ON THE FRAGILITY OF INTERPOLATION

ANDRZEJ TARLECKI

Abstract. We study a version of the Craig interpolation theorem formulated in the framework of the theory of institutions. This formulation proved crucial in the development of a number of key results concerning foundations of software specification and formal development. We investigate preservation of interpolation properties under institution extensions by new models and sentences. We point out that some interpolation properties remain stable under such extensions, even if quite arbitrary new models and sentences are permitted. We give complete characterisations of such situations for institution extensions by new models, by new sentences, as well as by new models and sentences, respectively.

§1. Introduction.

The *Craig interpolation* theorem [13] states that when an implication $\varphi \Rightarrow \psi$ between premise φ and conclusion ψ holds then there is an interpolant θ built using the symbols the premise and the conclusion have in common (i.e., built on the intersection of the signatures of φ and of ψ, respectively) that witnesses this implication, that is, such that both $\varphi \Rightarrow \theta$ and $\theta \Rightarrow \psi$ hold. This is one of the fundamental properties of the classical first-order logic, with numerous consequences and links with other key properties developed in the framework of classical model theory [12].

In the area of foundations of system specification and formal development, interpolation proved indispensable for a number of most fundamental features of various approaches. This was perhaps first pointed out in [28], where it was used to ensure composability of subsequent implementation steps (later refined in various forms of so-called modularisation theorem [49, 50]). Perhaps better known is the work on module algebra [3], where the interpolation theorem was used to obtain crucial distributive laws for their export operator ([35] joined the two threads later). The standard by now proofs of completeness of proof calculi for consequences of structured specification rely on interpolation [6, 11] (in fact, no “good” sound and complete such proof calculus may exist without an appropriate interpolation property for the underlying logic [41]). These and further results concerning completeness of various reasoning systems necessary in the process of reliable software development involve interpolation explicitly, but the same idea that showing properties of a union of a number of extensions of a basic theory must rely on some form of interpolation (perhaps disguised as the *Robinson consistency* theorem [36]) is omnipresent in both practical and foundational aspects of computing.

Received September 9, 2023.

2020 *Mathematics Subject Classification.* Primary 03B70, 03C40, 03C95, 68N30, 68Q65.

Key words and phrases. interpolation, institutions, institutional abstract model theory, specification theory.

0022-4812/00/0000-0000
DOI:10.1017/jsl.2024.19

Applications of logic in computer science face the problem of dealing with a multitude of logical systems. This follows from the real needs of practical software development, based on the multitude of application areas as well as of practically important programming paradigms, features, and languages. In the area of software specification, this led to various attempts to abstract away from the details of a specific logical system in use. Such an independence of the foundations for software specification from the particulars of the underlying logic has been successfully achieved by relying on the concept of an institution, introduced by Goguen and Burstall as a formalisation of the notion of a logical system [26] (see, for instance, [40] for an exhaustive account of such ideas, with further examples in the development of specification formalisms such as CASL [1]).

Independently of these applications, it has been realised quite early that institutions offer a framework for developing a very abstract version of model theory, going beyond what has been studied within abstract model theory following [2]. This was noted in [42] and expanded in many crucial directions by Diaconescu and his group; Diaconescu's monograph [15] offers an extensive overview of this work, with later developments scattered through numerous articles (see, e.g., [18] and the references therein).

In particular, in the institutional model theory the interpolation property is formulated so that it can be studied (and used) for logical systems departing considerably from the first-order logic. This was put forward in [42], but we use here a still more refined formulation of interpolation given in [14, 38]. This formulation of the interpolation property uses logical entailment (rather than implication), sets of sentences (rather than individual sentences) and, most crucially, works over arbitrary commutative squares of signature morphisms (rather than over union/intersection squares only). Consequently, it caters for instance for logical systems where one lacks compactness, conjunction and other classical connectives, and even the concept of the set of symbols used in a formula and intersection/union of signatures may not be directly available. Indeed, the key point of many of the applications mentioned above is the need to abstract away from signature inclusions and deal with interpolation properties with respect to other signature morphisms. For instance, non-injective signature morphisms are of practical importance when parameterised specifications with the standard pushout-style parameter passing are considered [21, 47]. Much subsequent work used this formulation, and included development of generic model-theoretic proof techniques to establish interpolation property for logical systems formalised as institutions satisfying a number of structural properties. This led to new results concerning various logical systems, as well as to studying interpolation in even more general context of non-standard entailment relations [7, 14, 17, 19, 23–25, 34].

The need for the use of many logical systems leads to the need for establishing their required properties, including the interpolation property we study here. Rather than establishing such properties for each system anew, it is desirable to ensure them in the course of systematic construction of new logics, perhaps along the lines aimed at for instance in [31, 32] or [8–10]. Typically, the new logics are linked with the original ones by institution (co)morphisms [26, 27]. An important line of research here was to clarify sufficient conditions on the institution (co)morphisms that allow interpolation properties to be "borrowed" from one institution by another [16, 23].

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

In this paper we address a perhaps more basic question that arises in this framework: namely, whether interpolation properties can be spoiled by extending a logic by new abstract models or sentences. Looking at the standard formulation of the Craig interpolation property, it seems that the answer is always positive: given a true implication, to spoil an interpolant for its premise and conclusion, just add a new abstract model that satisfies the premise but not the interpolant, or a new abstract model that satisfies the interpolant but not the conclusion, thus spoiling the required entailment between the premise and the interpolant, or between the interpolant and the conclusion, respectively. This should work, except for the trivial cases when the signature of the premise includes or is included in the signature of the conclusion. At a closer look though, in the framework where one considers arbitrary morphisms between the signatures involved, when we add new models for the signature of the premise or for the signature of the conclusion, new models for their union signature may emerge (as reducts w.r.t. some signature morphisms of the models added explicitly) and ruin the implication between the premise and the conclusion one starts with.

We explore the consequences of this phenomena, and characterise exactly the situations where interpolation is stable under extensions of the institution. Equivalently, looking at the other side of this coin, we characterise the situations where new models or sentences may spoil the interpolation property. More precisely: we consider separately institution extensions where only new models, only new sentences, and both new models and sentences, respectively, are permitted. In each of these three cases complete characterisations are given, formulating necessary and sufficient conditions for a commutative square of signature morphisms under which no such institution extension may spoil interpolation properties over this square.

Similar characterisations are then derived for a natural finitary version of the Craig interpolation property, where the interpolant set of sentences is required to be finite (or, somewhat more generally, of a bounded cardinality) for any sets of premises and conclusions similarly bounded.

Finally, we study the so-called *Craig–Robinson* (or *parameterised*) *interpolation*, which is in general stronger than the Craig interpolation and is in fact needed in many applications, for instance in the area of software specifications and development [15, 20, 40]. Similar complete characterisations are obtained, with an interesting difference concerning the characterisation of commutative squares of signature morphisms that admit interpolation in any extension of the institution by new models and sentences, when a certain symmetry in the role of the premise and conclusion signatures, present in the classical Craig interpolation, breaks down.

§2. Institutions.

2.1. Notational preliminaries. For any function $f: X \to Y$, given a set $X' \subseteq X$, $f(X') = \{f(x) \mid x \in X'\} \subseteq Y$ is the *image* of X' w.r.t. f, and for $Y' \subseteq Y$, $f^{-1}(Y') = \{x \in X \mid f(x) \in Y'\}$ is the *coimage* of Y' w.r.t. f.

Throughout the paper we freely use the basic notions from category theory (*category*, *functor*, *natural transformation*, *pushout*, etc.). Composition in any category is denoted by ";" (semicolon) and written in the diagrammatic order. For instance, $f: A \to B$ is a *retraction* if for some $g: B \to A$ we have $g;f = id_B$, and

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

$f : A \to B$ is a *coretraction* (or *section*) if for some $g : B \to A$ we have $f;g = id_A$. The collection of objects of any category $\mathbf{K}$ is written as $|\mathbf{K}|$. The category of sets is denoted by **Set**, and the (quasi-)category of classes (or discrete categories) by **Class**.

2.2. Institutions. In the area of foundations of software specification and development [40] it is standard by now to abstract away from the details of the logical system in use, relying on the formalisation of a logical system as an *institution* [26]. An institution $\mathbf{I}$ consists of:

- a category $\mathbf{Sig_I}$ of *signatures*;
- a functor $\mathbf{Sen_I}: \mathbf{Sig_I} \to \mathbf{Set}$, giving a set $\mathbf{Sen_I}(\Sigma)$ of Σ-*sentences* for each signature $\Sigma \in |\mathbf{Sig_I}|$;
- a functor $\mathbf{Mod_I}: \mathbf{Sig_I}^{op} \to \mathbf{Class}$, giving a class (or a discrete category)[1] $\mathbf{Mod_I}(\Sigma)$ of Σ-*models* for each signature $\Sigma \in |\mathbf{Sig_I}|$; and
- a family $\langle \models_{\mathbf{I},\Sigma} \subseteq \mathbf{Mod_I}(\Sigma) \times \mathbf{Sen_I}(\Sigma)\rangle_{\Sigma \in |\mathbf{Sig_I}|}$ of *satisfaction relations*

such that the *reducts* $\mathbf{Mod_I}(\sigma): \mathbf{Mod_I}(\Sigma') \to \mathbf{Mod_I}(\Sigma)$ of models and *translations* $\mathbf{Sen_I}(\sigma): \mathbf{Sen_I}(\Sigma) \to \mathbf{Sen_I}(\Sigma')$ of sentences induced any signature morphism $\sigma: \Sigma \to \Sigma'$ preserve the satisfaction relation, that is, for any $\varphi \in \mathbf{Sen_I}(\Sigma)$ and $M' \in \mathbf{Mod_I}(\Sigma')$ the following *satisfaction condition* holds:

$$M' \models_{\mathbf{I},\Sigma'} \mathbf{Sen_I}(\sigma)(\varphi) \quad \text{iff} \quad \mathbf{Mod_I}(\sigma)(M') \models_{\mathbf{I},\Sigma} \varphi.$$

The subscripts $\mathbf{I}$ and Σ are typically omitted. For any signature morphism $\sigma: \Sigma \to \Sigma'$, the translation $\mathbf{Sen}(\sigma): \mathbf{Sen}(\Sigma) \to \mathbf{Sen}(\Sigma')$ is often denoted by $\sigma: \mathbf{Sen}(\Sigma) \to \mathbf{Sen}(\Sigma')$, and the reduct $\mathbf{Mod}(\sigma): \mathbf{Mod}(\Sigma') \to \mathbf{Mod}(\Sigma)$ by $\_|_\sigma: \mathbf{Mod}(\Sigma') \to \mathbf{Mod}(\Sigma)$. For instance, combining this with the notation for image and coimage, for $\Phi \subseteq \mathbf{Sen}(\Sigma)$, $\sigma(\Phi) = \{\sigma(\varphi) \mid \varphi \in \Phi\} \subseteq \mathbf{Sen}(\Sigma')$, and for $\mathcal{M} \subseteq \mathbf{Mod}(\Sigma)$, $\mathcal{M}|_\sigma^{-1} = \{M' \in \mathbf{Mod}(\Sigma') \mid M'|_\sigma \in \mathcal{M}\} \subseteq \mathbf{Mod}(\Sigma')$, and the satisfaction condition may be re-stated as: $M' \models \sigma(\varphi)$ iff $M'|_\sigma \models \varphi$.

For any signature Σ, the satisfaction relation extends naturally to sets of Σ-sentences and classes of Σ-models. For any set $\Phi \subseteq \mathbf{Sen}(\Sigma)$, the *class of models of* Φ is $Mod(\Phi) = \{M \in \mathbf{Mod}(\Sigma) \mid M \models \Phi\}$ (such classes of models are called *definable*), and for any class $\mathcal{M} \subseteq \mathbf{Mod}(\Sigma)$, the *theory of* $\mathcal{M}$ is $Th(\mathcal{M}) = \{\varphi \in \mathbf{Sen}(\Sigma) \mid \mathcal{M} \models \varphi\}$. The latter notation is also used for the *theory generated by a set of sentences*: for $\Phi \subseteq \mathbf{Sen}(\Sigma)$, $Th(\Phi) = Th(Mod(\Phi))$.

As usual, each satisfaction relation determines (semantic) entailment between sets of sentences: $\Phi \subseteq \mathbf{Sen}(\Sigma)$ *entails* $\Psi \subseteq \mathbf{Sen}(\Sigma)$ (or Ψ *is a consequence of* Φ), written $\Phi \models \Psi$, when $\Psi \subseteq Th(\Phi)$. The satisfaction condition implies that the semantic entailment is preserved under translation along signature morphisms: for any $\sigma: \Sigma \to \Sigma'$, if $\Phi \models \Psi$ then $\sigma(\Phi) \models \sigma(\Psi)$. If the opposite implication holds as well, i.e., $\Phi \models \Psi$ iff $\sigma(\Phi) \models \sigma(\Psi)$ for all $\Phi, \Psi \subseteq \mathbf{Sen}(\Sigma)$, we say that $\sigma: \Sigma \to \Sigma'$ is *conservative*. In particular, if the reduct $\_|_\sigma: \mathbf{Mod}(\Sigma') \to \mathbf{Mod}(\Sigma)$ is surjective then $\sigma: \Sigma \to \Sigma'$ is conservative.[2]

[1] We disregard here model morphisms, which are crucial in many applications of the notion of institution [15, 40], but for the purposes of this paper are irrelevant.

[2] The terminology varies; some authors use the term "conservative" for signature morphism that induce surjective reducts [27]. The more permissive definition used here seems closer to the standard definition of a conservative theory interpretation [12].

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

We typically decorate the names for institution components and for other derived notions by primes, indices, etc., to identify the institution they refer to, and rely on this convention whenever the institution is clear from the context. So, for instance, $\mathbf{Mod}_1$ is the model functor in an institution $\mathbf{I}_1$, $\models'$ is the satisfaction relation (and entailment) in $\mathbf{I}'$, etc.

Examples of institutions abound, see, for instance, [15, 40] for detailed definitions of many standard and not so standard logical systems formalised as institutions. Here, let us just sketch three standard examples.

Example 2.1. The institution **FO** of (many-sorted) *first-order logic* has signatures that consist of a set of sort names, a set of operation names with an arity (given as a finite sequence of sort names) and a result sort indicated for each operation name, and a set of predicate names with an arity indicated for each predicate name. We consider finite signatures only, with all symbols taken from a predefined (infinite) vocabulary, which makes the category of signatures small. Terms are built from variables by "formal application" of operation names respecting their arities and result sorts (constants are nullary operations). Then atomic formulae are predicate "applications" to tuples of terms of the sorts indicated by the predicate arities, and first-order formulae are built from those using the usual Boolean connectives (including nullary false) and quantification. First-order sentences are closed formulae (i.e., formulae with no free occurrences of variables). We assume that in each sentence variables of different sorts are distinct. First-order models consist of many-sorted carrier sets (one set for each sort name), functions to interpret operation names and relations to interpret predicate names, in accordance with the indicated arities and result sorts. Satisfaction of first-order sentences in first-order models so built is defined as usual. Finally, signature morphisms map sort names to sort names, operation names to operation names and predicate names to predicate names preserving their arities and result sorts. For any such morphism, translation of sentences is defined by renaming sorts (for variables), operation and predicate names as indicated by the morphism, and reducts of models are defined by interpreting each symbol of the source signature as the symbol the signature morphism maps it to is interpreted in the argument model. The satisfaction condition holds, and this indeed defines an institution [26]. We will assume that all carrier sets in first-order models are nonempty. The variant of first-order logic where empty carrier sets are allowed in models will be denoted by $\mathbf{FO}_\emptyset$.[3] Another variant is the institution $\mathbf{FO}_{\mathbf{EQ}}$ of first-order logic with equality, where we have a binary equality predicate for each sort, interpreted as the identity relation in all models.

Example 2.2. The institution **EQ** of (many-sorted) *equational logic* may be defined as the restriction of the institution $\mathbf{FO}_{\mathbf{EQ}}$ of first-order logic with equality to the signatures with no predicates other than equalities (models are usually called algebras then), and sentences are limited to universally quantified equalities. Again, $\mathbf{EQ}_\emptyset$ is the variant of **EQ** where empty carriers are permitted (see [15, 40] for a more explicit definition).

[3] The distinction between **FO** and $\mathbf{FO}_{\mathbf{EQ}}$ does not matter much, since the (non-)emptiness of the carrier of any sort may be captured by a logical sentence. However, this is in contrast with equational logic, sketched in Example 2.2, where the same distinction is crucial and leads to different properties of the logic (see, for instance, Example 3.1 and [45]).

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

Example 2.3. The institution **PL** of *propositional logic* may be viewed as a restriction of the institution of first-order logic to signatures with no sort names (and hence no operation names and nullary predicates only). More explicitly, **PL** has finite sets of *propositional variables* as signatures, with signature morphisms being arbitrary functions between those sets. Propositional sentences are built from propositional variables using the usual Boolean connectives (with obvious translations under functions renaming propositional variables). Models over a signature are given as subsets of this signature (consisting of the propositional variables that are satisfied in the model) with reducts w.r.t. signature morphisms given as their coimage. With the usual satisfaction of propositional sentences in such models, the satisfaction condition is easy to check.

In the above sample institutions **FO**, **EQ**, and **PL** all injective signature morphisms induce surjective reducts, and so are conservative. This need not be the case for non-injective morphisms. However, in $\mathbf{FO}_\emptyset$ in $\mathbf{EQ}_\emptyset$, the variants of **FO** and of **EQ** where empty carriers of some sorts are permitted in models, not all injective signature morphisms are conservative.

In the examples above, and in many other standard cases, all the signatures, sentences, and models are quite familiar, and link with many intuitions and implicit assumptions. We should stress though that when exploiting the generality of the concept and working with an arbitrary institution, such connotations should be dropped. All the entities involved (signatures, their morphisms, sentences, models, satisfaction relations) are considered entirely abstract, with completely unknown structure and properties. It is perhaps surprising how far one can go with developments of the foundations for software specification [40] and an abstract version of model theory [15] in such an abstract setting.

2.3. Extending institutions by models and sentences. We introduce two basic ways of extending institutions, by adding new "abstract" models, and new "abstract" sentences, respectively. The definitions are shaped after the definition of *constraints* in [26, 40]. The basic observation is that when a new sentence is to be added to the set of sentences over a signature, with some predefined notion of satisfaction in the institution models, it must also be "fitted" to other signatures to mimic its translation along signature morphisms with this signature as a source. Hence, together with each new sentence, we also add its "formal translations" along signature morphisms. Then, the satisfaction of the formal translations so added is determined by the satisfaction condition. Similarly, when we want to add new models to the class of models over a signature—apart from the new models themselves, we must also add their "formal reducts".

Consider an arbitrary institution $\mathbf{I} = \langle \mathbf{Sig}, \mathbf{Sen}, \mathbf{Mod}, \langle \models_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|} \rangle$.

Suppose that for each signature we are given a set of (new) "sentences" with predefined satisfaction relation in **I**-models, which may be organised as a signature-indexed family of sets with relations between the model classes and these sets: $\mathcal{NS} = \langle \mathcal{NS}_\Sigma, \models^{\mathcal{NS}}_\Sigma \subseteq \mathbf{Mod}(\Sigma) \times \mathcal{NS}_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|}$.[4]

[4] We disregard foundational problems that may arise here: in general the collection $\mathbf{Sen}^+(\Sigma)$ defined below may turn out to be a proper class (not a set). One way around this is to work with a more general notion of institution, where classes (rather than sets) of sentences over any signature are allowed.

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

We define the *extension of* **I** *by sentences* $\mathcal{NS}$ to be the institution $\mathbf{I}^+ = \langle \mathbf{Sig}, \mathbf{Sen}^+, \mathbf{Mod}, \langle \models^+_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|} \rangle$, where for $\Sigma \in |\mathbf{Sig}|$, $\mathbf{Sen}^+(\Sigma) = \mathbf{Sen}(\Sigma) \cup \{ \lceil \tau(\varphi') \rceil \mid \varphi' \in \mathcal{NS}_{\Sigma'}, \tau \colon \Sigma' \to \Sigma \}$.[5] Then for $M \in \mathbf{Mod}(\Sigma)$, $M \models^+_\Sigma \varphi$ iff $M \models_\Sigma \varphi$ for $\varphi \in \mathbf{Sen}(\Sigma)$, and for $\varphi' \in \mathcal{NS}_{\Sigma'}, \tau \colon \Sigma' \to \Sigma$, we define $M \models^+_\Sigma \lceil \tau(\varphi') \rceil$ to hold iff $M\big|_\tau \models^{\mathcal{NS}}_{\Sigma'} \varphi'$. Finally, for any signature morphism $\sigma \colon \Sigma \to \Sigma''$, $\mathbf{Sen}^+(\sigma)(\varphi) = \mathbf{Sen}(\sigma)(\varphi)$ for $\varphi \in \mathbf{Sen}(\Sigma)$, and for $\varphi' \in \mathcal{NS}_{\Sigma'}$, $\tau \colon \Sigma' \to \Sigma$, we define $\mathbf{Sen}^+(\sigma)(\lceil \tau(\varphi') \rceil) = \lceil (\tau;\sigma)(\varphi') \rceil$.

This defines an institution, where for $\Sigma \in |\mathbf{Sig}|$, the new sentences $\varphi \in \mathcal{NS}_\Sigma$ are present as $\lceil id_\Sigma(\varphi) \rceil$. Clearly, such an extension does not affect semantic entailments between sets of sentences of the original institution.

Institution extensions by new sentences compose in the following sense: if $\mathbf{I}^{++}$ is an extension by new sentences of an extension $\mathbf{I}^+$ of **I** by new sentences then $\mathbf{I}^{++}$ is an extension of **I** by sentences (the union of the sets of new sentences added in each step should be used for each signature). Note also that **I** is its own extension by (the empty set of) new sentences.

Suppose then that for each signature we are given a class of (new) "models" with predefined satisfaction relation for **I**-sentences, organised as a signature-indexed family of classes with relations between these classes and the sets of sentences: $\mathcal{NM} = \langle \mathcal{NM}_\Sigma, \models^{\mathcal{NM}}_\Sigma \subseteq \mathcal{NM}_\Sigma \times \mathbf{Sen}(\Sigma) \rangle_{\Sigma \in |\mathbf{Sig}|}$.

Then we define the *extension of* **I** *by models* $\mathcal{NM}$ to be the institution $\mathbf{I}^+ = \langle \mathbf{Sig}, \mathbf{Sen}, \mathbf{Mod}^+, \langle \models^+_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|} \rangle$, where for $\Sigma \in |\mathbf{Sig}|$, $\mathbf{Mod}^+(\Sigma) = \mathbf{Mod}(\Sigma) \cup \{ \lceil M'\big|_\tau \rceil \mid M' \in \mathcal{NM}_{\Sigma'}, \tau \colon \Sigma \to \Sigma' \}$.[6] Then for $\varphi \in \mathbf{Sen}(\Sigma)$, $M \models^+_\Sigma \varphi$ iff $M \models_\Sigma \varphi$ for $M \in \mathbf{Mod}(\Sigma)$, and for $M' \in \mathcal{NM}_{\Sigma'}$, $\tau \colon \Sigma \to \Sigma'$, we define $\lceil M'\big|_\tau \rceil \models^+_\Sigma \varphi$ to hold iff $M' \models^{\mathcal{NM}}_{\Sigma'} \tau(\varphi)$. Finally, for any signature morphism $\sigma \colon \Sigma'' \to \Sigma$, $\mathbf{Mod}^+(\sigma)(M) = M\big|_\sigma$ for $M \in \mathbf{Mod}(\Sigma)$, and for $M' \in \mathcal{NS}_{\Sigma'}$, $\tau \colon \Sigma \to \Sigma'$, we define $\mathbf{Mod}^+(\sigma)(\lceil M'\big|_\tau \rceil) = \lceil M'\big|_{\sigma;\tau} \rceil$.

