearly, DA is effectively Friedman-reflexive. We note that DA is a sub-theory of $\mathsf{PA}^{\ddagger \ddagger}$ 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 $A \mapsto \Box A$, where $\rho(\Box A) \leq n$, for all $A$. It is easy to see that in case $V$ is a sequential restrictedly Friedman-reflexive theory, then, for any $N : V \upharpoonright S^1_2$, there is a formula $C(x)$ such that, for all $A$, we have $V \vdash \Box A \leftrightarrow C(A)$, 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 10.7.** 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 $\rho(S^1_2) = k_1$. Suppose $I_p$ is the first logarithmic cut in the sequence. And let $k_2$ be the maximum of the $\rho$-complexities of the $S_i$ for $i < p$. Finally, let $k_3$ be the complexity of a standard $\Sigma^0_1$-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 $\Sigma^0_1$-sound for every consistent extension of $S^1_2$ with complexity $\leq k$. The existence of such a cut is guaranteed by Theorem 10.2. We may assume that $n > p$. Let

$$A := S^1_2 + \{ \neg S_i \mid i < p \text{ and } S_i \text{ is false}$$. 

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


We claim that \( A + \neg C(\forall B^\forall) \) is consistent. Suppose it were not. Then, we would have \( A \vdash C(\forall B^\forall) \). It follows that \( DA + A \vdash \square B \), and, hence, \((DA + A) \triangleright B\). Since, \( A \) locally cut-interprets \( PA \updownarrow A \) and, hence, \( DA + A \), we find that \( A \triangleright B \). *Quod non*, by the usual no-interpretation version of G2.

By the special property of \( I_n \), it follows that

\[
V := A + \neg C(\forall B^\forall) + \{ \neg S^i_n \mid i \geq n \text{ and } S_i \text{ is false} \}
\]

is consistent. By Theorem \( \text{A.1} \) we have \( S^1_2 \vdash \neg \text{true}(S) \to \neg S^i_n \), for any \( \Sigma^1 \) sentence \( S \). It follows that \( V \) extends \( DA \). Hence, \( V \) is Friedman-reflexive with \( \square \) as selection function and \( V \vdash \neg \square B \). Since \( V \) proves every true \( \Pi^1 \) sentence, including \( \square_B \top \), on \( I_n \), we find \( V \triangleright B \) and, hence \( V \vdash \square B \). A contradiction.

\[\Box\]

**Open Question 10.8.** 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 \( K : A \triangleright DA \), such that, for no \( D(x) \) in the \( A \) language, we have, for all \( B \) in the \( A \)-language, \( A \vdash D(\forall B^\forall) \leftrightarrow \square_{K.A} B \)?

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 \( \text{10.7} \) uses special features of \( DA \). So, the proof does not generalise in an obvious way, to a proof of a positive answer to (iv). \( \bigcirc \)

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

**Theorem 10.9.** Suppose \( U \) is consistent, Friedman-reflexive, sequential and RE. Then, any axiomatisation of \( U \) must have axioms of \( \rho \)-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 \( N : U \triangleright S^1_2 \). Let \( A \) be any consistent finitely axiomatised theory such that \( U \nvdash A \). We find that \( U \nvdash \square A \). So, \( U + \neg \square A \) is consistent.

We have, for all \( N \)-cuts \( I \), that \( (U + \square_A^I \top) \triangleright A \). So, \( U + \square_A^I \top \vdash \square A \). It follows that \( U + \neg \square A \vdash \square_A \bot \). So, by Theorem \( \text{A.2} \) we find that \( \square_A \bot \) is true and, thus, that \( A \) is inconsistent. *Quod non*. \( \Box \)

So, e.g., neither \( \text{PRA} \), nor \( \Pi^1_n \), nor \( \text{ACA}_0 \) is Friedman-reflexive.

**Remark 10.10.** We note that we could, alternatively, have framed the proof of Theorem \( \text{10.9} \) as follows. Theorem \( \text{A.2} \) tells us that each consistent finite extension of \( U \) tolerates \( S^1_2 \) plus all true \( \Pi^1 \) sentences and, hence, is polyglottic. On the other hand, clearly, there is an \( A \) such that \( U \nvdash A \) and so \( U + \neg \square A \) is consistent. By Theorem \( \text{5.2} \) this last theory is not polyglottic. A contradiction. \( \bigcirc \)

**Example 10.11.** Since, Peano Corto is both reflexive and mutually locally interpretable with \( S^1_2 \), we find that \( PA \updownarrow \bowtie U(S^1_2) \). Also, \( U(S^1_2) \) is restricted and, hence, not Friedman-reflexive. Ergo, Friedman-reflexivity is not preserved by mutual interpretability. \( \bigcirc \)
We have the following insight.

**Theorem 10.12.** 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 the property demanded for $A$ and the property demanded $U$ 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 $K : U \vdash A$ and $M : A \vdash U$. Consider any finitely axiomatised $B$ such that $A \nvdash B$. For example, we could take $B := (S_2^1 + \Diamond A \top)$.

We have (a) $A \vdash (A + \neg M) B$ is consistent, since otherwise $A \vdash (A + \neg M) B \vdash (U + B) \nvdash B$.

Quod non. We have:

$$(U + \Diamond^M B) \vdash (A + \Diamond^M B) \vdash (U + B) \vdash B.$$

So, $U + \Diamond^M B \vdash B$. Ergo, $A + \Diamond^M B \vdash \Diamond^M B$, and hence, $A + \neg \Diamond^M B \vdash \neg \Diamond^M B$. It follows that (b) $(M \circ K) : (A + \neg \Diamond^M B) \vdash (A + \neg \Diamond^M B)$, where $(M \circ K)$ is the interpretation based on the same translation as $M \circ K$.

Suppose $N : A \vdash S_1^2$. Let $I$ be any $N$-cut in $A$. We have

$$(U + \Diamond^I B \top) \vdash (A + \Diamond^I B \top) \vdash (S_2^1 + \Diamond B \top) \vdash B.$$

So $U + \Diamond^I B \top \vdash B$. It follows that $U + \neg B \vdash \Box^I B \bot$, and, hence, we have $A + \neg \Diamond^M B \vdash \Box^I B \bot$.

