EXTENSION & INTERPRETABILITY

ALBERT VISSER

Abstract. In this paper we study the combined structure of the relations of theory-extension and interpretability between theories for the case of finitely axiomatised theories. We focus on two main questions.

The first is the matter of definability of salient notions in terms of the structure. We show, for example, that local tolerance, locally faithful interpretability and the finite model property are definable over the structure.

The second question is how to think about 'good' properties of theories that are independent of implementation details and of 'bad' properties that do depend on implementation details. Our degree structure is suitable to study this contrast, since one of our basic relations, to wit theory-extension, is dependent on implementation details and the other relation, interpretability, is not. Nevertheless, we can define new good properties using bad ones. We introduce a new notion of sameness of theories *i-bisimilarity* that is second-order definable over our structure. We define a notion of *goodness* in terms of this relation. We call this notion *being fine*. We illustrate that some intuitively good properties, like being a complete theory, are not fine.

1. Introduction

The structure of extensions and interpretations combines two important structures: the structure of theory-extensions for finitely axiomatized theories, which is a disjoint sum of Lindenbaum algebras. Lindenbaum and the degrees of interpretability of finitely axiomatised theories. Sentence algebras are studied e.g. in [Han75] and [Mye89]. The degrees of interpretability for are studied in e.g. [Šve78], [Lin79], [Lin84b], [Lin84a], [Ben86], [Lin02], [Fri07], [Vis14a], [Vis14b].

The main reason for considering the double structure in this paper is our focus on questions of concrete first- and second-order definability over the structure. We are interested in the question of definability of such notions as *sequentiality* and *faithful interpretability*.

We zoom in on the structure of consistent, finitely axiomatized theories with as relations extension-in-the-same-language and parameter-free, multi-dimensional, piecewise interpretability. As will become apparent from the paper, this structure seems to be best suited to the treatment of *local* notions. We show, example, that locally faithful interpretability, local tolerance and the finite model property are definable in the structure. With respect to faithful interpretability and sequentiality,

---

Date: July 27, 2015.

2000 *Mathematics Subject Classification.* 03B30, 03F25, 03F30.

Key words and phrases. degrees of interpretability, Lindenbaum algebras, arithmetic, sequentiality.

I thank Johan van Benthem for his comments and questions. I am grateful to Volodya Shavrukov for his comments and help with the references. I thank Victor Pambuccian for pointing me to Szczerba's work. I thank Clemens Grabmayer, Vincent van Oostrom and Freek Wiedijk for their help with a LaTeX problem.

we have the partial negative result that they are not definable in a certain fragment of the first-order language over the structure. The question of definability in the full language remains open.

The study of interpretability is, in part, about the escape from the tyranny of signature. Specific choices for the language are implementation artifacts introduced because, after all is said an done, we have to do things one way or another. Good mathematical properties of theories should be independent of these arbitrary choices. In this light, the study of the combined structures of extension and interpretability seems to be a retrograde movement since the notions of extension-in-the-same-language is implementation-dependent. Upon closer inspection, this bug turns out to be a feature: precisely because we have both an implementation-independent notion (interpretability) and an implementation-dependent notion (extension) in one framework, we can study how the implementation-dependent notion can function as an auxiliary to define implementation-independent concepts.

To get a grip on implementation-independence, we need a notion of sameness of theories. I think there is a best notion of sameness, to wit *bi-interpretability*. This notion preserves many good mathematical properties of theories that are as diverse as finite axiomatisability, decidability and κ-categoricity. To study bi-interpretability in the context of our framework, however, is at this stage on bridge too far. So, in the paper we replace bi-interpretability by a cruder notion, to wit (*local*) *i-bisimilarity*. We study the concept of 'good property' derived from i-bisimilarity. We call this idea of goodness: *being fine*. We illustrate what i-bisimilarity in the context of our degree structure can and cannot do. E.g., we show that the notions of complete theory, of sequential theory and of faithful interpretability are not fine, where they are definitely good when measured by our best notion of sameness: bi-interpretability.

2. Preliminaries

In this section, we give basic dfinitions and discuss some basic facts.

2.1. **Signatures, Formulas, Theories.** Signatures will be finite in this paper. For any signature Σ, we take $\mathcal{L}_\Sigma$ to be the set of sentences of that signature. We assign a finitely axiomatized theory $\mathcal{I}_\Sigma$ to Σ. This is the theory of identity for Σ including the axiom $\exists x\, x = x$.

We will use A, B, C, ... as variables ranging over consistent finitely axiomatized theories of finite signature. We take the theory of identity $\mathcal{I}_\Sigma$ as part of predicate logic CQC_Σ of signature Σ. Still we insist that $\mathcal{I}_\Sigma$ is *also* syntactically present in a theory. The reason for this somewhat strange stipulation is that we allow identity to be translated to some other relation than identity. Hence, we need the axioms of identity to be interpreted.[1]

We use ϕ, ψ as ranging over sentences. We suppose that theories and sentences have built-in signature, so that we can read out the signature from the sentence. So, e.g. $\exists x\, P(x)$ is a different sentence depending on whether it occurs in a language

[1]The unnaturality is caused by the fact that, on the one hand, we treat identity as *a logical relation*, and, on the other hand, unlike other logical relations, we do not translate it to itself. The truly coherent way of proceeding would be to work without identity as part of the logic and with a free logic. This way would greatly simplify things. In this paper, however, we stick to the traditional framework and do the sometimes boring homework to treat identity.

of signature with only the predicate symbol P or whether there is e.g. a second binary predicate symbol Q.

A basic relation between theories is extension. By this we mean: extension-in-the-same-language. More precisely B is an extension A when $\Sigma_A = \Sigma_B$ and $B \vdash A$. We write $A \subseteq B$ for: B extends A in the same language.

A second basic relation is interpretability or $\lhd$. We will explain this in Subsection 2.2 and in Appendix A.

A special theory we will consider is the arithmetical theory S^1_2. See e.g. [Bus86] or [HP93] for an explanation. The signature of the language of S^1_2 is $\mathcal{A}$.

2.2. **Translations.** A translation maps a language to a language. The basic idea is very simple: the translation commutes with the logical connectives. In spite of the apparent simplicity. there is some work to do to give a full and correct definition of a translation. There are two reasons. The first is that we have to get nasty details concerning renaming of variables out of the way. The second is that we want to add a number of features. The idea in this paper is to ignore the problem of the variables. It is clear that it can be taken care of in some appropriate way. We will however have a lot to say in Appendix A about the extra features. The features are these:

- Our translations do not necessarily translate identity to identity. Identity may go to a formula that is intended to represent an equivalence relation.
- Our translations are relativised. We will relativise the translations of the quantifiers to prespecified domains.
- Our translations are more dimensional. This means that one object in the translated language may be represented as a sequence of objects in the translating language.
- We allow piecewise translations. This means that our domains can be built up from various pieces. These pieces may be of different dimensions. Moreover, even if they are of the same dimension, a sequence of elements shared by two pieces may, in the context of the first piece, represent a different object of the translated language than the object represented by the sequence in the context of the second piece.
- We may have parameters in our translations. The main development of the paper is executed for the parameter-free case. In Appendix D, we sketch how to adapt it for the case with parameters.

The notions of translation yields the notion of interpretation. If we have theories A and B, then A interprets B or $A \rhd B$ if $A \vdash B^\tau$.

We give many detail in Appendix A on how translations are defined and how they work.

2.3. **The Structure.** We will study the structure $\mathbb{EI}$ of extension and interpretability. Its objects are the finitely axiomatized consistent theories A, B, ... The structure has two relations $\subseteq$ and $\lhd$. Here $\subseteq$ is a partial ordering, $\lhd$ is a partial pre-ordering and $\subseteq$ is a subrelation of $\lhd$.

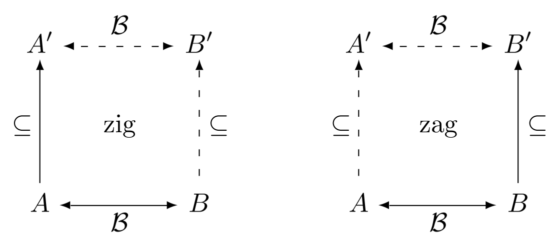
In the main body of the paper we take $\lhd$ to stand for: parameter-free, piecewise, multi-dimensional interpretability. We discuss the matter of adapting the framework to translations with parameters in Appendix D.

We will be interested in properties $P(A_0, \ldots, A_{n-1})$ of finitely axiomatised consistent theories that can be formulated in terms of the structure. In various ways these properties can be divided in good and bad properties. Given a notion of sameness for theories $\sim$ we say that P is $\sim$-invariant if P preserves $\sim$, in other words, if $A_0 \sim B_0, \ldots, A_{n-1} \sim B_{n-1}$ and $P(A_0, \ldots, A_{n-1})$, then $P(B_0, \ldots, B_{n-1})$.

I believe in the intuitive thesis that the right notion of sameness between theories is bi-interpretability. In other words, preservation of $\sim_{\mathsf{bii}}$ is true goodness. Bi-interpretability is more flexible than the very strict notion of synonymy, but still preserves many mathematically interesting properties like categoricity and decidability. There are two more salient notions iso-congruence and sentential or elementary congruence. See Appendix A.7. All of these notions have an important property that mutual interpretability and mutual faithful interpretability lack. Suppose $A \sim B$, where $\sim$ is synonymy or bi-interpretability or iso-congruence or sentential congruence. Suppose we add a new axiom ϕ to A. Then, we can add a matching axiom ψ to B such that $(A + \phi) \sim (B + \psi)$.[2] Of course, we also have a converse: if we add a ψ to B, there is a matching ϕ for A. Thus $\sim$ is a bisimulation with respect to $\subseteq$. There is a crudest notion below mutual interpretability or $\equiv$ that has this property, to wit *i-bisimilarity.*[3]

An i-bisimulation is a relation $\mathcal{B}$ on $\mathbb{EI}$ such that (i) $\mathcal{B}$ is a subrelation of $\equiv$ and (ii) $\mathcal{B}$ is a bisimulation w.r.t. $\subseteq$. We remind the reader that $\mathcal{B}$ is a bisimulation w.r.t. $\subseteq$ iff, it has both the forward or zig property and the backward or zag property: if $A \mathrel{\mathcal{B}} B$, then

zig: for all $A' \supseteq A$, there is a $B' \supseteq B$, such that $A' \mathrel{\mathcal{B}} B'$;
zag: for all $B' \supseteq B$, there is an $A' \supseteq A$, such that $A' \mathrel{\mathcal{B}} B'$.



The theories A and B are i-bisimilar or $A \approx B$ iff, there an i-bisimulation $\mathcal{B}$, such that $A \mathrel{\mathcal{B}} B$. We note that $\approx$ is the maximal i-bisimulation. Since i-bisimulations contain the identity relations and are closed under converse and composition, we find that $\approx$ is an equivalence relation.

Warning: If we would consider the structure of arbitrary consistent theories with our two relations, the notion of i-bisimularity would be different. It is easy to give examples of two finitely axiomatized theories A and B that are i-bisimilar in the sense of $\mathbb{EI}$ but not i-bisimilar in the sense of the corresponding structure with possibly infinite theories.

We see that the notion of i-bisimularity has a Σ^1_1-description over $\mathbb{EI}$. We call a property *fine* if it is $\approx$-invariant. We note that all fine properties are also $\sim$-invariant where $\sim$ is synonymy, bi-interpretability, iso-congruence and sentential

[2]For a proof of this claim, see Appendix A.7.6.

[3]The notion should really be called: *local i-bisimilarity.* However, since this paper is about local notions, we omit the 'local'.

equivalence. The converse does not hold: as we will see *being a complete theory* is not fine but it is $\sim$-invariant for sentential congruence and, *ipso facto*, for all more refined equivalence relations.

Open Question 2.1. We define $A \equiv_0 B$ iff $A \equiv B$ and $A \equiv_{n+1} B$ iff $A \equiv B$ and, for all $A' \supseteq A$, there is a $B' \supseteq B$, such that $A' \equiv_n B'$, and for all $B' \supseteq B$, there is a $A' \supseteq A$, such that $A' \equiv_n B'$. Clearly, $\equiv_{n+1}$ is a subrelation of $\equiv_n$, and $\approx$ is a subrelation of the intersection $\equiv_\omega$ of the $\equiv_n$. Are all these inclusions strict? ❑

Remark 2.2. In stead of $\subseteq$ we could take $\subsetneq$ as the second relation of our framework. Let the corresponding equivalence relation be $\approx^+$. We can easily see that $\approx^+$ is strictly contained in $\approx$. ❑

We introduce a fragment $\mathcal{P}$ of first-order formulas over $\mathbb{EI}$ such that every formula ϕ in $\mathcal{P}$ defines a fine property. We define $\exists A' \supseteq A\ \ldots$ by $\exists A'\,(A \supseteq A' \wedge \ldots)$ and we define $\forall A' \supseteq A\ \ldots$ by the formula $\forall A'\,(A \supseteq A' \to \ldots)$. Here we assume that A' and A are distinct variables. (Officially, we allow that A and A' are the same. In this case, we translate e.g. $\exists A \supseteq A\ \ldots$ to $\exists A^\star\,(A^\star \supseteq A \wedge (\ldots)[A := A^\star])$, where $A^\star$ is a fresh variable.) We define the fragment $\mathcal{F}$ by:

- $V ::= A \mid B \mid C \mid \ldots$
- $F ::= \bot \mid \top \mid V \lhd V \mid \neg F \mid (F \wedge F) \mid (F \vee F) \mid (F \to F) \mid$
  $\forall V\, F \mid \exists V\, F \mid \forall V \supseteq V\ \ F \mid \exists V \supseteq V\ \ F$

We claim that every formula in $\mathcal{F}$ defines a fine relation. The proof is by induction of the formulas in $\mathcal{F}$. We treat the case of the reverse bounded existential quantifier. Suppose $\phi(A, \vec{C}\,)$ defines a fine relation. Here the $\vec{C}$ cover all occurrences of free variables in ϕ except A and A represents all occurrences of the free variable A. We claim that the relation given by $\exists A \supseteq B\ \ \phi(A, \vec{C}\,)$ is also fine. Here B may be one of the $\vec{C}$. We assume that A and B are distinct. It is clearly sufficient to consider the case where we replace one occurrence of D by D' with $D \approx D'$, since more complicated replacements can be constructed as a number of such simpler replacements. Each such replacement is unproblematic as long as we do not replace the first occurrence of B. Suppose we replace the first occurrence of B by B', where $B \approx B'$. For any $A \supseteq B$, we can find $A' \supseteq B'$ with $A \approx A'$. Hence, by the induction hypothesis, $\phi(A, \vec{C})$ iff $\phi(A', \vec{C})$. It follows that $\exists A \supseteq B\ \ \phi(A, \vec{C}\,)$ iff $\exists A' \supseteq B'\ \ \phi(A', \vec{C}\,)$ and by α-conversion, $\exists A \supseteq B\ \ \phi(A, \vec{C}\,)$ iff $\exists A \supseteq B'\ \ \phi(A, \vec{C}\,)$.

Open Question 2.3. Is every fine property definable in $\mathcal{F}$? ❑

2.4. **Salient Notions.** We introduce a number of salient notions relevant to our framework. These notions are generalisations and localisations of the notions introduced by Per Lindström and Giorgi Japaridze for extensions of PA. See [JdJ98, Section 11]. Because of our restriction to finitely axiomatized theories our framework seems to be primarily suitable for the study of *local* versions of the various notions. The first two concepts, to wit the logic Λ^θ_A of A and admissible inference over A, will play no essential role in this paper. We just add them to make our list reasonably complete.

- $\Lambda^\Theta_A := \{\phi \in \mathcal{L}_\Theta \mid \text{for all } \tau : \Theta \to \Sigma_A \text{ such that } A \vdash \mathcal{I}^\tau_\Theta \text{ we have } A \vdash \phi^\tau\}$.
  Λ^Θ_A is the logic of A for Θ.
- $B\ \mathrel{\vert\!\sim}_A \phi$ iff, $\Sigma_B = \Sigma_\phi$ and, for all $\tau : \Sigma_B \to \Sigma_A$, if $A \vdash B^\tau$, then $A \vdash \phi^\tau$.
  The relation $\mathrel{\vert\!\sim}_A$ represents admissible inference over A.

- A *interprets* B or $A \rhd B$ iff, for some τ, $A \vdash B^\tau$.
- A *weakly interprets* B or A *tolerates* B or $A \Vdash B$ iff, for some translation τ, we have that $A + B^\tau$ is consistent.
- A is *locally tolerant* iff $A \Vdash B$, for each B.
- A is *essentially locally tolerant* iff, for each $A' \supseteq A$, we have $A' \Vdash B$, for each B.
- A *co-interprets* B or $A \blacktriangleright B$ if, for some τ, for all ϕ, if $B \vdash \phi^\tau$, then $A \vdash \phi$.[4]
- A *locally co-interprets* B or $A \blacktriangleright_{\mathsf{loc}} B$ if, for all ϕ, there is a τ such that, if $B \vdash \phi^\tau$, then $A \vdash \phi$.
- A *faithfully interprets* B or $A \rhd_{\mathsf{faith}} B$ iff, there is a τ, such that, for all ϕ, we have:
  $B \vdash \phi$ iff $A \vdash \phi^\tau$.
- A *locally faithfully interprets* B or $A \rhd_{\mathsf{lofa}} B$ iff, for all ϕ with $B \nvdash \phi$, there is a τ such that $A \vdash B^\tau$ and $A \nvdash \phi^\tau$.

3. Characterisations

Our salient local notions have various characterisations. In this section we collect those characterisations that are 'theory-free', i.e. for which we do not need results like the Interpretation Existence Theorem. More theoretically involved characterisations will be treated in Section 6.

3.1. The Logic of a Theory, Admissibility and Interpretability.

We give various characterisations of the logic of a theory, of admissible rules and of interpretability. We repeat the relevant definitions.

- $\Lambda_A^\Theta := \{\phi \in \mathcal{L}_\Theta \mid \text{for all } \tau : \Theta \to \Sigma_A \text{ such that } A \vdash \mathcal{I}_\Theta^\tau \text{ we have } A \vdash \phi^\tau\}$.
- $B \;\mid\!\sim_A \phi$ iff, $\Sigma_B = \Sigma_\phi$ and, for all $\tau : \Sigma_B \to \Sigma_A$, if $A \vdash B^\tau$, then $A \vdash \phi^\tau$.
- A *interprets* B iff, for some τ, $A \vdash B^\tau$.

Theorem 3.1. *We have:*
1. $\Lambda_A^\Theta = \{\phi \in \mathcal{L}_\Theta \mid \mathcal{I}_\Theta \;\mid\!\sim \phi\}$.
2. $\phi \in \Lambda_A^\Theta$ *iff,* $\phi \in \mathcal{L}_\Theta$, *and, for all* $\tau \colon \Theta \to \Sigma_A$, *we have* $A \vdash (\bigwedge \mathcal{I}_\Theta \to \phi)^\tau$.
3. $\phi \notin \Lambda_A^\Theta$ *iff* $A \Vdash (\mathcal{I}_\Theta \wedge \neg \phi)$.