This defines an institution, where for $\Sigma \in |\mathbf{Sig}|$, the new models $M \in \mathcal{NM}_\Sigma$ are present as $\lceil M\big|_{id_\Sigma} \rceil$. Clearly, such an extension mail spoil some of the semantic entailments between sets of sentences of the original institution: for $\Sigma \in |\mathbf{Sig}|$, $\Phi, \Psi \subseteq \mathbf{Sen}(\Sigma)$ if $\Phi \models^+ \Psi$ then $\Phi \models \Psi$ but the opposite may fail in general (this is in contrast with institution extensions by sentences).

Institutions extensions by new models compose in the following sense: if $\mathbf{I}^{++}$ is an extension by new models of an extension $\mathbf{I}^+$ of **I** by new models then $\mathbf{I}^{++}$ is an extension of **I** by models (the union of the classes of new models added in each step should be used for each signature). Note also that **I** is its own extension by (the empty class of) new models.

In the rest of this paper we will use the above constructions presenting new sentences $\mathcal{NS}$ and new models $\mathcal{NM}$ somewhat informally, avoiding much of the notational burden. In particular, we will disregard the formal distinction between

Alternatively, we may assume that **Sig** is small, or that it is locally small and $\mathcal{NS}_\Sigma \neq \emptyset$ for a set of signatures Σ only.

[5] $\lceil \tau(\varphi') \rceil$ is our syntax for the sentence $\varphi' \in \mathcal{NS}_{\Sigma'}$ formally "fitted" by $\tau \colon \Sigma' \to \Sigma$ to the signature Σ; we assume that no sentences of the form $\lceil \tau(\varphi') \rceil$ are present in **I**.

[6] $\lceil M'\big|_\tau \rceil$ is our syntax for the model $M' \in \mathcal{NM}_{\Sigma'}$ formally "fitted" by $\tau \colon \Sigma \to \Sigma'$ to the signature Σ; we assume that no models of the form $\lceil M'\big|_\tau \rceil$ are present in **I**.

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

$\varphi \in \mathcal{NS}_\Sigma$ and $\lceil id_\Sigma(\varphi)\rceil$, as well as between $M \in \mathcal{NM}_\Sigma$ and $\lceil M|_{id_\Sigma}\rceil$. For $\Sigma \in |\mathbf{Sig}|$, we may also define the satisfaction relations $\models_\Sigma^{\mathcal{NS}}$ indirectly by defining $Mod^+(\varphi) \subseteq \mathbf{Mod}(\Sigma)$ for each $\varphi \in \mathcal{NS}_\Sigma$ (then for $M \in \mathbf{Mod}(\Sigma)$, $M \models_\Sigma^{\mathcal{NS}} \varphi$ iff $M \in Mod^+(\Sigma)$), and $\models_\Sigma^{\mathcal{NM}}$ by defining $Th^+(M) \subseteq \mathbf{Sen}(\Sigma)$ for each $M \in \mathcal{NM}_\Sigma$ (then for $\varphi \in \mathbf{Sen}(\Sigma)$, $M \models_\Sigma^{\mathcal{NM}} \varphi$ iff $\varphi \in Th^+(M)$).

EXAMPLE 2.4. We may define an extension of the institution **PL** of propositional logic (see Example 2.3) by sentences, adding for each signature Σ a new sentence $\mathtt{even}_\Sigma$, with the satisfaction relation extended so that $M \models^+ \mathtt{even}_\Sigma$ if M contains an even number of propositional variables (an even number of propositional variables holds in M). In the resulting extension $\mathbf{PL}^+$ defined as above, for a signature morphism (which is a function between the sets of propositional variables) $\sigma\colon \Sigma \to \Sigma'$, $\mathbf{Sen}^+(\sigma)(\mathtt{even}_\sigma)$ is $\lceil\sigma(\mathtt{even}_\Sigma)\rceil$, which is distinct from $\mathtt{even}_{\Sigma'}$. Indeed, putting $\mathbf{Sen}^+(\sigma)(\mathtt{even}_\Sigma) = \mathtt{even}_{\Sigma'}$ would violate the satisfaction condition for some σ.

EXAMPLE 2.5. We may also define an extension of the institution **PL** of propositional logic by models, adding for each signature Σ and Σ-model M, a new model $\widetilde{M}$, where the satisfaction of propositional sentences in $\widetilde{M}$ is defined by interpreting propositional connectives as usual, but the truth of all occurrences of propositional variables is determined separately for each occurrence, from left to right, and after each occurrence the values of all propositional variables are "swapped" (from true to false and vice versa). Thus, for instance the sentence $p \wedge q$ holds in $\widetilde{M}$ if $p \in M$ and $q \notin M$, and $p \vee p$ holds in any model $\widetilde{M}$. In the resulting extension $\mathbf{PL}^+$, for any signature Σ and $M \in \mathbf{Mod}(\Sigma)$, for any signature morphism $\sigma\colon \Sigma' \to \Sigma$, $\widetilde{M}|_\sigma$ (that is, $\mathbf{Mod}^+(\sigma)(\widetilde{M})$) and $\widetilde{M|_\sigma}$ are distinct Σ'-models, even though one may easily check that they satisfy exactly the same propositional sentences.

2.4. Institution morphisms. There are a number of standard notions to capture relationships between different institutions, with institution morphisms [26] and comorphisms [27] (plain maps [29] or representations [43]) perhaps the most common.

Let $\mathbf{I} = \langle \mathbf{Sig}, \mathbf{Sen}, \mathbf{Mod}, \langle\models_\Sigma\rangle_{\Sigma\in|\mathbf{Sig}|}\rangle$ and $\mathbf{I}' = \langle \mathbf{Sig}', \mathbf{Sen}', \mathbf{Mod}', \langle\models'_{\Sigma'}\rangle_{\Sigma'\in|\mathbf{Sig}'|}\rangle$ be institutions. An *institution morphism* $\mu\colon \mathbf{I} \to \mathbf{I}'$ consists of:

- a functor $\mu^{Sig}\colon \mathbf{Sig} \to \mathbf{Sig}'$,
- a natural transformation $\mu^{Sen}\colon \mu^{Sig};\mathbf{Sen}' \to \mathbf{Sen}$, i.e., a family of functions $\mu_\Sigma^{Sen}\colon \mathbf{Sen}'(\mu^{Sig}(\Sigma)) \to \mathbf{Sen}(\Sigma)$ natural in $\Sigma \in |\mathbf{Sig}|$, and
- a natural transformation $\mu^{Mod}\colon \mathbf{Mod} \to (\mu^{Sig})^{op};\mathbf{Mod}'$, i.e., a family of functions $\mu_\Sigma^{Mod}\colon \mathbf{Mod}(\Sigma) \to \mathbf{Mod}'(\mu^{Sig}(\Sigma))$ natural in $\Sigma \in |\mathbf{Sig}|$

such that for any signature $\Sigma \in |\mathbf{Sig}|$, $\varphi' \in \mathbf{Sen}'(\mu^{Sig}(\Sigma))$, and $M \in \mathbf{Mod}(\Sigma)$, $M \models_\Sigma \mu_\Sigma^{Sen}(\varphi')$ iff $\mu_\Sigma^{Mod}(M) \models'_{\mu^{Sig}(\Sigma)} \varphi'$ (this is referred to as the *satisfaction condition* for μ).

To simplify the notation, all three components of an institution morphism μ are typically denoted by μ as well, omitting the superscripts whenever they are clear from the context.

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

It follows that semantic entailment is preserved by translation under institution morphisms: for any signature $\Sigma \in |\mathbf{Sig}|$ and sets of sentences $\Phi', \Psi' \subseteq \mathbf{Sen}'(\mu(\Sigma))$, if $\Phi' \models' \Psi'$ then $\mu_\Sigma(\Phi') \models \mu_\Sigma(\Psi')$. Moreover, if the translation of models $\mu_\Sigma \colon \mathbf{Mod}(\Sigma) \to \mathbf{Mod}'(\mu(\Sigma))$ is surjective then the opposite implication holds as well, that is, $\Phi' \models' \Psi'$ iff $\mu_\Sigma(\Phi') \models \mu_\Sigma(\Psi')$.

For instance, there is an obvious institution morphisms from the institution **FO** of first-order logic to the institution **PL** of propositional logic (removing from signatures everything but nullary predicates). For further examples of institution morphisms spelled out in detail we refer to [15, 40].

Throughout this paper we deal with a special case of institution morphisms that leave the signature category intact, that is, where the signature functor is the identity. This also allows us to disregard institution comorphisms, since in this case the two notions are essentially the same (institution morphisms from **I** to **I**$'$ with the identity signature functor coincide with comorphisms from **I**$'$ to **I** with the identity signature functor).

An institution morphism $\mu \colon \mathbf{I} \to \mathbf{I}'$ is *logically trivial* if it is the identity on signatures and surjective on sentences and models, that is, $\mathbf{Sig}' = \mathbf{Sig}$ and $\mu^{Sig} = id_{\mathbf{Sig}}$, and for all signatures $\Sigma \in |\mathbf{Sig}|$, the functions $\mu_\Sigma \colon \mathbf{Sen}'(\Sigma) \to \mathbf{Sen}(\Sigma)$ and $\mu_\Sigma \colon \mathbf{Mod}(\Sigma) \to \mathbf{Mod}'(\Sigma)$ are surjective.

PROPOSITION 2.6. *Logically trivial institution morphisms identify only sentences and models that are logically equivalent, that is, if an institution morphism $\mu \colon \mathbf{I} \to \mathbf{I}'$ is logically trivial then for any signature $\Sigma \in |\mathbf{Sig}|$:*

1. *for any **I**$'$-sentences $\varphi', \psi' \in \mathbf{Sen}'(\Sigma)$, if $\mu_\Sigma(\varphi') = \mu_\Sigma(\psi')$ then for all **I**$'$-models $M' \in \mathbf{Mod}'(\Sigma)$, $M' \models' \varphi'$ iff $M' \models' \psi'$;*
2. *for any **I**-models $M, N \in \mathbf{Mod}(\Sigma)$, if $\mu_\Sigma(M) = \mu_\Sigma(N)$ then for all **I**-sentences $\varphi \in \mathbf{Sen}(\Sigma)$, $M \models \varphi$ iff $N \models \varphi$.*

PROOF. Follows by the satisfaction condition for $\mu \colon \mathbf{I} \to \mathbf{I}'$ and surjectivity of $\mu_\Sigma \colon \mathbf{Mod}(\Sigma) \to \mathbf{Mod}'(\Sigma)$ and $\mu_\Sigma \colon \mathbf{Sen}'(\Sigma) \to \mathbf{Sen}(\Sigma)$:

1. Suppose $\varphi = \mu_\Sigma(\varphi') = \mu_\Sigma(\psi')$. Since $\mu_\Sigma \colon \mathbf{Mod}(\Sigma) \to \mathbf{Mod}'(\Sigma)$ is surjective, for any $M' \in \mathbf{Mod}'(\Sigma)$ there is $M \in \mathbf{Mod}(\Sigma)$ such that $\mu_\Sigma(M) = M'$. Hence, by the satisfaction condition for $\mu \colon \mathbf{I} \to \mathbf{I}'$, $M' \models' \varphi'$ iff $M \models \varphi$ iff $M' \models' \psi'$.
2. Similarly, suppose $\mu_\Sigma(M) = \mu_\Sigma(N) = M'$. Since $\mu_\Sigma \colon \mathbf{Sen}'(\Sigma) \to \mathbf{Sen}(\Sigma)$ is surjective, for any $\varphi \in \mathbf{Sen}(\Sigma)$ there is $\varphi' \in \mathbf{Sen}'(\Sigma)$ such that $\mu_\Sigma(\varphi') = \varphi$. Hence, by the satisfaction condition for $\mu \colon \mathbf{I} \to \mathbf{I}'$, $M \models \varphi$ iff $M' \models' \varphi'$ iff $N \models \varphi$. $\dashv$

Special institution morphisms relate institutions with their extensions by new sentences and by new models, respectively, introduced in Section 2.3.

Let $\mathbf{I}^+_{\mathcal{NS}}$ be the extension of institution $\mathbf{I} = \langle \mathbf{Sig}, \mathbf{Sen}, \mathbf{Mod}, \langle \models_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|} \rangle$ by sentences $\mathcal{NS} = \langle \mathcal{NS}_\Sigma, \models^{\mathcal{NS}}_\Sigma \subseteq \mathbf{Mod}(\Sigma) \times \mathcal{NS}_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|}$, as defined in Section 2.3. Then there is an obvious institution morphism $\mu_{\mathcal{NS}} \colon \mathbf{I}^+_{\mathcal{NS}} \to \mathbf{I}$, where $\mu^{Sig}_{\mathcal{NS}}$ and $\mu^{Mod}_{\mathcal{NS}}$ are identities (the former is the identity functor on **Sig**, the latter is the identity natural transformation on $\mathbf{Mod} \colon \mathbf{Sig}^{op} \to \mathbf{Class}$), and for $\Sigma \in |\mathbf{Sig}|$, $(\mu^{Sen}_{\mathcal{NS}})_\Sigma \colon \mathbf{Sen}(\Sigma) \to \mathbf{Sen}^+_{\mathcal{NS}}(\Sigma)$ are inclusions. Somewhat ambiguously, we refer to this institution morphism as the extension of **I** by $\mathcal{NS}$ as well.

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

Similarly, let $\mathbf{I}^+_{\mathcal{NM}}$ be the extension of $\mathbf{I} = \langle \mathbf{Sig}, \mathbf{Sen}, \mathbf{Mod}, \langle \models_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|} \rangle$ by models $\mathcal{NM} = \langle \mathcal{NM}_\Sigma, \models^{\mathcal{NM}}_\Sigma \subseteq \mathcal{NM}_\Sigma \times \mathbf{Sen}(\Sigma) \rangle_{\Sigma \in |\mathbf{Sig}|}$, as defined in Section 2.3. There is an obvious institution morphism $\mu_{\mathcal{NM}} \colon \mathbf{I} \to \mathbf{I}^+_{\mathcal{NM}}$, where $\mu^{Sig}_{\mathcal{NM}}$ and $\mu^{Sen}_{\mathcal{NM}}$ are identities, and for $\Sigma \in |\mathbf{Sig}|$, $(\mu^{Mod}_{\mathcal{NM}})_\Sigma \colon \mathbf{Mod}(\Sigma) \to \mathbf{Mod}^+_{\mathcal{NM}}(\Sigma)$ are inclusions. We also refer to this institution morphism as the extension of $\mathbf{I}$ by $\mathcal{NM}$.

Institution morphisms compose in the obvious, component-wise manner [26].

Proposition 2.7. *Consider institutions* $\mathbf{I}' = \langle \mathbf{Sig}, \mathbf{Sen}', \mathbf{Mod}', \langle \models'_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|} \rangle$ *and* $\mathbf{I}'' = \langle \mathbf{Sig}, \mathbf{Sen}'', \mathbf{Mod}'', \langle \models''_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|} \rangle$ *with a common signature category, and an institution morphism* $\mu \colon \mathbf{I}' \to \mathbf{I}''$ *with* $\mu^{Sig} = id_{\mathbf{Sig}}$. *Then for some institution* $\mathbf{I}$, *extension* $\mathbf{I}^+_{\mathcal{NS}}$ *of* $\mathbf{I}$ *by new sentences, extension* $\mathbf{I}^+_{\mathcal{NM}}$ *of* $\mathbf{I}$ *by new models, and logically trivial institution morphisms* $\mu' \colon \mathbf{I}' \to \mathbf{I}^+_{\mathcal{NS}}$ *and* $\mu'' \colon \mathbf{I}^+_{\mathcal{NM}} \to \mathbf{I}''$ *we have* $\mu = \mu';\mu_{\mathcal{NS}};\mu_{\mathcal{NM}};\mu''$:

$$\underbrace{\mathbf{I}' \xrightarrow{\mu'} \mathbf{I}^+_{\mathcal{NS}} \xrightarrow{\mu_{\mathcal{NS}}} \mathbf{I} \xrightarrow{\mu_{\mathcal{NM}}} \mathbf{I}^+_{\mathcal{NM}} \xrightarrow{\mu''} \mathbf{I}''}_{\mu}$$

Proof. First, define $\mathbf{I} = \langle \mathbf{Sig}, \mathbf{Sen}'', \mathbf{Mod}', \langle \models_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|} \rangle$, where for $\Sigma \in \mathbf{Sig}$, $M' \in \mathbf{Mod}'(\Sigma)$ and $\varphi'' \in \mathbf{Sen}''(\Sigma)$, we define $M' \models_\Sigma \varphi''$ to hold iff $M' \models'_\Sigma \mu_\Sigma(\varphi'')$, or equivalently (by the satisfaction condition for μ) iff $\mu_\Sigma(M') \models''_\Sigma \varphi''$. This indeed defines an institution, since the satisfaction condition for $\mathbf{I}$ follows from the satisfaction condition for $\mathbf{I}'$ and naturality of μ^{Sen} (or the satisfaction condition for $\mathbf{I}''$ and naturality of μ^{Mod}).

Consider "new" sentences $\mathcal{NS} = \langle \mathcal{NS}_\Sigma, \models^{\mathcal{NS}}_\Sigma \subseteq \mathbf{Mod}'(\Sigma) \times \mathcal{NS}_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|}$, where for $\Sigma \in |\mathbf{Sig}|$, $\mathcal{NS}_\Sigma = \mathbf{Sen}'(\Sigma) \setminus \mu_\Sigma(\mathbf{Sen}''(\Sigma))$ and $\models^{\mathcal{NS}}_\Sigma$ is the restriction of $\models'_\Sigma$ to $\mathcal{NS}_\Sigma$. Let $\mathbf{I}^+_{\mathcal{NS}}$ be the extension of $\mathbf{I}$ by sentences $\mathcal{NS}$, as defined in Section 2.3, with the institution morphism $\mu_{\mathcal{NS}} \colon \mathbf{I}^+_{\mathcal{NS}} \to \mathbf{I}$ defined above.[7] Then define the institution morphism $\mu' \colon \mathbf{I}' \to \mathbf{I}^+_{\mathcal{NS}}$ to be the identity on signatures and models, with $\mu'_\Sigma \colon \mathbf{Sen}^+_{\mathcal{NS}}(\Sigma) \to \mathbf{Sen}'(\Sigma)$, for $\Sigma \in |\mathbf{Sig}|$, defined as $\mu_\Sigma \colon \mathbf{Sen}''(\Sigma) \to \mathbf{Sen}'(\Sigma)$ on $\mathbf{Sen}''(\Sigma) \subseteq \mathbf{Sen}^+_{\mathcal{NS}}(\Sigma)$, and for $\tau \colon \Sigma' \to \Sigma$ in $\mathbf{Sig}$ and $\varphi' \in \mathcal{NS}_{\Sigma'} \subseteq \mathbf{Sen}'(\Sigma')$, $\mu'_\Sigma(\lceil \tau(\varphi') \rceil) = \mathbf{Sen}'(\tau)(\varphi') \in \mathbf{Sen}'(\Sigma)$.

The translations of sentences so defined are indeed natural in Σ: for any $\sigma \colon \Sigma_1 \to \Sigma_2$, we have to check that $\mu'_{\Sigma_1};\mathbf{Sen}'(\sigma) = \mathbf{Sen}^+_{\mathcal{NS}}(\sigma);\mu'_{\Sigma_2}$ as functions from $\mathbf{Sen}^+_{\mathcal{NS}}(\Sigma_1)$ to $\mathbf{Sen}'(\Sigma_2)$. For sentences in $\mathbf{Sen}''(\Sigma_1)$ this follows directly from the naturality of μ^{Sen}. For sentences of the form $\lceil \tau(\varphi') \rceil \in \mathbf{Sen}^+_{\mathcal{NS}}(\Sigma_1)$, where $\tau \colon \Sigma' \to \Sigma_1$ and $\varphi' \in \mathcal{NS}_{\Sigma'}$, we have

$$\begin{aligned}\mathbf{Sen}'(\sigma)(\mu'_{\Sigma_1}(\lceil \tau(\varphi') \rceil)) = \mathbf{Sen}'(\sigma)(\mathbf{Sen}'(\tau)(\varphi')) = \mathbf{Sen}'(\tau;\sigma)(\varphi') = {}&\\ \mu'_{\Sigma_2}(\lceil (\tau;\sigma)(\varphi') \rceil) = \mu'_{\Sigma_2}(\mathbf{Sen}^+_{\mathcal{NS}}(\sigma)(\lceil \tau(\varphi') \rceil)).&\end{aligned}$$

To check the satisfaction condition for μ', consider $\Sigma \in |\mathbf{Sig}|$, $M' \in \mathbf{Mod}'(\Sigma)$ and $\varphi \in \mathbf{Sen}^+_{\mathcal{NS}}(\Sigma)$. We have to show that $M' \models'_\Sigma \mu'_\Sigma(\varphi)$ iff $M' \models^+_{\mathcal{NS},\Sigma} \varphi$. For $\varphi \in \mathbf{Sen}''(\Sigma)$, this follows from the satisfaction condition for μ and our definitions: $M' \models'_\Sigma \mu'_\Sigma(\varphi)$ is then the same as $M' \models'_\Sigma \mu_\Sigma(\varphi)$, which is equivalent to $\mu_\Sigma(M') \models''_\Sigma \varphi$, which in turn defines $M' \models_{\mathbf{I},\Sigma} \varphi$ and $M' \models^+_{\mathcal{NS},\Sigma} \varphi$. For φ of the form $\lceil \tau(\varphi') \rceil$, where $\tau \colon \Sigma' \to \Sigma$ and $\varphi' \in \mathcal{NS}_{\Sigma'}$, this follows as well, since $M' \models'_\Sigma \mu'_\Sigma(\lceil \tau(\varphi') \rceil)$

[7] Footnote 4 applies here as well if needed.

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

coincides with $M' \models'_{\Sigma} \mathbf{Sen}'(\tau)(\varphi')$, which is the same as $M' \models^{\mathcal{NS}}_{\Sigma} \mathbf{Sen}'(\tau)(\varphi')$, which in turn defines $M' \models^{+}_{\mathcal{NS},\Sigma} \lceil \tau(\varphi') \rceil$.

Consider now "new" models $\mathcal{NM} = \langle \mathcal{NM}_{\Sigma}, \models^{\mathcal{NM}}_{\Sigma} \subseteq \mathcal{NM}_{\Sigma} \times \mathbf{Sen}''(\Sigma) \rangle_{\Sigma \in |\mathbf{Sig}|}$, where for $\Sigma \in |\mathbf{Sig}|$, $\mathcal{NM}_{\Sigma} = \mathbf{Mod}''(\Sigma) \setminus \mu_{\Sigma}(\mathbf{Mod}'(\Sigma))$ and $\models^{\mathcal{NM}}_{\Sigma}$ is the restriction of $\models''_{\Sigma}$ to $\mathcal{NM}_{\Sigma}$. Let $\mathbf{I}^{+}_{\mathcal{NM}}$ be the extension of $\mathbf{I}$ by models $\mathcal{NM}$, as defined in Section 2.3, with institution morphism $\mu_{\mathcal{NM}} \colon \mathbf{I} \to \mathbf{I}^{+}_{\mathcal{NM}}$ defined above. Then let the institution morphism $\mu'' \colon \mathbf{I}^{+}_{\mathcal{NM}} \to \mathbf{I}''$ be the identity on signatures and sentences, with $\mu''_{\Sigma} \colon \mathbf{Mod}^{+}_{\mathcal{NM}}(\Sigma) \to \mathbf{Mod}''(\Sigma)$, for $\Sigma \in |\mathbf{Sig}|$, defined as $\mu_{\Sigma} \colon \mathbf{Mod}'(\Sigma) \to \mathbf{Mod}''(\Sigma)$ on $\mathbf{Mod}'(\Sigma) \subseteq \mathbf{Mod}^{+}_{\mathcal{NM}}(\Sigma)$, and for $\tau \colon \Sigma \to \Sigma'$ in $\mathbf{Sig}$ and $M' \in \mathcal{NM}_{\Sigma'} \subseteq \mathbf{Mod}''(\Sigma)$, $\mu''_{\Sigma}(\lceil M'|_{\tau} \rceil) = \mathbf{Mod}''(\tau)(M')$. By similar arguments as for $\mu' \colon \mathbf{I}' \to \mathbf{I}^{+}_{\mathcal{NS}}$, it follows that $\mu''_{\Sigma} \colon \mathbf{Mod}^{+}_{\mathcal{NM}}(\Sigma) \to \mathbf{Mod}''(\Sigma)$, $\Sigma \in |\mathbf{Sig}|$, are natural in Σ, and the satisfaction condition holds for μ''.

