Definability and Interpolation

Model-theoretic investigations

ILLC Dissertation Series DS-2001-05

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Plantage Muidergracht 24
1018 TV Amsterdam
phone: +31-20-525 6051
fax: +31-20-525 5206
e-mail: illc@wins.uva.nl
homepage: http://www.illc.uva.nl/

Definability and Interpolation

Model-theoretic investigations

Academisch Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof.dr. J.J.M. Franse
ten overstaan van een door het college voor
promoties ingestelde commissie, in het openbaar
te verdedigen in de Aula der Universiteit
op woensdag 20 juni 2001, te 12.00 uur

door

Eva Hoogland

geboren te Amsterdam.

Promotor: Prof.dr. D.H.J. de Jongh
Co-promotor: Dr. Y. Venema
Faculteit der Natuurwetenschappen, Wiskunde en Informatica
Universiteit van Amsterdam
Plantage Muidergracht 24
1018 TV Amsterdam

This research was supported by the Netherlands Organization for Scientific Research (NWO).

Printed and bound by PrintPartners Ipskamp

ISBN: 90–5776–067–3

to Harrie

Contents

Acknowledgments

The past few years have been a period of growth for me, both in the personal and the professional sphere. This is not the appropriate place to discuss my personal life, but I would like to take the opportunity to sketch my mathematical development and thank those people who played an important role in it. I am greatly indebted to all of the people mentioned below.

The single most important experience has been a shift in what mathematics meant to me. Looking back, I think this change was brought about in 1996 during my stay at the Logic Graduate School in Budapest. Before, the world of mathematics had been my private hide-away. A beautiful place where I could retire without anyone to intrude. However, from the examples set by Ágnes Kurucz, Judit Madarász, István Németi, Ildikó Sain and András Simon I learned that mathematics can be a joint activity. I was delighted. In this way, doing research is like going on an adventure with close friends: you are close enough to sometimes finish each other's train of thought while a moment later you may take one another into completely unknown territory. Intimacy of a kind I rarely ever experienced. Upon my return to Amsterdam I decided to become a logician.

During my PhD, logic became more and more a means of communication. This has been foremost thanks to Carlos Areces, Wim Blok, Ramon Jansana, Dick de Jongh, Maarten Marx, Martin Otto and Yde Venema, all of whom I had the pleasure to work with. But also the teaching allowed me to pass on interesting ideas and stir people's curiosity. So did the conferences I attended and the articles I wrote. Over the years, the world of mathematics has become a meeting-place.

Hand in hand with the above development, my self-assurance increased. This has for a large part been due to my supervisors, Dick de Jongh and Yde Venema. From the very beginning they expressed their confidence in me by giving me full reins. This has meant a lot to me.

Finally, I would like to thank Rosalie Iemhoff, Joost Joosten and Maarten Marx. The *third floor* would have been a lot less colorful if it weren't for them. Besides that, they also each offered me a home where I felt truly welcome. A special word of thanks to Carlos Areces who literally always rendered his assistance whenever I called for it.

Amsterdam | Eva Hoogland
April, 2001.

Chapter 1

Introduction

Outline of the chapter

In this introduction we give a brief overview of the dissertation. We also mention the origins of the various chapters.

1.1 What to expect from this dissertation?

In this thesis we study definability and interpolation. These are properties of logics such as compactness or decidability that have been established as yardsticks by which to measure the behavior of logics. What do they look like? In a slogan, the *Beth (definability) property* states that implicit definability equals explicit definability. These notions will be explained in full detail in the thesis. The gist is that implicit definability is a semantic concept whereas explicit definability is a syntactic phenomenon. To say that the two forms of definability coincide (as the Beth property does) may therefore be regarded as an indication that there is a good balance between syntax and semantics of a logic.

Proving that a given logic S has the Beth property usually proceeds by way of proving the *interpolation property* for S. This property requires that any validity $\varphi \to \psi$ has an *interpolant*. That is, there exists a formula ϑ in the common language of φ, ψ, such that $\varphi \to \vartheta$ and $\vartheta \to \psi$ are again validities. Apart from its connection with definability, interpolation is also an interesting notion in itself which points to a well-behaved deductive system.