We have shown: (c) for all $N$-cuts $I$, we have $A + \neg \Diamond^M B \vdash \Box^I B \bot$. 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 $A \gg U$. Then, we can find a finitely axiomatised sequential $U_0 \subseteq U$, such that $A \gg U_0$ and, hence $U_0 \gg U$, contradicting the first case.

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

We note Theorem 10.12 implies that there is no Friedman-reflexive and sequential $U$ such that $S_2^1 \gg U$. The reason is that any sequential theory interprets $S_2^1$. Of course, $S_2^1$ does interpret the theory of the ordering of the natural numbers. However, that only gives the trivial interpreter logic that proves $\bot$. So, one might wonder whether there is a more interesting base for $S_2^1$. We will discuss some hopeful signs that there is in Section 11.

10.4. **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.
10.4.1. Soundness. Combining Theorems 6.3 and 10.12 we find:

**Theorem 10.13.** Let \( \langle A, U \rangle \) be an FM-frame. Suppose that one of \( A, U \) is sequential. Then, \( \Lambda_{U}^{A} \) 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., \( A \vdash \Box \top \leftrightarrow \neg \Box \top \). The Gödel Fixed Point Lemma plays a role in our proof of this fact, but it is still not a direct application. We use the Fixed Point Lemma in a Rosser-style argument in the proof of Theorem A.3. Can we prove it more directly? We do not 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 10.14.** Suppose \( U \) is sequential and effectively Friedman-reflexive. Let \( N : U \triangleright S_{1}^{2} \). We define \( \mathfrak{I} \) w.r.t. \( N \). Suppose \( A(x) \) is a boolean combination of \( \Sigma_{1}^{0} \)-formulas with just \( x \) free. Then, there is a \( B \) in the \( U \)-language, such that \( U^{e} \vdash B \leftrightarrow A^{3}(\Box B ) \).

**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 \( \Sigma_{1}^{0} \)-sentences \( S \) to \( S' \), where \( S' \) is equivalent over \( U^{e} \) to \( S^{2} \). We can lift this function to Boolean combinations of \( \Sigma_{1}^{0} \)-sentences. Let’s say the result is \( G \).

Suppose \( A(x) \) is a boolean combination of \( \Sigma_{1}^{0} \)-formulas with just \( x \) free. We write \( A(G(x)) \) for the result of replacing each \( \Sigma_{1}^{0} \)-component \( Sx \) of the Boolean combination by \( \exists y, z, u (G_{0}xyz \land S_{0}yu) \), where \( G_{0}xyz \) is a \( \Delta_{1}^{0} \)-formula such that \( \exists z G_{0}xyz \) represents the graph of \( G \) and \( S_{0}yu \) is a \( \Delta_{0}^{0} \)-formula such that \( Sx \) is (equivalent to) \( \exists u S_{0}yu \).

We can find a \( C \) such that \( E \mathfrak{A} \vdash C \leftrightarrow A(G(\bar{C}t)) \), 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 we as we did for the function \( G \), so that \( C \) is again a boolean combination of \( \Sigma_{1}^{0} \)-sentences. Let \( B := G(C) \). Then,

\[
U^{e} \vdash (C \leftrightarrow A(G(\bar{C}t)))^{3} \\
\vdash C^{3} \leftrightarrow (A(G(\bar{C}t)))^{3} \\
\vdash B \leftrightarrow (A(\bar{C}B))^{3}
\]

**Theorem 10.15** (Alternative Proof for Löb’s Principle in the effective Case). Suppose \( U \) is sequential and effectively Friedman-reflexive. Let \( K : A \triangleright U \) be an FM-interpretation. Then, \( \Lambda_{U}^{F} \) proves Löb’s Principle.

**Proof.** Suppose \( U \) is sequential and effectively Friedman-reflexive. Let \( K : A \triangleright U \) be an FM-interpretation. Consider any \( B \) in the \( A \)-language. By Theorem 10.14 we can find a \( D \) such that \( U^{e} \vdash D \leftrightarrow \Delta_{4}^{0}(D^{K} \rightarrow B) \). Thus, \( U \vdash D \leftrightarrow \Box_{4}(D^{K} \rightarrow B) \). Setting \( E := D^{K} \), we find \( A \vdash E \leftrightarrow \Box_{4}(E \rightarrow B) \). So, we have one variant of the Löb Fixed Point in \( A \). Since we have \( K4 \), Löb’s Principle follows.
10.4.2. **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 $\alpha$ range over $0,1,\ldots,\infty$. We define $\mathbf{1} := \bot$, $\mathbf{1}^{k+1} := \mathbf{1} \mathbf{1}$, and $\mathbf{1}^\infty := \top$ and, similarly, for $\Box$. Suppose $K : A \vdash U$ is an FM-interpretation. Let $d(K)$ be the smallest $\alpha$ such that $A \vdash \mathbf{1}_A^\alpha \bot$.

**Theorem 10.16** (Solovay’s Theorem for sequential effectively Friedman-reflexive bases). Suppose that $K : A \vdash U$ is an FM-interpretation and that $U$ is effectively Friedman-reflexive and sequential. Then, $\Lambda^*_K = \text{GL} + \Box d(K) \bot$.

We will prove Solovay’s Theorem by verifying the conditions for it given in [dJJM91]. See also [Vis15].

The idea of de Jongh, Jumelet and Montagna is that Solovay’s embedding result can be verified in an extension of the modal logic $\text{GL}$ enriched with certain fixed points. We introduce the logic $\text{R}^-$ of Guaspari and Solovay. See [GS79]. The language of $\text{R}^-$ is given by:

- $\alpha ::= p_0 | p_1 | \ldots$
- $\varphi ::= \alpha | \bot | \top | \neg \varphi | \Box \varphi | (\varphi \land \varphi) | (\varphi \lor \varphi) | (\varphi \rightarrow \varphi) | \Box \varphi < \Box \varphi | \Box \varphi \leq \Box \varphi$

The logic $\text{R}^-$ is axiomatised by the axioms and rules of GL (for the extended language) plus the following axioms.