Proof. Claim (1) is trivial.

We prove Claim (2). Suppose $\phi \in \Lambda_A^\Theta$ and $\tau \colon \Theta \to \Sigma_A$. Let $\tau_0 \colon \Theta \to \Sigma_A$ be defined as follows. The translation τ_0 is 1-dimensional, $\delta_\tau(x) := \top$, and τ sends any predicate symbol including identity to $\top$.[5] We define $\tau^\star := \tau\langle \mathcal{I}_\Theta^\tau \rangle \tau_0$.[6] Clearly, $A \vdash \mathcal{I}_\Theta^{\tau^\star}$. Hence, $A \vdash \phi^{\tau^\star}$. We may conclude that $A \vdash (\bigwedge \mathcal{I}_\Theta \to \phi)^\tau$. The converse is immediate.

Claim (3) is immediate from Claim (2). ❑

The characterisation of Theorem 3.1(2) is very useful. We will use it without mentioning that we apply the theorem.

---

[4] In our formulations we will always implicitly assume that the ϕ we quantify over are of the right signature. In this case the hidden assumption is that $\Sigma_\phi = \Sigma_A$.

[5] It would be more natural to use a 0-dimensional translation, but I wanted to avoid any suspicion that our argument depends on some special etheric feature of the notion of translation.

[6] $\tau\langle \mathcal{I}_\Theta^\tau \rangle \tau_0$ is the translation that is τ if $\mathcal{I}_\Theta^\tau$ and that is τ_0 otherwise. See Appendices A.2 and A.6 for the official definitions.

Theorem 3.2. *$A \rhd B$ iff $B \not\vdash\!\sim_A \bot$.*

Proof. We have:

$$\begin{aligned} B \not\vdash\!\sim_A \bot &\Leftrightarrow \exists\tau\,(A \vdash B^\tau \text{ and } A \nvdash \bot^\tau) \\ &\Leftrightarrow A \rhd B \end{aligned}$$

❑

Theorem 3.3. *Suppose B and ϕ have signature Θ. We have:*

$$B \vdash\!\sim^{\Theta}_{A} \phi \quad \Leftrightarrow \quad (B \vdash\!\sim^{\Theta}_{A} \bot \text{ or } B \vdash_{\Lambda^{\Theta}_{A}} \phi).$$

Proof. The right-to-left direction is trivial. We prove left-to-right. Suppose we have $B \vdash\!\sim^{\Theta}_{A} \phi$. In case we have $B \vdash\!\sim^{\Theta}_{A} \bot$, we are done. Otherwise there is a τ_0 such that $A \vdash B^{\tau_0}$. Consider any translation $\tau : \Theta \to \Sigma_U$. Let $\tau^\star := \tau\langle B^\tau\rangle\tau_0$. We clearly have $A \vdash B^{\tau^*}$ and, hence, $A \vdash \phi^{\tau^*}$. It follows that $A + B^\tau \vdash \phi^\tau$. ❑

3.2. **Weak Interpretability and Local Cointerpretability.** We first repeat the definitions of weak interpretability and of local cointerpretability.

- $A \Rsh B$ iff, for some translation τ, we have that $A + B^\tau$ is consistent.
- $A \blacktriangleright_{\mathsf{loc}} B$ iff, for all ϕ, there is a τ such that, if $B \vdash \phi^\tau$, then $A \vdash \phi$.

Theorem 3.4. *$A \Rsh B$ iff $\exists C \supseteq A\ \ C \rhd B$.*

Proof. Suppose $A \Rsh B$. Then for some τ, the theory $A + B^\tau$ is consistent. Hence $A \subseteq C := A + B^\tau$ and $C \rhd B$.

Conversely, suppose $A \subseteq C$ and $C \rhd B$. Let τ witness the interpretability of B in C. Then certainly $A + B^\tau$ is consistent. ❑

We note that the formula $\exists C \supseteq A\ \ C \rhd B$ is in $\mathcal{F}$ and hence weak interpretability is a fine relation. Our next order of business is to show that local cointerpretability can also be defined in the fragment $\mathcal{F}$. Local cointerpretability is preservation of the tolerated. Thus, local cointerpretability is a fine relation.

Theorem 3.5. *$A \blacktriangleright_{\mathsf{loc}} B$ iff, for all C, if $A \Rsh C$, then $B \Rsh C$.*

Proof. Suppose $A \blacktriangleright_{\mathsf{loc}} B$. Consider any C and suppose $A \Rsh C$. Then, for some σ, we have $A \nvdash (\neg C)^\sigma$. Taking $\phi := \neg C^\sigma$ in the definition of $A \blacktriangleright_{\mathsf{loc}} B$, we find, for some τ, that $B \nvdash (\neg C)^{\sigma\tau}$. It follows that $B \Rsh C$.

Suppose that, for all C, if $A \Rsh C$, then $B \Rsh C$. Consider any ϕ. We take $C := (\mathcal{I}_{\Sigma_A} \wedge \neg\phi)$. Suppose $A \nvdash \phi$, then $A \nvdash (\mathcal{I}_{\Sigma_A} \to \phi)$, since $A \vdash I_{\Sigma_A}$. Hence, $A \Rsh C$. It follows that $B \Rsh C$, so, for some τ, we have $B \nvdash (\mathcal{I}_{\Sigma_A} \to \phi)^\tau$, and, *a fortiori*, $B \nvdash \phi^\tau$. ❑

We can write Theorem 3.5 a bit differently. If we define $[\![A]\!] := \{C \mid A \Rsh C\}$. Then, $A \blacktriangleright_{\mathsf{loc}} B$ iff $[\![A]\!] \subseteq [\![B]\!]$.

Here is another characterisation in $\mathcal{F}$ that is easy to remember.

Theorem 3.6. *$A \blacktriangleright_{\mathsf{loc}} B$ iff, for all $A' \supseteq A$, there is a $B' \supseteq B$, such that $B' \rhd A'$.*

Proof. We use Theorem 3.5. Suppose that, for all C, if $A \Rsh C$, then $B \Rsh C$. Suppose $A' \supseteq A$. Then, the identity interpretation witnesses that $A \Rsh A'$. It follows that $B \Rsh A'$. Hence, for some τ, we have $B' := B + (A')^\tau$ is consistent. Clearly $B' \supseteq B$ and $B' \rhd A'$.

Conversely suppose that, for all $A' \supseteq A$, there is a $B' \supseteq B$, such that $B' \rhd A'$. Suppose ρ witnesses that $A \Rsh C$. Then $A' := A + C^\rho$ is consistent. It follows that there is a $B' \supseteq B$ such that $B' \rhd A'$. Suppose $B' \rhd A'$ is witnessed by σ. We may conclude that $B + C^{\rho\sigma}$ is consistent, so $B \Rsh C$. ❑

We prove a number of further equivalents of cointerpretability.

Theorem 3.7. *The following are equivalent:*

1. $A \blacktriangleright_{\mathsf{loc}} B$.
2. *For all* Θ, *we have:* $\Lambda^\Theta_B \subseteq \Lambda^\Theta_A$.
3. $\Lambda^{\Sigma_A}_B \subseteq \Lambda^{\Sigma_A}_A$.
4. $\Lambda^{\Sigma_A}_B \subseteq A$.

Proof. We prove: (1) ⇒ (2). Suppose $A \blacktriangleright_{\mathsf{loc}} B$. Let Θ be any signature. Consider any $\phi \in \Lambda^\Theta_B$. So, for all $\sigma : \Theta \to \Sigma_B$, $B \vdash (\mathcal{I}_\Theta \to \phi)^\sigma$. Consider any $\rho : \Theta \to \Sigma_A$. Let τ be the witness of $A \blacktriangleright_{\mathsf{loc}} B$ for $(\mathcal{I}_\Theta \to \phi)^\rho$. We have $(\tau \circ \rho) : \Theta \to \Sigma_B$ and hence $B \vdash (\mathcal{I}_\Theta \to \phi)^{\rho\tau}$. It follows that $A \vdash (\mathcal{I}_\Theta \to \phi)^\rho$. Thus, ϕ in Λ^Θ_A.

We get (2) ⇒ (3) by universal instantiation.

We get (3) to (4), since, trivially $\Lambda^{\Sigma_A}_A \subseteq A$.

We prove (4) ⇒ (1). Suppose $\Lambda^{\Sigma_A}_A \subseteq A$. We note that $A \blacktriangleright_{\mathsf{loc}} B$ iff, for all ϕ, if (for all $\tau : \Sigma_B \to \Sigma_A$, we have $B \vdash \phi^\tau$), then $A \vdash \phi$.

Suppose, for all $\tau : \Sigma_B \to \Sigma_A$, we have $B \vdash \phi^\tau$. Then, *a fortiori*, for all $\tau : \Sigma_B \to \Sigma_A$, we have $B \vdash \mathcal{I}_{\Sigma_B} \to \phi^\tau$. Thus, $\phi \in \Lambda^{\Sigma_A}_B$. Hence, $A \vdash \phi$. ❑

Remark 3.8. We consider the following to notions.

- $A \blacktriangleright_{\mathsf{mod}} B$ iff, there is a τ such that, for every model $\mathcal{M}$ of A, there is a model $\mathcal{N}$ of B, such that $\mathcal{M}$ is elementarily equivalent to $\widetilde{\tau}(\mathcal{N})$.
- $A \blacktriangleright_{\mathsf{lomo}} B$ iff, for every model $\mathcal{M}$ of A, there is a τ and there is a model $\mathcal{N}$ of B, such that $\mathcal{M}$ is elementarily equivalent to $\widetilde{\tau}(\mathcal{N})$.

It is easily seen that $\blacktriangleright_{\mathsf{mod}}$ coincides with $\blacktriangleright$. Also, if $A \blacktriangleright_{\mathsf{lomo}} B$, then $A \blacktriangleright_{\mathsf{loc}} B$. The converse direction, however, fails. We have, for example, $\mathsf{EQ} \blacktriangleright_{\mathsf{loc}} \mathbb{1}$. On the other hand, $\mathsf{EQ} \not\blacktriangleright_{\mathsf{lomo}} \mathbb{1}$.

Finally, note that, if we replace elementary equivalence by isomorphism in the above definitions we obtain two further notions that are *prima facie* different. ❑

3.3. **(Locally) Faithful Interpretability.** We remind the reader that:

- $A \rhd_{\mathsf{lofa}} B$ iff,
  for all ϕ such that $B \nvdash \phi$, there is a τ, such that $A \vdash B^\tau$ and $A \nvdash \phi^\tau$.

We start with an immediate observation.

Theorem 3.9. $A \rhd_{\mathsf{lofa}} B$ *iff, for all* ϕ, *if* $B \mathrel{\sim_A} \phi$, *then* $B \vdash \phi$.

Proof. We have:

$$\begin{aligned} A \rhd_{\mathsf{lofa}} B &\Leftrightarrow \forall \phi\, (B \nvdash \phi \Rightarrow \exists \tau\, (A \vdash B^\tau \text{ and } A \nvdash \phi^\tau)) \\ &\Leftrightarrow \forall \phi\, (\forall \tau\, (A \vdash B^\tau \Rightarrow A \vdash \phi^\tau) \Rightarrow B \vdash \phi) \\ &\Leftrightarrow \forall \phi\, (B \mathrel{\sim_A} \phi \Rightarrow B \vdash \phi) \end{aligned}$$ ❑

The following theorem shows that $A \rhd_{\mathsf{lofa}} B$ is definable in $\mathcal{F}$.

Theorem 3.10. $A \rhd_{\mathsf{lofa}} B$ *iff* ($A \rhd B$ *and* $B \blacktriangleright_{\mathsf{loc}} A$).

Proof. We use the characterisation of Theorem 3.6.

Suppose $A \rhd_{\mathsf{lofa}} B$. By putting $\phi := \bot$ in the definition of $\rhd_{\mathsf{lofa}}$, we see that $A \rhd B$.

Consider any $B' \supseteq B$. Suppose B' is axiomatized by ϕ. Since B' is consistent, it follows that $B \nvdash \neg\,\phi$. So there is a τ such that $A \vdash B^\tau$ and $A \nvdash (\neg\,\phi)^\tau$. We may conclude that $A \rhd B$ and that $A' := A + \phi^\tau$ is consistent. Clearly, $A' \rhd B'$. We may conclude $B \blacktriangleright_{\mathsf{loc}} A$.

Suppose $A \rhd B$ and $B \blacktriangleright_{\mathsf{loc}} A$. Let τ_0 witness $A \rhd B$. Suppose $B \nvdash \phi$. If follows that $B' := B + \neg\,\phi$ is consistent. Then, for some $A' \supseteq A$ we have $A' \rhd B'$. Let τ be the witness of $A' \rhd B'$. We take $\tau^\star := \tau\langle (B')^\tau \rangle \tau_0$. Clearly, $A \vdash B^{\tau^\star}$. Suppose $A \vdash \phi^{\tau^\star}$. Then, $(A + (B')^\tau) \vdash \phi^\tau$, since under the assumption $(B')^\tau$, the translations τ and $\tau^\star$ coincide. On the other hand, $(A + (B')^\tau) \vdash \neg\,\phi^\tau$, so $(A + (B')^\tau) \vdash \bot$. But $(A + (B')^\tau) \subseteq A'$ and A' is consistent. So we have a contradiction. Ergo, $A \nvdash \phi^{\tau^\star}$. ❑

We note that Theorem 3.10 in combination with the various characterisations of $\blacktriangleright_{\mathsf{loc}}$ gives us many alternate characterisations for $\rhd_{\mathsf{lofa}}$.

The nice decoupling into interpretability and cointerpretability also works in the non-local case —at least as long as the theories we consider are finite.

Theorem 3.11. *We have $A \rhd_{\mathsf{faith}} B$ iff ($A \rhd B$ and $B \blacktriangleright A$).*

Proof. From-left-to-right is easy. We treat from-right-to-left. Suppose τ_0 witnesses $A \rhd B$ and τ_1 witnesses $B \blacktriangleright A$. Let $\tau^\star := \tau_1 \langle B^{\tau_1} \rangle \tau_0$. Clearly $A \vdash B^{\tau^\star}$. Suppose $A \vdash \phi^{\tau^\star}$. Then $A + B^{\tau_1} \vdash \phi^{\tau_1}$. It follows that $B \vdash (B \to \phi)$, so $B \vdash \phi$. ❑

Remark 3.12. We pick up the thread of Remark 3.8. We define $A \rhd_{\mathsf{lomo}} B$ iff $A \rhd B$ and $B \blacktriangleright_{\mathsf{lomo}} A$. We note that $\lhd_{\mathsf{faith}}$ is a sub-relation of $\lhd_{\mathsf{lomo}}$ and $\lhd_{\mathsf{lomo}}$ is a sub-relation of $\lhd_{\mathsf{lofa}}$.

On the other hand, we have $\mathsf{EQ} \lhd_{\mathsf{lofa}} \mathbb{1}$, but $\mathsf{EQ} \not\lhd_{\mathsf{lomo}} \mathbb{1}$. Let C be the theory of linear dense orderings without endpoints. It is easy to see that $\mathsf{EQ} \lhd_{\mathsf{lomo}} C$. On the other hand, since C is a complete theory, $\mathsf{EQ} \not\lhd_{\mathsf{faith}} C$.[7] So both inclusions are strict.

The notion of faithfulness one obtains by combining interpretability with model theoretic cointerpretability with isomorphism was studied by Leslaw Szczerba in [Szc76]. ❑

3.4. **Local Tolerance.** We remind the reader of two definitions.

- A is locally tolerant iff $\forall B\; A \Rsh B$.
- A is essentially locally tolerant iff $\forall B \supseteq A\; \forall C\; B \Rsh C$.

We give two important characterisations of local tolerance.

Theorem 3.13. *The following are equivalent:*

i. A is locally tolerant.
ii. For all B, if $A \rhd B$, then $A \rhd_{\mathsf{lofa}} B$.

---

[7]This insight is trivial since we do not work with parameters: parameter-free faithful interpretability preserves completeness from the interpreting theory to the interpreted theory. The insight still holds when we do allow parameters. In this case one uses that, in the unique countable model of C (modulo isomorphism), the parameters only have finitely many constellations modulo automorphism.

iii. $A \rhd_{\mathsf{lofa}} \mathsf{CQC}_2$*, where* CQC_2 *is predicate logic for a binary relation symbol.*

These characterisations are the local versions of well-known characterisations of tolerance. See e.g. [Vis05]. In Section 6, we will add a further characterisation.

Proof. (i) $\Rightarrow$ (ii). Suppose A is locally tolerant and $A \rhd B$. Let τ_0 witness $A \rhd B$. Suppose $B \nvdash \phi$. Let $B' := B + \neg\,\phi$. We have $A \Rsh B'$. Let τ_1 be the witness of $A \Rsh B'$. We define $\tau^\star := \tau_1\langle (B')^{\tau_1}\rangle\tau_0$. Clearly, $A \vdash B^{\tau^\star}$. Suppose $A \vdash \phi^{\tau^\star}$. In that case, we must have $A + (B')^{\tau_1} \vdash \bot$. Quod non.

(ii) $\Rightarrow$ (iii). This is immediate from the fact that $A \rhd \mathsf{CQC}_2$.

(iii) $\Rightarrow$ (i). We can do this in two ways. First, we can use a theorem from [Hod93] that any theory B is bi-interpretable with a theory $\widetilde{B}$ in the signature of one binary relation. Since $\mathsf{CQC}^2 \nvdash \neg\widetilde{B}$, it follows that $A + (\widetilde{B})^\tau$ is consistent, for some τ. So, $A \Rsh \widetilde{B}$, and, hence, $A \Rsh B$.

Alternatively, we can use the fact that we have an interpretation of S^1_2 in adjunctive set theory AS, which is a theory in one binary relation. Say this is witnessed by ν. We want a Σ_1-sound translation here: say, when we interpret AS in the hereditarily finite sets, then ν gives us an isomorphic copy of the natural numbers. Consider any theory B. We easily show that $A \Rsh (\mathsf{AS} + (\mathsf{S}^1_2 + \mathsf{con}(B))^\nu)$ —noting that the theory that is tolerated is true in the hereditarily finite sets. Since, $(\mathsf{AS} + (\mathsf{S}^1_2 + \mathsf{con}(B))^\nu) \rhd (\mathsf{S}^1_2 + \mathsf{con}(B))$. and $(\mathsf{S}^1_2 + \mathsf{con}(B)) \rhd B$, it follows that $A \Rsh B$. ❑

Remark 3.14. The above result matches precisely with what we know about tolerance. A theory A is *tolerant* iff, for all infinite theories X, we have $A \Rsh X$. Alternatively, A is tolerant iff it tolerates S^1_2 plus all true Π_1-sentences. We have: The following are equivalent:

i. A is tolerant.
ii. For all B, if $A \rhd B$, then $A \rhd_{\mathsf{faith}} B$.
iii. $A \rhd_{\mathsf{faith}} \mathsf{CQC}_2$.