It is easy now to check directly that indeed $\mu = \mu';\mu_{\mathcal{NS}};\mu_{\mathcal{NM}};\mu''$. ⊣

§3. Interpolation.

3.1. Classical interpolation. The Craig interpolation theorem [13] states that if an implication between two first-order formulae $\varphi \Rightarrow \psi$ holds then there is a formula θ that uses only the symbols common to φ and ψ such that both $\varphi \Rightarrow \theta$ and $\theta \Rightarrow \psi$ hold; θ is then called an *interpolant* for φ and ψ. This is one of the key properties of first-order logic, with numerous applications, including simpler proofs of similarly famous and important results like the Robinson consistency [36] and Beth definability [4] theorems. The original proof in [13] relied on proof-theoretic arguments, even though many of the applications (as well as some later proofs) of the result have been model-theoretic in nature. The interpolation property has been investigated (and proved or disproved) for many standard extensions (and fragments) of first-order logic [48] as well as for other logical systems, for instance for various modal and intuitionistic logics [22].

The above statement of the interpolation property implicitly involves the following union/intersection square of signatures:

$$\begin{tikzcd}
 & \boxed{\varphi \Rightarrow \psi} & \\
 & \Sigma_p \cup \Sigma_c & \\
\boxed{\varphi \Rightarrow \theta}\ \Sigma_p \arrow[ur] & & \Sigma_c\ \boxed{\theta \Rightarrow \psi} \arrow[ul] \\
 & \Sigma_p \cap \Sigma_c \arrow[ul] \arrow[ur] & \\
 & \boxed{\theta} &
\end{tikzcd}$$

where Σ_p and Σ_c are (first-order) signatures for φ and ψ, respectively, and the arrows indicate signature inclusions.

As recalled in Section 1, interpolation proved indispensable for many foundational aspects of computer science and software engineering, in particular, in the foundations of software specification and development [40]. However, the classical formulation of Craig's interpolation for many applications in this area requires some generalisations, which perhaps do not bring much new insight for this property in the framework of first-order logic, but may be important when other logical systems are considered.

To begin with, the use of implication should be replaced by entailment. Then, we should deal with entailments between sets of sentences, rather than between individual sentences (strictly speaking, this is needed for the premise φ and especially for the interpolant θ—for notational symmetry, we do this for the conclusion ψ as well). Both these generalisations are irrelevant for first-order logic, where implication captures semantic entailment, and a set of sentences in the premise of each single-conclusion entailment may always be replaced by a single sentence (since we have finite conjunctions and the logic is compact[8]). However, for instance, working in equational logic we have no implication available, and an interpolant cannot be always expressed as a single equation—even though the interpolation property holds if a set of equations is permitted as an interpolant [37].

Perhaps most importantly, for instance in applications where parameterised specifications and their "pushout-style" instantiations [21] are involved, we have to go beyond union/intersection squares of signatures and beyond inclusions to relate the signatures. More general classes of signature squares are needed, with non-injective signature morphisms necessary to capture for instance morphisms from the formal to actual parameters, used to "fit" the latter into the mould given by the former. Typically in applications at least pushouts of signature morphism are involved, sometimes additionally restricted to indicated classes of morphisms permitted at the "bottom-left" and "bottom-right" of the squares, respectively [6, 15, 34, 49, 50]. However, for the purposes of this paper we will consider interpolation properties for an arbitrary commuting square of signature morphisms.

The above remarks lead to a general definition of the interpolation property in an arbitrary institution.

3.2. Interpolation in an institution. Throughout the rest of this paper, we consider an institution $\mathbf{I} = \langle \mathbf{Sig}, \mathbf{Sen}, \mathbf{Mod}, \langle \models_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|} \rangle$, and study interpolation properties over the following commutative square $(*)$ of signature morphisms:[9]

$$\begin{tikzcd}
 & \Sigma_u & \\
\Sigma_p \arrow[ur, "\sigma_{pu}"] & (*) & \Sigma_c \arrow[ul, "\sigma_{cu}"'] \\
 & \Sigma_i \arrow[ul, "\sigma_{ip}"] \arrow[ur, "\sigma_{ic}"'] &
\end{tikzcd}$$

[8] An institution $\mathbf{I} = \langle \mathbf{Sig}, \mathbf{Sen}, \mathbf{Mod}, \langle \models_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|} \rangle$ is *compact* if for any signature $\Sigma \in |\mathbf{Sig}|$, set $\Phi \subseteq \mathbf{Sen}(\Sigma)$ of Σ-sentences and Σ-sentence $\varphi \in \mathbf{Sen}(\Sigma)$, whenever $\Phi \models \varphi$ then $\Phi_0 \models \varphi$ for some finite $\Phi_0 \subseteq \Phi$.

[9] To help memorising the notation: p for *premise*, c for *conclusion*, u for *union,* and i for *intersection* (or *interpolant*).

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

Let $\Phi \subseteq \mathbf{Sen}(\Sigma_p)$ and $\Psi \subseteq \mathbf{Sen}(\Sigma_c)$ be such that $\sigma_{pu}(\Phi) \models_{\Sigma_u} \sigma_{cu}(\Psi)$. An *interpolant for* Φ *and* Ψ (over diagram $(*)$) is a set $\Theta \subseteq \mathbf{Sen}(\Sigma_i)$ of Σ_i-sentences such that $\Phi \models_{\Sigma_p} \sigma_{ip}(\Theta)$ and $\sigma_{ic}(\Theta) \models_{\Sigma_c} \Psi$.

$$\begin{tikzcd}
 & & \boxed{\sigma_{pu}(\Phi) \models \sigma_{cu}(\Psi)} & & \\
 & & \Sigma_u & & \\
\boxed{\Phi \models \sigma_{ip}(\Theta)} & \Sigma_p \arrow[ur, "\sigma_{pu}"] & & \Sigma_c \arrow[ul, "\sigma_{cu}"'] & \boxed{\sigma_{ic}(\Theta) \models \Psi} \\
 & & \Sigma_i \arrow[ul, "\sigma_{ip}"] \arrow[ur, "\sigma_{ic}"'] & & \\
 & & \boxed{\Theta} & &
\end{tikzcd}$$

To simplify some further statements, if $\sigma_{pu}(\Phi) \not\models_{\Sigma_u} \sigma_{cu}(\Psi)$ then we say that any set $\Theta \subseteq \mathbf{Sen}(\Sigma_i)$ is an interpolant for Φ and Ψ (over diagram $(*)$).

A commutative square $(*)$ of signature morphisms *admits interpolation* if all sets $\Phi \subseteq \mathbf{Sen}(\Sigma_p)$ and $\Psi \subseteq \mathbf{Sen}(\Sigma_c)$ such that $\sigma_{pu}(\Phi) \models_{\Sigma_u} \sigma_{cu}(\Psi)$ have an interpolant.

Example 3.1. In the institution **FO** of first-order logic, as well as any of its variants mentioned in Example 2.1, if the square $(*)$ is a pushout and at least one of $\sigma_{ip} \colon \Sigma_i \to \Sigma_p$, $\sigma_{ic} \colon \Sigma_i \to \Sigma_c$ is injective on sorts then $(*)$ admits interpolation; otherwise interpolation may fail for $(*)$ (see [7]). In the institution **EQ** of equational logic if the square $(*)$ is a pushout and $\sigma_{ic} \colon \Sigma_i \to \Sigma_c$ is injective then $(*)$ admits interpolation; otherwise interpolation may fail for $(*)$, and in $\mathbf{EQ}_{\emptyset}$, where empty carriers are permitted, interpolation may fail even for intersection/union squares of signatures (see [45]). In the institution **PL** of propositional logic, all pushouts admit interpolation.

It is well known that the interpolation property of a logical system is fragile. When the logic is strengthened or weakened, when new models or sentences are added, the interpolation property may easily be spoiled. Clearly, this may happen when entirely new signatures are added, with new models and sentences over them. Therefore, we will consider the category of signatures to be fixed, and consider only such extensions of institutions that preserve it.

Throughout the rest of the paper we study in some detail how the interpolation property may be spoiled by adding new models or sentences. This will be done from a "local" perspective, for particular commutative squares of signature morphisms, as well as for particular interpolants.

We say that an interpolant $\Theta \subseteq \mathbf{Sen}(\Sigma_i)$ for $\Phi \subseteq \mathbf{Sen}(\Sigma_p)$ and $\Psi \subseteq \mathbf{Sen}(\Sigma_c)$ (over diagram $(*)$) is *stable* under extensions of the institution by models if for every extension $\mathbf{I}^+$ of $\mathbf{I}$ by new models, Θ is an interpolant for Φ and Ψ in $\mathbf{I}^+$; otherwise we say that the interpolant Θ is *fragile*.

While adding new models may spoil existing interpolants, it cannot create new non-trivial ones: in all extensions $\mathbf{I}^+$ of $\mathbf{I}$ by new models, if $\sigma_{pu}(\Phi) \models^+_{\Sigma_u} \sigma_{cu}(\Psi)$ then any interpolant for Φ and Ψ in $\mathbf{I}^+$ is an interpolant for Φ and Ψ in $\mathbf{I}$. Adding

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

new sentences cannot spoil a particular interpolant, but may spoil interpolation property for a given diagram (there may be no interpolant for new sentences, or sets of sentences that contain them), and may create new interpolants (involving some new sentences).

To begin with, we identify some special cases where interpolation is ensured and is stable under any extension of the institution.

3.3. Interpolants may be stable.

LEMMA 3.2. *Consider the diagram* $(*)$ *of signature morphisms.*

1. *If* $\mathbf{Sen}(\sigma_{ip})\colon \mathbf{Sen}(\Sigma_i) \to \mathbf{Sen}(\Sigma_p)$ *is surjective and* $\sigma_{cu}\colon \Sigma_c \to \Sigma_u$ *is conservative then* $(*)$ *admits interpolation.*
2. *If* $\mathbf{Sen}(\sigma_{ic})\colon \mathbf{Sen}(\Sigma_i) \to \mathbf{Sen}(\Sigma_c)$ *is surjective and* $\sigma_{pu}\colon \Sigma_p \to \Sigma_u$ *is conservative then* $(*)$ *admits interpolation.*

PROOF. Let $\Phi \subseteq \mathbf{Sen}(\Sigma_c)$ and $\Psi \in \mathbf{Sen}(\Sigma_c)$ be such that $\sigma_{pu}(\Phi) \models \sigma_{cu}(\Psi)$.

1. Suppose $\mathbf{Sen}(\sigma_{ip})\colon \mathbf{Sen}(\Sigma_i) \to \mathbf{Sen}(\Sigma_p)$ is surjective and $\sigma_{cu}\colon \Sigma_c \to \Sigma_u$ is conservative. Consider $\Theta = \sigma_{ip}^{-1}(\Phi) \subseteq \mathbf{Sen}(\Sigma_i)$. First, since $\Phi = \sigma_{ip}(\Theta)$, we have $\Phi \models \sigma_{ip}(\Theta)$. Then, since $(*)$ commutes, $\sigma_{pu}(\Phi) = \sigma_{pu}(\sigma_{ip}(\Theta)) = \sigma_{cu}(\sigma_{ic}(\Theta))$, and so $\sigma_{cu}(\sigma_{ic}(\Theta)) \models \sigma_{cu}(\Psi)$. Hence $\sigma_{ic}(\Theta) \models \Psi$ by conservativity of σ_{cu}. Thus Θ is an interpolant for Φ and Ψ.
2. Suppose $\mathbf{Sen}(\sigma_{ic})\colon \mathbf{Sen}(\Sigma_i) \to \mathbf{Sen}(\Sigma_c)$ is surjective and $\sigma_{pu}\colon \Sigma_p \to \Sigma_u$ is conservative. Consider $\Theta = \sigma_{ic}^{-1}(\Psi) \subseteq \mathbf{Sen}(\Sigma_i)$. Then $\Psi = \sigma_{ic}(\Theta)$, and so $\sigma_{ic}(\Theta) \models \Psi$. Moreover, $\sigma_{pu}(\sigma_{ip}(\Theta)) = \sigma_{cu}(\sigma_{ic}(\Theta)) = \sigma_{cu}(\Psi)$, and so $\sigma_{pu}(\Phi) \models \sigma_{pu}(\sigma_{ip}(\Theta))$, which implies $\Phi \models \sigma_{ip}(\Theta)$ by conservativity of σ_{pu}. Thus Θ is an interpolant for Φ and Ψ. ⊣

A trivial special case here is when σ_{ip} and σ_{cu}, or σ_{ic} and σ_{pu}, are isomorphisms, which can be further refined as follows:

COROLLARY 3.3. *Consider the diagram* $(*)$ *of signature morphisms. If:*

1. $\sigma_{ip}\colon \Sigma_i \to \Sigma_p$ *is a retraction and* $\sigma_{cu}\colon \Sigma_c \to \Sigma_u$ *is a coretraction, or*
2. $\sigma_{ic}\colon \Sigma_i \to \Sigma_c$ *is a retraction and* $\sigma_{pu}\colon \Sigma_p \to \Sigma_u$ *is a coretraction,*

then $(*)$ *admits interpolation.*

PROOF. Follows by Lemma 3.2, since signature morphisms that are retractions induce surjective translations of sentences, and signature morphisms that are coretractions induce surjective reduct functions on model classes, and so are conservative. ⊣

This shows that if the signature morphisms in $(*)$ satisfy the premises of Corollary 3.3 then the diagram enjoys a stable interpolation property, which cannot be spoiled by any institution extension that leaves the category of signatures unchanged! No matter how we add new models or sentences, the interpolation property is ensured by the properties of the signature morphisms involved and the implied properties of the translations of sentences and reducts of models they induce in the institution and in any of its extensions.

The conditions stated in Corollary 3.3 are in fact quite strong and in many practical situations do not depart too far from the trivial case when Σ_p is (up

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

to isomorphism) included in Σ_c or vice versa. Namely, when the diagram $(*)$ is a pushout then condition 1 implies that $\sigma_{cu}\colon \Sigma_c \to \Sigma_u$ is an isomorphism, and condition 2 implies that $\sigma_{pu}\colon \Sigma_p \to \Sigma_u$ is an isomorphism. Dually, when $(*)$ is a pullback then condition 1 implies that $\sigma_{ip}\colon \Sigma_i \to \Sigma_p$ is an isomorphism, and condition 2 implies that $\sigma_{ic}\colon \Sigma_i \to \Sigma_c$ is an isomorphism.

Somewhat similarly, interpolation is preserved and reflected by logically trivial institution morphisms:

PROPOSITION 3.4. *Let $\mu\colon \mathbf{I} \to \mathbf{I}'$ be a logically trivial institution morphism. Diagram $(*)$ in the category of signatures admits interpolation in $\mathbf{I}$ iff it admits interpolation in $\mathbf{I}'$.*

PROOF. Since for each signature $\Sigma \in |\mathbf{Sig}|$, $\mu_\Sigma\colon \mathbf{Mod}(\Sigma) \to \mathbf{Mod}'(\Sigma)$ is surjective, for any sets of $\mathbf{I}'$-sentences $\Phi', \Psi' \subseteq \mathbf{Sen}'(\Sigma)$, $\Phi' \models' \Psi'$ iff $\mu_\Sigma(\Phi') \models \mu_\Sigma(\Psi')$ (by the remark after the definition of institution morphism in Section 2.4). Moreover, since $\mu_\Sigma\colon \mathbf{Sen}'(\Sigma) \to \mathbf{Sen}(\Sigma)$ is surjective, for any sets of $\mathbf{I}$-sentences $\Phi, \Psi \subseteq \mathbf{Sen}(\Sigma)$, $\Phi = \mu_\Sigma(\mu_\Sigma^{-1}(\Phi))$ and $\Psi = \mu_\Sigma(\mu_\Sigma^{-1}(\Psi))$, so that $\Phi \models \Psi$ iff $\mu_\Sigma^{-1}(\Phi) \models' \mu_\Sigma^{-1}(\Psi)$.

Thus, for $\Phi' \subseteq \mathbf{Sen}'(\Sigma_p)$ and $\Psi' \subseteq \mathbf{Sen}'(\Sigma_c)$, if $\mu_{\Sigma_p}(\Phi')$ and $\mu_{\Sigma_c}(\Psi')$ have an interpolant $\Theta \subseteq \mathbf{Sen}(\Sigma_i)$ in $\mathbf{I}$ then $\mu_{\Sigma_i}^{-1}(\Theta)$ is an interpolant for Φ' and Ψ' in $\mathbf{I}'$. Similarly, for $\Phi \subseteq \mathbf{Sen}(\Sigma_p)$ and $\Psi \subseteq \mathbf{Sen}(\Sigma_c)$, if $\mu_{\Sigma_p}^{-1}(\Phi)$ and $\mu_{\Sigma_c}^{-1}(\Psi)$ have an interpolant $\Theta' \subseteq \mathbf{Sen}'(\Sigma_i)$ in $\mathbf{I}'$ then $\mu_{\Sigma_i}(\Theta')$ is an interpolant for Φ and Ψ in $\mathbf{I}$. ⊣

Propositions 2.7 and 3.4 imply that institution extensions by new models and by new sentences are of primary importance for our study of the fragility of interpolation.

§4. Spoiling an interpolant by new models.

Recall that we study interpolation over a commutative square of signature morphisms $(*)$ in an institution $\mathbf{I} = \langle \mathbf{Sig}, \mathbf{Sen}, \mathbf{Mod}, \langle \models_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|} \rangle$. Throughout this section, let $\Phi \subseteq \mathbf{Sen}(\Sigma_c)$ and $\Psi \subseteq \mathbf{Sen}(\Sigma_c)$ be such that $\sigma_{pu}(\Phi) \models \sigma_{cu}(\Psi)$, and let $\Theta \subseteq \mathbf{Sen}(\Sigma_i)$ be an interpolant for Φ and Ψ in $\mathbf{I}$.

LEMMA 4.1. *Suppose that there exists a set of Σ_p-sentences $\Phi^\bullet \supseteq \Phi$ such that $\sigma_{ip}(\Theta) \not\subseteq \Phi^\bullet$ and for all signature morphisms $\tau\colon \Sigma_u \to \Sigma_p$, if $\tau(\sigma_{pu}(\Phi)) \subseteq \Phi^\bullet$ then $\tau(\sigma_{cu}(\Psi)) \subseteq \Phi^\bullet$. Then the interpolant Θ for Φ and Ψ is not stable under extensions of $\mathbf{I}$ by models.*

PROOF. Let $\mathbf{I}^+$ be the extension of $\mathbf{I}$ by a new Σ_p-model M (and its reducts $\lceil M|_\tau \rceil \in \mathbf{Mod}^+(\Sigma)$ for $\tau\colon \Sigma \to \Sigma_p$, see Section 2.3), with $Th^+(M) = \Phi^\bullet$.

Then for all models $K \in \mathbf{Mod}^+(\Sigma_u)$, if $K \models^+ \sigma_{pu}(\Phi)$ then $K \models^+ \sigma_{cu}(\Psi)$: this clearly holds for $K \in \mathbf{Mod}(\Sigma_u)$. For new models of the form $K = \lceil M|_\tau \rceil$ with $\tau\colon \Sigma_u \to \Sigma_p$, if $\lceil M|_\tau \rceil \models^+ \sigma_{pu}(\Phi)$ then $M \models^+ \tau(\sigma_{pu}(\Phi))$, that is $\tau(\sigma_{pu}(\Phi)) \subseteq \Phi^\bullet$, which by the assumptions implies $\tau(\sigma_{cu}(\Psi)) \subseteq \Phi^\bullet$. Hence $M \models^+ \tau(\sigma_{cu}(\Psi))$ and so $\lceil M|_\tau \rceil \models^+ \sigma_{cu}(\Psi)$. This shows $\sigma_{pu}(\Phi) \models^+ \sigma_{cu}(\Psi)$.

However, $M \not\models^+ \sigma_{ip}(\Theta)$ (since $\sigma_{ip}(\Theta) \not\subseteq \Phi^\bullet$) and so $\Phi \not\models^+ \sigma_{ip}(\Theta)$, which proves that Θ is not an interpolant for Φ and Ψ in $\mathbf{I}^+$. ⊣

The key property of the set $\Phi^\bullet$ used in the above lemma is that it cannot be used to separate $\sigma_{pu}(\Phi)$ from $\sigma_{cu}(\Psi)$ via any morphism $\tau\colon \Sigma_u \to \Sigma_p$. More formally, for any signatures $\Sigma, \Sigma' \in |\mathbf{Sig}|$, we say that $\Upsilon \subseteq \mathbf{Sen}(\Sigma)$ *never separates* $\Phi' \subseteq \mathbf{Sen}(\Sigma')$ *from* $\Psi' \subseteq \mathbf{Sen}(\Sigma')$ when for all morphisms $\tau\colon \Sigma' \to \Sigma$, if $\tau(\Phi') \subseteq \Upsilon$ then $\tau(\Psi') \subseteq \Upsilon$.

Lemma 4.2. *For any sets $\Phi \subseteq \mathbf{Sen}(\Sigma)$ of Σ-sentences and $\Phi', \Psi' \subseteq \mathbf{Sen}(\Sigma')$ of Σ'-sentences, there is the least set $[\Phi' \overset{\Sigma'}{\underset{\Sigma}{\leadsto}} \Psi'](\Phi) \subseteq \mathbf{Sen}(\Sigma)$ of Σ-sentences that includes Φ and never separates Φ' from Ψ'.*

Proof. Consider the set $\mathcal{E}$ of all sets $\Upsilon \subseteq \mathbf{Sen}(\Sigma)$ such that $\Phi \subseteq \Upsilon$ and for all signature morphisms $\tau\colon \Sigma' \to \Sigma$, if $\tau(\Phi') \subseteq \Upsilon$ then $\tau(\Psi') \subseteq \Upsilon$. $\mathcal{E}$ is nonempty and is closed under intersection. Then $[\Phi' \overset{\Sigma'}{\underset{\Sigma}{\leadsto}} \Psi'](\Phi) = \bigcap \mathcal{E}$. $\dashv$

Corollary 4.3. *If $\sigma_{ip}(\Theta) \not\subseteq [\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_p}{\leadsto}} \sigma_{cu}(\Psi)](\Phi)$ then the interpolant Θ for Φ and Ψ is not stable under extensions of* $\mathbf{I}$ *by models.*

Proof. Directly from Lemma 4.1, with $\Phi^\bullet = [\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_p}{\leadsto}} \sigma_{cu}(\Psi)](\Phi)$. $\dashv$

Lemma 4.4. *Suppose that there exists a set of Σ_c-sentences $\Psi^\circ \supseteq \sigma_{ic}(\Theta)$ such that $\Psi \not\subseteq \Psi^\circ$ and for all signature morphisms $\tau\colon \Sigma_u \to \Sigma_c$, if $\tau(\sigma_{cu}(\Phi)) \subseteq \Psi^\circ$ then $\tau(\sigma_{pu}(\Psi)) \subseteq \Psi^\circ$. Then the interpolant Θ for Φ and Ψ is not stable under extensions of* $\mathbf{I}$ *by models.*

Proof. Let $\mathbf{I}^+$ be the extension of $\mathbf{I}$ by a new Σ_c-model N (and its reducts $\lceil N|_\tau \rceil \in \mathbf{Mod}^+(\Sigma)$ for $\tau\colon \Sigma \to \Sigma_c$), with $Th^+(N) = \Psi^\circ$.