- **R1.** $\vdash \Box \varphi \leq \Box \psi \rightarrow \Box \varphi$
- **R2.** $\vdash (\Box \varphi \leq \Box \psi \land \Box \psi \leq \Box \chi) \rightarrow \Box \varphi \leq \Box \chi$
- **R3.** $\vdash \Box \varphi < \Box \psi \leftrightarrow (\Box \varphi \leq \Box \psi \land \Box \psi \leq \Box \varphi)$
- **R4.** $\vdash \Box \varphi \rightarrow (\Box \varphi \leq \Box \psi \lor \Box \psi \leq \Box \varphi)$
- **R5.** $\vdash \Box \varphi \leq \Box \psi \rightarrow (\Box \varphi \leq \Box \psi)$
- **R6.** $\vdash \Box \varphi < \Box \psi \rightarrow (\Box \varphi < \Box \psi)$

Now consider a finite Kripke model $K$ of $\text{GL}$ 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 $\ell_i$, for $i < n$, to the language of $\text{R}^-$ 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.** $\vdash \ell_j \leftrightarrow (\Box \neg \ell_i \land \bigwedge_{j > i} \ell_j \land \bigwedge_{j \parallel i} \bigvee_{k \leq i, k \parallel j} \Box \neg \ell_k < \Box \neg \ell_j)$
- **fp2.** For $i \neq j$, we have $\vdash \Box \neg \ell_i \leq \Box \neg \ell_j \rightarrow \Box \neg \ell_i < \Box \neg \ell_j$.

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

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

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

- We define $\Box_A B < \Box_A C$ by $(\Delta_A B < \Delta_A C)^I$, where $I$ is a cut such that $U^* \vdash (\Delta_A B < \Delta_A C)^I \leftrightarrow (\Delta_A B < \Delta_A C)^3$.
- We define $\Box_{K,A} B < \Box_{K,A} C$ by $(\Box_A B < \Box_A C)^K$.
- We define $\Box_A B \leq \Box_A C$ and $\Box_{K,A} B \leq \Box_{K,A} C$ similarly.
We extend the notion of translation of the modal language to the richer vocabulary in the obvious way.

**Theorem 10.17.** Suppose $K : A \triangleright U$ is an FM-interpretation, where $U$ is sequential. Let $N : U \triangleright S_2^1$. Then, we have $R^-$ for all $K,A$-translations of the modal language.

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

$$\text{DA}^e \vdash (\Delta_A B \leq \Delta_A C \rightarrow \Delta_A B)^3.$$ 

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

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

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

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

$$U^e \vdash (\Delta_A B < \Delta_A C)^I \leftrightarrow (\Delta_A B < \Delta_A C)^I.$$ 

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

$$U^e \vdash (\Delta_A B \leq \Delta_A C \rightarrow \Delta_A(\Delta_A B \leq \Delta_A C)^IK)^3.$$ 

And, thus,

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

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

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

To provide the $\ell_i$ we need an extension of Theorem 10.14.

**Theorem 10.18.** Suppose $U$ is sequential and effectively Friedman-reflexive. Let $N : U \triangleright S_2^1$. We define $\ell$ w.r.t. $N$. Suppose we have formulas $A_0(x_0, \ldots, x_{k-1})$, for $i < k$, where each $A_i$ is a boolean combination of $\Sigma^0_1$-formulas with just $x_0, \ldots, x_{k-1}$ free. Then, for $i < k$, there are $B_i$ in the $U$-language, such that

$$A \vdash B_i \leftrightarrow A_0^3(\Box B_0^3, \ldots, \Box B_{k-1}^3),$$

for each $i < k$.

The proof of the Theorem is the similar to the proof of Theorem 10.14 using the Multiple Fixed Point Lemma as it is available in $\text{EA}$.

In the proof of Theorem 10.19 we need the assumptions that we work with a single conclusion system and that the fixed point construction is a usual construction.

**Theorem 10.19.** Suppose $K : A \triangleright U$ is an FM-interpretation, where $U$ is sequential. Suppose $U$ is effectively Friedman-reflexive. Let $N : U \triangleright S_2^1$. Let $K$ be a Kripke model for $\text{GL}$ 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. $A \vdash L_i \leftrightarrow (\Box_{K,A} \neg L_i \land \bigwedge_{j<i} (\Box_{K,A} L_j \land \bigwedge_{i<j} (\bigvee_{k \leq i, k \| i} \Box_{K,A} \neg L_k < \Box_{K,A} \neg L_j))$.

b. For $i \neq j$, we have $A \vdash \Box_{K,A} \neg L_i \leq \Box_{K,A} \neg L_j \rightarrow \Box_{K,A} \neg L_i < \Box_{K,A} \neg L_j$. 

Proof. Using Theorem 10.18 we can find sentences $\lambda_i$ such that:

$$U^e \vdash \lambda_i \leftrightarrow (i = i \land A \neg \lambda_i^K \land \bigwedge_{j > i} (A \lambda_j^K \land \bigwedge_{k \leq i, k || j} (A \neg \lambda_k^K < A \neg \lambda_j^K)).$$

Hence,

$$U \vdash \lambda_i \leftrightarrow (A \neg \lambda_i^K \land \bigwedge_{j > i} (A \lambda_j^K \land \bigwedge_{k \leq i, k || j} (A \neg \lambda_k^K < A \neg \lambda_j^K)).$$

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

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

$$EA \vdash A \neg L_i \leq A \neg L_j \rightarrow A \neg L_i < A \neg L_j.$$

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 $\lambda_i$ are pairwise distinct. This is because the first conjunct of $\Lambda_i$ has the form $(i = i)^I$, for some $N$-cut $I$. $\square$

Solovay’s Theorem now follows from the results of [dJJM91]. We have the following corollary.

**Corollary 10.20.** Suppose $K : A \triangleright U$ is a faithful FM-interpretation and $A$ is consistent and $U$ is effectively Friedman-reflexive, sequential, and polyglottic. Then, $\Lambda_{fr}^K = GL$.

*Proof.* We assume the conditions of the corollary. Clearly $A + \Box_K A \top$ is consistent. Suppose $A + \Box_K A \top$ is consistent. So, by polyglotticity, $U + \Box_A \Box_K A \top$ is consistent. But, then, by faithfulness, $A + \Box_{K,A} A \top$ is consistent. It follows that $d(K) = \infty$. $\square$

**Corollary 10.21.** Suppose $(A,U)$ is an FM-frame, $A$ is consistent and sequential, and $U$ is RE, effectively Friedman-reflexive, sequential, and polyglottic. Then, $\Lambda_{fr}^{A,U} = GL$.