The reader is referred to [Vis05] for explanation and proofs. ❑

Finally, we remind the reader that $A \blacktriangleleft_{\mathsf{loc}} B$ iff $[\![A]\!] \supseteq [\![B]\!]$, where we define $[\![C]\!] := \{D \mid C \Rsh D\}$. Hence, a theory is locally tolerant iff it is minimal w.r.t. $\blacktriangleleft_{\mathsf{loc}}$. Thus, we have:

Theorem 3.15. *A is locally tolerant iff, for all B, we have $A \blacktriangleleft_{\mathsf{loc}} B$.*

In Subsection 5.2, we will see a characterisation of the maximal elements of $\blacktriangleleft_{\mathsf{loc}}$: these are precisely the theories with the finite model property.

4. Disjoint Sum is a fine Operation

The disjoint sum of two theories is more or less what you expect it to be: make the signatures disjoint, relativise both theories to newly introduced domains and take the union. The disjoint sum is introduced and discussed in Appendix A.8. In this section, we show that $\boxplus$ preserves $\approx$.

Theorem 4.1. *$\boxplus$ is a fine operation.*

Proof. Suppose $A_0 \approx B_0$ and $A_1 \approx B_1$. We want to show that $(A_0 \boxplus A_1) \approx (B_0 \boxplus B_1)$.

We define $C \mathrel{\mathcal{B}} D$ iff, for some $\phi_{00}, \cdots, \phi_{0(k-1)}$ in the language of A_0, and some $\phi_{10}, \cdots, \phi_{1(k-1)}$ in the language of A_1, and some $\psi_{00}, \cdots, \psi_{0(k-1)}$ in the language of B_0, and some $\psi_{10}, \cdots, \psi_{1(k-1)}$ in the language of B_1, we have $C = ((A_0 \boxplus A_1) + \bigvee_{i<k}(\phi_{0i} \wedge \phi_{1i}))$ and $D = ((B_0 \boxplus B_1) + \bigvee_{i<k}(\psi_{0i} \wedge \psi_{1i}))$ and, for all $j < 2$ and $i < k$, we have $(A_j + \phi_{ji}) \approx (B_j + \psi_{ji})$.

We first show that $(A_0 \approx B_0) \mathrel{\mathcal{B}} (A_1 \approx B_1)$. This is immediate from $A_0 \approx B_0$ and $A_1 \approx B_1$, by taking $k := 1$ and $\phi_{00} := \top$, $\phi_{10} := \top$, $\psi_{00} := \top$ and $\psi_{10} := \top$.

We check that $C \mathrel{\mathcal{B}} D$ implies $C \equiv D$. Suppose $C \mathrel{\mathcal{B}} D$. We have witnessing ϕ_{ji} and ψ_{ji}. We have $(A_j + \phi_{ji}) \approx (B_j + \psi_{ji})$, and, hence, $(A_j + \phi_{ji}) \equiv (B_j + \psi_{ji})$. Suppose τ_{ji} witnesses $(A_j + \phi_{ji}) \rhd (B_j + \psi_{ji})$. Then $\tau_i^\star := \tau_{0i} \boxplus \tau_{1i}$ witnesses

$$\begin{aligned} ((A_0 \boxplus A_1) + (\phi_{0i} \wedge \phi_{1i})) &= ((A_0 + \phi_{0i}) \boxplus (A_1 + \phi_{1i})) \\ &\rhd ((B_0 + \psi_{0i}) \boxplus (B_1 + \phi_{1i})) \\ &= ((B_0 \boxplus B_1) + (\psi_{0i} \wedge \psi_{1i})) \end{aligned}$$

Let:

$$\nu := \tau_0^\star \langle \phi_{00} \wedge \phi_{10} \rangle (\tau_1^\star \langle \phi_{01} \wedge \phi_{11} \rangle (\cdots \langle \phi_{0(k-2)} \wedge \phi_{1(k-2)} \rangle \tau_{k-1}^\star) \cdots).$$

Then ν witnesses:

$$C = ((A_0 \boxplus A_1) + \bigvee_{i<k} (\phi_{0i} \wedge \phi_{1i})) \rhd ((B_0 \boxplus B_1) + \bigvee_{i<k} (\psi_{0i} \wedge \psi_{1i})) = D.$$

Similarly, in the other direction.

We check the forward property. The backward property is, of course, analogous. Suppose $C \mathrel{\mathcal{B}} D$ and let, as before ϕ_{ji} and ψ_{ji} be the witnessing formulas. Suppose $C' \supseteq C$. Since $C' \supseteq (A_0 \boxplus A_1)$, we have by Theorem A.6, that $C = (A_0 \boxplus A_1) + \bigvee_{\ell<n}(\chi_{0\ell} \wedge \chi_{1\ell})$, where the $\chi_{0\ell}$ are in the A_0-language and the $\chi_{1\ell}$ are in the A_1-language. Since, $\bigvee_{i<k}(\phi_{0i} \wedge \phi_{1i})$ is also in C, we have by propositional logic:

$$C' = (A_0 \boxplus A_1) + \bigvee_{i<k, \ell<n} ((\phi_{0i} \wedge \chi_{0\ell}) \wedge (\phi_{1i} \wedge \chi_{1\ell})).$$

By our assumption, we have $(A_j + \phi_{ji}) \approx (B_j + \psi_{ji})$. Moreover, we have $(A_j + \phi_{ji}) \subseteq (A_j + (\phi_{ji} \wedge \chi_{j\ell}))$. Hence, by the forward property for $\approx$, there is a $\nu_{1i\ell}$ such that:

$$(B_j + \psi_{ji}) \subseteq (B_j + \nu_{ji\ell}) \text{ and } (A_j + (\phi_{ji} \wedge \chi_{j\ell})) \approx (B_j + \nu_{1i\ell}).$$

Thus, we can take $D' := (B_0 \boxplus B_1) + \bigvee_{i<k,\ell<n}(\nu_{0i\ell} \wedge \nu_{1i\ell})$. Clearly, we have $C' \mathrel{\mathcal{B}} D'$. ❑

5. Faithful Interpretability and Locally Faithful Interpretability

Locally faithful interpretability looks to me like a fairly natural notion. Still I have not seen it formulated before. In the present section, we provide a few basic insights concerning the new notion. We give a separating example to show that locally faithful interpretability is different from faithful interpretability. Finally, we briefly indicate the connection of faithful interpretability and locally faithful interpretability with the forward property.

5.1. **Decidability.** The relation $\rhd_{\mathsf{lofa}}$ preserves decidability, i.e., if $A \rhd_{\mathsf{lofa}} B$ and A is decidable, then B is decidable. Suppose $A \rhd_{\mathsf{lofa}} B$ and A is decidable. Is ϕ a theorem of B? On the positive side we enumerate the theorems of B until we find a proof of ϕ; on the negative side we run through translations $\tau_i : \Sigma_B \to \Sigma_A$ and decide whether $A \vdash B^{\tau_i}$ and $A \nvdash \phi^{\tau_i}$. Eventually, one of the parallel processes must yield an answer. We note that the argument also works when we allow parameters.

Unfortunately, our argument fails when we switch to infinitely axiomatised theories. So, in the infinite case, ordinary faithful interpretability seems to be in better shape.

Open Question 5.1. Is there an example of $U \rhd_{\mathsf{lofa}} V$ (or even $A \rhd_{\mathsf{lofa}} V$), where U is decidable and V is not? ❏

We note that it follows that decidability is a fine property.

Open Question 5.2. Is decidability first-order or even second-order definable in $\mathbb{E}\mathbb{I}$? ❏

5.2. **The Finite Model Property.** A basic property of theories is the finite model property or FMP. A theory A has the finite model property iff, for every ϕ with $A \nvdash \phi$, there is a finite model $\mathcal{M}$ such that $\mathcal{M} \models A$ and $\mathcal{M} \models \neg\,\phi$.

There are many theories with the finite model property that also have an infinite model. The theory of pure identity EQ or CQC_0 is one example of these. Another example is the theory of discrete linear orderings with endpoints.

Theorem 5.3. *The theory of discrete linear orderings with endpoints has the finite model property.*

The proof is given in Appendix C. For a similar result see [VS00].

Theorem 5.4. *The following are equivalent.*

i. *A has the finite model property.*
ii. $\mathbb{1} \approx A$.
iii. $\mathbb{1} \equiv_{\mathsf{lofa}} A$.
iv. *$\mathbb{1}$ is mutually locally cointerpretable with A.*

Proof. We need to simple observations. First, any A with the finite model property is mutually interpretable with $\mathbb{1}$. Any theory interprets $\mathbb{1}$. Suppose $\mathcal{M}$ is a finite model of A. Since, we allow piecewise interpretations, we can transform our finite model into an interpretation of A in $\mathbb{1}$.

Secondly, the finite model property is preserved over $\subseteq$. Suppose A has the finite model property. Consider any $B \supseteq A$ and suppose $B \nvdash \phi$. Then, $A \nvdash B \to \phi$. So there is a finite model $\mathcal{M}$ of A such that $\mathcal{M} \models \neg\,(B \to C)$, so $\mathcal{M} \models B$ and $\mathcal{M} \models \neg\,\phi$. We may conclude that B has the finite model property.

(i) $\Rightarrow$ (ii). We define $D\, \mathcal{B}\, E$ iff $D = \mathbb{1}$ and $E \supseteq A$. By the above observations, $\mathcal{B}$ is a i-bisimulation. Hence $\mathbb{1} \approx A$.

(ii) $\Rightarrow$ (iii). Since $\rhd_{\mathsf{lofa}}$ is a fine relation and $\rhd_{\mathsf{lofa}}$ is reflexive, it is immediate, from $\mathbb{1} \approx A$, that $\mathbb{1} \equiv_{\mathsf{lofa}} A$.

(iii) $\Rightarrow$ (iv). From $A \equiv_{\mathsf{lofa}} \mathbb{1}$, we have that A and $\mathbb{1}$ are mutually locally cointerpretable.

(iv) $\Rightarrow$ (i). Suppose $A \blacktriangleright_{\mathsf{loc}} \mathbb{1}$. Suppose $A \nvdash \phi$. Then $A \Rsh (A + \neg\phi)$. It follows that $\mathbb{1} \Rsh (A + \neg\,\phi)$. Hence $\mathbb{1} + (A + \neg\,\phi)^\tau$ is consistent, for some τ. Since, $\mathbb{1}$ is complete it follows that $\mathbb{1} \vdash (A + \neg\,\phi)^\tau$. Thus, τ defines a finite internal model of $A + \neg\,\phi$ in the unique model of $\mathbb{1}$. Thus, A has the finite model property. ❑

It follows that the finite model property is a fine, $\mathcal{F}$-definable, property.

Example 5.5. Let $A := (\forall x\,\forall y\, x = y \vee \bigwedge \mathsf{S}^1_2)$. We have $\mathbb{1} \rhd A$, but not $\mathbb{1} \rhd_{\mathsf{lofa}} A$, since A is not decidable. This gives us a first separating example between $\rhd_{\mathsf{lofa}}$ and $\rhd$. ❑

Example 5.6. The theory $\mathbb{1}$ is complete and the theory EQ is not complete and $\mathbb{1} \approx \mathsf{EQ}$. Ergo, completeness is not a fine property. On the other hand, completeness is preserved by sentential congruence and, *a fortiori*, by more refined notions of sameness like bi-interpretability. Thus, the example of completeness shows us the limitations of our framework: completeness is good but not fine. On the other hand, the defect also comes with an advantage: the nice characterisation of the finite model property. ❑

We end this subsection with the observation that the theories with the finite model property are precisely the theories that are maximal with respect to $\blacktriangleleft_{\mathsf{loc}}$. So:

Theorem 5.7. *A has the finite model property iff, for all B, we have* $A \blacktriangleright_{\mathsf{loc}} B$.

5.3. **Separating Examples.** In this section, we provide separating examples between a number of salient notions

Example 5.8. *i-Bisimilarity does not imply mutual faithful interpretability.* We have seen that $\mathbb{1} \approx \mathsf{EQ}$. However, it is impossible that $\mathbb{1} \rhd_{\mathsf{faith}} \mathsf{EQ}$, since for any τ such that $\mathbb{1} \vdash \mathcal{I}^\tau_0$, necessarily there is an n such that $\mathbb{1}$ proves the τ-translation of *there are at most n elements*. So, faithful interpretability is not fine. ❑

Open Question 5.9. Is faithful interpretability first-order or even second-order definable over $\mathbb{E}\mathbb{I}$? ❑

Remark 5.10. Picking up the thread of Remark 2.2, we note that Example 5.8 does not work for $\approx^+$. So, one might hope that $\approx^+$ does imply mutual faithful interpretability. In Appendix C, we show, however, that the theory of discrete linear orderings with endpoints does bear the relation $\approx^+$ to EQ, but EQ does not faithfully interpret the theory of discrete linear orderings with endpoints. ❑

Example 5.11. *Mutual faithful interpretability does not imply i-Bisimilarity.* Let A be sequential. It is easy to see that the theories A and $A \boxplus A$ are mutually faithfully interpretable and, *a fortiori*, mutually locally faithfully interpretable. By Theorem 6.1, there are $A_0 \supseteq A$ and $A_1 \supseteq A$ such that $A_0 \not\rhd A_1$ and $A_1 \not\rhd A_0$. We have: $(A \boxplus A) \subseteq (A_0 \boxplus A_1)$. Suppose there were an $A' \supseteq A$ with $A' \equiv (A_0 \boxplus A_1)$. Then, since A' is sequential and, hence, connected,[8] it follows that $A_i \rhd A'$, for some i. But this would give us: $A_i \rhd A' \rhd (A_0 \boxplus A_1) \rhd A_{1-i}$. *Quod non.* We may conclude that $A \not\approx (A \boxplus A)$. ❑

We show that mutual faithful interpretability is incomparable to $\approx$.

Example 5.12. *Mutual faithful interpretability is not mutual locally faithful interpretability and mutual locally faithful interpretability is not i-bisimilarity.* We note that both $\equiv_{\mathsf{faith}}$ and $\approx$ are subrelations of $\equiv_{\mathsf{lofa}}$ Hence, Example 5.8 also separates $\equiv_{\mathsf{lofa}}$ and $\equiv_{\mathsf{faith}}$. Example 5.11 separates $\equiv_{\mathsf{lofa}}$ and $\approx$. ❑

[8]See Section 7.

Example 5.13. *Interpretability does not imply locally faithful interpretability.* Let A be the theory of dense linear orderings without end-points and let B be $\mathsf{CQC}_{\mathcal{A}}$, i.e. predicate logic for the signature of arithmetic. Clearly, we have $A \rhd B$. On the other hand, for any interpretation τ such that $A \vdash B^\tau$, we have $A \vdash (\neg \bigwedge \mathsf{Q})^\tau$, where Q is Robinson's Arithmetic, since A is decidable and Q is essentially undecidable. Thus, $A \not\rhd_{\mathsf{lofa}} B$. ❑

Example 5.14. *Local faithful interpretability does not imply faithful interpretability.* We already showed that $\mathbb{1} \rhd_{\mathsf{lofa}} \mathsf{EQ}$ but $\mathbb{1} \not\rhd_{\mathsf{faith}} \mathsf{EQ}$.

It might be thought that our example leans on very specific features of piecewise or many-dimensional interpretations. However, we can improve our example in order to get the witnessing interpretations for $A \rhd_{\mathsf{lofa}} B$ one-dimensional and without pieces. We take A the theory of a linear discrete ordering with an initial point, i.e., the theory of the ordering of the natural numbers. This is a complete theory. We take B the theory of discrete linear order with endpoints. Clearly, $A \rhd_{\mathsf{lofa}} B$ using only one-dimensional interpretations without parameters. However, since A is complete, there can be no faithful interpretation of A in B.[9] ❑

5.4. **The Forward Property.** Consider the following characterisation.

- $A \lhd_{\mathsf{lofa}} B$ iff $A \lhd B$ and, for all $A' \supseteq A$, there is a $B' \supseteq B$, such that $A' \lhd B'$.

We note that this characterisation is reminiscent of the forward property. It is so to speak the first step towards the forward property. It turns out that faithful interpretability does have the forward property.

Theorem 5.15. *The relation $\lhd_{\mathsf{faith}}$ has the forward property.*

Proof. Suppose $K : A \lhd_{\mathsf{faith}} B$ and $A \subseteq A'$. We claim that $A' \lhd_{\mathsf{faith}} (B + (A')^{\tau_K})$. Clearly, $K' : A' \lhd (B + (A')^{\tau_K})$, where K' is based on τ_K. Suppose $B + (A')^{\tau_K} \vdash \phi^{\tau_K}$. Then by faithfulness: $A' = A + A' \vdash \phi$. Hence, K' is faithful. We note that it also follows that $B + (A')^{\tau_K}$ is consistent. So we can take $B' := (B + (A')^{\tau_K})$. ❑

We define the obvious analogues of $\equiv_n$ and $\approx$ for the forward property.

- A relation $\mathcal{S}$ is an *i-simulation* iff (i) $\mathcal{S}$ is a subrelation of $\lhd$ and (ii) $\mathcal{S}$ has the forward or zig property: if $A \mathrel{\mathcal{S}} B$ and $A' \supseteq A$, then, there is a $B' \supseteq B$, such that $A \mathrel{\mathcal{S}} B$.
- B simulates A, or A is simulated by B, or $A \precsim B$ iff, there is a simulation $\mathcal{S}$ such that $A \mathrel{\mathcal{S}} B$.
- $A \cong B$ iff $A \precsim B$ and $A \succsim B$.
- $A \lhd_0^\circ B$ iff $A \lhd B$, and $A \lhd_{n+1}^\circ B$ iff $A \lhd B$ and, for all $A' \supseteq A$, there is a $B' \supseteq B$, such that $A' \lhd_n^\circ B'$.
- $A \lhd_\omega^\circ B$ iff, for all n, we have $A \lhd_n^\circ B$.
- $A \equiv_n^\circ B$ iff $A \lhd_n^\circ B$ and $A \rhd_n^\circ B$.

Trivially all these notions are fine.

We note that $\approx$ is a subrelation of $\cong$. In Section 7, we will provide an example that $\cong$ and $\approx$ do not coincide.

[9] We note that this argument does not work when we allow parameters. If we allow parameters a possible example is that the successor part of Q locally faithfully interprets the theory of a discrete ordering with endpoints but that the successor part of Q does not faithfully interpret the theory of a discrete ordering with endpoints.

Every $\lhd^\circ_{n+1}$ is a subrelation of $\lhd^\circ_n$. Also $\lhd^\circ_1$ is $\lhd_{\mathsf{lofa}}$. The relation $\precsim$ is a subrelation of $\lhd^\circ_\omega$.

Open Question 5.16. 1. Is the descending $\lhd^\circ_n$ hierarchy strict?
2. Is $\lhd^\circ_\omega$ strictly contained in $\precsim$? ❑

We note that $\mathsf{EQ} \precsim \mathbb{1}$, but not $\mathsf{EQ} \lhd_{\mathsf{faith}} \mathbb{1}$. Hence, $\lhd_{\mathsf{faith}}$ is strictly contained in $\precsim$.