Then for all models $K \in \mathbf{Mod}^+(\Sigma_u)$, if $K \models^+ \sigma_{pu}(\Phi)$ then $K \models^+ \sigma_{cu}(\Psi)$: this clearly holds for $K \in \mathbf{Mod}(\Sigma_u)$. For new models of the form $K = \lceil N|_\tau \rceil$ with $\tau\colon \Sigma_u \to \Sigma_c$, if $\lceil N|_\tau \rceil \models^+ \sigma_{pu}(\Phi)$ then $N \models^+ \tau(\sigma_{pu}(\Phi))$, that is, $\tau(\sigma_{pu}(\Phi)) \subseteq \Psi^\circ$. By the assumptions this implies $\tau(\sigma_{pu}(\Psi)) \subseteq \Psi^\circ$. Hence $N \models^+ \tau(\sigma_{cu}(\Psi))$, and so $\lceil N|_\tau \rceil \models^+ \sigma_{pu}(\Psi)$. This shows $\sigma_{pu}(\Phi) \models^+ \sigma_{cu}(\Psi)$.

However, $N \models \sigma_{ic}(\Theta)$ (since $\sigma_{ic}(\Theta) \subseteq \Psi^\circ$), while $N \not\models \Psi$ (since $\Psi \not\subseteq \Psi^\circ$). Hence $\sigma_{ic}(\Theta) \not\models^+ \Psi$, which shows that Θ is not an interpolant for Φ and Ψ in $\mathbf{I}^+$. $\dashv$

Corollary 4.5. *If $\Psi \not\subseteq [\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_c}{\leadsto}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta))$ then the interpolant Θ for Φ and Ψ is not stable under extension of* $\mathbf{I}$ *by models.*

Proof. By Lemma 4.4, with $\Psi^\circ = [\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_c}{\leadsto}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta))$. $\dashv$

Corollaries 4.3 and 4.5 present sufficient conditions which ensure that a particular interpolant may be spoiled under an extension of the institution by new models. In fact, these conditions jointly are also necessary:

Theorem 4.6. *The interpolant Θ for Φ and Ψ is stable under extensions of* $\mathbf{I}$ *by models if and only if the following conditions hold*:

1. $\sigma_{ip}(\Theta) \subseteq [\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_p}{\leadsto}} \sigma_{cu}(\Psi)](\Phi)$, *and*
2. $\Psi \subseteq [\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_c}{\leadsto}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta))$.

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

Proof. The "only if" part follows by Corollaries 4.3 and 4.5, by contraposition. For the "if" part: assume that the interpolant Θ for Φ and Ψ is not stable under extensions of $\mathbf{I}$ by models, and let $\mathbf{I}^+$ be an extension of $\mathbf{I}$ by models such that Θ is not an interpolant for Φ and Ψ in $\mathbf{I}^+$, that is, we have $\sigma_{pu}(\Phi) \models^+ \sigma_{cu}(\Psi)$, but $\Phi \not\models^+ \sigma_{ip}(\Theta)$ or $\sigma_{ic}(\Theta) \not\models^+ \Psi$.

1. If $\Phi \not\models^+ \sigma_{ip}(\Theta)$ then for some model $M \in \mathbf{Mod}^+(\Sigma_p)$, $M \models^+ \Phi$ and $M \not\models^+ \sigma_{ip}(\Theta)$. Then $\Phi \subseteq Th^+(M)$ and $\sigma_{ip}(\Theta) \not\subseteq Th^+(M)$. Moreover, $Th^+(M)$ never separates $\sigma_{pu}(\Phi)$ from $\sigma_{cu}(\Psi)$, since if for some $\tau\colon \Sigma_u \to \Sigma_p$, $\tau(\sigma_{pu}(\Phi)) \subseteq Th^+(M)$ and $\tau(\sigma_{cu}(\Psi)) \not\subseteq Th^+(M)$, then $M|_\tau \models^+ \sigma_{pu}(\Phi)$ and $M|_\tau \not\models^+ \sigma_{cu}(\Psi)$, which contradicts $\sigma_{pu}(\Phi) \models^+ \sigma_{cu}(\Psi)$. It follows now that $[\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_p}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\Phi) \subseteq Th^+(M)$, and so $\sigma_{ip}(\Theta) \not\subseteq [\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_p}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\Phi)$.
2. If $\sigma_{ic}(\Theta) \not\models^+ \Psi$ then for some model $N \in \mathbf{Mod}^+(\Sigma_c)$, $N \models^+ \sigma_{ic}(\Theta)$ and $N \not\models^+ \Psi$. Then $\sigma_{ic}(\Theta) \subseteq Th^+(N)$ and $\Psi \not\subseteq Th^+(N)$. Moreover, $Th^+(N)$ never separates $\sigma_{pu}(\Phi)$ from $\sigma_{cu}(\Psi)$, since if for some $\tau\colon \Sigma_u \to \Sigma_c$, $\tau(\sigma_{pu}(\Phi)) \subseteq Th^+(N)$ and $\tau(\sigma_{cu}(\Psi)) \not\subseteq Th^+(N)$, then $N|_\tau \models^+ \sigma_{pu}(\Phi)$ and $N|_\tau \not\models^+ \sigma_{cu}(\Psi)$, contradicting $\sigma_{pu}(\Phi) \models^+ \sigma_{cu}(\Psi)$. It follows now that $[\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_c}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta)) \subseteq Th^+(N)$, so $\Psi \not\subseteq [\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_c}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta))$. ⊣

The above theorem gives precise conditions that ensure stability of a particular interpolant under extensions of the institution by new models. Of course, this also yields a precise characterisation of specific interpolation properties that can be spoiled by adding new abstract models. It should be stressed that the conditions in use a purely "syntactic"—they do not refer to the semantic properties of the sets of sentences involved, so in particular, they depend on a specific syntactic form of the sentences, and the conclusions may vary when sentences considered are replaced by semantically equivalent sentences that are of a different syntactic form.

Example 4.7. Consider a trivial example in the institution $\mathbf{PL}$ of propositional logic. In the diagram $(*)$, let $\Sigma_p = \{p, r\}$, $\Sigma_c = \{p, q\}$, $\Sigma_u = \Sigma_p \cup \Sigma_c = \{r, p, q\}$, $\Sigma_i = \Sigma_p \cap \Sigma_c = \{p\}$, and the four signature morphisms are inclusions.

Let φ be $r \wedge p$ and ψ be $p \vee q$. Clearly, $\varphi \models \psi$, and φ and ψ have a number of distinct interpolants in $\mathbf{PL}$.[10]

One interpolant for φ and ψ is p (since clearly $r \wedge p \models p$ and $p \models p \vee q$). Consider $\mathbf{PL}$-model $M = \{r\} \in \mathbf{Mod}_{\mathbf{PL}}(\Sigma_p)$. Let $\mathbf{PL}^+$ be an extension of $\mathbf{PL}$ by a new Σ_p-model $\widetilde{M}$ (with interpretation of propositional sentences "swapping" the valuation of propositional variables, as in Example 2.5). Then $\widetilde{M} \models^+ r \wedge p$ while $\widetilde{M} \not\models^+ p$, and so p is not an interpolant for φ and ψ in $\mathbf{PL}^+$. In fact, it is easy to check that $\Phi^\bullet = \{\varphi \in \mathbf{Sen}_{\mathbf{PL}}(\Sigma_p) \mid \widetilde{M} \models^+ \varphi\}$ satisfies the premises of Lemma 4.1.

Moreover, one can easily calculate $[r \wedge p \overset{\Sigma_u}{\underset{\Sigma_p}{\rightsquigarrow}} p \vee q](r \wedge p) \subseteq \mathbf{Sen}_{\mathbf{PL}}(\Sigma_p)$: there are exactly two morphisms $\tau, \tau'\colon \Sigma_u \to \Sigma_p$ such that $\tau(r \wedge p) = \tau'(r \wedge p) = r \wedge p$, namely they both map r to r and p to p, and then map q to any of the symbols in

[10] When convenient, we write φ for $\{\varphi\}$, relying on the context to impose such identification of a sentence with the one-element set that contains it.

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

Σ_p, say, $\tau(q) = p$ and $\tau'(q) = r$. Consequently, $[r \wedge p \overset{\Sigma_u}{\underset{\Sigma_p}{\rightsquigarrow}} q \vee p](r \wedge p) = \{r \wedge p, r \vee p, p \vee p\}$ (since no morphism from Σ_u to Σ_p maps $r \wedge p$ into $\{p \vee r, p \vee p\}$). Thus, by Corollary 4.3, any interpolant for φ and ψ other than $p \vee p$ may be spoiled by extending **PL** by new models.

Indeed, $p \vee p$ is an interpolant for φ and ψ (since of course $r \wedge p \models p \vee p$ and $p \vee p \models p \vee q$). Consider **PL**-model $N = \{q\} \in \mathbf{Mod}_{\mathbf{PL}}(\Sigma_c)$. Let now $\mathbf{PL}^+$ be the extension of **PL** by a new Σ_c-model $\widetilde{N}$ (with interpretation of propositional sentences "swapping" the valuation of propositional variables, as in Example 2.5). Then $\widetilde{N} \models^+ p \vee p$ while $\widetilde{N} \not\models^+ p \vee q$, which shows that $p \vee p$ is not an interpolant for φ and ψ in $\mathbf{PL}^+$. In fact, since no morphism from Σ_u to Σ_c maps $r \wedge p$ to $p \vee p$, we have $[r \wedge p \overset{\Sigma_u}{\underset{\Sigma_c}{\rightsquigarrow}} p \vee q](p \vee p) = \{p \vee p\} \subseteq \mathbf{Sen}_{\mathbf{PL}}(\Sigma_c)$, and so it also follows directly from Corollary 4.5 that in some extension of **PL** by new models $p \vee p$ is not an interpolant for φ and ψ.

Summing up: none of the interpolants for φ and ψ in **PL** is stable under extension of **PL** by new models.

Let now φ' be $(p \vee r) \wedge (p \vee \neg r)$ and ψ' be $(p \vee q) \wedge (p \vee \neg q)$. Clearly, $(p \vee r) \wedge (p \vee \neg r) \models (p \vee q) \wedge (p \vee \neg q)$. Perhaps the most obvious interpolant for φ' and ψ' is p (since $(p \vee r) \wedge (p \vee \neg r) \models p$ and $p \models (p \vee q) \wedge (p \vee \neg q)$). This interpolant, however, is fragile: it may be spoiled by extending **PL** by new models. Namely, reasoning similarly as above, we can calculate:

$$\begin{aligned}[(p \vee r) \wedge (p \vee \neg r) \overset{\Sigma_u}{\underset{\Sigma_p}{\rightsquigarrow}} (p \vee q) \wedge (p \vee \neg q)]((p \vee r) \wedge (p \vee \neg r)) = \\ \{(p \vee r) \wedge (p \vee \neg r), (p \vee p) \wedge (p \vee \neg p)\} \subseteq \mathbf{Sen}_{\mathbf{PL}}(\Sigma_p).\end{aligned}$$

Thus, by Corollary 4.3, p is not an interpolant for φ' and ψ' in an extension of **PL** by new models.

Another interpolant for φ' and ψ' in **PL** is $(p \vee p) \wedge (p \vee \neg p)$ (which in **PL** is semantically equivalent to p). Since

$$(p \vee p) \wedge (p \vee \neg p) \in [(p \vee r) \wedge (p \vee \neg r) \overset{\Sigma_u}{\underset{\Sigma_p}{\rightsquigarrow}} (p \vee q) \wedge (p \vee \neg q)]((p \vee r) \wedge (p \vee \neg r)),$$

Corollary 4.3 cannot be used here to conclude that this interpolant gets spoiled in an extension of **PL** by new models. Moreover,

$$\begin{aligned}[(p \vee r) \wedge (p \vee \neg r) \overset{\Sigma_u}{\underset{\Sigma_c}{\rightsquigarrow}} (p \vee q) \wedge (p \vee \neg q)]((p \vee p) \wedge (p \vee \neg p)) = \\ \{(p \vee p) \wedge (p \vee \neg p), (p \vee q) \wedge (p \vee \neg q)\} \subseteq \mathbf{Sen}_{\mathbf{PL}}(\Sigma_c).\end{aligned}$$

Consequently, Corollary 4.5 does not apply here either.

Theorem 4.6 implies that $(p \vee p) \wedge (p \vee \neg p)$ is an interpolant for φ' and ψ' stable under extensions of **PL** by new models.

§5. Spoiling interpolation by new models. As in the previous section, consider institution $\mathbf{I} = \langle \mathbf{Sig}, \mathbf{Sen}, \mathbf{Mod}, \langle \models_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|} \rangle$, commutative square of signature morphisms $(*)$, and sets of sentences $\Phi \subseteq \mathbf{Sen}(\Sigma_p)$ and $\Psi \in \mathbf{Sen}(\Sigma_c)$ such that $\sigma_{pu}(\Phi) \models \sigma_{cu}(\Psi)$. Theorem 4.6 gives the exact characterisation of interpolants that are stable under extensions of **I** by new models. Of course, this also characterises interpolants

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

that are fragile. In this section we characterise situations where all interpolants for the premise Φ and conclusion Ψ may be spoiled at once when the institution is extended by new models.

Corollary 5.1. *Define* $\Phi^\bullet = [\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_p}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\Phi) \subseteq \mathbf{Sen}(\Sigma_p)$ *and* $\Psi^\circ = [\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_c}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\sigma_{ic}(\sigma_{ip}^{-1}(\Phi^\bullet))) \subseteq \mathbf{Sen}\Sigma_c$. *If* $\Psi \nsubseteq \Psi^\circ$ *then there is an extension* $\mathbf{I}^+$ *of* $\mathbf{I}$ *by models such that there is no interpolant for* Φ *and* Ψ *in* $\mathbf{I}^+$.

Proof. Let $\mathbf{I}^+$ be the extension of $\mathbf{I}$ by a new Σ_p-model M (and its reducts $\lceil M|_\tau \rceil \in \mathbf{Mod}^+(\Sigma)$ for $\tau \colon \Sigma \to \Sigma_p$), with $Th^+(M) = \Phi^\bullet$, and a new Σ_c-model N (and its reducts $\lceil N|_\tau \rceil \in \mathbf{Mod}^+(\Sigma)$ for $\tau \colon \Sigma \to \Sigma_c$), with $Th^+(N) = \Psi^\circ$.

Then $\sigma_{pu}(\Phi) \models^+ \sigma_{cu}(\Psi)$—the corresponding arguments in the proofs of Lemmas 4.1 and 4.4 work here as well.

Consider now any set $\Theta \subseteq \mathbf{Sen}(\Sigma_i)$. If $\sigma_{ip}(\Theta) \nsubseteq \Phi^\bullet$ then $M \not\models^+ \sigma_{ip}(\Theta)$, but $M \models^+ \Phi$ (since $\Phi \subseteq \Phi^\bullet$), and so $\Phi \not\models^+ \sigma_{ip}(\Theta)$. Otherwise $\Theta \subseteq \sigma_{ip}^{-1}(\Phi^\bullet)$, so $\sigma_{ic}(\Theta) \subseteq \sigma_{ic}(\sigma_{ip}^{-1}(\Phi^\bullet)) \subseteq \Psi^\circ$. Hence $N \models^+ \sigma_{ic}(\Theta)$ and since $N \not\models^+ \Psi$, $\sigma_{ic}(\Theta) \not\models^+ \Psi$. Thus, no set $\Theta \subseteq \mathbf{Sen}(\Sigma_i)$ is an interpolant for Φ and Ψ in $\mathbf{I}^+$. ⊣

The converse of Corollary 5.1 does not hold, since the conclusion follows as well when we limit our attention to consequences of Φ, rather than arbitrary sentences in $\Phi^\bullet = [\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_p}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\Phi)$.

To avoid repetition, for the rest of this section let

$$\Theta^* = \sigma_{ip}^{-1}([\sigma_{pu}(\Phi) \overset{\Sigma_p}{\underset{\Sigma_u}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\Phi) \cap Th(\Phi))$$

(more explicitly: $\Theta^* = \{\theta \in \mathbf{Sen}(\Sigma_i) \mid \sigma_{ip}(\theta) \in [\sigma_{pu}(\Phi) \overset{\Sigma_p}{\underset{\Sigma_u}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\Phi), \Phi \models \sigma_{ip}(\theta)\}$).

Lemma 5.2. *If* $\Psi \nsubseteq [\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_c}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta^*))$ *then no interpolant for* Φ *and* Ψ *is stable under extensions of* $\mathbf{I}$ *by models.*

Proof. Consider an interpolant $\Theta \subseteq \mathbf{Sen}(\Sigma_i)$ for Φ and Ψ in $\mathbf{I}$.

If $\Theta \nsubseteq \Theta^*$ then $\sigma_{ip}(\Theta) \nsubseteq [\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_p}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\Phi)$, since $\Phi \models \sigma_{ip}(\Theta)$. Therefore, by Corollary 4.3, the interpolant Θ for Φ and Ψ is not stable under extensions of $\mathbf{I}$ by models.

Otherwise $\Theta \subseteq \Theta^*$. Then we have $\Psi \nsubseteq [\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_c}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta))$ since $[\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_c}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta)) \subseteq [\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_c}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta^*))$. Hence, by Corollary 4.5, the interpolant Θ for Φ and Ψ is not stable under extensions of $\mathbf{I}$ by models. ⊣

The thesis of Lemma 5.2 seems weaker that that of Corollary 5.1—but only superficially so:

Lemma 5.3. *If no interpolant for* Φ *and* Ψ *is stable under extensions of* $\mathbf{I}$ *by models then there is an extension* $\mathbf{I}^+$ *of* $\mathbf{I}$ *by models such that* Φ *and* Ψ *have no interpolant in* $\mathbf{I}^+$.

Proof. Let $\mathcal{E}$ be the family of all interpolants $\Theta \subseteq \mathbf{Sen}(\Sigma_i)$ for Φ and Ψ in $\mathbf{I}$. For each $\Theta \in \mathcal{E}$, let $\mathbf{I}^\Theta$ be an extension of $\mathbf{I}$ by models such that Θ is not an interpolant for Φ and Ψ in $\mathbf{I}^\Theta$. Without loss of generality we may assume that the new models added in $\mathbf{I}^\Theta$ are distinct for $\Theta \in \mathcal{E}$, i.e., model classes $\mathbf{Mod}^\Theta(\Sigma) \setminus \mathbf{Mod}(\Sigma)$, for $\Theta \in \mathcal{E}$, $\Sigma \in |\mathbf{Sig}|$, are mutually disjoint. Define $\mathbf{I}^+$ to be the extension of $\mathbf{I}$ by models such that for $\Sigma \in |\mathbf{Sig}|$, $\mathbf{Mod}^+(\Sigma) = \bigcup_{\Theta \in \mathcal{E}} \mathbf{Mod}^\Theta(\Sigma)$ with the satisfaction relation inherited from the appropriate $\mathbf{I}^\Theta$, $\Theta \in \mathcal{E}$. Then $\sigma_{pu}(\Phi) \models^+ \sigma_{cu}(\Psi)$, since this holds in every $\mathbf{I}^\Theta$, $\Theta \in \mathcal{E}$. Moreover, none of $\Theta \in \mathcal{E}$ is an interpolant for Φ and Ψ in $\mathbf{I}^+$, since it is not an interpolant for Φ and Ψ in $\mathbf{I}^\Theta$. Consequently, there is no interpolant for Φ and Ψ in $\mathbf{I}^+$. ⊣

Corollary 5.4. *If $\Psi \not\subseteq [\sigma_{pu}(\Phi) \underset{\Sigma_c}{\overset{\Sigma_u}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta^*))$ then there is an extension $\mathbf{I}^+$ of $\mathbf{I}$ by models such that there is no interpolant for Φ and Ψ in $\mathbf{I}^+$.*

Proof. Directly by Lemmas 5.2 and 5.3. ⊣

Theorem 5.5. *There is an interpolant for Φ and Ψ in every extension of $\mathbf{I}$ by models if and only if $\Psi \subseteq [\sigma_{pu}(\Phi) \underset{\Sigma_c}{\overset{\Sigma_u}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta^*))$ and $\sigma_{ic}(\Theta^*) \models \Psi$.*

Proof. For the “if” part: by definition of Θ^*, we have $\Phi \models \sigma_{ip}(\Theta^*)$, and so if $\sigma_{ic}(\Theta^*) \models \Psi$ then Θ^* is an interpolant for Φ and Ψ in $\mathbf{I}$. Moreover, since $\sigma_{ip}(\Theta^*) \subseteq [\sigma_{pu}(\Phi) \underset{\Sigma_p}{\overset{\Sigma_u}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\Phi)$, by Theorem 4.6, if $\Psi \subseteq [\sigma_{pu}(\Phi) \underset{\Sigma_c}{\overset{\Sigma_u}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta^*))$ then Θ^* is an interpolant for Φ and Ψ in every extension of $\mathbf{I}$ by models.

For the “only if” part: if there is an interpolant for Φ and Ψ in every extension of $\mathbf{I}$ by models then, by contrapositive of Lemma 5.3, there is an interpolant $\Theta \subseteq \mathbf{Sen}(\Sigma_i)$ for Φ and Ψ in $\mathbf{I}$ that is stable under extensions of $\mathbf{I}$ by models. Therefore, by Theorem 4.6, $\Psi \subseteq [\sigma_{pu}(\Phi) \underset{\Sigma_c}{\overset{\Sigma_u}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta))$ and $\sigma_{ip}(\Theta) \subseteq [\sigma_{pu}(\Phi) \underset{\Sigma_p}{\overset{\Sigma_u}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\Phi)$. Together with $\Phi \models \sigma_{ip}(\Theta)$, the latter implies $\Theta \subseteq \Theta^*$. Thus $\sigma_{ic}(\Theta) \subseteq \sigma_{ic}(\Theta^*)$, hence $\Psi \subseteq [\sigma_{pu}(\Phi) \underset{\Sigma_c}{\overset{\Sigma_u}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta^*))$, and since $\sigma_{ic}(\Theta) \models \Psi$, we also have $\sigma_{ic}(\Theta^*) \models \Psi$—which completes the proof. ⊣

Example 5.6. Recall Example 4.7. As argued there, every interpolant for $r \wedge p$ and $p \vee q$ in $\mathbf{PL}$ is fragile. Consequently, by Lemma 5.3, there is an extension of $\mathbf{PL}$ by models in which $r \wedge p$ and $p \vee q$ have no interpolant. Let us also check how Theorem 5.5 works here:

As in Example 4.7, $[r \wedge p \underset{\Sigma_p}{\overset{\Sigma_u}{\rightsquigarrow}} p \vee q](r \wedge p) = \{r \wedge p, p \vee r, p \vee p\}$. Then, applying the notation Θ^* as defined above for the case at hand, $\Theta^* = \{p \vee p\}$. Recalling another argument in Example 4.7, $[r \wedge p \underset{\Sigma_c}{\overset{\Sigma_u}{\rightsquigarrow}} p \vee q](\Theta^*) = \{p \vee p\}$, and so $p \vee q \notin [r \wedge p \underset{\Sigma_c}{\overset{\Sigma_u}{\rightsquigarrow}} p \vee q](\Theta^*)$. Thus, by Theorem 5.5, it is not the case that in every extension of $\mathbf{PL}$ by models there is an interpolant for $r \wedge p$ and $p \vee q$.