*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 [Smo85a] or [Vis05]. $\square$

10.5. **Relation between Logics.** We consider the functor $\text{Seq}$ of Remark 9.7. Let $\gamma_V$ be the interpretation based on relativisation to $D$.

**Theorem 10.22.** Suppose $U$ is sequential and Friedman-reflexive and $K : A \triangleright U$. Then, $\Lambda_{fr}^{A\circ K} \subseteq \Lambda_{fr}^K$.

*Proof.* The theorem is a direct application of Theorem 6.12 with $j_A$ in the role of $\eta_A$. $\square$

11. **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.
11.1. **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 7 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 $S^1_2$, $E_A$, and $I\Sigma_1$\footnote{Similarly, the finite axiomatisation of predicative comprehension over a pair theory does involve various implementation details.}. 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 $S^1_2 + \Diamond S^1_2 \top$. However, neither $S^1_2 + \Diamond S^1_2 \top$ itself nor the Henkin interpretation that we employ is coordinate-free.
- Peano Corto is interpretable in $E_A$, since $E_A$ interprets $S^1_2 + \Diamond S^1_2 \top$ on a super-logarithmic cut. Here $E_A$ is coordinate-free (ignoring to the worry articulated above) but the only interpretation that we know of is not.
- By Vaught’s Theorem (see \cite{Vau67, Vis12b}), 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 $E_A$. The interpretation of $E_A$ in $W$ can be taken to be coordinate-free: we have $E_A$ 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 $E_A$ and Peano Corto is $E_A \not\models (PA^{++} + \Diamond E_A)$, which is perfectly fine, and, similarly, for the fact that the frame-logic of $\langle E_A, PA^{++} \rangle$ is $GL$. 

Remark 11.1. We consider the following theory IIA (Initial Isomorphism Arithmetic). We start with \( S_1^2 \). We expand the language with second-order variables and we expand the theory with predicative comprehension obtaining \( PC(S_1^2) \). Now we expand the language with a new unary predicate \( J \) and a binary predicate \( F \). We add an axiom \( \forall I (\text{cut}(I) \rightarrow J \subseteq I) \) and an axiom stating that \( F \) is an isomorphism between \( ID \) and \( J \).

By Theorem 5.9 of [Vis14a], 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.

Since \( PC(S_1^2) \) is mutually interpretable with \( S_1^2 + S_1^2 \top \) and \( S_1^2 + S_1^2 \top \) is mutually interpretable with \( EA \), we find that IIA interprets \( EA \).

Open Question 11.2. 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 \( \Sigma_0^1 \)-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 \( I\Sigma_n \)?

iii. Do we have \( EA \vdash IIA \)?

11.2. 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) \( EA \not\vdash (S_1^2 + \Diamond_{EA} \top) \). However, we do have \( EA \vdash (S_1^2 + \forall_{EA} \top) \). Our version of G2 with base \( PA^{\downarrow \downarrow} \) is (b) \( EA \not\vdash (PA^{\downarrow \downarrow} + \Diamond_{EA} \top) = (PA^{\downarrow \downarrow} + \forall_{EA} \top) \). We see that in (a) the base theory is weaker and the consistency statement stronger. In (b) is is the other way around. Both (a) and (b) follow from a version of G2 due to Pudlák: (c) \( EA \not\vdash \forall(EA) := (S_1^2 + \{\Diamond_{n,EA} \top \mid n \in \omega \}) \). We suspect that the version of (b) with \( PA^{\downarrow \downarrow} \) replaced by \( DA \) does not directly follow from (c). However, this depends on a negative answer to Question 10.8(i).

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

11.3. 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 \( S_1^2 \).

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

References

[Boo93] G. Boolos. The logic of provability. Cambridge University Press, Cambridge, 1993.
[Bus86] S.R. Buss. Bounded Arithmetic. Bibliopolis, Napoli, 1986.

\[^{17}\text{We could even take } PA^- \text{ as starting point rather than } S_1^2, \text{ so removing all doubts regarding coordinate freedom. In fact we need only add closure under } \omega_1 \text{ to the } PA^- \text{-based version of IIA to obtain precisely IIA. Moreover, we could add } \omega_1 \text{ using a new symbol and adding appropriate recursion axioms, thus even avoiding sequence coding and the like.}\]
[dJJM91] D.H.J. de Jongh, M. Jumelet, and F. Montagna. On the proof of Solovay’s theorem. *Studia Logica*, 50:51–70, 1991.

[Fe60] S. Feferman. Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, 49:35–92, 1960.

[Fri21] 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.

[GS79] D. Guaspari and R.M. Solovay. Rosser sentences. *Annals of Mathematical Logic*, 16:81–99, 1979.

[HL20] V. Halbach and G. Leigh. The road to paradox: A guide to syntax, truth, and modality. To appear with Cambridge University Press, 2020.

[HP93] P. Hájek and P. Pudlák. *Metamathematics of First-Order Arithmetic*. Perspectives in Mathematical Logic. Springer, Berlin, 1993.

[Kra87] J. Krajíček. A note on proofs of falsehood. *Archiv für Mathematische Logik und Grundlagenforschung*, 26(1):169–176, 1987.

[Pud85] P. Pudlák. Cuts, consistency statements and interpretations. *The Journal of Symbolic Logic*, 50(2):423–441, 1985.

[PV21] P. Pakhomov and A. Visser. Finitely axiomatized theories lack self-comprehension. *arXiv preprint*, arXiv:2109.02548, 2021.

[PPZ20] F. Pakhomov and A. Zapryagaev. Multi-dimensional interpretations of Presburger Arithmetic in itself. *Journal of Logic and Computation*, 30(8):1681–1693, 09 2020.

[Sho58] J.R. Shoenfield. Degrees of formal systems 1. *The Journal of Symbolic Logic*, 23(4):389–392, 1958.

[Smo85a] C. Smoryński. Nonstandard models and related developments. In L.A. Harrington, M.D. Morley, A. Scedrov, and S.G. Simpson, editors, *Harvey Friedman’s Research on the Foundations of Mathematics*, pages 179–229. North Holland, Amsterdam, 1985.