In Appendix A.7.6 we show that faithful retractions in the categories INT_i, for $i \leq 3$, have the forward property. So we could define $A \sqsubseteq B$ as the maximal simulation that is a subrelation of $\equiv$. This relation would contain the faithful retractions of INT_3 (and, *a fortiori*, the faithful retractions of INT_i for $i \leq 3$). We note that the induced equivalence relation of $\sqsubseteq$ is precisely $\approx$.

6. Arithmetic

In the present section, we collect a number of characterisations and results connected with the arithmetical theory S^1_2.

6.1. **Incomparable Theories.**

We need a sufficient store of incomparable extensions of given finitely axiomatised theories. The following theorem provides these.

Theorem 6.1. *Suppose A and B both tolerate S^1_2. Then, there are $A^\star \supseteq A$ and $B^\star \supseteq B$, that are incomparable w.r.t. $\lhd$, i.e., $A^\star \not\rhd B^\star$ and $B^\star \not\rhd A^\star$.*

Proof. Suppose τ witnesses that A tolerates S^1_2 and ν witnesses that B tolerates S^1_2. We take $A' := A + (\mathsf{S}^1_2)^\tau$ and $B' := B + (\mathsf{S}^1_2)^\nu$. By the Gödel Fixed Point Lemma, we find R such that:

$$\mathsf{EA} \vdash R \leftrightarrow ((A' + R^\tau) \rhd (B' + \neg R^\nu)) \leq ((B' + \neg R^\nu) \rhd (A' + R^\tau)).$$

We take $A^\star := A' + R^\tau$ and $B^\star := B' + \neg R^\nu$. Suppose $A^\star \rhd B^\star$. It follows that R or $R^\perp$. In case we have R, we find, by Σ_1-completeness, that $A' \rhd \bot$. Quod non. If we have $R^\perp$, it follows by the fixed point equation that $(B' + \neg R^\tau) \rhd (A' + R^\nu)$. By Σ_1-completeness, we have $B' \rhd \bot$. Quod non. We may conclude that $A^\star \not\rhd B^\star$.

The proof that $B^\star \not\rhd A^\star$ is similar. ❑

6.2. **Characterisations.**

We can connect our previous characterisations to arithmetical ones using two basic insights:

I. $(\mathsf{S}^1_2 + \mathsf{con}_n(A)) \rhd A$, where $n \geq \rho(A)$.
II. If A is sequential, then $A \rhd (\mathsf{S}^1_2 + \mathsf{con}_n(A))$, where $n \geq \rho(A)$.

In (I) $\mathsf{con}_n(A)$ refers to consistency for n-provability. We only allow n-proofs, i.e. proofs involving formulas of complexity $\leq n$. Our complexity measure is *depth of quantifier alternations*. The translation that realises the interpretation is the Henkin interpretation η.

We have the following basic insight.

Theorem 6.2. *Suppose A is sequential. Then, $A \Rsh B$ iff $A \Rsh (\mathsf{S}^1_2 + \mathsf{con}_{\rho(B)}(B))$.*

Proof. Suppose A is sequential and $A \Rsh B$. Then $A + B^\tau$ is consistent, for some τ. But $A + B^\tau$ is sequential, hence it interprets $\mathsf{S}^1_2 + \mathsf{con}_{\rho(B)}(B)$. We can see this by noting that for any $n \geq \rho(A + B^\tau)$, we have $(A + B^\tau) \rhd (\mathsf{S}^1_2 + \mathsf{con}_n(A + B^\tau))$. On the other hand, for sufficiently large n, we have $\mathsf{S}^1_2 \vdash \Box_{B,\rho(B)}\bot \to \Box_{A+B^\tau,n}\bot$. So it follows that $(A + B^\tau) \rhd (\mathsf{S}^1_2 + \mathsf{con}_{\rho(B)}B)$. Thus, we may conclude that $A \Rsh (\mathsf{S}^1_2 + \mathsf{con}_{\rho(B)}B)$.

The other direction is immediate by Basic Insight (I). ❑

If we are interested in local tolerance, we do not need the assumption of sequentiality.

Theorem 6.3. *A is locally tolerant iff, for all true Π_1-sentences P, we have $A \Vdash (\mathsf{S}^1_2 + P)$.*

Proof. From-left-to-right is just specialisation. From-right-to-left, we may conclude $A \Vdash B$, from the fact that $A \Vdash (\mathsf{S}^1_2 + \mathsf{con}_n(B))$. ❑

We can characterise interpretability as local Π_1-conservativity in case either the target theory or the source theory is sequential.

Theorem 6.4. *Suppose B is sequential. Then, $A \rhd B$ iff, for all Π_1-sentences P, iff $B \rhd (\mathsf{S}^1_2 + P)$, then $A \rhd (\mathsf{S}^1_2 + P)$.*

Proof. Suppose $A \rhd B$. Then, trivially, for all Π_1-sentences P, iff $B \rhd (\mathsf{S}^1_2 + P)$, then $A \rhd (\mathsf{S}^1_2 + P)$.

Conversely, suppose, for all Π_1-sentences P, iff $B \rhd (\mathsf{S}^1_2 + P)$, then $A \rhd (\mathsf{S}^1_2 + P)$.

Suppose B is sequential. Then $B \rhd (\mathsf{S}^1_2 + \mathsf{con}_{\rho(B)}(B))$ and, hence, $A \rhd (\mathsf{S}^1_2 + \mathsf{con}_{\rho(B)}(B)) \rhd B$. ❑

We remind the reader of the Friedman Characterisation.

Theorem 6.5. *Suppose A is sequential. Then,*
$A \rhd B$ iff $(\mathsf{EA} + \mathsf{con}_{\rho(A)}(A)) \supseteq (\mathsf{EA} + \mathsf{con}_{\rho(B)}(B))$.

Here EA is Elementary Arithmetic or $\mathrm{I}\Delta_0 + \mathsf{Exp}$. In the context of this paper the following characterisation is relevant.

Theorem 6.6. *Suppose A is sequential. The following are equivalent:*

a. $A \rhd B$.
b. For all $n \geq \rho(B)$, there is an $m \geq \rho(A)$, such that $(\mathsf{S}^1_2 + \mathsf{con}_m(A)) \supseteq (\mathsf{S}^1_2 + \mathsf{con}_n(B))$.
c. There is an $m \geq \rho(A)$, such that $(\mathsf{S}^1_2 + \mathsf{con}_m(A)) \supseteq (\mathsf{S}^1_2 + \mathsf{con}_{\rho(B)}(B))$.

Proof. (a) $\Rightarrow$ (b). Suppose τ witnesses $A \rhd B$. We can use τ to transform, in the context of S^1_2, an n-inconsistency proof of B into an m-inconsistency proof of A, where m is roughly $n + \rho(\tau)$. Here $\rho(\tau)$ is the maximum of $\rho(\delta_\tau)$ and the $\rho(P_\tau)$. The main point is that $\rho(\phi^\tau)$ will be $\rho(\phi) + \rho(\tau)$ plus some fixed standard overhead.

(b) $\Rightarrow$ (c). This is just specialisation.

(c) $\Rightarrow$ (a). We have: $A \rhd (\mathsf{S}^1_2 + \mathsf{con}_m(A)) \supseteq (\mathsf{S}^1_2 + \mathsf{con}_{\rho(B)}(B)) \rhd B$. ❑

6.3. **Local (In)tolerance.** We start with a question.

Open Question 6.7. Let's say that a theory is *self-confident* iff it is mutually interpretable with a sequential theory. Does every locally tolerant theory have a self-confident extension? If not, does every tolerant theory have a self-confident extension? ❑

It is easy to give examples of theories that are locally intolerant. For example, every decidable theory is locally intolerant since it does not tolerate S^1_2. It is unknown whether there is a theory that tolerates S^1_2 but still is locally intolerant.

Open Question 6.8. Is there a locally intolerant theory A with $A \upharpoonright \mathsf{S}^1_2$? ❑

In this subsection, we take a small step in thinking about this question by showing that, if a theory A is locally intolerant, then it has an extension B such that $\mathsf{S}^1_2 \vdash_{\Lambda_B} \Box_{B,\rho(B)}\bot$. So, B believes in its own restricted inconsistency in the strongest possible way.

Suppose, A is locally intolerant. This tells us that there exists a false Σ_1-sentence S, such that, for all $\tau : \mathcal{A} \to \Sigma_A$, we have $A+(\mathsf{S}^1_2)^\tau \vdash S^\tau$. In other words, $\mathsf{S}^1_2 \vdash_{\Lambda^{\mathcal{A}}_A} S$.

We show that there is a $B \supseteq A$, such that, for all $\tau : \mathcal{A} \to \Sigma_A$, we have $B + (\mathsf{S}^1_2)^\tau \vdash \Box^\tau_{B,\rho(B)}\bot$. In other words, $\mathsf{S}^1_2 \vdash_{\Lambda_B} \Box_{B,\rho(B)}\bot$.

In case, S^1_2 is not interpretable in A, we are easily done, taking $B := A$. Suppose ν witnesses $\mathsf{S}^1_2 \lhd A$. We find R such that $\mathsf{S}^1_2 \vdash R \leftrightarrow S \leq \Box_{A,m} R^\nu$. Here we take m to be $\mathsf{max}(\rho(A), \rho(R)) + 1$. We note that the complexity of R only depends on the complexity of S plus some fixed constant derived from the complexity of the provability predicate and the overhead of the fixed point construction.

Consider any $\tau : \mathcal{A} \to \Sigma_A$. We work in $\alpha := A + (\mathsf{S}^1_2)^\tau$. We allow α to be inconsistent. We take a cut I of τ, such that $\alpha \vdash \forall x \in I\; 2^{2^x} \in \delta_\tau$ and $\alpha \vdash (\mathsf{T}^1_2)^I$. We have, *ex hypothesi*, $\alpha \vdash S^I$. Hence, $\alpha \vdash (R \vee R^\perp)^I$. We have, by verifiable Σ_1-completeness, using that I is sufficiently 'deep', $\alpha \vdash R^I \to \Box^\tau_{A,m} R^\nu$ and, by the fixed-point equation, $\alpha \vdash (R^\perp)^I \to \Box^I_{A,m} R^\nu$. Hence, $A + (\mathsf{S}^1_2)^\tau \vdash \Box^\tau_{A,m} R^\nu$.

We take $B := A + \neg R^\nu$. If B would be inconsistent, we would have $A \vdash R^\nu$. Hence, by cut-elimination, $A \vdash_m R^\nu$. So, for some standard p, we may conclude that $A \vdash \underline{p}\ \mathsf{wit}\ \Box^\nu_{A,m} R^\nu$. It follows that $A \vdash \bigvee_{q \leq p}(\underline{q}\ \mathsf{wit}\ S)^\nu$. Since $\neg S$, we have $\bigwedge_{q \leq p} \neg\, q\ \mathsf{wit}\ S$, and, hence, by Σ_1-completeness, $A \vdash \bigwedge_{q\leq p}(\neg\, \underline{q}\ \mathsf{wit}\ S)^\nu$. So, A is inconsistent. *Quod non.*

Clearly, for any τ, we have $B + (\mathsf{S}^1_2)^\tau \vdash \Box^\tau_{B,\rho(B)}\bot$, noting that $\rho(B) \geq m$.

Our result shows that a sequential A must be locally tolerant. If a sequential A were locally intolerant, then there would be a sequential B that proves $\Box_{B,\rho(B)}\bot$ for all interpretations of S^1_2 in B. But we know there is an interpretation of S^1_2 on which we have $\mathsf{con}_{\rho(B)}(B)$.[10] This is a weaker result than the result of Harvey Friedman and, independently, Jan Krajíček. See [Smo85], [Kra87] and [Vis05], but the point of our exercise was not precisely reproving that earlier theorem.

7. Sequential Theories

In the present section we discuss a number of properties of sequential theories visible over our framework. Suppose A is sequential. We have the following *Basic Insights.*

a. For every B there is a sequential C with $C \rhd B$.
b. If $A \subseteq B$, then B is sequential.
c. A is connected, i.e., suppose $A \lhd B \boxplus C$, then $A \lhd B$ or $A \lhd C$.
d. If $A \equiv B$, then B is tolerant, and, *a fortiori*, locally tolerant.
e. If B is sequential and $A \lhd B$, then there is an $A' \supseteq A$ such that $A' \equiv B$.

Ad Basic Insight (a): This can be seen in two ways. We can form $C := \mathsf{seq}(B)$ as follows. We add a unary predicate $\triangle$ and a binary predicate $\in$ to the language

[10]This argument does not work for self-confident theories, where a theory C is self-confident if $C \rhd (\mathsf{S}^1_2 + \mathsf{con}_{\rho(C)}(C))$. So, one may wonder whether it can be adapted to accommodate these?

of B. We relativise B to $\triangle$ and we add AS. Alternatively, we can take $C := (\mathsf{S}^1_2 + \mathsf{con}_{\rho(B)}(B))$.

Ad Basic Insight (b): This is a triviality.

Ad Basic Insight (c): This is a non-trivial result. It was first proved by Pavel Pudlák in [Pud83]. An essentially different proof was given by Alan Stern in [Ste89]. For a discussion of the significance of this result: see [MPS90].

Ad Basic Insight (d): Tolerance simply means that for any possible infinite theory X, there is a τ such that $B + X^\tau$ is consistent. Alternatively, tolerance means that B faithfully interprets predicate logic in the language with one binary predicate.

The fact that any sequential theory is tolerant was proved by Harvey Friedman (see [Smo85]) and, independently, Jan Krajíček (see [Kra87]). The strengthening involving mutual interpretablity was noted in [Vis05].

The property of being mutually interpretable with a sequential theory can be given the following form. A theory A is *self-confident* if $A \rhd (\mathsf{S}^1_2 + \mathsf{con}_{\rho(A)}(A))$. It is easy to see that the self-conident theories are precisely the theories that are mutually interpretable with a sequential theory. (*Warning:* This last result holds only in the finitely axiomatised case.) Thus, (d) tells us that all self-confident theories are locally tolerant, and we even know that all such theories are tolerant. We note that self-confidence is definitely a fine notion. Example 5.11 will illustrate that, while sequentiality is upwards closed under $\subseteq$, self-confidence is not upwards closed under $\subseteq$.

Ad Basic Insight (e). This insight is one of the central results of [Vis14b].

Basic Insight (e) can be connected to our framework in an interesting way.

Theorem 7.1. *Suppose A and B are sequential, then $A \approx B$ iff $A \equiv B$.*

Proof. From-left-to-right, is immediate, since $\equiv$ is fine. We prove the right-to-left direction. Let $\mathcal{B}$ be $\equiv$ restricted to the sequential theories. Suppose A and B are sequential and $A \equiv B$ and $A \subseteq A'$. Then, $A' \rhd B$. Then, by (d), there is a $B' \supseteq B$ such that $A' \equiv B'$. Thus, we have the zig-property. Similarly, we have the zag-property. Trivially, $\mathcal{B}$ is a subrelation of $\equiv$. We may conclude that $\mathcal{B}$ is a i-bisimulation. ❑

Example 7.2. It is easily seen that, for any theory A, we have $A \approx (A \boxplus \mathbb{1})$. Since, $\mathbb{1} \approx \mathsf{EQ}$, it follows that $A \approx (A \boxplus \mathsf{EQ})$. It is easy to see that $A \boxplus \mathsf{EQ}$ cannot be sequential. Thus, we have, for sequential A, that $A \approx (A \boxplus \mathsf{EQ})$ and $A \boxplus \mathsf{EQ}$ is not sequential. Hence, sequentiality is not fine. ❑

Example 7.3. We note that, for sequential A, we have $A \equiv (A \boxplus A)$. It follows that $A \equiv_{\mathsf{faith}} (A \boxplus A)$.[11] By Theorem 6.1, there are A_0 and A_1 extending A such that $A_0 \not\rhd A_1$ and $A_1 \not\rhd A_0$. So, $(A_0 \boxplus A_1) \supseteq (A \boxplus A)$, Since any A' extending A is sequential by (b), no such A' can be mutually interpretable with $(A_0 \boxplus A_1)$ by (c). So, $A \not\approx (A \boxplus A)$. Thus, there are theories B that are mutually faithfully interpretable with a sequential theory A, but that are not i-bisimilar to it. ❑

Open Question 7.4. Are the sequential theories closed under sentential congruence? And, if not, are they closed under iso-congruence? We already know that the sequential theories are closed under bi-interpretability. ❑

[11] This is immediate by the Friedman-Krajíček result (see [Smo85], [Kra87], [Vis05]).

We have the following theorem.

Theorem 7.5. *Suppose B is sequential. Then, $A \precsim B$ iff $A \lhd B$.*

Proof. We define $C \mathcal{S} D$ iff $C \lhd D$ and D is sequential. We show that $\mathcal{S}$ has the forward property. Suppose $C \mathcal{S} D$. Then $C \lhd D$ and D is sequential. Let $C' \supseteq C$. Since B is locally tolerant, there is a $B' \supseteq B$ such that $B' \rhd C'$.

Alternatively, we know by the results of [Vis05], that $A \lhd B$ iff $A \lhd_{\mathsf{faith}} B$. Theorem 5.15 tells us that $\lhd_{\mathsf{faith}}$ is a subrelation of $\precsim$, which is in its turn a subrelation of $\lhd$. ❑

We note that in the first proof of Theorem 7.5 we just used the fact that sequential theories are essentially locally tolerant. So, in fact, we have proved: if B is essentially locally tolerant, then $A \precsim B$ iff $A \lhd B$.

Example 7.6. Suppose A is sequential. We have already seen that $A \not\approx (A \boxplus A)$. On the other hand, we have $A \rhd (A \boxplus A)$, and, hence $A \succsim (A \boxplus A)$. It is clear that the mapping $A \mapsto (A \boxplus A)$ has the forward property. Hence $A \cong (A \boxplus A)$.

Alternatively, we could simply note that, by Theorem 7.5, $A \cong B$ iff $A \equiv B$, for sequential A. Moreover, by the results of [Vis05], we have that $A \equiv B$ iff $A \equiv_{\mathsf{faith}} B$, for sequential A. So we can apply Example 7.3. ❑

Let $[A]_\sim$ be the equivalence class of $\sim$. Let SEQ be the class of sequential theories. Our knowledge at this point is summarised by the following theorem.

Theorem 7.7. *Suppose A is sequential Then:*

$$([A]_\equiv \cap \mathsf{SEQ}) = ([A]_\approx \cap \mathsf{SEQ}) \subsetneq [A]_\approx \subsetneq [A]_\cong = [A]_{\equiv_{\mathsf{faith}}} = [A]_{\equiv_{\mathsf{lofa}}} = [A]_\equiv.$$

Proof. Suppose A is sequential. Let's number the claims of the theorem.

$$([A]_\equiv \cap \mathsf{SEQ}) \overset{(1)}{=} ([A]_\approx \cap \mathsf{SEQ}) \overset{(2)}{\subsetneq} [A]_\approx \overset{(3)}{\subsetneq} [A]_\cong \overset{(4)}{=} [A]_{\equiv_{\mathsf{faith}}} \overset{(5)}{=} [A]_{\equiv_{\mathsf{lofa}}} \overset{(6)}{=} [A]_\equiv.$$

Claim 1 is Theorem 7.1. The non-identity in Claim 2 is Example 7.2. The non-identity in Claim 3 is by Example 7.6. The identities 5 and 6 follow since by the results of [Vis05], we have $A \equiv B$ iff $A \equiv_{\mathsf{faith}} B$ in case A is sequential. Moreover, $\equiv_{\mathsf{lofa}}$ is between $\equiv_{\mathsf{faith}}$ and $\equiv$. Finally, by Theorem 7.5 in combination with identities 5 and 6, we have identity 4. ❑

Remark 7.8. We can do a bit of reverse meta-mathematics and rederive earlier insights from Theorem 7.1.