Looking now at the interpolants for $(p \vee r) \wedge (p \vee \neg r)$ and $(p \vee q) \wedge (p \vee \neg q)$, as in Example 4.7, we have

$$[(p \vee r) \wedge (p \vee \neg r) \underset{\Sigma_p}{\overset{\Sigma_u}{\rightsquigarrow}} (p \vee q) \wedge (p \vee \neg q)]((p \vee r) \wedge (p \vee \neg r)) = \{(p \vee r) \wedge (p \vee \neg r), (p \vee p) \wedge (p \vee \neg p)\}.$$

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

Therefore, again applying the notation Θ^* for the current case, $\Theta^* = \{(p \vee p) \wedge (p \vee \neg p)\}$, and then:

$$[(p \vee r) \wedge (p \vee \neg r) \overset{\Sigma_u}{\underset{\Sigma_c}{\leadsto}} (p \vee q) \wedge (p \vee \neg q)](\Theta^*) = \\ \{(p \vee q) \wedge (p \vee \neg q), (p \vee p) \wedge (p \vee \neg p)\},$$

which contains $(p \vee q) \wedge (p \vee \neg q)$. Since $(p \vee p) \wedge (p \vee \neg p) \models (p \vee q) \wedge (p \vee \neg q)$, by Theorem 5.5, $(p \vee r) \wedge (p \vee \neg r)$ and $(p \vee q) \wedge (p \vee \neg q)$ have an interpolant in every extension of **PL** by models. Indeed, in Example 4.7 we argued independently that $(p \vee p) \wedge (p \vee \neg p)$ is such an interpolant.

§6. Spoiling interpolation by new sentences. As before, we study interpolation in an institution $\mathbf{I} = \langle \mathbf{Sig}, \mathbf{Sen}, \mathbf{Mod}, \langle \models_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|} \rangle$ over a commutative square of signature morphisms $(*)$.

Changes to a logical system and its properties that may arise when new sentences are introduced are in no sense dual to those resulting from extending the logical system by new models. In particular, new sentences do not modify entailments between the sentences of the original system, so they never spoil existing interpolants for old sentences. However, on the one hand, new sentences (over the premise and conclusion signatures) may lead to new entailments $\sigma_{pu}(\Phi) \models^+ \sigma_{cu}(\Psi)$ with no interpolant for Φ and Ψ (when Φ or Ψ involve new sentences). On the other hand, adding appropriate new sentences (over the interpolant signature) may restore (or establish) the interpolation property (with new interpolants involving new sentences).

The first rough idea (see, for instance, the semantic characterisation of interpolation in [15]) is that to spoil interpolation for the diagram $(*)$, we look for a class $\mathcal{K} \subseteq \mathbf{Mod}(\Sigma_i)$ that is not definable in $\mathbf{I}$, and then build an extension $\mathbf{I}^+$ of $\mathbf{I}$ by new sentences $\varphi \in \mathbf{Sen}^+(\Sigma_p)$ and $\psi \in \mathbf{Sen}^+(\Sigma_c)$ such that $Mod^+(\varphi) = \mathcal{K}|_{\sigma_{ip}}^{-1}$ and $Mod^+(\psi) = \mathcal{K}|_{\sigma_{ic}}^{-1}$. It follows then that $\sigma_{pu}(\varphi) \models^+ \sigma_{cu}(\psi)$, and it may seem that there should be no interpolant for φ and ψ (since such an interpolant would have to define $\mathcal{K}$). However, the latter need not be true in general.

One technical nuance is that a set $\Theta \subseteq \mathbf{Sen}^+(\Sigma_i)$ of sentences may then be an interpolant for φ and ψ even if $Mod^+(\Theta) \neq \mathcal{K}$, namely when $\mathcal{K} \subsetneq Mod^+(\Theta)$ and no model in $Mod^+(\Theta) \setminus \mathcal{K}$ has a σ_{ic}-expansion to a model in $\mathbf{Mod}(\Sigma_c)$.

EXAMPLE 6.1. In the institution $\mathbf{EQ}_\emptyset$ of equational logic (with empty carriers permitted) consider the diagram $(*)$, where Σ_i has two sorts s, t and constants $a, b \colon t$, Σ_p extends it by a unary operation $f \colon s \to t$, Σ_c extends Σ_i by a constant $c \colon s$, $\Sigma_u = \Sigma_p \cup \Sigma_c$, and the signature morphisms are inclusions. Let $\Phi = \{\forall x{:}s.\, f(x) = a, \forall x{:}s.\, f(x) = b\} \subseteq \mathbf{Sen}_{\mathbf{EQ}_\emptyset}(\Sigma_p)$ and $\Psi = \{a = b\} \subseteq \mathbf{Sen}_{\mathbf{EQ}_\emptyset}(\Sigma_c)$. Then $\Phi \models_{\Sigma_u} \Psi$ but $Mod(\Phi)|_{\sigma_{ip}} \nsubseteq Mod(\Psi)|_{\sigma_{ic}}$ (since $Mod(\Phi)$ contains models with the carrier of sort s empty, while $Mod(\Psi)$ does not). However, $\Theta = \{\forall x{:}s.\, a = b\} \subseteq \mathbf{Sen}_{\mathbf{EQ}_\emptyset}(\Sigma_i)$ is an interpolant for Φ and Ψ.

Another technicality is that the strong requirement $Mod^+(\varphi) = \mathcal{K}|_{\sigma_{ip}}^{-1}$ may be weakened to $Mod^+(\varphi)|_{\sigma_{ip}} = \mathcal{K}$. Similarly, at the conclusion side, it is enough to assume that all σ_{ic}-expansions of the models in $\mathcal{K}$ are in $Mod(\psi)$, $\mathcal{K}|_{\sigma_{ic}}^{-1} \subseteq Mod(\psi)$, or equivalently, no model in $\mathcal{K}$ is a σ_{ic}-reduct of a Σ_c-model outside $Mod(\psi)$, $\mathcal{K} \subseteq \mathbf{Mod}(\Sigma_i) \setminus ((\mathbf{Mod}(\Sigma_c) \setminus Mod(\psi))|_{\sigma_{ic}})$. We may also permit a gap between $Mod^+(\varphi)|_{\sigma_{ip}}$ and $\mathbf{Mod}(\Sigma_i) \setminus ((\mathbf{Mod}(\Sigma_c) \setminus Mod(\psi))|_{\sigma_{ic}})$ as long as no definable class separates them.

Most importantly though, new sentences over signatures Σ_p and Σ_c may result in new Σ_i-sentences as well (as translations of the added sentences), and some Σ_i-model classes that are not definable in $\mathbf{I}$ may become definable in $\mathbf{I}^+$.

The following notion will be used to take care of this: for any signature $\Sigma \in |\mathbf{Sig}|$ and collection $\mathcal{F} = \{\langle \Sigma_j, \mathcal{M}_j \rangle \mid \Sigma_j \in |\mathbf{Sig}|, \mathcal{M}_j \subseteq \mathbf{Mod}(\Sigma_j), j \in \mathcal{J}\}$,[11] we say that a class $\mathcal{M} \subseteq \mathbf{Mod}(\Sigma)$ of Σ-models is *definable in* $\mathbf{I}$ *from* $\mathcal{F}$ if for a family of signature morphisms $\tau_l \colon \Sigma_{j_l} \to \Sigma$, where $j_l \in \mathcal{J}$, $l \in \mathcal{L}$, and a set $\Phi \subseteq \mathbf{Sen}(\Sigma)$ of Σ-sentences we have $\mathcal{M} = \bigcap_{l \in \mathcal{L}} \mathcal{M}_{j_l}|^{-1}_{\tau_{j_l}} \cap Mod(\Phi)$.

Lemma 6.2. *If there are classes of models* $\mathcal{M} \subseteq \mathbf{Mod}(\Sigma_p)$ *and* $\mathcal{N} \subseteq \mathbf{Mod}(\Sigma_c)$ *such that*:

1. $\mathcal{M}|^{-1}_{\sigma_{pu}} \subseteq \mathcal{N}|^{-1}_{\sigma_{cu}}$ *and*
2. *no class of models* $\mathcal{K} \subseteq \mathbf{Mod}(\Sigma_i)$ *such that* $\mathcal{M}|_{\sigma_{ip}} \subseteq \mathcal{K}$ *and* $\mathcal{K}|^{-1}_{\sigma_{ic}} \subseteq \mathcal{N}$ *is definable in* $\mathbf{I}$ *from* $\{\langle \Sigma_p, \mathcal{M} \rangle, \langle \Sigma_c, \mathcal{N} \rangle\}$,

then there is an extension $\mathbf{I}^+$ *of* $\mathbf{I}$ *by new sentences such that the diagram* $(*)$ *does not admit interpolation.*

Proof. Let $\mathbf{I}^+$ extend $\mathbf{I}$ by the following new sentences: Σ_p-sentence φ (and its translations $\lceil \tau(\varphi) \rceil \in \mathbf{Sen}^+(\Sigma)$ for $\tau \colon \Sigma_p \to \Sigma$) such that $Mod^+(\varphi) = \mathcal{M}$, and Σ_c-sentence ψ (and its translations $\lceil \tau(\psi) \rceil \in \mathbf{Sen}^+(\Sigma)$ for $\tau \colon \Sigma_c \to \Sigma$) such that $Mod^+(\psi) = \mathcal{N}$. Then $\sigma_{pu}(\varphi) \models^+ \sigma_{cu}(\psi)$, since $Mod^+(\sigma_{pu}(\varphi)) = \mathcal{M}|^{-1}_{\sigma_{pu}} \subseteq \mathcal{N}|^{-1}_{\sigma_{cu}} = Mod^+(\sigma_{cu}(\psi))$.

Suppose that there is an interpolant $\Theta^+ \subseteq \mathbf{Sen}^+(\Sigma_i)$ for φ and ψ in $\mathbf{I}^+$. By the construction of $\mathbf{I}^+$, $\Theta^+ = \Theta \cup \{\lceil \tau_l(\varphi) \rceil \mid \tau_l \colon \Sigma_p \to \Sigma_i, l \in \mathcal{L}_p\} \cup \{\lceil \tau_l(\psi) \rceil \mid \tau_l \colon \Sigma_c \to \Sigma_i, l \in \mathcal{L}_c\}$, where $\Theta \subseteq \mathbf{Sen}(\Sigma_i)$ (and $\mathcal{L}_p$ and $\mathcal{L}_c$ are disjoint). This means that $\mathcal{K} = Mod^+(\Theta^+)$ is definable in $\mathbf{I}$ from $\{\langle \Sigma_p, \mathcal{M} \rangle, \langle \Sigma_c, \mathcal{N} \rangle\}$.

However, $\varphi \models^+ \sigma_{ip}(\Theta^+)$, hence $\mathcal{M} \subseteq Mod^+(\sigma_{ip}(\Theta^+)) = \mathcal{K}|^{-1}_{\sigma_{ip}}$ and so $\mathcal{M}|_{\sigma_{ip}} \subseteq \mathcal{K}$. Moreover, $\sigma_{ic}(\Theta^+) \models^+ \psi$, and so $\mathcal{K}|^{-1}_{\sigma_{ic}} = Mod^+(\sigma_{ic}(\Theta^+)) \subseteq \mathcal{N}$—which yields a contradiction. ⊣

Theorem 6.3. *There is an extension* $\mathbf{I}^+$ *of* $\mathbf{I}$ *by new sentences in which the diagram* $(*)$ *does not admit interpolation if and only if there are classes of models* $\mathcal{M} \subseteq \mathbf{Mod}(\Sigma_p)$ *and* $\mathcal{N} \subseteq \mathbf{Mod}(\Sigma_c)$ *such that*:

1. $\mathcal{M}|^{-1}_{\sigma_{pu}} \subseteq \mathcal{N}|^{-1}_{\sigma_{cu}}$ *and*
2. *no class of models* $\mathcal{K} \subseteq \mathbf{Mod}(\Sigma_i)$ *such that* $\mathcal{M}|_{\sigma_{ip}} \subseteq \mathcal{K}$ *and* $\mathcal{K}|^{-1}_{\sigma_{ic}} \subseteq \mathcal{N}$ *is definable in* $\mathbf{I}$ *from* $\{\langle \Sigma_p, \mathcal{M} \rangle, \langle \Sigma_c, \mathcal{N} \rangle\}$.

Proof. The "if" part is Lemma 6.2.

For the "only if" part: consider an extension $\mathbf{I}^+$ of $\mathbf{I}$ by new sentences, and let $\Phi^+ \subseteq \mathbf{Sen}^+(\Sigma_p)$ and $\Psi^+ \subseteq \mathbf{Sen}^+(\Sigma_c)$ be such that $\sigma_{pu}(\Phi^+) \models^+ \sigma_{cu}(\Psi^+)$ but there is

[11] $\mathcal{J}$ is a set of indices that "name" the elements of $\mathcal{F}$; we introduce such sets of indices whenever convenient.

no interpolant for Φ^+ and Ψ^+ in $\mathbf{I}^+$. Put $\mathcal{M} = Mod^+(\Phi^+)$ and $\mathcal{N} = Mod^+(\Psi^+)$. Clearly, $\mathcal{M}|^{-1}_{\sigma_{pu}} \subseteq \mathcal{N}|^{-1}_{\sigma_{cu}}$.

Suppose there is a class of models $\mathcal{K} \subseteq \mathbf{Mod}(\Sigma_i)$ such that $\mathcal{M}|_{\sigma_{ip}} \subseteq \mathcal{K}$ and $\mathcal{K}|^{-1}_{\sigma_{ic}} \subseteq \mathcal{N}$ that is definable in $\mathbf{I}$ from $\{\langle \Sigma_p, \mathcal{M}\rangle, \langle \Sigma_c, \mathcal{N}\rangle\}$. Then there are Σ_i-sentences $\Theta \subseteq \mathbf{Sen}(\Sigma_i)$ and signature morphisms $\tau_l \colon \Sigma_p \to \Sigma_i$, $l \in \mathcal{L}_p$, and $\tau_l \colon \Sigma_c \to \Sigma_i$, $l \in \mathcal{L}_c$, such that $\mathcal{K} = \bigcap_{l \in \mathcal{L}_p} \mathcal{M}|^{-1}_{\tau_l} \cap \bigcap_{l \in \mathcal{L}_c} \mathcal{N}|^{-1}_{\tau_l} \cap Mod(\Theta)$. Put $\Theta^+ = \Theta \cup \bigcup_{l \in \mathcal{L}_p} \tau_l(\Phi^+) \cup \bigcup_{l \in \mathcal{L}_c} \tau_l(\Psi^+) \subseteq \mathbf{Sen}^+(\Sigma_i)$. Then $Mod^+(\Theta^+_i) = \mathcal{K}$, and Θ^+ is an interpolant for Φ^+ and Ψ^+ in $\mathbf{I}^+$—which yields a contradiction. ⊣

Example 6.4. Consider an example in the institution $\mathbf{FO_{EQ}}$ of first-order logic with equality. Let all the signatures in the diagram $(*)$ extend Σ_i, which has exactly one sort Nat, constant $0 \colon Nat$ and operation $s \colon Nat \to Nat$. In addition, Σ_p has $bop \colon Nat \times Nat \to Nat$ and Σ_c has $\_+\_ \colon Nat \times Nat \to Nat$. Finally, $\Sigma_u = \Sigma_p \cup \Sigma_c$, and all four signature morphisms in $(*)$ are inclusions.

Let $\mathcal{M} \subseteq \mathbf{Mod}(\Sigma_p)$ be the class of all models with the carrier set freely generated by 0 and s (where each element is the value of exactly one of the terms of the form $s^n(0)$). Let then $\mathcal{N} \subseteq \mathbf{Mod}(\Sigma_c)$ be the class of models that satisfy the following implication:

$$\psi \equiv (\forall x, y{:}Nat.\, x + 0 = x \wedge x + s(y) = s(x + y)) \Rightarrow \forall x, y{:}Nat.\, x + y = y + x.$$

Let $\mathbf{FO^+_{EQ}}$ be the extension of $\mathbf{FO_{EQ}}$ by a new Σ_p-sentence φ (and its formal translations) such that $Mod^+(\varphi) = \mathcal{M}$.[12] No new Σ_c-sentence is added, since $\mathcal{N}$ is already definable in $\mathbf{FO_{EQ}}$. Clearly, $\mathcal{M}|^{-1}_{\sigma_{pu}} \subseteq \mathcal{N}|^{-1}_{\sigma_{cu}}$, and so $\sigma_{pu}(\varphi) \models^+ \sigma_{cu}(\psi)$.

However, no class of models $\mathcal{K} \subseteq \mathbf{Mod}(\Sigma_i)$ that is definable by first-order sentences excludes non-standard models of natural numbers (with "infinitary" elements). Moreover, there is no signature morphism from Σ_p to Σ_i. Therefore, if $\mathcal{M}|_{\sigma_{ip}} \subseteq \mathcal{K} \subseteq \mathbf{Mod}(\Sigma_i)$ and $\mathcal{K}$ is definable in $\mathbf{FO_{EQ}}$ from $\{\langle \Sigma_p, \mathcal{M}\rangle\}$ then $\mathcal{K}|^{-1}_{\sigma_{ic}} \not\models^+ \psi$ (addition does not have to commute on "infinitary" arguments). Consequently, φ and ψ have no interpolant in $\mathbf{FO^+_{EQ}}$.

However, if we remove the additional operation bop from the signature Σ_p (and replace it by a unary operation $uop \colon Nat \to Nat$) the situation becomes quite different. We have then a (unique) signature morphism $\tau \colon \Sigma_p \to \Sigma_i$, and the sentence $\lceil \tau(\varphi) \rceil \in \mathbf{Sen}^+(\Sigma_i)$ defines up to isomorphism the standard model of natural numbers, and therefore is an interpolant for φ and ψ.

For institutions like $\mathbf{PL}$, where all classes of models are definable, it might seem that all commutative squares of signature morphisms admit interpolation, and no extension by sentences may spoil this property. However, this need not be the case, since in general, in an arbitrary institution, for classes of models $\mathcal{M} \subseteq \mathbf{Mod}(\Sigma_p)$ and $\mathcal{N} \subseteq \mathbf{Mod}(\Sigma_c)$ such that $\mathcal{M}|^{-1}_{\sigma_{pu}} \subseteq \mathcal{N}|^{-1}_{\sigma_{cu}}$ the inclusion $\mathcal{M}|_{\sigma_{ip}} \subseteq \mathbf{Mod}(\Sigma_i) \setminus ((\mathbf{Mod}(\Sigma_c) \setminus \mathcal{N})|_{\sigma_{ic}})$ may fail, and then no class $\mathcal{K} \subseteq \mathbf{Mod}(\Sigma_i)$ satisfies $\mathcal{M}|_{\sigma_{ip}} \subseteq \mathcal{K}$ and $\mathcal{K}|^{-1}_{\sigma_{ic}} \subseteq \mathcal{N}$.

[12] For instance, using Standard ML [30] notation, φ might be written as `datatype` $Nat = 0 \mid s$ `of` Nat.

Example 6.5. In the institution **PL** consider the diagram $(*)$ where $\Sigma_p = \{p\}$, $\Sigma_c = \{q\}$, $\Sigma_u = \{r\}$, and $\Sigma_i = \emptyset$ (this determines the four signature morphisms as well). Note that $\mathbf{Sen}(\Sigma_i)$ is non-empty (it contains for instance false, $\neg\mathsf{false}$, etc.) and $\mathbf{Mod}(\Sigma_i) = \{\emptyset\}$, where $\emptyset$ is the empty Σ_i-model. Putting $\mathcal{M} = \{\{p\}\}$ and $\mathcal{N} = \{\{q\}\}$, we have $\mathcal{M}|^{-1}_{\sigma_{pu}} = \{\{r\}\} = \mathcal{N}|^{-1}_{\sigma_{cu}}$, but $\mathcal{M}|_{\sigma_{ip}} = \{\emptyset\} \not\subseteq \mathbf{Mod}(\Sigma_i) \setminus ((\mathbf{Mod}(\Sigma_c) \setminus \mathcal{N})|_{\sigma_{ic}})$, since $\{\emptyset\}|^{-1}_{\sigma_{ic}} = \{\emptyset, \{q\}\} \not\subseteq \mathcal{N}$. Indeed, there is no interpolant for p and q, even though $\sigma_{pu}(p) = r = \sigma_{cu}(q)$.

The diagram $(*)$ admits *weak amalgamation* if for all models $M \in \mathbf{Mod}(\Sigma_p)$ and $N \in \mathbf{Mod}(\Sigma_c)$ such that $M|_{\sigma_{ip}} = N|_{\sigma_{ic}}$ there is a model $K' \in \mathbf{Mod}(\Sigma_u)$ such that $K'|_{\sigma_{pu}} = M$ and $K'|_{\sigma_{cu}} = N$. The diagram $(*)$ admits *amalgamation* if such a model $K' \in \mathbf{Mod}(\Sigma_u)$ is always unique. This is a standard property used extensively in “institutional” foundations of software specifications [39, 40]. Amalgamation (and hence weak amalgamation) holds for pushouts in all the sample institutions and their variants we defined in Examples 2.1–2.3; it fails though for some non-pushout diagrams.

Lemma 6.6. *Suppose that the diagram* $(*)$ *admits weak amalgamation. Then for all classes of models* $\mathcal{M} \subseteq \mathbf{Mod}(\Sigma_p)$ *and* $\mathcal{N} \subseteq \mathbf{Mod}(\Sigma_c)$, $\mathcal{M}|^{-1}_{\sigma_{pu}} \subseteq \mathcal{N}|^{-1}_{\sigma_{cu}}$ *implies* $(\mathcal{M}|_{\sigma_{ip}})|^{-1}_{\sigma_{ic}} \subseteq \mathcal{N}$.

Proof. Let $M \in \mathcal{M}$, and let $N \in \mathbf{Mod}(\Sigma_c)$ be a σ_{ic}-expansion of $M|_{\sigma_{ip}}$, i.e., $N|_{\sigma_{ic}} = M|_{\sigma_{ip}}$. By the weak amalgamation property we have $K' \in \mathbf{Mod}(\Sigma_u)$ such that $K'|_{\sigma_{pu}} = M$ and $K'|_{\sigma_{cu}} = N$. Then $K' \in \mathcal{M}|^{-1}_{\sigma_{pu}} \subseteq \mathcal{N}|^{-1}_{\sigma_{cu}}$, and so $N = K'|_{\sigma_{cu}} \in \mathcal{N}$. ⊣

Corollary 6.7. *If the diagram* $(*)$ *admits weak amalgamation and each class of* Σ_i*-models is definable then the diagram* $(*)$ *admits interpolation in every extension of the institution* **I** *by new sentences.*

Proof. Directly from Lemma 6.6 and Theorem 6.3. ⊣

It turns out that the weak amalgamation property is also a necessary condition in the above corollary, in a strong sense:

Corollary 6.8. *If the diagram* $(*)$ *does not admit weak amalgamation then it does not admit interpolation in some extension of the institution by new sentences, nor in its further extensions by new sentences.*

Proof. Consider $M \in \mathbf{Mod}(\Sigma_p)$ and $N \in \mathbf{Mod}(\Sigma_c)$ such that $M|_{\sigma_{ip}} = N|_{\sigma_{ic}}$, but there is no model $K' \in \mathbf{Mod}(\Sigma_u)$ such that $K'|_{\sigma_{pu}} = M$ and $K'|_{\sigma_{cu}} = N$. Then the classes $\mathcal{M} = \{M\} \subseteq \mathbf{Mod}(\Sigma_p)$ and $\mathcal{N} = \mathbf{Mod}(\Sigma_c) \setminus \{N\} \subseteq \mathbf{Mod}(\Sigma_c)$ satisfy the requirements 1 ($\mathcal{M}|^{-1}_{\sigma_{pu}} \subseteq \mathcal{N}|^{-1}_{\sigma_{cu}}$) and 2 (since $(\mathcal{M}|_{\sigma_{ip}})|^{-1}_{\sigma_{ic}} \not\subseteq \mathcal{N}$) in Lemma 6.2, and so indeed, as in the proof of Lemma 6.2, interpolation over $(*)$ fails in the extension $\mathbf{I}^+$ of $\mathbf{I}$ by new sentences $\varphi \in \mathbf{Sen}^+(\Sigma_p)$ and $\psi \in \mathbf{Sen}^+(\Sigma_c)$ with $Mod^+(\varphi) = \mathcal{M}$ and $Mod^+(\psi) = \mathcal{N}$. Moreover, since there is no class $\mathcal{K} \subseteq \mathbf{Mod}(\Sigma_i)$ such that $\mathcal{M}|_{\sigma_{ip}} \subseteq \mathcal{K}$ and $\mathcal{K}|^{-1}_{\sigma_{ic}} \subseteq \mathcal{N}$, no further extension of $\mathbf{I}^+$ by new sentences may create an interpolant for φ and ψ. ⊣

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

Theorem 6.9. *Assume that each class of Σ_i-models is definable. Then the diagram* $(*)$ *admits interpolation in every extension of the institution* $\mathbf{I}$ *by new sentences if and only if it admits weak amalgamation.*

Proof. The "if" part is Corollary 6.7; the "only if" part follows by Corollary 6.8. $\dashv$

If we disregard foundational issues (see footnote 4) and extend the institution by enough new sentences to make all classes of Σ_i-models definable (in general this may require a proper class of sentences though) then in such an extension of the institution by new sentences the diagram $(*)$ admits interpolation provided it admits weak amalgamation.