[Smo85b] C. Smoryński. *Self-Reference and Modal Logic*. Universitext. Springer, New York, 1985.

[Vau62] R.L. Vaught. On a theorem of Cobham concerning undecidable theories. In E. Nagel, P. Suppes, and A. Tarski, editors, *Logic, Methodology and Philosophy of Science. Proceedings of the 1960 International Congress*, pages 14–25. Stanford University Press, Stanford, 1962.

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

[Vis92] A. Visser. An inside view of EXP. *The Journal of Symbolic Logic*, 57(1):131–165, 1992.

[Vis93] A. Visser. The unprovability of small inconsistency. *Archive for Mathematical Logic*, 32(4):275–298, 1993.

[Vis05] A. Visser. Faith & Falsity: a study of faithful interpretations and false $\Sigma^0_1$-sentences. *Annals of Pure and Applied Logic*, 131(1–3):103–131, 2005.

[Vis11] A. Visser. Can we make the Second Incompleteness Theorem coordinate free? *Journal of Logic and Computation*, 21(4):543–560, 2011.

[Vis12a] A. Visser. The Second Incompleteness Theorem and bounded interpretations. *Studia Logica*, 100(1-2):399–418, 2012. doi: 10.1007/s11225-012-9385-z.

[Vis12b] A. Visser. Vaught’s theorem on axiomatizability by a scheme. *Bulletin of Symbolic Logic*, 18(3):382–402, 2012.

[Vis13] A. Visser. What is sequentiability? In P. Cégielski, Ch. Cornaros, and C. Dimitracopoulos, editors, *New Studies in Weak Arithmetics*, volume 211 of *CSLI Lecture Notes*, pages 229–269. CSLI Publications and Presses Universitaires du Pôle de Recherche et d’Enseignement Supérieur Paris-est, Stanford, 2013.

[Vis14a] J. Visser. Peano Corso and Peano Basso: A study of local induction in the context of weak theories. *Mathematical Logic Quarterly*, 60(1–2):92–117, 2014.

[Vis14b] A. Visser. Why the theory $\mathbb{R}$ is special. In Neil Tennant, editor, *Foundational Adventures. Essays in honour of Harvey Friedman*, pages 7–23. College Publications, UK, 2014. Originally published online by Templeton Press in 2012. See http://foundationaladventures.com/.

[Vis15] A. Visser. The arithmetics of a theory. *Notre Dame Journal of Formal Logic*, 56(1):81–119, 2015.
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 we use in the paper is that predicate logic can interpret the theory of any finite model. The reader is referred to our paper $[Vis19]$ for a quick introduction to the details. The idea of a piece-wise interpretation is explained in $[Vis17]$ and in $[Vis14b]$.

We will use depth-of-quantifier-alternations as our measure $\rho$ of complexity of formulas. The notion is worked out in detail in $[Vis19]$.

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 $\Box, B$ for the formalised statement that $B$ is provable from the theory with axiom-set represented by $\alpha$. In case $A$ is a finitely axiomatised theory, we write $\Box_\alpha B$ for $\Box_{A_0} B$, where $\alpha_0 = \bigvee_{i<n} x = [\lceil A_i \rceil]$, where $A_0, \ldots, A_{n-1}$ are the axioms of $A$. We write $\Diamond$ for $\neg \Box \neg$.

- We write $\Box_{n, A} B$ for the formalised statement that $B$ is provable from $A$ using only statements of $\rho$-complexity $\leq n$, where $\rho$ measures the depth of quantifier alternations. We will usually implicitly assume that $n \geq \rho(A)$.

- We write $\Box_{\rho(A \rightarrow B), A} B$.

- 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 $\exists K : U \triangleright V$, etcetera, We write $U \bowtie V$ for $U$ and $V$ are mutually interpretable. In other words, $U \bowtie V$ iff $U \triangleright V$ and $U \triangleleft V$. In the context of a category of interpretations, we will use $V \xrightarrow{K} U$, for $K : V \triangleleft U$.

- $\mathcal{E} := INT^{+}_\uparrow$ is the category of theories and interpretations, where two interpretations $K,K' : V \rightarrow U$ are the same iff, for all $V$-sentences $A$, we have $U \vdash A^K \leftrightarrow A^{K'}$.

---

$[Vis16]$ A. Visser. The second incompleteness theorem: Reflections and ruminations. In G"odel’s Disjunction: The Scope and Limits of Mathematical Knowledge, pages 67–90. Oxford University Press, 2016.

$[Vis17]$ A. Visser. On Q. Soft Computing, 21(1):39–56, 2017.

$[Vis18]$ A. Visser. The interpretation existence lemma. In Feferman on Foundations, number 13 in Outstanding Contributions to Logic, pages 101–144, New York, 2018. Springer.

$[Vis19]$ A. Visser. The small-is-very-small principle. Mathematical Logic Quarterly, 65(4):453–478, 2019.

$[Vis21]$ A. Visser. Extension and interpretability. In Mathematics, Logic and their Philosophies, volume 49 of Logic, Epistemology, and the Unity of Science, pages 53–92. Springer Nature, 2021.

$[WP87]$ A.J. Wilkie and J.B. Paris. On the scheme of induction for bounded arithmetic formulas. Annals of Pure and Applied Logic, 35:261–302, 1987.
• $E_{UV}$ is the interpretation based on the identical translation that witnesses $V \subseteq U$.
• $U$ and $V$ are sententially congruent or elementary congruent iff they are isomorphic in $E$.
• A theory $U$ is restricted if, for some $m$ all its axioms are of $\rho$-complexity $\leq m$.
• Let an interpretation $N : S_2 \triangleleft U$ be given. A definable $N$-cut in a number theory $U$ is given by a formula $I$ such that $U \vdash I$ is a subclass of $N$ and is closed under $0, S, +, \times$ and $\omega$ in $N$ 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 external results that we employ in the paper. Let true be a standard $\Sigma^0_1$-truth predicate. The following theorem is a direct consequence of the estimate of the transformation of witnesses, when we move from $S$ to true$(S)$, for $\Sigma^0_1$-sentences $S$. See [HP93, 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 $S_2$. Then,

a. $S_2 \vdash S \rightarrow \text{true}(\langle S \rangle)$.

b. $S_2 \vdash S \rightarrow \square_{n,S_2} S$, for sufficiently large $m$.