We note that the fact that sequential theories are locally tolerant is immediate from Theorem 7.1. Consider any sequential A. Our theory is mutually interpretable with $\mathsf{S}^1_2 + \mathsf{con}_{\rho(A)}(A)$. Hence, $A \approx (\mathsf{S}^1_2 + \mathsf{con}_{\rho(A)}(A))$. By Theorem 6.3, it is clear that $\mathsf{S}^1_2 + \mathsf{con}_{\rho(A)}(A)$ is locally tolerant. Since local tolerance is a fine property, we find that A is locally tolerant. I do not see how we to derive (d) in full from Theorem 7.1 without self-referential arguments.

We show that Basic Insight (e) follows from Theorem 7.1. Suppose A and B are sequential and $A \lhd B$. By Theorem 6.6, there is an $m \geq \rho(B)$, such that

$$(\mathsf{S}^1_2 + \mathsf{con}_m(B)) \supseteq (\mathsf{S}^1_2 + \mathsf{con}_{\rho(A)}(A)).$$

Since $A \approx (\mathsf{S}^1_2 + \mathsf{con}_{\rho(A)}(A))$ and $B \approx (\mathsf{S}^1_2 + \mathsf{con}_m(B))$, we are easily done. ❑

References

[Ben86] C. Bennet. *On some orderings of extensions of arithmetic*. Department of Philosophy, University of Göteborg, 1986.

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

[dB65a] K. L. de Bouvère. Logical synonymy. *Indagationes Mathematicae*, 27:622–629, 1965.

[dB65b] K. L. de Bouvère. Synonymous Theories. In J.W. Addison, L. Henkin, and A. Tarski, editors, *The Theory of Models, Proceedings of the 1963 International Symposium at Berkeley*, pages 402–406. North Holland, Amsterdam, 1965.

[Fri07] H. Friedman. Interpretations according to Tarski. This is one of the 2007 Tarski Lectures at Berkeley. The lecture is available at `http://www.math.osu.edu/~friedman.8/pdf/Tarski1,052407.pdf`, 2007.

[Han75] William Hanf. The boolean algebra of logic. *Bulletin of the American Mathematical Society*, 81(3):587–589, 1975.

[Hod93] W. Hodges. *Model theory*. Encyclopedia of Mathematics and its Applications, vol. 42. Cambridge University Press, Cambridge, 1993.

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

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

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

[Lin79] P. Lindström. Some results on interpretability. In F.V. Jensen, Mayoh B.H., and Moller K.K., editors, *Proceedings of the 5th Scandinavian Logic Symposium 1979*, pages 329–361, Aalborg, 1979. Aalborg University Press.

[Lin84a] P. Lindström. On faithful interpretability. In M.M. et al Richter, editor, *Computation and Proof Theory*, pages 279–288. Springer Verlag, Berlin, Heidelberg, New York, 1984.

[Lin84b] P. Lindström. On partially conservative sentences and interpretability. *Proceedings of the American Mathematical Society*, 91:436–443, 1984.

[Lin02] P. Lindström. *Aspects of Incompleteness*, volume 10 of *Lecture Notes in Logic*. ASL / A.K. Peters, Natick, Massachusetts, 2002.

[Mac71] S. MacLane. *Categories for the Working Mathematician*. Number 5 in Graduate Texts in Mathematics. Springer, New York, 1971.

[MPS90] J. Mycielski, P. Pudlák, and A.S. Stern. *A lattice of chapters of mathematics (interpretations between theorems)*, volume 426 of *Memoirs of the American Mathematical Society*. AMS, Providence, Rhode Island, 1990.

[Mye89] D de Myers. Lindenbaum-Tarski algebras. In *Handbook of Boolean algebras*, volume 3, pages 1167–1196. Elsevier Science Publishers, 1989.

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

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

[Ste89] A.S. Stern. Sequential theories and infinite distributivity in the lattice of chapters. *The Journal of Symbolic Logic*, 54:190–206, 1989.

[Šve78] V. Švejdar. Degrees of interpretability. *Commentationes Mathematicae Universitatis Carolinae*, 19:789–813, 1978.

[Szc76] L.W. Szczerba. Interpretability and categoricity. *Bulletin de l'Académie Polonaise des Sciences, série des sciences math., astr. et phys.*, 24(5):309–312, 1976.

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

[Vis13] A. Visser. What is sequentiality? 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'Enseingement Supérieur Paris-est, Stanford, 2013.

[Vis14a] A. Visser. Why the theory 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/`.

[Vis14b] Albert Visser. Interpretability degrees of finitely axiomatized sequential theories. *Archive for Mathematical Logic*, 53(1-2):23–42, 2014.

[VS00] N.K. Vereshchagin and A. Shen. Languages and calculi. In *MCCME*, volume 2 of *Lectures on Mathematical Logic and the Theory of Algorithms*. 2000. In Russian.

Appendix A. Basics

In this appendix, we provide detailed definitions of translations, interpretations and morphisms between interpretations.

A.1. **Theories and Provability.** Theories in this paper are one-sorted theories of first order predicate logic of finite relational signature. We take identity to be a logical constant. Our official signatures are relational, however, via the term-unwinding algorithm, we can also accommodate signatures with functions. For most purposes in the present paper a theory can be identified with a deductively closed set of sentences of the given language. The exception is the few places where we use Rosser style arguments.

Our main focus will be on finitely axiomatised theories, but in this appendix, we will develop the material also for the infiniltely axiomatised case.

We will sometimes use the modal notation $\Box_A\phi$ for $\mathsf{prov}_A(\ulcorner\underline{\phi}\urcorner)$. We will also consider *restricted provability*. This is provability where we restrict the formulas occurring in the proof to formulas of complexity n, for some given m. Our measure of complexity is *depth of quantifier alternations*. This measure is defined officially as follows: $\rho := \rho_\exists$, where:

- $\rho_\exists(A) := \rho_\forall(A) = 1$, if A is atomic.
- $\rho_\exists(\neg B) := \rho_\forall(B)$, $\rho_\forall(\neg B) := \rho_\exists(B)$.
- $\rho_\exists(B \wedge C) := \mathsf{max}(\rho_\exists(B), \rho_\exists(C))$, $\rho_\forall(B \wedge C) := \mathsf{max}(\rho_\forall(B), \rho_\forall(C))$.
- $\rho_\exists(B \vee C) := \mathsf{max}(\rho_\exists(B), \rho_\exists(C))$, $\rho_\forall(B \vee C) := \mathsf{max}(\rho_\forall(B), \rho_\forall(C))$.
- $\rho_\exists(B \to C) := \mathsf{max}(\rho_\forall(B), \rho_\exists(C))$, $\rho_\forall(B \to C) := \mathsf{max}(\rho_\exists(B), \rho_\forall(C))$.
- $\rho_\exists(\exists v\, B) := \rho_\exists(B)$, $\rho_\forall(\exists v\, B) := \rho_\exists(B) + 1$.
- $\rho_\exists(\forall v\, B) := \rho_\forall(B) + 1$, $\rho_\forall(\forall v\, B) := \rho_\forall(B)$.

We write $\Box_{A,n}\phi$ for provability restricted to formulas ψ with $\rho(\psi) \leq n$.

A.2. **Translations.** Translations are the heart of our interpretations. In fact, they are often confused with interpretations, but we will not do that officially. In practice it is often convenient to conflate an interpretation and its underlying translation.

To formulate the notion of translation it is pleasant to allow in the target language lambda terms of the form $\lambda x_0 \ldots x_{n-1} \cdot \phi(x_0, \ldots, x_{n-1})$, where A is a formula. We will call a term of this form an n-term. We think of such terms modulo α-conversion as is usual in λ-calculus.

As a start. we define more-dimensional, one-sorted, one-piece relative translations without parameters. We will later indicate how to modify the definition to get two extra features: interpretations with parameters and piecewise interpretations.

Let Σ and Θ be one-sorted signatures. A translation $\tau : \Sigma \to \Theta$ is given by a triple $\langle m, \delta, F\rangle$. Here δ will be a closed m-term. The mapping F associates to each relation symbol R of Σ with arity n a closed $m \times n$-term of signature Θ.

We demand that predicate logic proves $F(R)(\vec{x}_0, \ldots, \vec{x}_{n-1}) \to (\delta(\vec{x}_0) \wedge \ldots \delta(\vec{x}_{n-1}))$. Of course, given any candidate $m \times n$-term $F(R)$ not satisfying the restriction, we can obviously modify it to satisfy the restriction.

We translate Σ-formulas to Θ-formulas as follows.

- $(R(x_0, \ldots, x_{n-1}))^\tau := F(R)(\vec{x}_0, \ldots, \vec{x}_{n-1})$.
  Here we demand that the sequences $\vec{x}_i$ are fully disjoint if the original variables are x_i are different.
  The single variable x_i of the source language needs to have no obvious connection with the sequence of variables $\vec{x}_i$ of the target language that represents it. We need some conventions to properly handle the association $x_i \mapsto \vec{x}_i$.[12]
- $(\cdot)^\tau$ commutes with the propositional connectives;
- $(\forall x\, \phi)^\tau := \forall \vec{x}\, (\delta(\vec{x}) \to \phi^\tau)$;
- $(\exists x\, \phi)^\tau := \exists \vec{x}\, (\delta(\vec{x}) \wedge \phi^\tau)$.

Here are some convenient conventions and notations.

- We write δ_τ for 'the δ of τ' and F_τ for 'the F of τ'.
- We write R_τ for $F_\tau(R)$.
- We write $\vec{x} \in \delta$ for: $\delta(\vec{x})$.

There are some natural operations on translations. The identity translation $\mathsf{id} := \mathsf{id}_\Theta$ is one-dimensional and it is defined by:

- $\delta_{\mathsf{id}}(x) := (x = x)$,
- $R_{\mathsf{id}}(\vec{x}) := R\vec{x}$.

We can compose relative translations as follows. Suppose τ is an m-dimensional translation from Σ to Θ, and ν is a k-dimensional translation from Θ to Ξ. We define the $m \times k$-dimensional interpretation $\tau\nu$ or $\nu \circ \tau$ as follows.

- We suppose that with the variable x we associate under τ the sequence $x_0, \ldots, x_{m-1}$ and under ν we send x_i to $\vec{x}_i$.
  $\delta_{\tau\nu}(\vec{x}_0, \ldots, \vec{x}_{m-1}) := (\delta_\nu(\vec{x}_0) \wedge \ldots \wedge \delta_\nu(\vec{x}_{m-1}) \wedge (\delta_\tau(x))^\nu)$,
- Let R be n-ary. Suppose that under τ we associate with x_i the sequence $x_{i,0}, \ldots, x_{i,m-1}$ and that under ν we associate with $x_{i,j}$ the sequence $\vec{x}_{i,j}$. We take:
  $$R_{\tau\nu}(\vec{x}_{0,0}, \ldots \vec{x}_{n-1,m-1}) = \delta_\tau(\vec{x}_{0,0}) \wedge \ldots \wedge \delta_\tau(\vec{x}_{n-1,m-1}) \ \wedge (R_\tau(x_{0,0}, \ldots x_{n-1,m-1}))^\nu.$$

We can make a disjunctive interpretation as follows. Suppose τ and ν are translations from Σ to Θ. We assume that τ is k-dimensional and ν is m-dimensional. Let ϕ be a Θ-sentence. We introduce a $\mathsf{max}(k, m)$-dimensional interpretation $\tau\langle\phi\rangle\nu$.

We first 'lift' one of the interpretations by padding to get the dimensions equal.[13] Suppose e.g. that $k < m$. Then we define:

[12]There are several ways of handling such conventions. First we can work with a fixed global association beteen the x_i and the $\vec{x}_i$. Secondly, we can make such an association local and carry it around as an extra argument of the translation. Thirdly, we can throw away the mechanism of using variable-names and work in a language that works with explicit links between places. Fourthly, we can sidestep the problem by working in many-sorted languages and, for every k, adding sequences of length k (of various sorts). This construction can be viewed as a representation of more dimensional interpretations as arrows in a Kleisli category. Regrettably, each way of proceeding needs some work and produces some awkwardness somewhere. We demand that the $\vec{x}_i$ are fully disjoint when the x_i are different. In this paper, we will assume that these details are taken care of by one strategy or by another.

[13]As we will see the padding can be avoided by using piecewise interpretations.

- $\delta_{\tau'}(\vec{x}\vec{z}) :\leftrightarrow \delta_\tau(\vec{x})$,
- $P_{\tau'}(\vec{x}_0\vec{z}_0, \cdots, \vec{x}_{n-1}\vec{z}_{n-1}) :\leftrightarrow P_\tau(\vec{x}_0, \cdots, \vec{x}_{n-1})$.

Here the dimension of the $\vec{z}$ is $m - k$.

Suppose the results of the padding operation are τ' and ν', where, of course, in case $k < m$, $\nu = \nu'$, etcetera. We define $\tau\langle\phi\rangle\nu$ as follows:

- $\delta_{\tau\langle\phi\rangle\nu}(\vec{x}) := ((\phi \wedge \delta_{\tau'}(\vec{x})) \vee (\neg\,\phi \wedge \delta_{\nu'}(\vec{x})))$.
- $R_{\tau\langle\phi\rangle\nu}(\vec{x}_0, \ldots, \vec{x}_{n-1}) :=$
  $$((\phi \wedge R_{\tau'}(\vec{x}_0, \ldots, \vec{x}_{n-1})) \vee (\neg\,\phi \wedge R_{\nu'}(\vec{x}_0, \ldots, \vec{x}_{n-1}))).$$
  Here the $\vec{x}$ are $\mathsf{max}(k, m)$-dimensional.

An m-dimensional translation τ *preserves identity* if

$$\vec{x} =_\tau \vec{y} :\leftrightarrow \bigwedge_{i<m} (\delta_\tau(x_i) \wedge \delta_\tau(y_i) \wedge x_i = y_i).$$

An m-dimensional translation τ *is unrelativized* if $\delta_\tau(\vec{x}) = \top$. An m-dimensional translation τ is *direct* if it is unrelativized and preserves identity. Note that all these properties are preserved by composition (modulo provable equivalence in predicate logic).

Consider a model $\mathcal{M}$ with domain M of signature Σ and k-dimensional translation $\tau : \Sigma \to \Theta$. Let $N := \{\vec{m} \in M^k \mid \mathcal{M} \models \delta_\tau\vec{m}\}$. Suppose N is not empty. Let E be the equivalence relation on N defined in $\mathcal{M}$ by $=_\tau$. Then τ specifies an internal model $\mathcal{N}$ of $\mathcal{M}$ with domain N/E and with $\mathcal{N} \models R([\vec{m}_0]_E, \ldots, [\vec{m}_{n-1}]_E)$ iff $\mathcal{M} \models R_\tau(\vec{m}_0, \ldots, \vec{m}_{n-1})$. We will write $\widetilde{\tau}(\mathcal{M})$ for the internal model of $\mathcal{M}$ given by τ. We treat the mapping $\tau, \mathcal{M} \mapsto \widetilde{\tau\mathcal{M}}$ as a partial function that is defined precisely if $\delta_\tau^{\mathcal{M}}$ is non-empty. Let Mod or $\widetilde{(\cdot)}$ be the function that maps τ to $\widetilde{\tau}$. We have:

$$\mathsf{Mod}(\tau \circ \rho)(\mathcal{M}) = (\mathsf{Mod}(\rho) \circ \mathsf{Mod}(\tau))(\mathcal{M}).$$

So, Mod behaves contravariantly.

A.3. **Relative Interpretations.** A translation τ supports a *relative interpretation* of a theory U in a theory V, if, for all U-sentences ϕ, we have $U \vdash \phi \Rightarrow V \vdash \phi^\tau$. Note that this automatically takes care of the theory of identity and assures us that δ_τ is inhabited. We will write $K = \langle U, \tau, V\rangle$ for the interpretation supported by τ. We write $K : U \to V$ for: K is an interpretation of the form $\langle U, \tau, V\rangle$. If M is an interpretation, τ_M will be its second component, so $M = \langle U, \tau_M, V\rangle$, for some U and V.

Par abus de langage, we write 'δ_K' for: δ_{τ_K}; 'R_K' for: R_{τ_K}; 'A^K' for: A^{τ_K}, etc. Here are the definitions of three central operations on interpretations.

- Suppose U has signature Σ. We define:
  $\mathsf{ID}_U : U \to U$ is $\langle U, \mathsf{id}_\Sigma, U\rangle$.
- Suppose $K : U \to V$ and $M : V \to W$. We define:
  $M \circ K : U \to W$ is $\langle U, \tau_M \circ \tau_K, W\rangle$.
- Suppose $K : U \to (V + \phi)$ and $M : U \to (V + \neg\,\phi)$. We define:
  $K\langle\phi\rangle M : U \to V$ is $\langle U, \tau_K\langle\phi\rangle\tau_M, V\rangle$.

It is easy to see that we indeed correctly defined interpretations between the theories specified.

A.4. **Global and Local Interpretability.** We can view interpretability as a generalisation of provability. When we take this stand point, we write:

- $U \rhd V$ (or $V \lhd U$) for: U interprets V (or V is interpretable in U).
- $U \equiv V$ for: U and V are mutually interpretable.

A closely related notion is local interpretability. We define

- U locally interprets V or $U \rhd_{\mathsf{loc}} V$ iff, for every finitely axiomatize subtheory V_0 of V we have $U \rhd V_0$.
- We write $V \lhd_{\mathsf{loc}} U$ and $U \equiv_{\mathsf{loc}} V$ with the obvious meanings.

If we want to stress the contrast between local and ordinary interpretability, we often call ordinary interpretability *global interpretability*. We will write $\rhd_{\mathsf{glob}}$, etcetera. The degrees of local interpretability are $\mathsf{DEG}_{\mathsf{loc}}$.

Example A.1. Let $\mathbb{2}$ be the theory in the language of identity that says that there are precisely two elements. Let INF be the theory in the language of identity that has for every n and axiom saying 'there are at least n elements. Then $\mathbb{2} \rhd_{\mathsf{loc}} \mathsf{INF}$ but $\mathbb{2} \not\rhd_{\mathsf{glob}} \mathsf{INF}$ ❑

A.5. **Adding Parameters.** We can add parameters in the obvious way. A translation τ with parameters is a tuple $\langle k, \alpha, m, \delta, F\rangle$. Here α is a k-term, δ is a $k+m$-term and $F(P)$ is a $(k+m\times n)$-term for n-ary P. The term α represents a k-dimensional parameter domain.