§7. Spoiling interpolation by new models and sentences. As so far, we study interpolation over a commutative diagram of signature morphisms $(*)$ in an institution $\mathbf{I} = \langle \mathbf{Sig}, \mathbf{Sen}, \mathbf{Mod}, \langle \models_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|} \rangle$, in this section addressing the possibilities of spoiling interpolation by extending the institution with new models and sentences.

An extension of an institution $\mathbf{I}$ by new models and sentences is an extension $\mathbf{I}^{+\!+}$ by new sentences of an extension $\mathbf{I}^{+}$ by new models of the institution $\mathbf{I}$.

The order of the extensions used above is irrelevant. For, let $\mathbf{I}^{+}$ be the extension of $\mathbf{I}$ by models $\mathcal{NM} = \langle \mathcal{NM}_\Sigma, \models_\Sigma^{\mathcal{NM}} \subseteq \mathcal{NM}_\Sigma \times \mathbf{Sen}(\Sigma) \rangle_{\Sigma \in |\mathbf{Sig}|}$, and $\mathbf{I}^{+\!+}$ be the extension of $\mathbf{I}^{+}$ by sentences $\mathcal{NS} = \langle \mathcal{NS}_\Sigma, \models_\Sigma^{\mathcal{NS}} \subseteq \mathbf{Mod}^{+}(\Sigma) \times \mathcal{NS}_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|}$ (see Section 2.3 for the definitions and notation). Then define $\mathbf{I}'$ as the extension of $\mathbf{I}$ by sentences $\mathcal{NS}' = \langle \mathcal{NS}_\Sigma, \models_\Sigma^{\mathcal{NS}'} \subseteq \mathbf{Mod}(\Sigma) \times \mathcal{NS}_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|}$, where $M \models_\Sigma^{\mathcal{NS}'} \varphi$ iff $M \models_\Sigma^{\mathcal{NS}} \varphi$ for $\Sigma \in |\mathbf{Sig}|$, $M \in \mathbf{Mod}(\Sigma)$ and $\varphi \in \mathcal{NS}_\Sigma$. Then $\mathbf{I}^{+\!+}$ coincides with the extension of $\mathbf{I}'$ by models $\mathcal{NM}' = \langle \mathcal{NM}_\Sigma, \models_\Sigma^{\mathcal{NM}'} \subseteq \mathcal{NM}_\Sigma \times \mathbf{Sen}'(\Sigma) \rangle_{\Sigma \in |\mathbf{Sig}|}$, where for $\Sigma \in |\mathbf{Sig}|$ and $M \in \mathcal{NM}_\Sigma$, $M \models_\Sigma^{\mathcal{NM}'} \varphi$ iff $M \models_\Sigma^{\mathcal{NM}} \varphi$ for $\varphi \in \mathbf{Sen}(\Sigma)$, and for $\tau \colon \Sigma' \to \Sigma$, $\varphi' \in \mathcal{NS}_{\Sigma'}$, $M \models_\Sigma^{\mathcal{NM}'} \lceil \tau(\varphi') \rceil$ iff $\lceil M|_\tau \rceil \models_{\Sigma'}^{\mathcal{NS}} \varphi'$.

Obviously, we have "sinks" and "sources" of institution morphisms that link institution $\mathbf{I}$ and its extension $\mathbf{I}^{+\!+}$ by models and sentences:

$$\begin{tikzcd} \mathbf{I} \arrow[r, "\mu_{\mathcal{NM}}"] & \mathbf{I}^{+} & \mathbf{I}^{+\!+} \arrow[l, "\mu_{\mathcal{NS}}"'] & & \mathbf{I} & \mathbf{I}' \arrow[l, "\mu_{\mathcal{NS}'}"'] \arrow[r, "\mu_{\mathcal{NM}'}"] & \mathbf{I}^{+\!+} \end{tikzcd}$$

However, in general there is no institution morphism between $\mathbf{I}$ and $\mathbf{I}^{+\!+}$. Their relationship can be captured by another kind of mapping between institutions, where sentences and models translate covariantly, called institution encodings [44] or forward institution morphisms [27] (used in an interesting way for instance in [5]).

Corollary 3.3 gives a sufficient condition that ensures that the Craig interpolation property over a diagram $(*)$ is stable under extensions of the institution by new models and sentences. The key result here is that this is also a necessary condition: if the conditions 1 and 2 stated in Corollary 3.3 fail for the diagram $(*)$ then in some extension of the institution by new models and sentences, the diagram $(*)$ does not admit interpolation.

Theorem 7.1. *The diagram* $(*)$ *admits interpolation in all extensions of* $\mathbf{I}$ *by new models and sentences if and only if at least one of the following conditions holds*:

1. $\sigma_{ip} \colon \Sigma_i \to \Sigma_p$ *is a retraction and* $\sigma_{cu} \colon \Sigma_c \to \Sigma_u$ *is a coretraction, or*
2. $\sigma_{ic} \colon \Sigma_i \to \Sigma_c$ *is a retraction and* $\sigma_{pu} \colon \Sigma_p \to \Sigma_u$ *is a coretraction.*

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

Proof. The “if” part follows by Corollary 3.3.

For the “only if” part, assume that conditions 1 and 2 do not hold. Let $\mathbf{I}^{+}$ be the extension of $\mathbf{I}$ by a new Σ_p-model M and a new Σ_c-model N (and their formal reducts) such that M and N do not satisfy any $\mathbf{I}$-sentences. Let then $\mathbf{I}^{++}$ be the extension of $\mathbf{I}^{+}$ by a new Σ_p-sentence φ and a new Σ_c-sentence ψ (and their formal translations) such that:

- $Mod^{++}(\varphi) = \{M\} \cup \{\lceil N|_{\tau_{pi};\sigma_{ic}} \rceil \mid \tau_{pi} \colon \Sigma_p \to \Sigma_i, \tau_{pi};\sigma_{ip} = id_{\Sigma_p}\}$.
- $Mod^{++}(\psi) = \{\lceil M|_{\sigma_{cu};\tau_{up}} \rceil \mid \tau_{up} \colon \Sigma_u \to \Sigma_p, \sigma_{pu};\tau_{up} = id_{\Sigma_p}\} \cup$
  $\{\lceil N|_{\tau_{cc}} \rceil \mid \tau_{cc} \colon \Sigma_c \to \Sigma_c, \tau_{cc} \neq id_{\Sigma_c}\}$.

We have then:

- $Mod^{++}(\sigma_{pu}(\varphi)) = Mod^{++}(\varphi)|^{-1}_{\sigma_{pu}} = \mathcal{M}^{\varphi} \cup \mathcal{N}^{\varphi}$, where
  $\mathcal{M}^{\varphi} = \{\lceil M|_{\tau_{up}} \rceil \mid \tau_{up} \colon \Sigma_u \to \Sigma_p, \sigma_{pu};\tau_{up} = id_{\Sigma_p}\}$.
  $\mathcal{N}^{\varphi} = \{\lceil N|_{\rho_{uc}} \rceil \mid \rho_{uc} \colon \Sigma_u \to \Sigma_c, \sigma_{pu};\rho_{uc} = \tau_{pi};\sigma_{ic} \colon \Sigma_p \to \Sigma_c,$
  $\tau_{pi} \colon \Sigma_p \to \Sigma_i, \tau_{pi};\sigma_{ip} = id_{\Sigma_p}\}$.
- $Mod^{++}(\sigma_{cu}(\psi)) = Mod^{++}(\psi)|^{-1}_{\sigma_{cu}} = \mathcal{M}^{\psi} \cup \mathcal{N}^{\psi}$, where
  $\mathcal{M}^{\psi} = \{\lceil M|_{\rho_{up}} \rceil \mid \rho_{up} \colon \Sigma_u \to \Sigma_p, \sigma_{cu};\rho_{up} = \sigma_{cu};\tau_{up} \colon \Sigma_c \to \Sigma_p,$
  $\tau_{up} \colon \Sigma_u \to \Sigma_p, \sigma_{pu};\tau_{up} = id_{\Sigma_p}\}$.
  $\mathcal{N}^{\psi} = \{\lceil N|_{\rho_{uc}} \rceil \mid \rho_{uc} \colon \Sigma_u \to \Sigma_c, \sigma_{cu};\rho_{uc} \neq id_{\Sigma_c}\}$.

Clearly, $\mathcal{M}^{\varphi} \subseteq \mathcal{M}^{\psi}$. Moreover, $\mathcal{N}^{\varphi} \subseteq \mathcal{N}^{\psi}$ when $\sigma_{ip} \colon \Sigma_i \to \Sigma_p$ is not a retraction (since then $\mathcal{N}^{\varphi} = \emptyset$) or $\sigma_{cu} \colon \Sigma_c \to \Sigma_u$ is not a coretraction (since then all $\rho_{uc} \colon \Sigma_u \to \Sigma_c$ satisfy $\sigma_{cu};\rho_{uc} \neq id_{\Sigma_c}$). However, under our assumptions, at least one of these conditions holds (since condition 1 above does not hold), so we have $Mod^{++}(\sigma_{pu}(\varphi)) \subseteq Mod^{++}(\sigma_{cu}(\psi))$, that is, $\sigma_{pu}(\varphi) \models^{++} \sigma_{cu}(\psi)$.

Suppose now that $\Theta \subseteq \mathbf{Sen}^{++}(\Sigma_i)$ is an interpolant for φ and ψ in $\mathbf{I}^{++}$. In particular, $\varphi \models^{++} \sigma_{ip}(\Theta)$ and so $M \models^{++} \sigma_{ip}(\Theta)$.

For $\mathbf{I}$-sentences $\theta \in \mathbf{Sen}(\Sigma_i)$, $M \not\models^{++} \sigma_{ip}(\theta)$, so Θ must not contain any “old” sentences $\theta \in \mathbf{Sen}(\Sigma_i)$. Hence all sentences in Θ are formal translations of φ or of ψ to the signature Σ_i.

Consider such a translation of φ, $\lceil \tau_{pi}(\varphi) \rceil \in \mathbf{Sen}^{++}(\Sigma_i)$, where $\tau_{pi} \colon \Sigma_p \to \Sigma_i$. If $\lceil \tau_{pi}(\varphi) \rceil \in \Theta$ then $M \models^{++} \sigma_{ip}(\lceil \tau_{pi}(\varphi) \rceil)$, and so $\tau_{pi};\sigma_{ip} = id_{\Sigma_p}$. It follows that $\lceil N|_{\tau_{pi};\sigma_{ic}} \rceil \models^{++} \varphi$, and so $N \models^{++} \sigma_{ic}(\lceil \tau_{pi}(\varphi) \rceil)$.

Consider now a translation of ψ, $\lceil \rho_{ci}(\psi) \rceil \in \mathbf{Sen}^{++}(\Sigma_i)$, where $\rho_{ci} \colon \Sigma_c \to \Sigma_i$. If $\lceil \rho_{ci}(\psi) \rceil \in \Theta$ then $M \models^{++} \sigma_{ip}(\lceil \rho_{ci}(\psi) \rceil)$. Therefore $\rho_{ci};\sigma_{ip} = \sigma_{cu};\tau_{up}$ for some $\tau_{up} \colon \Sigma_u \to \Sigma_c$ such that $\sigma_{pu};\tau_{up} = id_{\Sigma_p}$. Then $\sigma_{pu} \colon \Sigma_p \to \Sigma_u$ is a retraction, and so $\sigma_{ic} \colon \Sigma_i \to \Sigma_c$ is not a coretraction (since condition 2 does not hold). Therefore, $\rho_{ci};\sigma_{ic} \neq id_{\Sigma_c}$, hence $\lceil N|_{\rho_{ci};\sigma_{ic}} \rceil \models^{++} \psi$, and so $N \models^{++} \sigma_{ic}(\lceil \rho_{ci}(\psi) \rceil)$.

Consequently, $N \models^{++} \sigma_{ic}(\Theta)$. But $N \not\models^{++} \psi$, hence $\sigma_{ic}(\Theta) \not\models^{++} \psi$.

This shows that no $\Theta \subseteq \mathbf{Sen}^{++}(\Sigma_i)$ is an interpolant for φ and ψ in $\mathbf{I}^{++}$ when conditions 1 and 2 do not hold. ⊣

§8. Bounded interpolation.

It may be argued that in practical applications the relevant sets of sentences considered in the definition of the interpolation property (premises, conclusions and, most crucially, interpolants) should be finite. In this section we show how the characterisation results concerning the fragility of interpolants and interpolation carry over to this case as well. We discuss this in a somewhat more general setting, allowing the "size" of the sets of sentences involved to be bounded by a suitable cardinal (rather than requiring them to be finite).

Let κ be a regular cardinal[13] —the finitary case mentioned above corresponds to $\kappa = \aleph_0$.

As so far, let $\mathbf{I} = \langle \mathbf{Sig}, \mathbf{Sen}, \mathbf{Mod}, \langle \models_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|} \rangle$ be an institution; we consider a commutative diagram $(*)$ in the category of signatures $\mathbf{Sig}$.

An interpolant $\Theta \subseteq \mathbf{Sen}(\Sigma_i)$ for $\Phi \subseteq \mathbf{Sen}(\Sigma_p)$ and $\Psi \subseteq \mathbf{Sen}(\Sigma_c)$ is κ-*bounded* if the cardinality of Θ is smaller than κ. A commutative square $(*)$ of signature morphisms *admits* κ-*bounded interpolation* if all sets $\Phi \subseteq \mathbf{Sen}(\Sigma_p)$ and $\Psi \subseteq \mathbf{Sen}(\Sigma_c)$ of cardinalities smaller than κ such that $\sigma_{pu}(\Phi) \models \sigma_{cu}(\Psi)$ have a κ-bounded interpolant.

A diagram $(*)$ may admit κ-bounded interpolation without admitting Craig interpolation (as defined in Section 3.2), and the opposite implication does not hold either (in compact institutions though, if $(*)$ admits Craig interpolation then it admits κ-bounded interpolation). Similarly, for $\kappa' < \kappa$, any κ'-bounded interpolant is κ-bounded, but a diagram $(*)$ may admit κ'-bounded interpolation without admitting κ-bounded interpolation, and the opposite implication does not hold either.

In Section 4 we discussed when particular interpolants may be spoiled by extending the institution by new models. The arguments and results there apply directly to the special situation when the interpolant is κ-bounded. In particular, Theorem 4.6 holds for κ-bounded interpolants as it is.

Section 5 culminates with Theorem 5.5, which in a way characterises the set $\Theta^* \subseteq \mathbf{Sen}(\Sigma_i)$ defined there as the largest possible interpolant for $\Phi \subseteq \mathbf{Sen}(\Sigma_p)$ and $\Psi \subseteq \mathbf{Sen}(\Sigma_c)$ stable under extensions of the institution by new models. For the bounded case we have to be able to choose an appropriate "sufficiently small" subset of Θ^*, otherwise the result and its proof carries over:

Theorem 8.1. *Consider* $\Phi \subseteq \mathbf{Sen}(\Sigma_p)$ *and* $\Psi \subseteq \mathbf{Sen}(\Sigma_c)$ *such that* $\sigma_{pu}(\Phi) \models \sigma_{cu}(\Psi)$. *Put* $\Theta^* = \sigma_{ip}^{-1}([\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_p}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\Phi) \cap Th(\Phi))$.

There is a κ-*bounded interpolant for* Φ *and* Ψ *in every extension of* $\mathbf{I}$ *by models if and only if for some* $\Theta^\circ \subseteq \Theta^*$ *of cardinality smaller than* κ, $\Psi \subseteq [\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_c}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta^\circ))$ *and* $\sigma_{ic}(\Theta^\circ) \models \Psi$.

Proof. For the "if" part: by definition of Θ^*, since by the assumption $\Theta^\circ \subseteq \Theta^*$, we have $\Phi \models \sigma_{ip}(\Theta^\circ)$. Together with $\sigma_{ic}(\Theta^\circ) \models \Psi$ this means that Θ° is an interpolant for Φ and Ψ in $\mathbf{I}$. Moreover, since $\sigma_{ip}(\Theta^\circ) \subseteq [\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_p}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\Phi)$ and $\Psi \subseteq [\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_c}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta^\circ))$, Theorem 4.6 implies that Θ° is an interpolant for Φ and Ψ in every extension of $\mathbf{I}$ by models. Of course, Θ° is κ-bounded by the assumption.

[13] An infinite cardinal κ is regular if the cardinality of the union of every set of cardinality smaller than κ of sets of cardinality smaller than κ is smaller than κ [33].

For the “only if” part, if there is a κ-bounded interpolant for Φ and Ψ in every extension of $\mathbf{I}$ by models then there is a κ-bounded interpolant $\Theta^\circ \subseteq \mathbf{Sen}(\Sigma_i)$ for Φ and Ψ in $\mathbf{I}$ that is stable under extensions of $\mathbf{I}$ by models—this follows by mimicking the proof of Lemma 5.3 with only κ-bounded interpolants considered. Therefore, by Theorem 4.6, $\Psi \subseteq [\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_c}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta^\circ))$ and $\sigma_{ip}(\Theta^\circ) \subseteq [\sigma_{pu}(\Phi) \overset{\Sigma_u}{\underset{\Sigma_p}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\Phi)$. Together with $\Phi \models \sigma_{ip}(\Theta^\circ)$, the latter implies $\Theta^\circ \subseteq \Theta^*$. Since we also have $\sigma_{ic}(\Theta^\circ) \models \Psi$—this completes the proof. ⊣

In the context of the κ-bounded interpolation property, we may additionally assume that the cardinalities of $\Phi \subseteq \mathbf{Sen}(\Sigma_p)$ and $\Psi \subseteq \mathbf{Sen}(\Sigma_c)$ are smaller than κ—this does not change the above result though.

To adapt the results of Section 6 to the κ-bounded interpolation, we first have to adjust some basic notions.

For any signature $\Sigma \in |\mathbf{Sig}|$, a class of models $\mathcal{M} \subseteq \mathbf{Mod}(\Sigma)$ is *κ-definable* in $\mathbf{I}$ if for a set $\Phi \subseteq \mathbf{Sen}(\Sigma)$ of cardinality smaller than κ, $\mathcal{M} = Mod(\Phi)$. Then, given a collection $\mathcal{F} = \{\langle \Sigma_j, \mathcal{M}_j \rangle \mid \Sigma_j \in |\mathbf{Sig}|, \mathcal{M}_j \subseteq \mathbf{Mod}(\Sigma_j), j \in \mathcal{J}\}$, $\mathcal{M} \subseteq \mathbf{Mod}(\Sigma)$ is *κ-definable in* $\mathbf{I}$ *from* $\mathcal{F}$ if for a set $\Phi \subseteq \mathbf{Sen}(\Sigma)$ of Σ-sentences of cardinality smaller than κ and a set $\mathcal{L}$ of cardinality smaller than κ with signature morphisms $\tau_l : \Sigma_{j_l} \to \Sigma$, $j_l \in \mathcal{J}$, $l \in \mathcal{L}$, we have $\mathcal{M} = \bigcap_{l \in \mathcal{L}} \mathcal{M}_{j_l}|^{-1}_{\tau_{j_l}} \cap Mod(\Phi)$.

The appropriate reformulation of Theorem 6.3 for the bounded interpolation is rather obvious now:

Theorem 8.2. *There is an extension* $\mathbf{I}^+$ *of* $\mathbf{I}$ *by new sentences in which the diagram* $(*)$ *does not admit κ-bounded interpolation if and only if there are classes of models* $\mathcal{M} \subseteq \mathbf{Mod}(\Sigma_p)$ *and* $\mathcal{N} \subseteq \mathbf{Mod}(\Sigma_c)$ *such that*:

1. $\mathcal{M}|^{-1}_{\sigma_{pu}} \subseteq \mathcal{N}|^{-1}_{\sigma_{cu}}$ *and*
2. *no class of models* $\mathcal{K} \subseteq \mathbf{Mod}(\Sigma_i)$ *such that* $\mathcal{M}|_{\sigma_{ip}} \subseteq \mathcal{K}$ *and* $\mathcal{K}|^{-1}_{\sigma_{ic}} \subseteq \mathcal{N}$ *is κ-definable in* $\mathbf{I}$ *from* $\{\langle \Sigma_p, \mathcal{M} \rangle, \langle \Sigma_c, \mathcal{N} \rangle\}$.

Proof. The proofs of Theorem 6.3 and of Lemma 6.2 essentially carry over to the present case (with κ-definability and sets of sentences of cardinality smaller than κ used in place of definability and arbitrary sets of sentences, respectively). ⊣

The links between the weak amalgamation and interpolation properties carry over to the bounded interpolation as well. In particular, Lemma 6.6 remains unaffected, and Corollary 6.8 holds for the κ-bounded interpolation. Moreover, Corollary 6.7 and Theorem 6.9 hold for the κ-bounded interpolation if we strengthen the requirement of definability of Σ_i-model classes to their κ-definability.

Interestingly, the final remark of Section 6 indicating that if the weak amalgamation property is assumed, the interpolation property may be ensured by extending the institution by (a possibly proper class of) new sentences, in the bounded case may be refined in a non-trivial way:

Theorem 8.3. *Assume that the category of signatures* $\mathbf{Sig}$ *is locally small. If the diagram* $(*)$ *admits weak interpolation then there is an extension* $\mathbf{I}^+$ *of* $\mathbf{I}$ *by new sentences such that the diagram* $(*)$ *admits κ-bounded interpolation in* $\mathbf{I}^+$.