(The number $m$ will be $\max(\rho(S_2), \rho(S))$ plus some constant for the overhead.)

The following theorem is a watered-down and inessentially modified version of Theorem 6.2. of [Vis19]. This theorem is an extension of earlier results independently obtained by Harvey Friedman (see [Smo85a]) and Jan Krajíček (see [Kra87]). See also [Vis93, Vis05].

**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 $N : S_2 \triangleleft U$. Then, from the data $m$, $K$, and $N$, we can effectively find an $N$-cut $I$ such that, for all $\Sigma^0_1$-sentences $S$, if $U \vdash S$, then $S$ is true.

The following theorem is Theorem 4.15 of [Vis05].

**Theorem A.3.** Suppose $A$ is a consistent, finitely axiomatised sequential theory. Let $U$ be RE. Suppose $A \bowtie U$. Then, there is an $N : S_2 \triangleleft U$, such that $N$ is $\Sigma^0_1$-sound. In other words, $U$ tolerates $S_2$ plus all true $\Pi^0_1$-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 $K : A \bowtie A$ and $N : A \bowtie S_2$. Then there is an $N$-cut $I$ such that $K \circ I$ is $\Sigma^0_1$-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 $N'$. \qed
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 theory V. We extend the signature of V with a unary predicate D and a binary predicate ∈. The axioms of 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. ηV is the interpretation of V in Seq(V) given by relativisation to D.

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

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

- M has two pieces s and o both of dimension 1.
- \( \delta_M^s \) is \( \delta_{K^*} \) and \( \delta_M^o \) is the complement of D.
- Identity between elements of different pieces is always \( \bot \). We have: \( x =^M_s y \) iff \( x \) and \( y \) are in \( \delta_M^s \) and \( x =_{K^*} y \), and \( x =^M_o y \) iff \( x \) and \( y \) are in \( \delta_M^o \) and \( x = y \).
- \( D_M^s \) is \( \delta_M^s \) and \( D_M^o \) is empty.
- Let \( P \) be an n-ary predicate of the V-language. Then \( P_{M}^{s \ldots o}(\vec{x}) \) tells us that each \( x_i \) is in \( \delta_M^s \) and that \( P_{K^*}(\vec{x}) \). If any variable in \( P(\vec{x}) \) is assigned a non-o piece, then \( P_M(\vec{x}) \) is false.
- We have:
  - \( x \in_M^s y \) and \( x \in_M^o y \) are false.
  - \( x \in_M^s y \) iff \( x \in \delta_M^s \), \( y \in \delta_M^s \), and there is an \( x' \) with \( x =_{K^*} x' \) and “\( (0, x') \in y' \). 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 . . .
  - \( x \in_M^s y \) iff \( x \in \delta_M^o, y \in \delta_M^o \), and “\( \langle 1, x \rangle \in y' \).”

It is easy to verify that the interpretation sketched here indeed is an interpretation of Seq(V) in W.

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

Open Question A.7. Does Seq or, some appropriate variant of it, lift to a functor on E?  

Theorem A.8. Suppose V is sequential. Then, Seq(V) \( \bowtie V \).

Proof. We have \( V \xrightarrow{\eta_V} \text{Seq}(V) \). Let \( \varepsilon^* \) be a V-formula that witnesses that V is sequential. Our interpretation \( M \) of Seq(V) in V looks as follows. We duplicate the domain by having two pieces s and o with domain \( x = x \). We let D correspond
to the elements in the $\sigma$-piece and have the predicates of $V$ on the $\sigma$-piece. We take $x \in^*_M y$ iff $x$ and $y$ are in the correct domains and “$(0, x) \in^* y$”. Similarly, $x \in^*_M y$ iff $x$ and $y$ are in the correct domains and “$(1, x) \in^* y$”. We set $\in^*_M$ to $\bot$ in the other cases.

**Appendix B. Proof of Theorems 6.11 and 6.12**

This appendix contains the proofs of Theorems 6.11 and 6.12 plus some discussion on an alternative more categorical formulation of Theorem 6.12.

**B.1. Proof of Theorem 6.11**

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 $K_0 \approx K_1$. We want to show that $U_1$ is Friedman-reflexive and that $\Lambda_{fr}^K K_0 = \Lambda_{fr}^K K_1$.

Since, $K_0 \approx K_1$, we can find $M, \tilde{M}, P, \tilde{P}$ such that the following diagram commutes.

\[
\begin{array}{ccc}
U_0 & \xrightarrow{K_0} & A_0 \\
\downarrow{M, \tilde{M}} & & \downarrow{P, \tilde{P}} \\
U_1 & \xrightarrow{K_1} & A_1
\end{array}
\]

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

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

**Lemma B.1.** $A_0 \vdash \Diamond_{K_0, A_0} B \leftrightarrow \Diamond_{K_1, A_1} B^P$ and $A_1 \vdash \Diamond_{K_1, A_1} D \leftrightarrow \Diamond_{K_0, A_0} D^P$.

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

\[ (†) \quad U_0 \vdash \Diamond_{(U_0), A_0} B \leftrightarrow \Diamond_{(U_1), A_1} B^P. \]

We have:

\[ U_0 + \Diamond_{(U_1), A_1} B^P \vdash U_1 + \Diamond_{(U_1), A_1} B^P \]

\[ \supseteq (A_1 + B^P) \]

It follows that

\[ (†) \quad U_0 + \Diamond_{(U_1), A_1} B^P \vdash \Diamond_{(U_0), A_0} B. \]