An interpretation with parameters is a tuple of the form $\langle U, \tau, V\rangle$, where we demand that $V \vdash \exists \vec{z}\, \alpha(\vec{z})$ and, for all ϕ, if $U \vdash \phi$, then $V \vdash \forall \vec{z}\, (\alpha(\vec{z}) \to \phi^{\tau,\vec{z}})$.

When we compose two translations σ and τ, we see that $\alpha_{\sigma\tau}(\vec{w}, \vec{z}_0, \cdots \vec{z}_{k-1})$ is

$$\alpha_\tau(\vec{w}) \wedge \bigwedge_{i<k_\sigma} \delta_\tau(\vec{z}_i) \wedge \alpha_\sigma^{\tau,\vec{w}}(\vec{z}_0, \cdots, \vec{z}_{k_\sigma-1})$$

We note that, in the presence of parameters, the function $\widetilde{K}$ associates *a class* of models of U to a model of V.

In Appendix A.6, we will briefly indicate how to adapt our treatment of parameters to the piecewise case.

A.6. **Piecewise Translations.** In this subsection we introduce piecewise translation. For more information on piecewise translations and interpretations see [Vis14a].

Before explaining what piecewise translations are, we state some of their advantages.

- Many constructions are conceptually cleaner when we use piecewise translations. Specifically, we avoid a lot of padding. As a consequence the heuristics for a construction is usually easier to grasp.
- The unnatural difference between one-element and at-least-two-element domains disappears.

We will show, in Appendix A.9, that, in case our interpreting theory proves that there are at least two elements, piecewise translations can be simulated by translations without pieces.

The idea of piecewise translations is that we can build up the domain from a number of pieces that may or may not be of the same dimension and that may or may not overlap. We first treat the case without parameters.

A piecewise translation is a tuple $\langle X, f, \delta, F\rangle$. Here X is a non-empty set of pieces and f is a function from X to ω. The function f gives us the arity of the domain associated to each piece. We use $\mathfrak{a}$, $\mathfrak{b}$, ... to range over pieces.

The term $\delta^{\mathfrak{a}}$ is a $f\mathfrak{a}$-term. Suppose P is n-ary. Let g be a function from $\{0,\ldots,n-1\}$ to X. Then $F^g(P)$ is a $(fg0 + fg1 + \cdots + fg(n-1))$-term.

Here are the clauses to lift our translation to the full language.

- Consider an n-ary predicate symbol P. Let j be a function from the set $\{0,\ldots,n-1\}$ to variables. Say $j(i) = x_i$. Suppose h is a function from $\{x_0,\ldots,x_{n-1}\}$ to X. (Here we allow that, for some i and j, the variables x_i and x_j are the same.) We define:
$$(R(x_0,\ldots,x_{n-1}))^{\tau,h} := F^{h\circ j}(R)(\vec{x}_0,\ldots,\vec{x}_{n-1}).$$
Here $\vec{x}_i$ has length $fh(x_i)$. We demand that the sequences $\vec{x}_i$ are fully disjoint if the original variables are x_i are different.
- Suppose h is a function from the free variables of $(\phi \wedge \psi)$ to pieces. Then, $(\phi\wedge\psi)^{\tau,h} = (\phi^{\tau,h\restriction \mathsf{FV}(\phi)} \wedge \psi^{\tau,h\restriction\mathsf{FV}(\psi)})$.
Similarly, for the other propositional connectives.
- $(\forall x\,\phi)^{\tau,h} := \bigwedge_{\mathfrak{a}\in X} \forall\vec{x}\,(\delta^{\mathfrak{a}}(\vec{x}) \to \phi^{\tau,h[x:=\mathfrak{a}]})$.
Similarly, for the existential quantifier.

We give three important examples of how piecewise interpretation works. A fourth example is given in Appendix A.8.

Let U be any theory and let $\mathsf{2}^+$ be the theory of two *named* elements, say c and d. We show that $U \rhd \mathsf{2}^+$. We write a for the function that assigns a to 0 and we write ab for the function that assigns a to 0 and b to 1. We write ε for the empty sequence. We define:

- $X := \{0,1\}$,
- $f(i) := 0$,
- $\delta^i(\varepsilon) := \top$,
- $F^{ij}(=)(\varepsilon,\varepsilon) := \top$, if $i = j$
$F^{ij}(=)(\varepsilon,\varepsilon) := \bot$, otherwise,
- $F^0(C)(\varepsilon) := \top$, $F^1(C)(\varepsilon) := \bot$,
- $F^0(D)(\varepsilon) := \bot$, $F^1(D)(\varepsilon) := \top$,

We leave the easy verification that we did indeed define an interpretation of $\mathsf{2}^+$ to the reader. We see that our interpretation is, in a sense, entirely independent of (the language of) U. We note that the interpretation would have worked even if we had started from a free logic. Thus, piecewise interpretation truly makes *creatio ex nihilo* possible.

We define the operation $\boxplus$ on theories as follows. The signature of $U_0 \boxplus U_1$ is the disjoint union of the signatures of U_0 and U_1, plus two new unary predicate $\triangle_0$ and $\triangle_1$. The axioms of $U_0 \boxplus U_1$ are:

- $P(x_0,\ldots,x_{n-1}) \to \bigwedge_{j<n} \triangle_i(x_j)$, if P is derived from the signature of U_i,
- the axioms of U_i relativized to $\triangle_i$,
- $\forall x\,(\triangle_0(x) \vee \triangle_1(x))$,
- $\forall x\,\neg\,(\triangle_0(x) \wedge \triangle_1(x))$.

Suppose τ_0 witnesses $V \rhd U_0$ and τ_1 witnesses $V \rhd U_1$. We construct a translation $\nu := [\tau_0, \tau_1]$ that witnesses $V \rhd (U_0 \boxplus U_1)$.

- $X^\nu := X^{\tau_0} \oplus X^{\tau_1} := (\{0\}\times X^{\tau_0}) \cup (\{1\}\times X^{\tau_1})$,

- $\delta^{\nu,(i,\mathfrak{a})} := \delta^{\tau_i,\mathfrak{a}}$,
- Suppose P is derived from an n-ary predicate of U_i. Then:
  $F^{\nu,h}(P) := F^{\tau_i,\pi_1\circ h}$ if $(\pi_0 \circ h)(k) = i$, for all $k < n$,
  $F^{\nu,h}(P) := \lambda x_0 \cdots x_{n-1} \cdot \bot$, otherwise.

Finally, we define the operation $\nu := \tau_0\langle\phi\rangle\tau_1$ using pieces.

- $X^\nu := X^{\tau_0} \oplus X^{\tau_1}$,
- $\delta^{\nu,(0,\mathfrak{a})}(\vec{x}) := (\phi \wedge \delta^{\tau_0,\mathfrak{a}}(\vec{x}))$,
  $\delta^{\nu,(1,\mathfrak{a})}(\vec{x}) := (\neg\,\phi \wedge \delta^{\tau_1,\mathfrak{a}}(\vec{x}))$,
- Suppose P is derived from an n-ary predicate of U_i. Then:
  $F^{\nu,h}(P)(\vec{x}_0,\cdots,\vec{x}_{n-1}) := (\phi \wedge P^{\pi_1\circ h}_{\tau_0}(\vec{x}_0,\cdots,\vec{x}_{n-1}))$, if $(\pi_0 \circ h)(k) = 0$, for all $k < n$,
  $F^{\nu,h}(P)(\vec{x}_0,\cdots,\vec{x}_{n-1}) := (\neg\,\phi \wedge P^{\pi_1\circ h}_{\tau_1}(\vec{x}_0,\cdots,\vec{x}_{n-1}))$, if $(\pi_0 \circ h)(k) = 1$, for all $k < n$,
  $F^{\nu,h}(P)(\vec{x}_0,\cdots,\vec{x}_{n-1}) := \bot$, otherwise.

The definition of *direct* for piecewise translations is simply that every piece has an unrelativised domain and as identity the pointwise identity of the components of the sequences representing the elements.

How to add parameters to piecewise translations. We can do it as follows. A piecewise translation with parameters is a tuple $\langle k, \alpha, X, f, \delta, F\rangle$. Here k is the dimension of the parameter domain, α is the parameter domain, X is a non-empty set of pieces and f is a function from X to ω. The function f gives us the arity of the domain associated to each piece. The term $\delta^{\mathfrak{a}}$ is a $f\mathfrak{a}+k$-term. Suppose P is n-ary. Let g be a function from $\{0,\ldots,n-1\}$ to X. Then $F^g(P)$ is a $(fg0 + fg1 + \cdots + fg(n-1) + k)$-term. The rest of the development is entirely as expected.

We could even go one better and also allow the parameter domain to be built from pieces. Then a translation would look as follows $\langle Y, g, \alpha, X, f, \delta, F\rangle$. This last way of proceeding would cohere optimally with the treatment in Appendix A.9.

A.7. **Five Categories.** We do not automatically get a category of theories and interpretations from the machinery we built up until now. For example, $\mathsf{ID}_U \circ \mathsf{ID}_U$ will not be strictly speaking identical with ID_U. We will obtain a category, when we divide out a suitable equivalence among interpretations. Below we will consider five kinds of equivalence that will give us five different categories. One important point of the categories is that isomorphism in each of them defines a salient notion of sameness of theories.

We treat our categories in the case where we do not have parameters nor pieces. To add these features, we have to adapt our definitions a bit. We sketch the idea of doing this via the co-Kleisli and the Kleisli construction in Appendix A.9.

A.7.1. *Provable equivalence of Interpretations.* Two interpretations are *provably equivalent* when the *target theory thinks they are the same.* Specifically, two interpretations $K, M : U \to V$ are provably equivalent if they have the same dimension, say m, and:

- $V \vdash \forall\vec{x}\,(\delta_K(\vec{x}) \leftrightarrow \delta_M(\vec{x}))$,
- $V \vdash \forall\vec{x}_0,\ldots,\vec{x}_{n-1}{\in}\delta_K\,(R_K(\vec{x}_0,\ldots,\vec{x}_{n-1}) \leftrightarrow R_M(\vec{x}_0,\ldots,\vec{x}_{n-1}))$.

Modulo this identification, the operations identity and composition give rise to a category INT_0, where the theories are objects and the interpretations arrows.

Isomorphism in this category is *synonymy* or *definitional equivalence*. This is the strictest notion of identity between theories in the literature. It was first introduced by Karel de Bouvère in [dB65a] and [dB65b].

Let MOD be the category with as objects classes of models, where the models in each class have the same signature, and as morphisms all functions between these classes. We define $\mathsf{Mod}(U)$ as the class of all models of U. Suppose $K : U \to V$. Then, $\mathsf{Mod}(K)$ is the function from $\mathsf{Mod}(V)$ to $\mathsf{Mod}(U)$ given by: $\mathcal{M} \mapsto \widetilde{K}(\mathcal{M}) := \widetilde{\tau_K}(\mathcal{M})$. It is clear that Mod is a *contravariant functor* from INT_0 to MOD.

A.7.2. *Definable Isomorphism of Interpretations.* For many applications provable equivalence is too strict. A better notions is provable isomorphism or *i-isomorphism.*

Consider $K, M : U \to V$. Suppose K is m-dimensional and M is k-dimensional. An i-isomorphism between interpretations $K, M : U \to V$ is given by an $m+k$-term F in the language of V. We demand that V verifies that "F is an isomorphism between K and M", or, equivalently, that, for each model $\mathcal{M}$ of V, the function $F^{\mathcal{M}}$ is an isomorphism between $\widetilde{K}(\mathcal{M})$ and $\widetilde{M}(\mathcal{M})$.

We spell out the syntactical definition of an i-isomorphism $F : K \Rightarrow M$.

- $V \vdash \vec{x}\ F\ \vec{y} \to (\vec{x} \in \delta_K \wedge \vec{y} \in \delta_M)$.
- $V \vdash (\vec{x} =_K \vec{u} \wedge \vec{u}\ F\ \vec{v} \wedge \vec{v} =_M \vec{y}) \to \vec{x}\ F\ \vec{y}$.
- $V \vdash \forall \vec{x} \in \delta_K\ \exists \vec{y} \in \delta_M\ \ \vec{x}\ F\ \vec{y}$.
- $V \vdash (\vec{x}_0 F \vec{y}_0 \wedge \ldots \vec{x}_{n-1} F \vec{y}_{n-1}) \to (R_K(\vec{x}_0, \ldots, \vec{x}_{n-1}) \leftrightarrow R_M(\vec{y}_0, \ldots, \vec{y}_{n-1}))$.

Here the last item includes identity in the role of R !

Two interpretations $K, M : U \to V$, are *i-isomorphic* iff there is an i-isomorphism between K and M. Wilfrid Hodges calls this notion: *homotopy.* See [Hod93], p222.

We can also define the notion of being i-isomorphic semantically. The interpretations $K, M : U \to V$, are *i-isomorphic* iff there is an F such that, for all V-models, $\mathcal{M}$, the relation $F^{\mathcal{M}}$ is an isomorphism between $\widetilde{K}(\mathcal{M})$ and $\widetilde{M}(\mathcal{M})$.

The default in this paper is that theories have finite signature: In this case we have a third characterisation. The interpretations $K, M : U \to V$, are *i-isomorphic* iff, for every V-model $\mathcal{M}$, there is an $\mathcal{M}$-definable isomorphism between $\widetilde{K}(\mathcal{M})$ and $\widetilde{M}(\mathcal{M})$. This characterisation follows by a simple compactness argument.

Clearly, if K and M are provably equivalent in the sense of the previous subsubsection, they will be i-isomorphic. The notion of i-isomorphism give rise to a category of interpretations modulo i-isomorphism. We call this category INT_1.

Isomorphism in INT_1 is *bi-interpretability.* Bi-interpretability is a very good notion of sameness that preserves such diverse properties as finite axiomatisability and κ-categoricity.

A.7.3. *Isomorphism.* Our third notion of sameness of the basic list is that K and M are the same if, for all models $\mathcal{M}$ of V, the internal models $\widetilde{K}(\mathcal{M})$ and $\widetilde{M}(\mathcal{M})$ are isomorphic. We will simply say that K and M are isomorphic. Clearly, i-isomorphism implies isomorphism. We call the associated category INT_2. Isomorphism in INT_2 is *iso-congruence.*

A.7.4. *Elementary Equivalence.* The fourth notion is to say that two interpretations K and M are the same if, for each $\mathcal{M}$, the internal models $\widetilde{K}(\mathcal{M})$ and $\widetilde{M}(\mathcal{M})$ are elementary equivalent. We will say that K and M are elementary equivalent.

By the Completeness Theorem, we easily see that this notion can be alternatively defined by saying that K is elementary equivalent to M iff, for all U-sentences A, we have $V \vdash A^K \leftrightarrow A^M$. It is easy to see that isomorphism implies elementary equivalence. We call the associated category INT_3. Isomorphism in INT_3 is *elementary congruence* or *sentential congruence*.

A.7.5. *Identity of Source and Target.* Finally, we have the option of abstracting away from the specific identity of interpretations completely, simply counting any two interpretations $K, M : U \to V$ the same. The associated category is INT_4. This is the structure of degrees of (global) interpretability $\mathsf{DEG}_{\mathsf{glob}}$. Isomorphism in INT_4 is *mutual interpretability.*

A.7.6. *Sections, faithful Retractions and Isomorphisms.* We remind the reader of the following. Consider a category $\mathcal{C}$. Suppose $f : x \to y$ and $g : y \to x$ and $g \circ f = \mathsf{id}_x$. In this case, we call f a *section* or *split monomorphism.* We call g a *retraction* or *split epimorphism.* The object x is in this situation a *retract* of y.

We have:

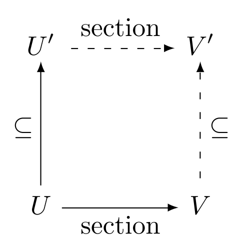
Theorem A.2. *Sections in* INT_i*, for* $i = 0, 1, 2, 3$ *are faithful interpretations.*

Proof. Since a section in INT_i for $i \leq 3$, is automatically a section in INT_3. It is sufficient to prove out claim for INT_3. Suppose $K : U \to V$ is a section with inverse $M : V \to U$. We have:

$$\begin{aligned} V \vdash \phi^{\tau_K} &\Rightarrow U \vdash \phi^{\tau_K \tau_M} \\ &\Rightarrow U \vdash \phi \end{aligned}$$

❑

The section relation has the forward or zig property w.r.t. theory-extension in INT_i, for $i = 0, 1, 2, 3$. This is illustrated by the following diagram.



Theorem A.3. *Let* $i \in \{0, 1, 2, 3\}$*. The section relation in* INT_i *has the forward or zig property with respect to theory extension.*

Proof. Suppose $K : U \to V$ is a section in INT_i. Let M be an inverse of K, so $M : V \to U$ and $M \circ K = \mathsf{ID}_U$ in INT_i. Suppose $U \subseteq U'$. We define

$$V' := \{\phi \in \mathsf{sent}_{\Sigma_V} \mid U' \vdash \phi^{\tau_M}\}.$$

Clearly, we have an interpretation $M' : V' \to U'$ based on τ_M. We have:

$$\begin{aligned} U' \vdash \psi \quad &\Rightarrow \quad U' \vdash \psi^{\tau_K \tau_M} \\ &\Rightarrow \quad V' \vdash \psi^{\tau_K} \end{aligned}$$

Hence there is an interpretation K' based on τ_K such that $K' : U' \to V'$.

We note that the further properties needed to be a retraction in one of our categories are trivially upwards preserved from U, K, M, V to U', K', M', V', since the corresponding interpretations are based on the same translations. E.g., in the case of INT_2, suppose $\mathcal{M}$ is a model of U'. Consider the inner model $\mathcal{M}^* := \widetilde{\tau}_K\widetilde{\tau}_M(\mathcal{M})$ Since $\mathcal{M}$ is a model of U, it follows that $\mathcal{M}^*$ is isomorphic to $\mathcal{M}$. ❑

The above construction has an important disadvantage: it does not *prima facie* preserve finite axiomatisation. Even if U, V and U' are finitely axiomatised, why should V' be finitely axiomatised? The next result fares better in this respect: faithful retractions also have the forward property and here finiteness is preserved.

Theorem A.4. *Let $i \in \{0, 1, 2, 3\}$. The relation of being a faithful retraction in INT_i has the forward or zig property with respect to theory extension. Moreover, the result is preserved when we restrict ourselves to finitely axiomatised theories.*

Proof. Suppose $K : U \to V$ is a faithful retraction in INT_i. Let M be an inverse of K, so $M : V \to U$ and $K \circ M = \mathsf{ID}_V$ in INT_i. Suppose $U \subseteq U'$. We define

$$V' := V + \{\phi^{\tau_K} \in \mathsf{sent}_{\Sigma_V} \mid U' \vdash \phi\}.$$

Clearly, we have an interpretation $K' : U' \to V'$ based on τ_K. Moreover, suppose $V' \vdash \phi^{\tau_K}$. Then, for some χ, we have $U \vdash \chi$ and $V + \chi^{\tau_K} \vdash \phi^{\tau_K}$. By the faithfulness of K, we find: $U + \chi \vdash \phi$. Ergo $U' \vdash \phi$. So K' is faithful.