Proof. Define institutions $\mathbf{I}_\alpha = \langle \mathbf{Sig}, \mathbf{Sen}_\alpha, \mathbf{Mod}, \langle \models^\alpha_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|} \rangle$ by transfinite induction as follows:

- $\mathbf{I}_0 = \mathbf{I}$,
- for any ordinal α, $\mathbf{I}_{\alpha+1}$ is the extension of $\mathbf{I}_\alpha$ by new Σ_i-sentences θ_Φ, one for each set $\Phi \subseteq \mathbf{Sen}_\alpha(\Sigma_p)$ of cardinality smaller than κ such that $Mod_\alpha(\Phi)|_{\sigma_{ip}}$ is not κ-definable in $\mathbf{I}_\alpha$, with $Mod_{\alpha+1}(\theta_\Phi) = Mod_\alpha(\Phi)|_{\sigma_{ip}}$,
- for any limit ordinal β, $\mathbf{I}_\beta = \langle \mathbf{Sig}, \mathbf{Sen}_\beta, \mathbf{Mod}, \langle \models^\beta_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|} \rangle$, where for $\Sigma \in |\mathbf{Sig}|$, $\mathbf{Sen}_\beta(\Sigma) = \bigcup_{\alpha<\beta} \mathbf{Sen}_\alpha(\Sigma)$ and $\models^\beta_\Sigma = \bigcup_{\alpha<\beta} \models^\alpha_\Sigma$, and for $\sigma \colon \Sigma \to \Sigma'$, $\mathbf{Sen}_\beta(\sigma) = \bigcup_{\alpha<\beta} \mathbf{Sen}_\alpha(\sigma)$.

By the construction, for any ordinal α and set $\Phi \subseteq \mathbf{Sen}_\alpha(\Sigma_p)$ of cardinality smaller than κ, the class $Mod_\alpha(\Phi)|_{\sigma_{ip}} \subseteq \mathbf{Mod}(\Sigma_i)$ is κ-definable in $\mathbf{I}_{\alpha+1}$.

Let γ be the initial (least) ordinal of cardinality κ. Then in $\mathbf{I}_\gamma$, for any set $\Phi \subseteq \mathbf{Sen}_\gamma(\Sigma_p)$ of cardinality smaller than κ, $Mod_\gamma(\Phi)|_{\sigma_{ip}}$ is κ-definable (which implies that $\mathbf{I}_{\gamma+1} = \mathbf{I}_\gamma$), since for any such set Φ we have that $\Phi \subseteq \mathbf{Sen}_\alpha(\Sigma_p)$ for some $\alpha < \gamma$. This holds for instance for $\alpha = \bigcup\{\delta < \gamma \mid \Phi \cap \mathbf{Sen}_\delta(\Sigma_p) \neq \emptyset\}$, since the cardinality of α, which is the union of a set of cardinality smaller than κ of ordinals of cardinality smaller than κ, is smaller than κ. (Note that this argument would not work if sets Φ of arbitrary cardinality were to be considered.) Consequently, $Mod_\gamma(\Phi)|_{\sigma_{ip}} = Mod_\alpha(\Phi)|_{\sigma_{ip}}$ is κ-definable in $\mathbf{I}_{\alpha+1}$, and so in $\mathbf{I}_\gamma$ as well.

Now, the thesis follows for $\mathbf{I}^+ = \mathbf{I}_\gamma$: for $\Phi \subseteq \mathbf{Sen}(\Sigma_p)$ and $\Psi \subseteq \mathbf{Sen}(\Sigma_c)$ of cardinalities smaller than κ, if $\sigma_{pu}(\Phi) \models \sigma_{cu}(\Psi)$ then $(Mod^+(\Phi)|_{\sigma_{ip}})|^{-1}_{\sigma_{ic}} \subseteq Mod^+(\Psi)$ by Lemma 6.6. Since $Mod^+(\Phi)|_{\sigma_{ip}}$ is κ-definable in $\mathbf{I}^+$, there is $\Theta \subseteq \mathbf{Sen}^+(\Sigma_i)$ of cardinality smaller than κ such that $Mod^+(\Theta) = Mod^+(\Phi)|_{\sigma_{ip}}$. It follows that Θ is a κ-bounded interpolant for Φ and Ψ. ⊣

The key result of Section 7, Theorem 7.1, applies for the bounded interpolation as it stands. The only adjustment needed is a small refinement in the proof of Lemma 3.2 (which may require an appropriate axiom of choice); the proof of Proposition 3.4 requires a similar adjustment:

Lemma 8.4. *Consider the diagram* $(*)$ *of signature morphisms.*

1. *If* $\mathbf{Sen}(\sigma_{ip}) \colon \mathbf{Sen}(\Sigma_i) \to \mathbf{Sen}(\Sigma_p)$ *is surjective and* $\sigma_{cu} \colon \Sigma_c \to \Sigma_u$ *is conservative then* $(*)$ *admits* κ*-bounded interpolation.*
2. *If* $\mathbf{Sen}(\sigma_{ic}) \colon \mathbf{Sen}(\Sigma_i) \to \mathbf{Sen}(\Sigma_c)$ *is surjective and* $\sigma_{pu} \colon \Sigma_p \to \Sigma_u$ *is conservative then* $(*)$ *admits* κ*-bounded interpolation.*

Proof. Adjusting the proof of Lemma 3.2: consider $\Phi \subseteq \mathbf{Sen}(\Sigma_c)$ and $\Psi \in \mathbf{Sen}(\Sigma_c)$ of cardinalities smaller than κ and such that $\sigma_{pu}(\Phi) \models \sigma_{cu}(\Psi)$.

1. Suppose $\mathbf{Sen}(\sigma_{ip}) \colon \mathbf{Sen}(\Sigma_i) \to \mathbf{Sen}(\Sigma_p)$ is surjective and $\sigma_{cu} \colon \Sigma_c \to \Sigma_u$ is conservative. Choose $\Theta \subseteq \sigma_{ip}^{-1}(\Phi) \subseteq \mathbf{Sen}(\Sigma_i)$ of the same cardinality as Φ such that $\sigma_{ip}(\Theta) = \Phi$. Trivially, $\Phi \models \sigma_{ip}(\Theta)$. Then $\sigma_{pu}(\Phi) = \sigma_{pu}(\sigma_{ip}(\Theta)) = \sigma_{cu}(\sigma_{ic}(\Theta))$, and so $\sigma_{cu}(\sigma_{ic}(\Theta)) \models \sigma_{cu}(\Psi)$. Hence $\sigma_{ic}(\Theta) \models \Psi$ by conservativity of σ_{cu}. Thus Θ is a κ-bounded interpolant for Φ and Ψ.
2. Suppose $\mathbf{Sen}(\sigma_{ic}) \colon \mathbf{Sen}(\Sigma_i) \to \mathbf{Sen}(\Sigma_c)$ is surjective and $\sigma_{pu} \colon \Sigma_p \to \Sigma_u$ is conservative. Choose $\Theta \subseteq \sigma_{ic}^{-1}(\Psi) \subseteq \mathbf{Sen}(\Sigma_i)$ of the same cardinality as Ψ such

that $\sigma_{ic}(\Theta) = \Psi$. Trivially, $\sigma_{ic}(\Theta) \models \Psi$. Moreover, $\sigma_{pu}(\sigma_{ip}(\Theta)) = \sigma_{cu}(\sigma_{ic}(\Theta)) = \sigma_{cu}(\Psi)$, and so $\sigma_{pu}(\Phi) \models \sigma_{pu}(\sigma_{ip}(\Theta))$, which implies $\Phi \models \sigma_{ip}(\Theta)$ by conservativity of σ_{pu}. Thus Θ is a κ-bounded interpolant for Φ and Ψ. ⊣

Now, Corollary 3.3 and Theorem 7.1 hold for κ-bounded interpolation: their proofs carry over relying on Lemma 8.4 in place of Lemma 3.2.

§9. Craig–Robinson interpolation. In many applications, in particular in the theory of structured specifications and modular software development, the Craig interpolation property turns out a bit too weak if the underlying institution does not enjoy some sufficient closure properties. What is needed is a stronger form of interpolation, where the entailments between the premise and the conclusion, and between the interpolant and the conclusion are to hold only in the context of an additional theory or specification, which may be viewed as an additional "parameter" for the interpolation property. This leads to the following definition, working again in an institution $\mathbf{I} = \langle \mathbf{Sig}, \mathbf{Sen}, \mathbf{Mod}, \langle \models_\Sigma \rangle_{\Sigma \in |\mathbf{Sig}|} \rangle$ over a commutative square $(*)$ of signature morphisms [15, 20, 40]:

For any sets of sentences $\Phi \subseteq \mathbf{Sen}(\Sigma_p)$ and $\Gamma, \Psi \subseteq \mathbf{Sen}(\Sigma_c)$ such that $\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma) \models_{\Sigma_u} \sigma_{cu}(\Psi)$, an *interpolant for* Φ *and* Ψ w.r.t. Γ (over diagram $(*)$) is a set $\Theta \subseteq \mathbf{Sen}(\Sigma_i)$ of Σ_i-sentences such that $\Phi \models_{\Sigma_p} \sigma_{ip}(\Theta)$ and $\sigma_{ic}(\Theta) \cup \Gamma \models_{\Sigma_c} \Psi$.

$$\begin{tikzcd}
 & & \boxed{\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma) \models \sigma_{cu}(\Psi)} & & \\
 & & \Sigma_u & & \\
\boxed{\Phi \models \sigma_{ip}(\Theta)} & \Sigma_p \arrow[ur, "\sigma_{pu}"] & & \Sigma_c \arrow[ul, "\sigma_{cu}"'] & \boxed{\sigma_{ic}(\Theta) \cup \Gamma \models \Psi} \\
 & & \Sigma_i \arrow[ul, "\sigma_{ip}"] \arrow[ur, "\sigma_{ic}"'] & & \\
 & & \boxed{\Theta} & &
\end{tikzcd}$$

The diagram $(*)$ *admits Craig–Robinson* (or *parameterised*) *interpolation* if for all $\Phi \subseteq \mathbf{Sen}(\Sigma_p)$ and $\Gamma, \Psi \subseteq \mathbf{Sen}(\Sigma_c)$ such that $\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma) \models \sigma_{cu}(\Psi)$, there is an interpolant $\Theta \subseteq \mathbf{Sen}(\Sigma_i)$ for Φ and Ψ w.r.t. Γ.

Clearly, the Craig interpolation property is a special case of the Craig–Robinson interpolation property where only the empty "parameter" set $\Gamma = \emptyset$ is considered—however, in general the latter property is stronger. This is true in spite of the fact that in institutions satisfying certain "closure" properties, the latter is implied by the former. For instance, if $\mathbf{I}$ *has infinitary implication*[14] then any diagram $(*)$ admits Craig interpolation in $\mathbf{I}$ if and only if it admits Craig–Robinson interpolation. The same holds if the institution is compact and has the usual binary implication, etc.

[14] That is: for any signature $\Sigma \in |\mathbf{Sig}|$, set of sentences $\Gamma \subseteq \mathbf{Sen}(\Sigma)$ and sentence $\psi \in \mathbf{Sen}(\Sigma)$, there is a sentence $\lceil \Gamma \Rightarrow \psi \rceil \in \mathbf{Sen}(\Sigma)$ such that for all models $M \in \mathbf{Mod}(\Sigma)$, $M \models \lceil \Gamma \Rightarrow \psi \rceil$ iff $M \not\models \Gamma$ or $M \models \psi$.

Consequently, in the institutions **FO** of first-order logic and **PL** of propositional logic when a commutative square of signature morphisms admits Craig interpolation then it admits Craig–Robinson interpolation as well. This is not the case for equational logic though:

Example 9.1. In the institution **EQ** of equational logic, consider the diagram $(*)$ where all signatures have a single sort s, and $\Sigma_i = \Sigma_c$ have constants $a, b, c, d \colon s$, while $\Sigma_p = \Sigma_u$ extends them by a unary operation $f \colon s \to s$ (and the signature morphisms are inclusions). Put $\Phi = \{f(a) = b, f(c) = d\}$, $\Gamma = \{a = c\}$, $\Psi = \{b = d\}$. Clearly, $\Phi \cup \Gamma \models \Psi$. However, there are no non-trivial consequences of Φ over the signature Σ_i (conditional equations are not in **EQ**), and so there is no interpolant for Φ and Ψ w.r.t. Γ. This shows that in the institution of equational logic even union-intersection squares of signature inclusions need not admit Craig–Robinson interpolation.

Although the results presented in the previous sections do not apply directly to the Craig–Robinson interpolation, the techniques introduced may be used to show pretty much similar facts. For instance:

Theorem 9.2. *Let* $\Phi \subseteq \mathbf{Sen}(\Sigma_p)$ *and* $\Gamma, \Psi \subseteq \mathbf{Sen}(\Sigma_c)$ *be sets of sentences such that* $\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma) \models \sigma_{cu}(\Psi)$. *An interpolant* $\Theta \subseteq \mathbf{Sen}(\Sigma_i)$ *for* Φ *and* Ψ *w.r.t.* Γ *is stable under extensions of* $\mathbf{I}$ *by models if and only if the following conditions hold*:

1. $\sigma_{ip}(\Theta) \subseteq [(\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma)) \overset{\Sigma_u}{\underset{\Sigma_p}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\Phi)$, *and*
2. $\Psi \subseteq [(\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma)) \overset{\Sigma_u}{\underset{\Sigma_c}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta) \cup \Gamma)$.

Proof. Following the pattern of the proof of Theorem 4.6 (and lemmas and corollaries it relies on):

For the "if" part, suppose that $\mathbf{I}^+$ is an extension of $\mathbf{I}$ by models such that Θ is not an interpolant for Φ and Ψ w.r.t. Γ in $\mathbf{I}^+$, that is, we have $\sigma_{pu}(\Phi) \models^+ \sigma_{cu}(\Psi)$, but $\Phi \not\models^+ \sigma_{ip}(\Theta)$ or $\sigma_{ic}(\Theta) \cup \Gamma \not\models^+ \Psi$.

1. If $\Phi \not\models^+ \sigma_{ip}(\Theta)$ then for some model $M \in \mathbf{Mod}^+(\Sigma_p)$, $M \models^+ \Phi$ and $M \not\models^+ \sigma_{ip}(\Theta)$. Then $\Phi \subseteq Th^+(M)$ and $\sigma_{ip}(\Theta) \not\subseteq Th^+(M)$. Moreover, $Th^+(M)$ never separates $\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma)$ from $\sigma_{cu}(\Psi)$. It follows that $[(\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma)) \overset{\Sigma_u}{\underset{\Sigma_p}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\Phi) \subseteq Th^+(M)$, which implies $\sigma_{ip}(\Theta) \not\subseteq [(\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma)) \overset{\Sigma_u}{\underset{\Sigma_p}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\Phi)$.
2. If $\sigma_{ic}(\Theta) \cup \Gamma \not\models^+ \Psi$ then for some model $N \in \mathbf{Mod}^+(\Sigma_c)$, $N \models^+ \sigma_{ic}(\Theta) \cup \Gamma$ and $N \not\models^+ \Psi$. Then $\sigma_{ic}(\Theta) \cup \Gamma \subseteq Th^+(N)$ and $\Psi \not\subseteq Th^+(N)$, and $Th^+(N)$ never separates $\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma)$ from $\sigma_{cu}(\Psi)$. It follows now that $[(\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma)) \overset{\Sigma_u}{\underset{\Sigma_c}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta)) \subseteq Th^+(N)$, which implies $\Psi \not\subseteq [(\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma)) \overset{\Sigma_u}{\underset{\Sigma_c}{\rightsquigarrow}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta))$.

For the "only if" part, let $\mathbf{I}^+$ be an extension of $\mathbf{I}$ by a new Σ_p-model M and a new Σ_c-model N (and their formal reducts) with:

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

- $Th^+(M) = [(\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma)) \overset{\Sigma_u}{\underset{\Sigma_p}{\leadsto}} \sigma_{cu}(\Psi)](\Phi)$,
- $Th^+(N) = [(\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma)) \overset{\Sigma_u}{\underset{\Sigma_c}{\leadsto}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta) \cup \Gamma)$.

In $\mathbf{I}^+$, we still have $\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma) \models^+ \sigma_{cu}(\Psi)$. However, if condition 1 fails then $M \not\models^+ \sigma_{ip}(\Theta)$, and so $\Phi \not\models^+ \sigma_{ip}(\Theta)$, and if condition 2 fails then $N \not\models^+ \Psi$, and so $\sigma_{ic}(\Theta) \cup \Gamma \not\models^+ \Psi$. In either case, Θ is not an interpolant for Φ and Ψ w.r.t. Γ in $\mathbf{I}^+$. ⊣

Theorem 9.3. *Consider* $\Phi \subseteq \mathbf{Sen}(\Sigma_p)$ *and* $\Gamma, \Psi \subseteq \mathbf{Sen}(\Sigma_c)$ *that satisfy* $\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma) \models \sigma_{cu}(\Psi)$. *Put* $\Theta^* = \sigma_{ip}^{-1}([(\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma)) \overset{\Sigma_u}{\underset{\Sigma_p}{\leadsto}} \sigma_{cu}(\Psi)](\Phi) \cap Th(\Phi))$.

There is an interpolant for Φ *and* Ψ *w.r.t.* Γ *in every extension of* $\mathbf{I}$ *by models if and only if* $\Psi \subseteq [(\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma)) \overset{\Sigma_u}{\underset{\Sigma_c}{\leadsto}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta^*))$ *and* $\sigma_{ic}(\Theta^*) \cup \Gamma \models \Psi$.

Proof. Following the pattern of the proof of Theorem 5.5:

For the "if" part, just notice that under the assumptions, Θ^* is an interpolant for Φ and Ψ w.r.t. Γ, and by Theorem 9.2 it is stable under extensions of $\mathbf{I}$ by new models.

For the "only if" part, if there is an interpolant for Φ and Ψ w.r.t. Γ in every extension of $\mathbf{I}$ by models then, reasoning similarly as in the proof of Lemma 5.3, there must be an interpolant $\Theta \subseteq \mathbf{Sen}(\Sigma_i)$ for Φ and Ψ w.r.t. Γ in $\mathbf{I}$ that is stable under extensions of $\mathbf{I}$ by models. Therefore, by Theorem 9.2:

- $\Psi \subseteq [(\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma)) \overset{\Sigma_u}{\underset{\Sigma_c}{\leadsto}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta) \cup \Gamma)$, and
- $\sigma_{ip}(\Theta) \subseteq [(\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma)) \overset{\Sigma_u}{\underset{\Sigma_p}{\leadsto}} \sigma_{cu}(\Psi)](\Phi)$.

Together with $\Phi \models \sigma_{ip}(\Theta)$, the latter implies $\Theta \subseteq \Theta^*$. Hence $\sigma_{ic}(\Theta) \subseteq \sigma_{ic}(\Theta^*)$, and so $\Psi \subseteq [(\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma)) \overset{\Sigma_u}{\underset{\Sigma_c}{\leadsto}} \sigma_{cu}(\Psi)](\sigma_{ic}(\Theta^*) \cup \Gamma)$. Since $\sigma_{ic}(\Theta) \cup \Gamma \models \Psi$, we also have $\sigma_{ic}(\Theta^*) \cup \Gamma \models \Psi$—which completes the proof. ⊣

Theorem 9.4. *There is an extension* $\mathbf{I}^+$ *of* $\mathbf{I}$ *by new sentences in which the diagram* $(*)$ *does not admit Craig–Robinson interpolation if and only if there are classes of models* $\mathcal{M} \subseteq \mathbf{Mod}(\Sigma_p)$ *and* $\mathcal{G}, \mathcal{N} \subseteq \mathbf{Mod}(\Sigma_c)$ *such that*:

1. $\mathcal{M}|^{-1}_{\sigma_{pu}} \cap \mathcal{G}|^{-1}_{\sigma_{cu}} \subseteq \mathcal{N}|^{-1}_{\sigma_{cu}}$ *and*
2. *no class of models* $\mathcal{K} \subseteq \mathbf{Mod}(\Sigma_i)$ *such that* $\mathcal{M}|_{\sigma_{ip}} \subseteq \mathcal{K}$ *and* $\mathcal{K}|^{-1}_{\sigma_{ic}} \cap \mathcal{G} \subseteq \mathcal{N}$ *is definable in* $\mathbf{I}$ *from* $\{\langle \Sigma_p, \mathcal{M} \rangle, \langle \Sigma_c, \mathcal{G} \rangle, \langle \Sigma_c, \mathcal{N} \rangle\}$.

Proof. Following the pattern of the proof of Theorem 6.3:

For the "if" part, let $\mathbf{I}^+$ be an extension of $\mathbf{I}$ by a new Σ_p-sentence φ and new Σ_c-sentences γ and ψ (and their formal translations) such that $Mod^+(\varphi) = \mathcal{M}$, $Mod^+(\gamma) = \mathcal{G}$ and $Mod^+(\psi) = \mathcal{N}$. Then, by assumption 1, we have $\{\sigma_{pu}(\varphi), \sigma_{cu}(\gamma)\} \models \sigma_{cu}(\psi)$. However, if there was an interpolant $\Theta^+ \subseteq \mathbf{Sen}^+(\Sigma_i)$ for φ and ψ w.r.t. γ then the class $Mod^+(\Theta^+) \subseteq \mathbf{Mod}(\Sigma_i)$ would be definable in $\mathbf{I}$ from $\{\langle \Sigma_p, \mathcal{M} \rangle, \langle \Sigma_c, \mathcal{G} \rangle, \langle \Sigma_c, \mathcal{N} \rangle\}$ and would satisfy $\mathcal{M}|_{\sigma_{ip}} \subseteq Mod^+(\Theta^+)$ and $Mod^+(\Theta^+)|^{-1}_{\sigma_{ic}} \cap \mathcal{G} \subseteq \mathcal{N}$, contradicting assumption 2.

For the "only if" part: consider any extension $\mathbf{I}^+$ of $\mathbf{I}$ by new sentences with $\Phi \subseteq \mathbf{Sen}^+(\Sigma_p)$ and $\Gamma, \Psi \subseteq \mathbf{Sen}^+(\Sigma_c)$ such that $\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma) \models^+ \sigma_{cu}(\Psi)$ but there is no interpolant for Φ and Ψ w.r.t. Γ in $\mathbf{I}^+$. Put $\mathcal{M} = Mod^+(\Phi)$, $\mathcal{G} = Mod^+(\Gamma)$ and $\mathcal{N} = Mod^+(\Psi)$. Clearly, condition 1 holds.

Suppose there is a class of models $\mathcal{K} \subseteq \mathbf{Mod}(\Sigma_i)$ such that $\mathcal{M}|_{\sigma_{ip}} \subseteq \mathcal{K}$ and $\mathcal{K}|^{-1}_{\sigma_{ic}} \cap \mathcal{G} \subseteq \mathcal{N}$ that is definable in $\mathbf{I}$ from $\{\langle \Sigma_p, \mathcal{M}\rangle, \langle \Sigma_c, \mathcal{G}\rangle, \langle \Sigma_c, \mathcal{N}\rangle\}$. This means that there are Σ_i-sentences $\Theta \subseteq \mathbf{Sen}(\Sigma_i)$ and signature morphisms $\tau_l : \Sigma_p \to \Sigma_i$, $l \in \mathcal{L}_p$, $\tau_l : \Sigma_c \to \Sigma_i$, $l \in \mathcal{L}'_c$, and $\tau_l : \Sigma_c \to \Sigma_i$, $l \in \mathcal{L}_c$, such that $\mathcal{K} = \bigcap_{l \in \mathcal{L}_p} \mathcal{M}|^{-1}_{\tau_l} \cap \bigcap_{l \in \mathcal{L}'_c} \mathcal{G}|^{-1}_{\tau_l} \cap \bigcap_{l \in \mathcal{L}_c} \mathcal{N}|^{-1}_{\tau_l} \cap Mod(\Theta)$.