By symmetry, we find $U_1 + \Diamond_{(U_0), A_0} D^Q \vdash \Diamond_{(U_1), A_1} D$. Substituting $B^P$ for $D$ gives $U_1 + \Diamond_{(U_0), A_0} B^{PQ} \vdash \Diamond_{(U_1), A_1} B^P$. By Theorem 6.13, we may replace provable equivalents under $\Diamond_{(U_0)}$, so: $U_1 + \Diamond_{(U_0), A_0} B \vdash \Diamond_{(U_1), A_1} B^P$. It follows that

\[ U_0 + \Diamond_{(U_0), A_0} B \vdash \Diamond_{(U_1), A_1} B^P \]

and, hence,

\[ $U_0 + \Diamond_{(U_0), A_0} B \vdash \Diamond_{(U_1), A_1} B^P$. \]
Combining (†) and (§), we find (‡). In its turn (†) gives us:

\[ A_0 \vdash \phi(\tau_0, K_0) \leftrightarrow \psi(\tau_0, K_0) \]
\[ A_0 \vdash \phi(\tau_0, K_0) \leftrightarrow \psi(\tau_0, K_0) \]
\[ A_0 \vdash \phi(\tau_0, K_0) \leftrightarrow \psi(\tau_0, K_0) \]
\[ A_0 \vdash \phi(\tau_0, K_0) \leftrightarrow \psi(\tau_0, K_0) \]

Suppose \( \tau_i \) 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 \( D \mapsto \rightarrow D \).

Lemma B.2. We have:

\[ A_0 \vdash \phi(\tau_0, K_0) \leftrightarrow \psi(\tau_0, K_0) \]
\[ A_0 \vdash \phi(\tau_0, K_0) \leftrightarrow \psi(\tau_0, K_0) \]
\[ A_0 \vdash \phi(\tau_0, K_0) \leftrightarrow \psi(\tau_0, K_0) \]
\[ A_0 \vdash \phi(\tau_0, K_0) \leftrightarrow \psi(\tau_0, K_0) \]

Proof. We prove the first conjunct. Our proof is by induction on \( \varphi \). In the atomic case we have:

\[ A_0 \vdash \phi(\tau_0, K_0) \leftrightarrow \psi(\tau_0, K_0) \]
\[ A_0 \vdash \phi(\tau_0, K_0) \leftrightarrow \psi(\tau_0, K_0) \]
\[ A_0 \vdash \phi(\tau_0, K_0) \leftrightarrow \psi(\tau_0, K_0) \]
\[ A_0 \vdash \phi(\tau_0, K_0) \leftrightarrow \psi(\tau_0, K_0) \]

The cases of the truth-functional connectives are simple. Finally, for the case of \( \Diamond \), we employ Lemma B.1

\[ A_0 \vdash \Diamond \psi(\tau_0, K_0) \leftrightarrow \Diamond \psi(\tau_0, K_0) \]
\[ A_0 \vdash \Diamond \psi(\tau_0, K_0) \leftrightarrow \Diamond \psi(\tau_0, K_0) \]
\[ A_0 \vdash \Diamond \psi(\tau_0, K_0) \leftrightarrow \Diamond \psi(\tau_0, K_0) \]
\[ A_0 \vdash \Diamond \psi(\tau_0, K_0) \leftrightarrow \Diamond \psi(\tau_0, K_0) \]

Finally, we prove the theorem. We have:

\[ A_0 \vdash \phi(\sigma, K_1) \Rightarrow A_0 \vdash \phi(\sigma, K_1) \]
\[ A_0 \vdash \phi(\sigma, K_1) \Rightarrow A_0 \vdash \phi(\sigma, K_1) \]

It follows that, if \( A_{K_1} \vdash \phi \), then \( A_{K_0} \vdash \phi \). By symmetry, we also have the other direction.

B.2. Proof of Theorem 6.12. Let \( F \) be an endofunctor of \( D \). Here we assume that \( F \) operates on concrete theories and interpretations. Suppose that:

A. \( F \) preserves finite axiomatisability.
B. For each theory \( V \), we have a faithful interpretation \( V \xrightarrow{\eta} F(V) \).
C. Suppose \( \Gamma \) is a set of sentences in the \( V \)-language. Then, \( F(V + \Gamma) = F(V) + \Gamma^\eta \).

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 \( F(V) \xrightarrow{M} V \), for all extensions \( V \) of \( U \) in the same language. Let \( A \) be finitely axiomatised and suppose \( U \xrightarrow{K} A \).

We show that \( A_{\eta A \circ K} \subseteq A_{K} \).
Proof. For some $P$, we have $(A + B) \xrightarrow{P} (U + \Box_A B)$. Let $M' := M_{U + \Box_A B}$. We have:

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

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

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

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

So, (b) $U \vdash \Box_{F(A)} B^{\eta_A} \leftrightarrow \Box_A B$.

We may conclude that:

$$(\dag) \quad F(A) \vdash \Box_{\eta_A \circ K, F(A)} B^{\eta_A} \leftrightarrow \Box_{\eta_A} B$$

Let $\sigma$ be any function from propositional variables to sentences of the $A$-language. To lighten our notational burdens a bit, we will confuse $\eta_A$ 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 $\Box$. Let $\varphi ::= \Box \psi$. Using $(\dag)$, we find:

$$F(A) \vdash \varphi(\eta_A \circ \sigma, \eta_A \circ K) \leftrightarrow \Box_{\eta_A \circ K, F(A)} \psi(\eta_A \circ \sigma, \eta_A \circ K) \leftrightarrow \Box_{\eta_A \circ K, F(A)}(\psi(\sigma, K))^{\eta_A} \leftrightarrow \Box_{\eta_A} \psi(\sigma, K) \leftrightarrow (\varphi(\sigma, K))^{\eta_A}$$

From the fact that $\eta_A$ is faithful, we now have:

$$(\ddag) \quad F(U) \vdash \varphi(\eta_A \circ \sigma, \eta_A \circ K) \leftrightarrow F(U) \vdash (\varphi(\sigma, K))^{\eta_A} \leftrightarrow U \vdash \varphi(\sigma, K)$$

Our theorem is immediate from $(\ddag)$.

Remark B.3. The conditions for Theorem 6.12 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 $\text{Seq}$ that is an endofunctor of $\mathbb{E}$.