Suppose $V' \vdash \psi$. Then, $V + \phi^{\tau_K} \vdash \psi$, for some ϕ such that $U' \vdash \phi$. It follows that $V + \phi^{\tau_K} \vdash \psi^{\tau_M\tau_K}$, since we are in INT_i with $i \leq 3$. By the faithfulness of K, we have $U + \phi \vdash \psi^{\tau_M}$. Hence, $U' \vdash \psi^{\tau_M}$. So there is an interpretation M' based on τ_M such that $M' : V' \to U'$.

We note that the further properties needed to be a retraction in one of our categories are upwards preserved from U, K, M, V to U', K', M', V', since the corresponding interpretations are based on the same translations.

Finally, if V and U' are finite, then, so is V'. Specifically, if $B := V$ and $A' := U'$, then $B' := V' := B + (A')^{\tau_K}$. ❑

Theorem A.5. *Let $i \in \{0, 1, 2, 3\}$. The relation of being isomorphic in INT_i is a bisimulation with respect to theory extension. Moreover, the result is preserved when we restrict ourselves to finitely axiomatised theories.*

Proof. We note that all relevant arrows are sections in INT_3 and, hence, faithful. We can use the proof of the previous theorem to prove the zig and the zag property. To see that the pairs of interpretations we found do indeed form isomorphisms, we note that the further properties needed to be an isomorphism in one of our categories are upwards preserved from U, K, M, V to U', K', M', V', since the corresponding interpretations are based on the same translations. ❑

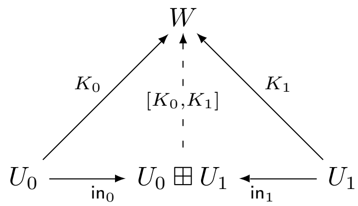
A.8. **Sums.** We define the operation $\boxplus$ on theories as follows. The signature of $U_0 \boxplus U_1$ is the disjoint union of the signatures of U_0 and U_1, plus two new unary predicate $\triangle_0$ and $\triangle_1$. The axioms of $U_0 \boxplus U_1$ are:

- $P(x_0, \ldots, x_{n-1}) \to \bigwedge_{i<n} \triangle_j(x_i)$, if P is derived from the signature of U_j,
- the axioms of U_j relativized to $\triangle_j$,
- $\forall x\, (\triangle_0(x) \vee \triangle_1(x))$,
- $\forall x\, \neg\, (\triangle_0(x) \wedge \triangle_1(x))$.

We treat identity as outside of the signature here. We have the ordinary theory of identity.

We note that $U_0 \boxplus U_1$ is synonymous with $U_1 \boxplus U_0$ and $(U_0 \boxplus U_1) \boxplus U_2$ is synonymous with $U_0 \boxplus (U_1 \boxplus U_2)$ and that both are synonymous with the ternary sum $\boxplus(U_0, U_1, U_2)$ which is defined in the obvious way using $\triangle_0$, $\triangle_1$ and $\triangle_2$.

We show that $\boxplus$ is the sum in the categories INT_i for $1 \leq i \leq 4$. We remind the reader of the sum diagram.



The arrows in_j interprets U_j in $U_0 \boxplus U_1$ by relativisation to $\triangle_j$. We note that, by our conventions we should take $x =_{\mathsf{in}_j} y$ iff $\triangle_j(x) \wedge \triangle_j(y) \wedge x = y$. The other predicate symbols do not need this addition.

Suppose $K_0 : U \to W$ and $K_1 : V \to W$. We define $M := [K_0, K-1] : (U_0 \boxplus U_1) \to W$. (This is one of those examples where piecewise interpretability is far more natural than the piece-free case.) We define:

- $X_{\tau_M} := X_{\tau_{K_0}} \oplus X_{\tau_{K_1}}$. To simplify in essentially we assume that $X_{\tau_{K_0}}$ and $X_{\tau_{K_1}}$ are already disjoint and we just take the union.
- Suppose $\mathfrak{a} \in X_{\tau_{K_j}}$. Then, $f_{\tau_M}(\mathfrak{a}) = f_{\tau_{K_j}}(\mathfrak{a})$.
- Suppose $\mathfrak{a} \in X_{\tau_{K_j}}$. Then, $\delta^{\mathfrak{a}}_{\tau_M} := \delta^{\mathfrak{a}}_{\tau_{K_j}}$.
- Suppose P is an n-ary predicate derived from an U_j-predicate that we also call P. Suppose h is a function from $\{0, \cdots, n-1\}$ to X_{τ_M}. Then, $P^h_{\tau_M} := P^h_{\tau_{K_j}}$, in case $h(s) \in X_{\tau_{K_j}}$, for all $s < n$, $P^h_{\tau_M} := \lambda\vec{x}_0 \cdots \vec{x}_{n-1} \cdot \bot$, otherwise. (For identity we need a slightly modified definition: it is the union of the two identities defined in the style of $P^h_{\tau_M}$.)
- $\triangle^{\mathfrak{a}}_{j,\tau_M} = \delta^{\mathfrak{a}}_{\tau_{K_j}}$ if $\mathfrak{a} \in X_{\tau_{K_j}}$; $\triangle^{\mathfrak{a}}_{j,\tau_M} = \lambda\vec{x} \cdot \bot$, otherwise.

Now we have defined $[K_0, K_1]$, it remains to verify commutation and uniqueness. We leave that to the reader.

There is an alternative construction of a sum $U_0 \oplus U_1$ in [MPS90] or [Ste89]. This alternative construction is for many purposes more convenient.

We have the following basic theorem.

Theorem A.6. *Consider the theory $W := U \boxplus V$. Consider any formula $A\vec{x}\vec{y}$ in the language of W. Then, there are formulas $B_i\vec{x}$ in the language of U and formulas $C_j\vec{y}$ in the language of V, such that $A\vec{x}\vec{y}$ is equivalent to a boolean combination of $B_i^{\mathsf{in}_0}\vec{x}$ and $C_j^{\mathsf{in}_1}\vec{y}$ in the theory $W + \bigwedge_k x_k : \triangle_0 + \bigwedge_\ell y_\ell : \triangle_1$.*

Proof. The proof of the theorem is by a simple induction on A. ❑

An important notion that is defined in terms of the notion of *sum* is *connectedness*. We say that a theory W is *connected* if, for any theories U and V, if $(U \boxplus V) \rhd_{\mathsf{loc}} W$, then $U \rhd_{\mathsf{loc}} W$ or $V \rhd_{\mathsf{loc}} W$. The following fundamental theorem is due to Pavel Pudlák in [Pud83]. It was reproved with a markedly different proof by Alan Stern in [Ste89]. For more context, see also [MPS90].

Theorem A.7. *Every sequential theory is connected.*

The notions of *sequentiality* is introduced in Appendix B.

A.9. **Parameters and Piecewise Interpretations Revisited.** Many features of interpretations can be added via the Kleisli and the co-Kleisli construction. See [Mac71]. We could, for example, start with a category for 1-dimensional direct interpretations and use the co-Kleisli construction to add domains, non-trivial identity and parameters. We can use the Kleisli construction to add more-dimensionality for each dimension m and to add the effect of piecewise interpretations.

Regrettably, the above idea has never been carefully worked out. Also, it is outside the scope of this appendix. We will restrict ourselves to specifying the functors between the relevant categories needed for the adjunctions that give rise to the co-Kleisli and Kleisli constructions. Then, we can reverse the conceptual order and start with the relevant endofunctor and natural transformations to obtain the enriched category.

A.9.1. *Parameters.* Let $\mathsf{INT}_1^{\mathsf{npar}}$ be INT_1 for interpretations without parameters. Let $\mathsf{INT}_1^{\mathsf{par}}$ be INT_1 for interpretations with parameters. Two interpretations with parameters $K : U \to V$ and $M : U \to V$ are isomorphic iff, there is, in V, a total and surjective relation G between the parameter domains α_K and α_M, and a relation F such that V proves that, whenever $\vec{z}\ G\ \vec{w}$, then we have $\vec{x}\ F_{\vec{z}\vec{w}}\ \vec{y}$ is an isomorphism between $K^{\vec{z}}$ and $M^{\vec{w}}$.

We define the functor $\Phi : \mathsf{INT}^{\mathsf{par}} \to \mathsf{INT}_1^{\mathsf{npar}}$ as follows. The signature of $\Phi(U)$ consists of all predicate symbols of the signature of U with their arities raised by 1 plus two unary predicate symbols $\triangle_0$ and $\triangle_1$. Here $\triangle_0$ represents the parameter domain and $\triangle_1$ represents the object domain. We translate the language of U as follows into the new language:

- $P^\circ(\vec{x}) := P(z, \vec{x})$. Here z is a fixed variable that is distinct from the x's.
- $(\cdot)^\circ$ commutes with the propositional connectives.
- $(\forall x\, \phi)^\circ := \forall x\, (\triangle_1(x) \to \phi^\circ)$.
  Similarly, for $\exists$.

The axioms of $\Phi(U)$ are:

- $\vdash \forall z\, (\triangle_0(z) \to \phi^\circ(z))$, where ϕ is an axiom of U.
- $\vdash \exists z\, \triangle_0(z)$.
- $\vdash \forall u\, (\triangle_0(u) \vee \triangle_1(u))$.
- $\vdash \forall u\, \neg\, (\triangle_0(u) \wedge \triangle_1(u))$.

Suppose $K : U \to V$ is an interpretation with parameters. We specify $\Phi(K)$.

- $X_{\Phi(K)} := X_K \cup \{*\}$, where $* \notin X_K$.
- $f_{\Phi(K)}(\mathfrak{a}) = f_K(\mathfrak{a}) + k_K$, if $\mathfrak{a} \in X_K$,
  $f_{\Phi(K)}(*) = k_K$.
- $\delta^{\mathfrak{a}}_{\Phi(K)}(\vec{z}, \vec{x}\,) := (\alpha(\vec{z}\,) \wedge \bigwedge_{i<f_K(\mathfrak{a})} \triangle_1(x_i) \wedge \delta^{\mathfrak{a},\circ}_{\Phi(K)}(\vec{z}, \vec{x}\,))$, if $\mathfrak{a} \in X_K$,
  $\delta^{*}_{\Phi(K)}(\vec{z}\,) := \alpha(\vec{z}\,)$.
- Let P be an n-ary U-predicate. Let $h'(j) := h(j+1)$, for $j < n$.
  $P^h_{\Phi(K)}(\vec{z}, \vec{x}_0, \cdots, \vec{x}_{n-1}\,) := (\bigwedge_{j<n+1} \delta^{h(j)}_{\Phi(K)}(\vec{z}, \vec{x}_j\,) \wedge P^{h',\circ}_K(\vec{z}, \vec{x}_0, \cdots, \vec{x}_{n-1}\,))$,
  in case $h(0) = *$ and $h(j) \in X_K$, for $j > 0$.
  $P^h_{\Phi(K)}(\vec{z}, \vec{x}_0, \cdots, \vec{x}_{n-1}\,) := \bot$, otherwise.
- $\triangle^h_0(\vec{z}\,) := \alpha(\vec{z})$ if $h(0) = *$,
  $\triangle^h_0(\vec{z}\,) := \bot$, otherwise.
- $\triangle^h_1(\vec{z}, \vec{x}\,) := \delta^{\mathfrak{a}}_{\Phi(K)}(\vec{z}, \vec{x}\,)$, if $h0 = *$ and $h1 = \mathfrak{a} \in X_K$,
  $\triangle^h_1(\vec{z}, \vec{x}\,) := \bot$, otherwise.

We define the embedding functor emb from $\mathsf{INT}_1^{\mathsf{npar}}$ to $\mathsf{INT}_1^{\mathsf{par}}$ as follows. Suppose $K : U \to V$ is an interpretations without parameters. We take $\mathsf{emb}(U) := U$. Moreover $\mathsf{emb}(K)$ will be given by $\langle 0, \top, X_K, f_K, \delta_K, F_K\rangle$. We now have that Φ is the left adjoint of emb.

The idea of the co-Kleisli construction is that we have an isomorphic copy of $\mathsf{INT}_1^{\mathsf{par}}$ inside $\mathsf{INT}_1^{\mathsf{npar}}$ by considering arrows $\Phi(U) \to V$.

A.9.2. *Pieces.* To simplify our sketch, we will treat the parameter-free case. Let $\mathsf{2}^+$ be the theory of two named elements. Let's say the named elements are 0 and 1. We define a functor $\Psi : \mathsf{INT}_1^{\mathsf{piece}} \to \mathsf{INT}_1^{\mathsf{npiece}}$. We take $\Psi(U) := U \boxplus \mathsf{2}^+$. Suppose K is a piecewise interpretation. We define $\Psi(K)$ as follows.

- Let s be the minimum number such that $2^s \geq |X_K|$. We take $m_{\Psi_K} := \mathsf{max}(\{f_K(\mathfrak{a}) \mid \mathfrak{a} \in X_K\}) + s$.
- The domain $\delta_{\Psi(K)}$ is specified as follows. We first number the pieces in X_K starting the count with 0. We represent each piece by a binary string of length s corresponding to its associated number. Say this string is $\sigma(\mathfrak{a})$. Let $\zeta(\mathfrak{a})$ be a sequence of $m_{\Psi(K)} - f_K(\mathfrak{a}) - s - 1$ zero's. For each piece $\mathfrak{a}$ in X_K, we add elements to the domain of the following form:
$$(\overbrace{0, 1, \cdots, 1, 0}^{\sigma(\mathfrak{a})}, \overbrace{d_0, \cdots, d_{f_K(\mathfrak{a})-1}}^{\text{element of } \delta_K^{\mathfrak{a}}}, \overbrace{0, 0 \cdots, 0, 0}^{\zeta(\mathfrak{a})}).$$
- $P_K(\vec{x}_0, \cdots, \vec{x}_{n-1})$, whenever each $\vec{x}_i$ is of the form $\sigma(\mathfrak{a}_i)\vec{d}_i\vec{\zeta}(\mathfrak{a}_i)$, where $\vec{d}_i$ is in $\delta_K^{\mathfrak{a}_i}$ and $P_K^h(\vec{d}_0, \cdots, \vec{d}_{n-1})$, where $h(i) := \mathfrak{a}_i$.

We note that our translation is not uniquely specified since it depends on the choice of the numbering of X_K. However, on the level of interpretations in $\mathsf{INT}_1^{\mathsf{npiece}}$, this choice is 'erased', since all choices give i-isomorphic interpretations.

In the other direction we define the functor π. We take $\pi(U) := U$. Moreover, $\pi(K)$ will be given by: $X_{\pi(K)} := \{0\}$, $f_{\pi(K)}(0) := m_K$, $\delta^0_{\pi(K)}(\vec{x}) := \delta_K(\vec{x})$, $P^h_{\pi(K)}(\vec{x}_0, \cdots, \vec{x}_{n-1}) := P_K(\vec{x}_0, \cdots, \vec{x}_{n-1})$.

We find that Ψ is a right adjoint of π. The Kleisli construction allows us to define an isomorphic copy of $\mathsf{INT}_1^{\mathsf{piece}}$ inside $\mathsf{INT}_1^{\mathsf{npiece}}$ by taking as arrows piece-free interpretations $K : U \to \Psi(V)$.

We note that if U proves that there are at least two elements, then U is bi-interpretable with $U \boxplus \mathsf{2}^+$. This means that as soon as the interpreting theory proves that there are at least two elements, piecewise interpretations contribute, in a sense, nothing new.

Appendix B. Sequential Theories

Sequential theories are theories of sequences where the possible length of the sequence is internally determined. The presence of sequences provides many good properties for such theories. For example, sequential theories are locally reflexive due to the presence of partial satisfaction predicates. We refer the reader to [Vis13] for more information about sequential and poly-sequential theories.

Even if the basic idea of sequentiality involves sequences and *ipso facto* numbers, sequentiality has a surprisingly simple definition. The theory AS is given by:

$\mathsf{AS1}$. $\vdash \exists x\, \forall y\, y \notin x$
$\mathsf{AS2}$. $\vdash \exists z\, \forall u\, (u \in z \leftrightarrow (u \in x \vee u = y))$

A theory is poly-sequential if it directly interprets AS. A theory is a sequential if it directly interprets AS via a 1-dimensional interpretation. Using these definitions, one may obtain the desired numbers and sequences by a substantial bootstrap.

Since direct interpretations are closed under composition, each theory that directly interprets a (poly-)sequential theory is itself a poly-sequential theory. Obviously, the identical embedding of a theory in an extension-in-the-same-language is direct. Ergo, being a (poly-)sequential theory is preserved under extension-in-the-same-language. Poly-sequentiality is also preserved under INT_1-retractions.

Theorem B.1. *Let U be a poly-sequential theory and suppose that V is a retraction in INT_1 of U. Then, V is a poly-sequential theory.*

Appendix C. Linear Discrete Orderings with Endpoints

We prove Theorem 5.3 that the theory of linear discrete orderings with endpoints has the finite model property.

Proof. Consider any models $\mathcal{A}$ and $\mathcal{B}$ of the theory of linear discrete orderings with endpoints. Say the minimal elements of our models are respectively a° and b° and the maximal elements are $a^\star$ and $b^\star$. Let:

$$\mathsf{d}(a,a') := |\{a'' \mid a \leq a'' < a' \text{ or } a' \leq a'' < a\}|.$$

Similarly, for $\mathsf{d}(b,b')$.

Let $\vec{a}$ and $\vec{b}$ be sequences of elements of $\mathcal{A}$, respectively $\mathcal{B}$. We write these sequences $a_1 \cdots a_{n-1}$ (they may be empty). We standardly add $a_0 := a^\circ$ and $a_n := a^\star$. Similarly, for $\vec{b}$. We define $\vec{a} \sim_n \vec{b}$ iff

i. $\vec{a}$ and $\vec{b}$ have the same length;
ii. $a_i < a_j$ iff $b_i < b_j$;
iii. $\mathsf{min}(\mathsf{d}(a_i,a_j),3^n) = \mathsf{min}(\mathsf{d}(b_i,b_j),3^n)$.

Here i and j range over $0,\cdots,n$. We define $\vec{a} \prec \vec{a}'$ iff $\vec{a}'$ is the result of inserting one new element in $\vec{a}$, so $\vec{a}$ is of the form $a_1 \cdots a_{i-1}a_i \cdots a_{k-1}$ and $\vec{a}'$ is of the form $a_1 \cdots a_{i-1}a'a_i \cdots a_{k-1}$. Similarly, for $\vec{b} \prec \vec{b}'$.

We show that $\sim_n$ is a bounded bisimulation w.r.t. $\prec$. This means that whenever $\vec{a} \prec \vec{a}'$ and $\vec{a} \sim_{n+1} \vec{b}$, then there is a $\vec{b}'$, such that $\vec{b} \prec \vec{b}'$ and $\vec{a}' \sim_n \vec{b}'$ (zig-property), and, vice versa, whenever $\vec{b} \prec \vec{b}'$ and $\vec{a} \sim_{n+1} \vec{b}$, then there is a $\vec{a}'$, such that $\vec{a} \prec \vec{a}'$ and $\vec{a}' \sim_n \vec{b}'$ (zag-property).