Put $\Theta^+ = \Theta \cup \bigcup_{l \in \mathcal{L}_p} \tau_l(\Phi) \cup \bigcup_{l \in \mathcal{L}'_c} \tau_l(\Gamma) \cup \bigcup_{l \in \mathcal{L}_c} \tau_l(\Psi) \subseteq \mathbf{Sen}^+(\Sigma_i)$. Then $Mod^+(\Theta^+) = \mathcal{K}$, and Θ^+ is an interpolant for Φ and Ψ w.r.t. Γ in $\mathbf{I}^+$—which yields a contradiction, proving condition 2. ⊣

Perhaps surprisingly, Theorem 7.1, Lemma 3.2, and Corollary 3.3 do not quite carry over. These results hint at a nice symmetry between the role of the premise and conclusion signatures the classical Craig interpolation in fact bears, in spite of its apparently asymmetrical formulation (this is also visible in the classical model theory through the equivalence between the Craig interpolation and Robinson consistency theorems, with the explicit symmetry in the formulation of the latter). This is lost for the Craig–Robinson interpolation: Example 9.1 shows that condition 2 in Lemma 3.2 does not entail Craig–Robinson interpolation property, and similarly in Corollary 3.3. However:

LEMMA 9.5. *If $\sigma_{ip} : \Sigma_i \to \Sigma_p$ is such that* $\mathbf{Sen}(\sigma_{ip}) : \mathbf{Sen}(\Sigma_i) \to \mathbf{Sen}(\Sigma_p)$ *is surjective and $\sigma_{cu} : \Sigma_c \to \Sigma_u$ is conservative then $(*)$ admits Craig–Robinson interpolation.*

PROOF. Let $\Phi \subseteq \mathbf{Sen}(\Sigma_c)$ and $\Gamma, \Psi \in \mathbf{Sen}(\Sigma_c)$ be such that $\sigma_{pu}(\Phi) \cup \sigma_{cu}(\Gamma) \models \sigma_{cu}(\Psi)$.

Consider $\Theta = \sigma_{ip}^{-1}(\Phi) \subseteq \mathbf{Sen}(\Sigma_i)$. First, since $\mathbf{Sen}(\sigma_{ip}) : \mathbf{Sen}(\Sigma_i) \to \mathbf{Sen}(\Sigma_p)$ is surjective, $\Phi = \sigma_{ip}(\Theta)$, and so $\Phi \models_{\Sigma_p} \sigma_{ip}(\Theta)$. Then, since $(*)$ commutes, $\sigma_{pu}(\Phi) = \sigma_{pu}(\sigma_{ip}(\Theta)) = \sigma_{cu}(\sigma_{ic}(\Theta))$, and so $\sigma_{cu}(\sigma_{ic}(\Theta)) \cup \sigma_{cu}(\Gamma) \models \sigma_{cu}(\Psi)$. Hence $\sigma_{ic}(\Theta) \cup \Gamma \models \Psi$ by conservativity of σ_{cu}. Thus Θ is an interpolant for Φ and Ψ w.r.t. Γ. ⊣

COROLLARY 9.6. *If $\sigma_{ip} : \Sigma_i \to \Sigma_p$ is a retraction and $\sigma_{cu} : \Sigma_c \to \Sigma_u$ is a coretraction then $(*)$ admits Craig–Robinson interpolation.*

PROOF. Follows by Lemma 9.5, as in the proof of Corollary 3.3. ⊣

Let's have a look at the opposite implication:

LEMMA 9.7. *If the diagram $(*)$ admits Craig–Robinson interpolation in all extensions of* $\mathbf{I}$ *by new sentences and models then $\sigma_{ip} : \Sigma_i \to \Sigma_p$ is a retraction.*

PROOF. Suppose that $\sigma_{ip} : \Sigma_i \to \Sigma_p$ is not a retraction, that is, there is no $\tau_{pi} : \Sigma_p \to \Sigma_i$ such that $\tau_{pi};\sigma_{ip} = id_{\Sigma_p}$.

Let $\mathbf{I}^+$ be the extension of $\mathbf{I}$ by a new Σ_p-model M and a new Σ_c-model N (and their formal reducts) such that M and N do not satisfy any $\mathbf{I}$-sentences. Let then $\mathbf{I}^{++}$ be the extension of $\mathbf{I}^+$ by a new Σ_p-sentence φ and new Σ_c-sentences γ and

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

ψ (and their formal translations) such that $Mod^{++}(\varphi) = \{M\}$, $Mod^{++}(\gamma) = \{N\}$, $Mod^{++}(\psi) = \emptyset$.

Since $Mod^{++}(\{\sigma_{pu}(\varphi), \sigma_{cu}(\gamma)\}) = \emptyset$, we have $\{\sigma_{pu}(\varphi, \sigma_{cu}(\gamma)\} \models^{++} \sigma_{cu}(\psi)$.

Suppose there is an interpolant $\Theta \subseteq \mathbf{Sen}^{++}(\Sigma_i)$ for φ and ψ w.r.t. γ. Then $\varphi \models^{++} \sigma_{ip}(\Theta)$, hence $M \models^{++} \sigma_{ip}(\Theta)$, and so:

- no **I**-sentences are in Θ;
- for $\rho_{pi} : \Sigma_p \to \Sigma_i$, $\lceil \rho_{pi}(\varphi) \rceil \notin \Theta$ since $\rho_{pi};\sigma_{ip} \neq id_{\Sigma_p}$, hence $\lceil M|_{\rho_{pi};\sigma_{ip}} \rceil \notin Mod^{++}(\varphi)$, and so $M \not\models^{++} \sigma_{ip}(\lceil \rho_{pi}(\varphi) \rceil)$;
- for $\rho_{ci} : \Sigma_c \to \Sigma_i$, $\lceil \rho_{ci}(\gamma) \rceil \notin \Theta$ and $\lceil \rho_{ci}(\psi) \rceil \notin \Theta$, since $\lceil M|_{\rho_{ci};\sigma_{ip}} \rceil \notin Mod^{++}(\gamma)$ and $\lceil M|_{\rho_{ci};\sigma_{ip}} \rceil \notin Mod^{++}(\psi)$, hence $M \not\models^{++} \sigma_{ip}(\lceil \rho_{ci}(\gamma) \rceil)$ and $M \not\models^{++} \sigma_{ip}(\lceil \rho_{ci}(\psi) \rceil)$.

Therefore $\Theta = \emptyset$. But $\gamma \not\models^{++} \psi$—which contradicts the assumption that Θ is an interpolant for φ and ψ w.r.t. γ. ⊣

LEMMA 9.8. *If the diagram* $(*)$ *admits Craig–Robinson interpolation in all extensions of* **I** *by new sentences and models then* $\sigma_{cu} : \Sigma_c \to \Sigma_u$ *is a coretraction.*

PROOF. Suppose that $\sigma_{cu} : \Sigma_c \to \Sigma_u$ is not a coretraction, that is, there is no $\tau_{uc} : \Sigma_u \to \Sigma_c$ such that $\sigma_{cu};\tau_{uc} = id_{\Sigma_c}$.

Let $\mathbf{I}^{+}$ be the extension of **I** by a new Σ_p-model M and a new Σ_c-model N (and their formal reducts) such that M and N do not satisfy any **I**-sentences. Let then $\mathbf{I}^{++}$ be the extension of $\mathbf{I}^{+}$ by a new Σ_p-sentence φ and new Σ_c-sentences γ and ψ (and their formal translations) such that:

- $Mod^{++}(\varphi) = \{M\} \cup \{\lceil N|_{\tau_{pi};\sigma_{ic}} \rceil \mid \tau_{pi} : \Sigma_p \to \Sigma_i, \tau_{pi};\sigma_{ip} = id_{\Sigma_p}\}$,
- $Mod^{++}(\gamma) = \{N\}$,
- $Mod^{++}(\psi) = \{\lceil N|_{\rho_{cc}} \rceil \mid \rho_{cc} \neq id_{\Sigma_c}\}$.

We have $Mod^{++}(\sigma_{cu}(\gamma)) = \{\lceil N|_{\tau_{uc}} \rceil \mid \sigma_{cu};\tau_{uc} = id_{\Sigma_c}\} = \emptyset$ since $\sigma_{cu} : \Sigma_c \to \Sigma_u$ is not a coretraction. Hence $\{\sigma_{pu}(\varphi, \sigma_{cu}(\gamma)\} \models^{++} \sigma_{cu}(\psi)$.

Suppose there is an interpolant $\Theta \subseteq \mathbf{Sen}^{++}(\Sigma_i)$ for φ and ψ w.r.t. γ. Then $\varphi \models^{++} \sigma_{ip}(\Theta)$, hence $M \models^{++}_{\Sigma_p} \sigma_{ip}(\Theta)$, and so:

- no **I**-sentences are in Θ;
- for $\rho_{pi} : \Sigma_p \to \Sigma_i$, if $\lceil \rho_{pi}(\varphi) \rceil \in \Theta$ then $M \models^{++} \sigma_{ip}(\lceil \rho_{pi}(\varphi) \rceil)$, hence $\rho_{pi};\sigma_{ip} = id_{\Sigma_p}$, which implies $\lceil N|_{\rho_{pi};\sigma_{ic}} \rceil \models^{++} \varphi$ and thus $N \models^{++} \sigma_{ic}(\lceil \rho_{pi}(\varphi) \rceil)$;
- for $\rho_{ci} : \Sigma_c \to \Sigma_i$, $\lceil \rho_{ci}(\gamma) \rceil \notin \Theta$ and $\lceil \rho_{ci}(\psi) \rceil \notin \Theta$, since $\lceil M|_{\rho_{ci};\sigma_{ip}} \rceil \notin Mod^{++}(\gamma)$ and $\lceil M|_{\rho_{ci};\sigma_{ip}} \rceil \notin Mod^{++}(\psi)$, hence $M \not\models^{++} \sigma_{ip}(\lceil \rho_{ci}(\gamma) \rceil)$ and $M \not\models^{++} \sigma_{ip}(\lceil \rho_{ci}(\psi) \rceil)$.

Therefore $N \models^{++} \sigma_{ic}(\Theta)$, but since we also have $N \models^{++} \gamma$ and $N \not\models^{++} \psi$, $\sigma_{ic}(\Theta) \cup \{\gamma\} \not\models^{++} \psi$—which contradicts the assumption that Θ is an interpolant for φ and ψ w.r.t. γ. ⊣

Summing up:

THEOREM 9.9. *The diagram* $(*)$ *admits Craig–Robinson interpolation in all extensions of* **I** *by new sentences and models if and only if* $\sigma_{ip} : \Sigma_i \to \Sigma_p$ *is a retraction and* $\sigma_{cu} : \Sigma_c \to \Sigma_u$ *is a coretraction.*

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

Proof. The “if” part is Corollary 9.6, and the “only if” part follows by Lemmas 9.7 and 9.8. ⊣

§10. Final remarks. In this paper we deal with a general interpolation property, recalling its formulation for an arbitrary logical system formalised as an institution. We study behaviour of interpolation properties over an arbitrary commutative square of signature morphisms under extensions of the institution by new models and sentences. We give an exact characterisation of the situations when a particular interpolant for a premise and a conclusion remains stable under institution extensions by new models (Theorem 4.6), or looking at this from the other side, when a particular interpolant for a premise and a conclusion is spoiled in some extension of the institution by new models. Another result (Theorem 5.5) gives sufficient and necessary conditions under which no interpolant for a given premise and conclusion may survive all extensions of the institution by new models, or turning to the positive view, when no extension by new models may spoil the interpolation property for a given premise and conclusion. Then we turn to institution extensions by new sentences, and give an exact characterisation of commutative squares of signature morphisms where adding new sentences may lead to the lack of interpolation (Theorem 6.3). Incidentally, we clarify here the role of the weak amalgamation property as a necessary condition without which interpolation fails if adding new sentences is permitted (Corollary 6.8). Finally, we give exact characterisation of commutative squares of signature morphisms where interpolation is ensured for any extension of the institution by new models and sentences (Theorem 7.1).

Then in Section 8 we argue that analogous characterisations hold for the stability under institution extensions by new models, by new sentences, and by new models and sentences, respectively, of bounded interpolation, where the size of the sets of sentences considered is bounded by some appropriate cardinal. We also show here that the weak amalgamation property makes it possible to extend the institution by new sentences so that the bounded interpolation property is ensured (Theorem 8.3). In particular, the results here cover finitary interpolation, where the interpolant sets of sentences are required to be finite (for finite sets of premises and conclusions).

Finally, in Section 9 we turn to the practically important Craig–Robinson (or parameterised) interpolation, where the conclusion is required to follow only when an additional “parameter” set of sentences over the signature of the conclusion is added to the premise and, respectively, to the interpolant. While the results concerning institution extensions by new models and institution extensions by new sentences carry over rather straightforwardly to this case, the final result concerning the stability of interpolation under institution extensions by new models and sentences differs and seems even stronger than the corresponding characterisation result for the standard Craig interpolation.

To avoid repetition, we refrain from studying in any detail a bounded version of Craig–Robinson interpolation—similar remarks and results as spelled out in Section 8 for bounded (Craig) interpolation would carry over.

In many applications, the class of signature morphisms and of their commutative squares for which the interpolation property is required does not cover all the possible morphisms. Typically, signature pushouts are of the utmost importance,

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

with further restrictions on the classes of morphisms used. In fact, this is necessary in many contexts, as many institutions involved (including the many-sorted first-order logic **FO** and equational logic **EQ**) simply do not admit interpolation for arbitrary signature pushouts. It would be interesting to check how such extra requirements on the signature morphisms involved interact with our characterisation theorems.

Acknowledgement. This is an essentially revised and extended version of [46]. Thanks to the anonymous CALCO'23 and JSL reviewers for a number of useful comments.

REFERENCES

[1] E. Astesiano, M. Bidoit, H. Kirchner, B. Krieg-Brückner, P. D. Mosses, D. Sannella, and A. Tarlecki, *CASL: The common algebraic specification language*. ***Theoretical Computer Science***, vol. 286 (2002), no. 2, pp. 153–196.

[2] J. Barwise, *Axioms for abstract model theory*. ***Annals of Mathematical Logic***, vol. 7 (1974), pp. 221–265.

[3] J. A. Bergstra, J. Heering, and P. Klint, *Module algebra*. ***Journal of the Association for Computing Machinery***, vol. 37 (1990), no. 2, pp. 335–372.

[4] E. W. Beth, *On Padoa's method in the theory of definition*. ***Indagationes Mathematicae (Proceedings)***, vol. 56 (1953), pp. 330–339.

[5] M. Bidoit and R. Hennicker, *Constructor-based observational logic*. ***Journal of Logic and Algebraic Programming***, vol. 67 (2006), nos. 1–2, pp. 3–51.

[6] T. Borzyszkowski, *Logical systems for structured specifications*. ***Theoretical Computer Science***, vol. 286 (2002), no. 2, pp. 197–245.

[7] ———, *Generalized interpolation in first-order logic*. ***Fundamenta Informaticae***, vol. 66 (2005), no. 3, pp. 199–219.

[8] C. Caleiro, P. Gouveia, and J. Ramos, *Completeness results for fibred parchments: Beyond the propositional base*, ***Recent Trends in Algebraic Development Techniques. Selected Papers from the 16th International Workshop on Algebraic Development Techniques*** (M. Wirsing, D. Pattinson, and R. Hennicker, editors), Lecture Notes in Computer Science, 2755, Springer, Cham, 2003, pp. 185–200.

[9] C. Caleiro, P. Mateus, J. Ramos, and A. Sernadas, *Combining logics: Parchments revisited*, ***Recent Trends in Algebraic Development Techniques. Selected Papers from the 15th Workshop on Algebraic Development Techniques Joint with the CoFI WG Meeting*** (M. Cerioli and G. Reggio, editors), Lecture Notes in Computer Science, 2267, Springer, Cham, 2001, pp. 48–70.

[10] C. Caleiro, A. Sernadas, and C. Sernadas, *Fibring logics: Past, present and future*, ***We Will Show Them! Essays in Honour of Dov Gabbay, Volume One*** (S. N. Artëmov, H. Barringer, A. S. d'Avila Garcez, L. C. Lamb, and J. Woods, editors), College Publications, 2005, pp. 363–388.

[11] M. V. Cengarle, ***Formal specifications with higher-order parameterization***, Ph.D. thesis, Ludwig-Maximilians-Universität München, Institut für Informatik, 1994.

[12] C.-C. Chang and H. Jerome Keisler, ***Model Theory***, third ed., North-Holland, Amsterdam, 1990.

[13] W. Craig, *Linear reasoning. A new form of the Herbrand–Gentzen theorem*, this Journal, vol. 22 (1957), no. 3, pp. 250–268.

[14] R. Diaconescu, *An institution-independent proof of Craig interpolation theorem*. ***Studia Logica***, vol. 77 (2004), no. 1, pp. 59–79.

[15] ———, ***Institution-Independent Model Theory***, ***Birkhäuser***, Basel, 2008.

[16] ———, *Borrowing interpolation*. ***Journal of Logic and Computation***, vol. 22 (2011), no. 3, pp. 561–586.

[17] ———, *Interpolation for predefined types*. ***Mathematical Structures in Computer Science***, vol. 22 (2012), no. 1, pp. 1–24.

[18] ———, *Three decades of institution theory*, ***Universal Logic: An Anthology*** (J.-Y. Béziau, editor), Birkhäuser, Basel, 2012, pp. 309–322.

[19] ———, *Generalised graded interpolation*. ***International Journal of Approximate Reasoning***, vol. 152 (2023), pp. 236–261 (English).

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

[20] T. Dimitrakos and T. S. E. Maibaum, *On a generalised modularization theorem*. ***Information Processing Letters***, vol. 74 (2000), nos. 1–2, pp. 65–71.

[21] H. Ehrig, H.-J. Kreowski, J. W. Thatcher, E. G. Wagner, and J. B. Wright, *Parameter passing in algebraic specification languages*. ***Theoretical Computer Science***, vol. 28 (1984), no. 1–2, pp. 45–81.

[22] D. M. Gabbay and L. Maksimova, ***Interpolation and Definability: Modal and Intuitionistic Logics***, Oxford University Press, Oxford, 2005.

[23] D. Găină, *Interpolation in logics with constructors*. ***Theoretical Computer Science***, vol. 474 (2013), pp. 46–59.

[24] ———, *Downward Löwenheim–Skolem theorem and interpolation in logics with constructors*. ***Journal of Logic and Computation***, vol. 27 (2015), no. 6, pp. 1717–1752.

[25] D. Găină and A. Popescu, *An institution-independent proof of the Robinson consistency theorem*. ***Studia Logica***, vol. 85 (2007), pp. 41–73.

[26] J. A. Goguen and R. M. Burstall, *Institutions: Abstract model theory for specification and programming*. ***Journal of the ACM***, vol. 39 (1992), no. 1, pp. 95–146.

[27] J. A. Goguen and G. Roşu, *Institution morphisms*. ***Formal Aspects of Computing***, vol. 13 (2002), nos. 3–5, pp. 274–307.

[28] T. S. E. Maibaum, M. R. Sadler, and P. A. S. Veloso, *Logical specification and implementation*, ***Foundations of Software Technology and Theoretical Computer Science*** (M. Joseph and R. Shyamasundar, editors), Springer, Berlin, 1984, pp. 13–30.

[29] J. Meseguer, *General logics*, ***Logic Colloquium '87*** (H.-D. Ebbinghaus, editor), North-Holland, Amsterdam, 1989, pp. 275–329.

[30] R. Milner, M. Tofte, R. Harper, and D. MacQueen, ***The Definition of Standard ML (Revised)***, MIT Press, Cambridge, 1997.

[31] T. Mossakowski, W. Pawłowski, D. Sannella, and A. Tarlecki, ***Parchments for CafeOBJ logics***, ***Specification, Algebra, and Software - Essays Dedicated to Kokichi Futatsugi*** (S. Iida, J. Meseguer, and K. Ogata, editors), Lecture Notes in Computer Science, 8373, Springer, Berlin, 2014, pp. 66–91.

[32] T. Mossakowski, A. Tarlecki, and W. Pawłowski, *Combining and representing logical systems using model-theoretic parchments*, ***Recent Trends in Data Type Specification. Selected Papers from the 12th International Workshop on Specification of Abstract Data Types*** (F. Parisi-Presicce, editor), Lecture Notes in Computer Science, 1376, Springer, Berlin, 1998, pp. 349–364.

[33] nLab, *Regular cardinal*, 2022. Available at https://ncatlab.org/nlab/show/regular+cardinal (accessed 15 August 2023].

[34] A. Popescu, T. F. Şerbănuţă, and G. Roşu, *A semantic approach to interpolation*. ***Theoretical Computer Science***, vol. 410 (2009), nos. 12–13, pp. 1109–1128.

[35] G. R. Renardel de Lavalette, *Interpolation in computing science: The semantics of modularization*. ***Synthese***, vol. 164 (2008), no. 3, pp. 437–450.

[36] A. Robinson, *A result on consistency and its application to the theory of definition*. ***Indagationes Mathematicae (Proceedings)***, vol. 59 (1956), pp. 47–58.

[37] P. H. Rodenburg, *A simple algebraic proof of the equational interpolation theorem*. ***Algebra Universalis***, vol. 28 (1991), pp. 48–51.

[38] G. Roşu and J. A. Goguen, *On equational Craig interpolation*. ***Journal of Universal Computer Science***, vol. 6 (2000), no. 1, pp. 194–200.

[39] D. Sannella and A. Tarlecki, *Specifications in an arbitrary institution*. ***Information and Computation***, vol. 76 (1988), nos. 2–3, pp. 165–210.

[40] ———, ***Foundations of Algebraic Specification and Formal Software Development***, Monographs in Theoretical Computer Science, An EATCS Series, Springer, Berlin, 2012.

[41] ———, *Property-oriented semantics of structured specifications*. ***Mathematical Structures in Computer Science***, vol. 24 (2014), no. 2, p. e240205.

[42] A. Tarlecki, *Bits and pieces of the theory of institutions*, ***Proceedings of the Tutorial and Workshop on Category Theory and Computer Programming*** (D. H. Pitt, S. Abramsky, A. Poigné, and D. E. Rydeheard, editors), Lecture Notes in Computer Science, 240, Springer, Berlin, 1986, pp. 334–360.

[43] ———, *Moving between logical systems*, ***Recent Trends in Data Type Specification. Selected Papers from the 11th Workshop on Specification of Abstract Data Types*** (M. Haveraaen, O. Owe, and O.-J. Dahl, editors), Lecture Notes in Computer Science, 1130, Springer, Berlin, 1996, pp. 478–502.

[44] ———, ***Towards heterogeneous specifications***, ***Frontiers of Combining Systems 2*** (D. Gabbay and M. de Rijke, editors), Studies in Logic and Computation, 7, Research Studies Press, Taunton, 2000, pp. 337–360.

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press

[45] ———, ***Some nuances of many-sorted universal algebra: A review***. ***Bulletin of the European Association for Theoretical Computer Science***, vol. 104 (2011), pp. 89–111.

[46] ———, *Interpolation is (not always) easy to spoil*, ***10th Conference on Algebra and Coalgebra in Computer Science (CALCO 2023)*** (Dagstuhl, Germany) (P. Baldan and V. de Paiva, editors), Leibniz International Proceedings in Informatics (LIPIcs), 270, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Wadern, 2023, pp. 8:1–8:19.

[47] J. W. Thatcher, E. G. Wagner, and J. B. Wright, *Data type specification: Parameterization and the power of specification techniques*. ***ACM Transactions on Programming Languages and Systems***, vol. 4 (1982), no. 4, pp. 711–732.

[48] J. Väänänen, *The Craig interpolation theorem in abstract model theory*. ***Synthese***, vol. 164 (2008), pp. 401–420.

[49] P. A. S. Veloso, *On pushout consistency, modularity and interpolation for logical specifications*. ***Information Processing Letters***, vol. 60 (1996), no. 2, pp. 59–66.

[50] P. A. S. Veloso and T. S. E. Maibaum, *On the modularization theorem for logical specifications*. ***Information Processing Letters***, vol. 53 (1995), no. 5, pp. 287–293.

INSTITUTE OF INFORMATICS
UNIVERSITY OF WARSAW
UL. BANACHA 2, 02-097 WARSAW, POLAND
E-mail: tarlecki@mimuw.edu.pl

https://doi.org/10.1017/jsl.2024.19 Published online by Cambridge University Press