We work in category $\mathbb{E}$ enriched with designated embedding arrows $V \xrightarrow{E_{VW}} W$ 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 $\bot \to F(V)$, then $\bot \to V$. 
C*. For each V, there is an interpretation V \stackrel{\eta_V}{\rightarrow} F(V) such that:

\[
\begin{array}{c}
W \xrightarrow{\eta_W} F(W) \\
\subseteq \xleftarrow{\subseteq} \\
V \xrightarrow{\eta_V} F(V)
\end{array}
\]

If V \subseteq W, then, E_{F(V)} \circ \eta_V = \eta_W \circ E_{V W}. Moreover, whenever E_{F(W) Z} \circ \eta_W = P \circ E_{V W}, then F(W) \subseteq Z and P = E_{F(W) Z} \circ \eta_W.

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

**APPENDIX C. 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 \seteq \{ \varphi \mid \forall X \exists \sigma \in X A \vdash \varphi(\sigma, K) \}.

* \Lambda is a local logic for \langle A, U \rangle if \Lambda is a subset of \Lambda^*_A U that contains K4 and is closed under modus ponens, necessitation, and substitution.

We have the following obvious facts.

**Theorem C.1.** Consider an FM-frame \langle A, U \rangle. Then, \Lambda^*_A U is closed under substitution and necessitation. Moreover, if \varphi is in \Lambda^*_A U and \psi is in \Lambda^*_A U, then (\varphi \land \psi) is in \Lambda^*_A U.

**Theorem C.2.** Consider an FM-frame \langle A, U \rangle. A local logic for \langle A, U \rangle is a logic. Moreover, every \varphi \in \Lambda^*_A U is contained in a minimal local logic containing it.

**Theorem C.3.** Consider an FM-frame \langle A, U \rangle. Let \Lambda be a local logic for \langle A, U \rangle. Then, the logic generated by \Lambda and \Lambda^*_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 \gg 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) \gg (W + C).

**Theorem C.4.** Consider an FM-frame \langle A, U \rangle. Then, \gg has the forward property for A, U iff S_4 is a local logic for \langle A, U \rangle.

**Proof.** Suppose \gg has the forward property for A, U. Consider any sentences B_0, ..., B_{n−1} in the language of A. Let B_0^*, ..., B_{n−1}^* enumerate all conjunctions of the form \bigwedge_{i<n} \pm B_i, where \pm B_i is either B_i or ¬B_i.

Consider any B_j^*. We have, for some C_j^* that (A + B_j^*) \gg (U + C_j^*). It follows that U \vdash C_j^* \rightarrow \Box_A B_j^*. Ergo, (A + B_j^*) \gg (U + \Box_A B_j^*). Let K_j be the witnessing interpretation. It follows that A + B_j^* \vdash \Box_{K_j, A} B_j^*. If B_i is unnegated in B_j^*, then A \vdash \Box_{K_j, A} B_j^* \rightarrow \Box_{K_j, A} B_i. If B_i is negated in B_j^*, then A + B_j^* \vdash B_i \rightarrow \Box_{K_j, A} B_i.

So, A + B_j^* \vdash \bigwedge_{i<n} (B_i \rightarrow \Box_{K_j, A} B_i)

Now let K := K_0(B_0^*)K_1(B_1^*)…). Because the B_j^* are mutually exclusive, we find that, for each j < k, we have A + B_j^* \vdash \bigwedge_{i<n} (B_i \rightarrow \Box_{K_j, A} B_i) and, thus
A ⊢ \bigwedge_{i<n} (B_i \rightarrow \Phi_{K,A} B_i). From this, it is immediate that \( p \rightarrow \Box p \) is in \( \Lambda_{A,U}^\star \).
Moreover, \( S4 \) is the logic generated by \( p \rightarrow \Box p \) over \( K4 \).

In the other direction, suppose \( S4 \) is a local logic for \( \langle A, U \rangle \). Consider any \( B \) in the \( A \)-language. For some \( K : A \triangleright U \), we have \( A + B \vdash \Phi_{K,A} B \). So, \( (A + B) \triangleright (U + \Phi_A B) \). Moreover, \( (U + \Phi_A B) \triangleright (A + B) \).

**Open Question C.5.** The second part of the proof of Theorem C.4 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 \( S4 \) and \( \Lambda_{A,U}^\star \) is a local logic for \( \langle A, U \rangle \).

**Appendix D. 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 5.19.

**Q2.** Can we find a more inspiring example of a theory with logic \( S4 \) than Example 6.8? Is it, perhaps, possible to find an FM-interpretation with interpreter logic precisely \( S4 \)? This is Question 6.9.

**Q3.** Suppose \( K, M : U \triangleleft A \) are FM-interpretations and \( \text{Th}(K) = \text{Th}(M) \). Do we have \( \Lambda_K^U = \Lambda_M^U \), or is there a counter-example? This is Question 6.16.

**Q4.** Is there an example of an FM-interpretation \( K : A \triangleright U \), where \( U \) is complete, with an interesting interpreter logic? This is Question 7.4.

**Q5.** Is there a consistent finitely axiomatised theory that is effectively Friedman-reflexive? This is Question 8.1.

**Q6.** 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 10.8(i).

**Q7.** Is there a finitely axiomatised \( A \) and \( K : A \triangleleft DA \), such that, for no \( D(x) \) in the \( A \)-language, we have, for all \( B \) in the \( A \)-language, \( A \vdash D(\downarrow B) \leftrightarrow \Phi_{K,A} B \)? Here the numerals are the \( K \)-numerals. This is Question 10.8(ii).

**Q8.** Is there an RE sequential theory that is Friedman-reflexive but not effectively so? This is Question 10.8(iii).

**Q9.** Suppose \( U \) is sequential and restrictedly (effectively) Friedman-reflexive. Does it follow that \( U \) is essentially sententially reflexive? This is Question 10.8(iv).

**Q10.** 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 \( \Sigma^0_1 \)-sentences by an unrestricted one that ranges over arbitrary formulas. This is Question 11.2(i).

**Q11.** Is there a coordinate-free specification of an interpretation of Peano Corto in \( EA \) or in some \( \Sigma_n \)? This is Question 11.2(ii).

**Q12.** Do we have \( EA \triangleright IA \)? This is Question 11.2(iii).

**Q13.** Does \( \text{Seq} \) or, some appropriate variant of it, lift to a functor on \( \mathbb{E} \)? This is Question A.7.

**Q14.** The second part of the proof of Theorem C.4 only uses a singleton set \( X \). We wonder whether that means that our approach may be simplified. This is Question C.5.