We prove the zig-property. Suppose $\vec{a} \prec \vec{a}'$ and $\vec{a} \sim_{n+1} \vec{b}$. In case a' is one of the a_i, this is trivial. So suppose a' is distinct from the a_i. In this case, there are i and j such that $a_i < a' < a_j$ and, for no a_s, we have $a_i < a_s < a_j$. (Here i,j,s range over $0,\ldots,n$.) If $\mathsf{d}(a_i,a_j) \leq 3^{n+1}$, we take b' with $b_i < b'$ and $\mathsf{d}(b_i,b') = d(a_i,a')$. Suppose $\mathsf{d}(a_i,a_j) > 3^{n+1}$. In case $\mathsf{d}(a_i,a') < 3^n$, we choose $b' > b_i$ with $\mathsf{d}(b_i,b') = \mathsf{d}(a_i,a')$. In case $\mathsf{d}(a',a_j) < 3^n$, we choose $b' < b_j$ with $\mathsf{d}(b',b_j) = \mathsf{d}(a',a_j)$. Otherwise, we choose $b' > a_i$ such that $\mathsf{d}(b_i,b') = 3^n$ (or any other b' such that $b_i < b' < b_j$ and $\mathsf{d}(b_i,b') \geq 3^n$ and $\mathsf{d}(b',b_j) \geq 3^n$). It is easy to see that b' as chosen satisfies the zig-property. By symmetry, we also have the zag-property.

Let $\nu(\phi)$ be the quantifier depth of ϕ. Using the above fact, is is easy to show that if $\nu(\phi) \leq n$ and $\vec{a} \sim_n \vec{b}$ that $\mathcal{A} \models \phi$ iff $\mathcal{B} \models \phi$.

Now suppose the theory of linear discrete orderings with endpoints does not prove ψ, where ψ is a sentence with $\nu(\psi) = \ell$. Then, there is a model $\mathcal{A}$ of the theory of linear discrete orderings with endpoints with $\mathcal{A} \not\models \psi$. If this model is finite we are done. Suppose $\mathcal{A}$ is infinite. Let $\mathcal{B}$ be the finite model given by the ordinal $3^\ell + 1$. Now $\varepsilon_{\mathcal{B}} \sim_\ell \varepsilon_{\mathcal{A}}$. Hence $\mathcal{B}$ is a model of the theory of linear discrete orderings with endpoints and $\mathcal{B} \not\models \psi$. So, we have the finite model property for the theory of linear discrete orderings with endpoints. ❑

We proceed to provide the promised separating example between $\approx^+$ and $\equiv_{\mathsf{faith}}$. This picks up the thread of Remarks 2.2 and 5.10.

Theorem C.1. *Let A be the theory of linear discrete orderings with endpoints. Then, $\mathsf{EQ} \not\rhd_{\mathsf{faith}} A$.*

Proof. Suppose, to obtain a contradiction that τ is a translation that witnesses that $\mathsf{EQ} \rhd_{\mathsf{faith}} A$. We define:

$$\uparrow n := \exists x_0 \cdots \exists x_{n-1} \bigwedge_{i<j<n} x_i \neq x_j.$$

Then, clearly, for fresh constants $\vec{c}$,

$$U := \mathsf{EQ} + \alpha_\tau(\vec{c}) + \{(\uparrow n)^{\tau,\vec{c}} \mid n \in \omega\}$$

is a consistent theory. Let $\mathcal{M}$ be a model of U. Since our model is infinite, we can replace τ modulo i-isomorphism by a piece-less m-dimensional translation τ^*.

A *type* is a pair (f, E), where f is a function from $m = \{0, \cdots, m-1\}$ to the parameters a_i and a fresh element $*$, and where E is an equivalence relation E on m. We demand that if $f(j) = c_i$, for any parameter c_i, then jEj' iff $f(j') = c_i$. An m-sequence $\vec{d}$ has type (f, E) iff whenever $f(j) = c_i$, then $d_j = c_i$, and $d_j = d_{j'}$ iff $f(j)Ef(j')$.

Consider two elements $\vec{a}$ and $\vec{b}$ of type (f, E) in $\delta^{\vec{c}}_{\tau^*}$. Since our domain is infinite there is a third element $\vec{d}$ of the same type such that the non-parameters in $\vec{d}$ are disjoint from both the non-parameters in $\vec{a}$ and in $\vec{b}$. There are involutive automorphisms α and β that fix the parameters such that $\alpha(\vec{a}) = \vec{d} = \beta(\vec{b})$. Suppose $\vec{a} <^{\vec{c}}_{\tau^*} \vec{d}$. Then, $\vec{d} = \alpha(\vec{a}) <^{\vec{c}}_{\tau^*} \alpha(\vec{d}) = \vec{a}$. *Quod impossibile.* So $\vec{a} =^{\vec{c}}_{\tau^*} \vec{d}$. Similarly, $\vec{b} =^{\vec{c}}_{\tau^*} \vec{d}$. Ergo, $\vec{a} =^{\vec{c}}_{\tau^*} \vec{b}$. It follows that, if $\vec{a}$ is in $\delta^{\vec{c}}_{\tau^*}$, then all elements of its type are equal in the sense of $=^{\vec{c}}_{\tau^*}$. But there are clearly only finitely many types. A contradiction. ❑

Theorem C.2. *Let A be the theory of a discrete ordering with endpoints. Then, $A \approx^+ \mathsf{EQ}$.*

Proof. Consider any sentence in the language of A. Suppose $\nu(A) = k$. Consider the model of A given by the ordinal $k^\star := 3^k + 1$. By the methods of the proof of theorem 5.3, we have: if $k^\star \models \phi$, then each model of cardinality $\geq k^\star$ models ϕ; if $k^\star \not\models \phi$, then each model of cardinality $\geq k^\star$ does not model ϕ. It follows that ϕ is equivalent with a statement purely in the language of identity. We define $B\,\mathcal{B}\,C$ iff, for some ψ in the language of identity $B = (A + \psi)$ and $C = (\mathsf{EQ} + \psi)$. It is easily seen that $\mathcal{B}$ is an i-bisimulation for $\subsetneq$. ❑

Appendix D. Extension meets Interpretability-with-Parameters

In this appendix, we show how to develop or framework using parameters. At some points we will want to compare the concepts with parameters with those without them. We will use e.g. $\rhd^*$ to designate the parameter-free version.

Remark D.1. In this appendix we are concerned with finitely axiomatised theories. For these theories a different treatment of parameters is possible. We do have a dimension for the parameters but no parameter domain. A theory B is interpretable in A, if, for some τ, we have $A \vdash \exists \vec{z}\, B^{\tau,\vec{z}}$. This is the treatment of parameters in e.g. [MPS90]. We can easily see that we can transform a translation τ with parameters in the sense of [MPS90] to a translation $\tau^\star$ in our sense by adding $\alpha_{\tau^\star}(\vec{z}\,) := B^{\tau,\vec{z}}$ to τ.

I have two reasons not to follow the style of [MPS90]. First, it is easier to see how to generalise the present treatment to the infinitely axiomatised case. Secondly, the reduction given above shows that $A \rhd B$ in the sense of [MPS90] iff $A \rhd B$ in the sense of our paper, but does the same work out also for e.g. faithful interpretability and other translation based notions? A closer look at these things would be needed.[14] ❑

Open Question D.2. Do translations in the sense of [MPS90] give us the same notions of faithful and locally faithful interpretability? ❑

The following abbreviation is useful:

- $[\phi]^\tau := (\exists \vec{z}\, \alpha_\tau(\vec{z}\,) \wedge \forall \vec{z}\, (\alpha_\tau(\vec{z}\,) \to \phi^{\tau,\vec{z}}))$.

We have:

Theorem D.3. $[[\phi]^\tau]^\nu \vdash [\phi]^{\tau\nu}$.

The proof is by patiently writing things out. We give extended definitions of a number of fundamental notions.

- $A \rhd B$ iff, for some τ, we have $A \vdash [B]^\tau$.
- $A \Rsh B$ iff, for some translation τ, we have that $A + [B]^\tau$ is consistent.
- A is *locally tolerant* iff $A \Rsh B$, for each B.
- A is *essentially locally tolerant* iff, for each $A' \supseteq A$, we have $A' \Rsh B$, for each B.
- $A \blacktriangleright_{\mathsf{loc}} B$ iff, for all ψ, if $A + \psi$ is consistent, then, there is a τ such that $B + [\psi]^\tau$ is consistent.
- $A \rhd_{\mathsf{faith}} B$ iff, there is a τ, such that, for all ϕ, $B \vdash \phi$ iff $A \vdash [\phi]^\tau$.
- $A \rhd_{\mathsf{lofa}} B$ iff, for all ϕ with $B \nvdash \phi$, there is a τ such that $A \vdash [B]^\tau$ and $A \nvdash [\phi]^\tau$.

We start with an elementary fact. Let Φ be the functor of Appendix A.9.1.

Theorem D.4. *i.* $A \rhd^* \Phi(A)$.
ii. $\Phi(A) \rhd_{\mathsf{faith}} A$.

Proof. Claim (i) is trivial by taking $\triangle_0$ to contain precisely one element. Let the interpretation we constructed in this way be Q_0 based on translation η_0.

For Claim (ii), we use the $\triangle_0$ as parameter set, $\triangle_1$ as domain and translate $P(\vec{x})$ as $(\bigwedge \triangle_1(x_i) \wedge P(z,\vec{x}))$. Let the interpretation we constructed by Q_1 based on η_1.

We confuse Q_0 and η_0 with there counterparts in the category of interpretations with parameters. Now it is clear that Q_1 is a section in $\mathsf{INT}_1^{\mathsf{par}}$ with inverse Q_0. Hence Q_1 is faithful. ❑

Theorem D.5. *We have:*

[14] The attentive reader will see, from the proof of Theorem D.10, that the ghost of the treatment of [MPS90] is still lingering.

1. *If* $A \rhd^* B$, *then* $A \rhd B$.
2. $A \rhd B$ *iff* $A \rhd^* \Phi(B)$.
3. *If* $A \Rsh^* B$, *then* $A \Rsh B$.
4. $A \Rsh B$ *iff* $A \Rsh^* \Phi(B)$.
5. *If* A *is locally tolerant without parameters, then* A *is locally tolerant.*
6. *If* A *is essentially locally tolerant without parameters, then* A *is essentially locally tolerant.*
7. *If* $A \blacktriangleright^*_{\mathsf{loc}} B$, *then* $A \blacktriangleright_{\mathsf{loc}} B$.
8. *If* $\Phi(A) \blacktriangleright^*_{\mathsf{loc}} B$, *then* $A \blacktriangleright_{\mathsf{loc}} B$.
9. *If* $A \rhd^*_{\mathsf{faith}} B$, *then* $A \rhd_{\mathsf{faith}} B$.
10. *If* $A \rhd^*_{\mathsf{faith}} \Phi(B)$, *then* $A \rhd_{\mathsf{faith}} B$.
11. *If* $A \rhd^*_{\mathsf{lofa}} B$, *then* $A \rhd_{\mathsf{lofa}} B$.
12. *If* $A \rhd^*_{\mathsf{lofa}} \Phi(B)$, *then* $A \rhd_{\mathsf{lofa}} B$.

Proof. All these claims are more or less trivial. We prove (7). Suppose $\Phi(A) \blacktriangleright^*_{\mathsf{loc}} B$. It follows that $\Phi(A) \blacktriangleright_{\mathsf{loc}} B$. On the other hand, we have $\Phi(A) \rhd_{\mathsf{faith}} A$ and, hence, $A \blacktriangleright_{\mathsf{loc}} \Phi(A)$. It follows that $A \blacktriangleright_{\mathsf{loc}} B$. ❑

Open Question D.6. Are there counterexamples to the converses of Theorem D.5 (3) and (5)–(12)? ❑

We turn to the characterisation of toleration.

Theorem D.7. $A \Rsh B$ *iff* $\exists C \supseteq A \ C \rhd B$.

Proof. Suppose $A \Rsh B$. Then for some τ, the theory $C := A + [B]^\tau$ is consistent. Hence $A \subseteq C \rhd B$.

Conversely, suppose $A \subseteq C$ and $C \rhd B$. Let τ witness the interpretability of B in C. Then certainly $A + [B]^\tau$ is consistent. ❑

We show that in our new context, local cointerpretability still is preservation of the tolerated.

Theorem D.8. $A \blacktriangleright_{\mathsf{loc}} B$ *iff, for all* C, *if* $A \Rsh C$, *then* $B \Rsh C$.

Proof. We will do the proof for the case without pieces. With pieces it is similar, but slightly more complicated.

Suppose $A \blacktriangleright_{\mathsf{loc}} B$. Consider any C and suppose $A \Rsh C$. This tells us that, for some σ, the theory $A + C^\sigma$ is consistent. It follows that, for some τ, the theory $B + [C^\sigma]^\tau$ is consistent. It follows that $B + [C]^{\sigma\tau}$ is consistent and, hence $B \Rsh C$.

Suppose that, for all C, if $A \Rsh C$, then $B \Rsh C$. Consider any ψ such that $A + \psi$ is consistent. We take $C := A + \psi$. Trivially, $A \Rsh C$. It follows that $B \Rsh C$, so, for some τ, we have that $B + [\psi]^\tau$ is consistent. ❑

As before we can reformulate our previous insight to: $A \blacktriangleright_{\mathsf{loc}} B$ iff $[\![A]\!] \subseteq [\![B]\!]$.

We turn to our second characterisation.

Theorem D.9. $A \blacktriangleright_{\mathsf{loc}} B$ *iff, for all* $A' \supseteq A$, *there is a* $B' \supseteq B$, *such that* $B' \rhd A'$.

Proof. We use Theorem D.8. Suppose that, for all C, if $A \Rsh C$, then $B \Rsh C$. Suppose $A' \supseteq A$. Then, the identity interpretation witnesses that $A \Rsh A'$. It follows that $B \Rsh A'$. Hence, for some τ, we have $B' := B + [A]^\tau$ is consistent. Clearly $B' \supseteq B$ and $B' \rhd A'$.

Conversely suppose that, for all $A' \supseteq A$, there is a $B' \supseteq B$, such that $B' \rhd A'$. Suppose ρ witnesses that $A \Rsh C$. Then, $A' := A + [C]^\rho$ is consistent. It follows that there is a $B' \supseteq B$ such that $B' \rhd A'$. Suppose $B' \rhd A'$ is witnessed by σ. We may conclude that $B + [[C]^\rho]^\sigma$ is consistent, and, hence that $B + [C]^{\rho\sigma}$ is consistent. So, $B \Rsh C$. ❑

We turn to the matter of local faith. We remind the reader that:

- $A \rhd_{\mathsf{lofa}} B$ iff, for all ϕ with $B \nvdash \phi$, there is a τ such that $A \vdash [B]^\tau$ and $A \nvdash [\phi]^\tau$.

Theorem D.10. *$A \rhd_{\mathsf{lofa}} B$ iff ($A \rhd B$ and $B \blacktriangleright_{\mathsf{loc}} A$).*

Proof. We use the characterisation of Theorem D.9.

Suppose $A \rhd_{\mathsf{lofa}} B$. By putting $\phi := \bot$ in the definition of $\rhd_{\mathsf{lofa}}$, we see that $A \rhd B$.

Consider any $B' \supseteq B$. Suppose B' is axiomatized by ϕ. Since B' is consistent, it follows that $B \nvdash \neg\phi$. So, there is a τ such that $A \vdash [B]^\tau$ and $A \nvdash [\neg\phi]^\tau$. We may now take $A' := A + \exists\vec{z}\,(\alpha_\tau(\vec{z}\,) \wedge \phi^{\tau,\vec{z}})$ or, equivalently, $A' := A + \neg[\neg\phi]^\tau$. We take τ' that is completely the same as τ but for the fact that $\alpha_{\tau'}(\vec{z}\,) := (\alpha_\tau(\vec{z}\,) \wedge \phi^{\tau,\vec{z}})$. We find: $A' \vdash [B \wedge \phi]^{\tau'}$. Thus, τ' witnesses that $A' \rhd B'$. We may conclude $B \blacktriangleright_{\mathsf{loc}} A$.

Suppose $A \rhd B$ and $B \blacktriangleright_{\mathsf{loc}} A$. Let τ_0 witness $A \rhd B$. Suppose $B \nvdash \phi$. If follows that $B' := B + \neg\phi$ is consistent. Then, for some $A' \supseteq A$ we have $A' \rhd B'$. Let τ be the witness of $A' \rhd B'$. We take $\tau^\star := \tau\langle[B']^\tau\rangle\tau_0$. We note that the disjunctive parameter domain of $\tau^\star$ needs to have as dimension the maximum of the dimensions of τ_0 and τ. We can obtain the desired effect by padding. Clearly, $A \vdash [B]^{\tau^\star}$.

Suppose $A \vdash [\phi]^{\tau^\star}$. Then, $A + [B']^\tau \vdash [\phi]^\tau$, since under the assumption $[B']^\tau$, the translations τ and $\tau^\star$ coincide. On the other hand, $A + [B']^\tau \vdash [\neg\phi]^\tau$, so $A + [B']^\tau \vdash [\bot]^\tau$. It follows that $A + [B']^\tau \vdash \bot$. But $(A + [B']^\tau) \subseteq A'$ and A' is consistent. So we have a contradiction. Ergo, $A \nvdash [\phi]^{\tau^\star}$. ❑

We treat the characterisations of local tolerance.

Theorem D.11. *The following are equivalent:*

i. A is locally tolerant.
ii. For all B, if $A \rhd B$, then $A \rhd_{\mathsf{lofa}} B$.
iii. $A \rhd_{\mathsf{lofa}} \mathsf{CQC}_2$, where CQC_2 is predicate logic for a binary relation symbol.

Proof. (i) $\Rightarrow$ (ii). Suppose A is locally tolerant and $A \rhd B$. Let τ_0 witness $A \rhd B$. Suppose $B \nvdash \phi$. Let $B' := B + \neg\phi$. We have $A \Rsh B'$. Let τ_1 be the witness of $A \Rsh B'$. We define $\tau^\star := \tau_1\langle[B']^{\tau_1}\rangle\tau_0$. Clearly, $A \vdash [B]^{\tau^\star}$. Suppose $A \vdash [\phi]^{\tau^\star}$. In that case, we must have $A + [B']^{\tau_1} \vdash [\bot]^{\tau_1}$, and, hence $A + [B']^{\tau_1} \vdash \bot$. Quod non.

(ii) $\Rightarrow$ (iii). This is immediate from the fact that $A \rhd \mathsf{CQC}_2$.

(iii) $\Rightarrow$ (i). We use a theorem from [Hod93] that any theory B is bi-interpretable with a theory $\widetilde{B}$ in the signature of one binary relation. Since $\mathsf{CQC}^2 \nvdash \neg\widetilde{B}$, it follows that $A + [\widetilde{B}]^\tau$ is consistent, for some τ. So, $A \Rsh \widetilde{B}$, and, hence, $A \Rsh B$. ❑

We note that also in the new setting we have:

Theorem D.12. *A is locally tolerant iff, for all B, we have $A \blacktriangleleft_{\mathsf{loc}} B$.*

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

E-mail address: a.visser@uu.nl