The objectives of this dissertation are fourfold. We successively

1. Provide "everything you always wanted to know about definability and interpolation but were afraid to ask."

2. Relate definability to the algebraic property of surjectiveness of epimorphisms.

3. Offer tools for proving and disproving definability theorems and interpolation theorems.

4. Present plenty of examples that show that the interpolation property is much stronger than the definability property. To this end, we do two detailed case studies, viz., of guarded fragments of first order logic and of interpretability logics.

A general picture of definability and interpolation The literature on definability and interpolation is of a fragmentary nature. Investigations tend to concentrate on particular logics, and results are scattered throughout the literature. What is missing is an easily accessible introduction to the subject that sketches the general picture. Chapter 2 has been written especially for this purpose.

Algebraic characterization of definability It is well-known that there is a close correspondence between logic and algebra, which can and has been exploited to transfer methods and results between these two fields. A prime and well-known example is formed by the variety of Boolean algebras which forms the algebraic counterpart of classical propositional logic. Actually, these algebras are named after G. Boole who was the first to study propositional logic from this algebraic perspective.

In general, a logic S can often be investigated by means of studying an appropriate class of algebras Mod^*S. How much information does this yield? For one thing, it may help to decide whether S has the interpolation property. For it turns out that for many logics S the question whether or not S has interpolation is answered as soon as you know whether or not Mod^*S has a certain algebraic property, viz. the amalgamation property. Since amalgamation has been well-investigated in the field of universal algebra, this last question may well have been answered. Other information on S that can be obtained by algebraic means concerns the compactness property, the deduction property, etc.

In this dissertation we give such algebraic characterizations of definability properties. Our main result states that, under mild conditions on the logic S, S has the Beth property iff Mod^*S has the property of surjectiveness of epimorphisms. As the latter property is well-known from the algebraic literature, our characterization is indeed useful. We supply plenty of applications to support this view, including applications to many-valued logics and relevance logics. Moreover, we show that the proof of our characterization of the Beth property is generally applicable in proving equivalences, mutatis mutandis, between all kinds of definability properties and surjectiveness of various kinds of epimorphisms. This gives us for example equally general characterizations of the weak Beth property and projective Beth property.

Tools for (dis)proving definability and interpolation How to prove and disprove definability theorems and interpolation theorems? In the literature many approaches can be found. In this dissertation we discuss several of them.[1] Besides the algebraic method described above, these include the following.

1. We extend a method for proving interpolation in modal logics (which resembles a Henkin-style completeness proof) to modal logics with an extra non-standard operator. We do this for the special case of interpretability logics, but the basic ideas are of a general nature.

2. We explore another technique known from modal logic in order to obtain results on interpolation for guarded fragments of first order logic. This technique involves bisimulations. Our findings include both positive and negative results.

The above methods yield results on interpolation. What about definability theorems? Usually, definability theorems are derived from interpolation theorems via a standard argument. But, as we will see, from the interpolation theorem for a

[1] Counting, for the moment, the algebraic method as a semantic approach, we confine ourselves to semantic methods. This is the obvious approach if one wishes to establish a negative result. On the other hand, positive results can be (and have been) obtained in abundance by syntactic means and so our adherence to the semantic side of the matter should be completely attributed to our personal liking.

given logic S we can infer much more than just the Beth definability theorem for S. This is illustrated in the case of interpretability logics where we derive the Beth property for *all* interpretability logics from the interpolation property for the (one) basic system IL. Putting it differently, we do not need the full strength of the interpolation theorem to derive the definability theorem. For example, our proof of the Beth theorem for the guarded fragment (which does not have the interpolation property) uses a limited form of interpolation. These results are not just simple applications of the aforementioned standard argument but require some extra effort.

Conclusion There is a widespread belief which lumps together the interpolation property and the Beth definability property. This common fallacy might be rooted in the fact that most (if not all) of the well-known logics have either both or neither of these two properties. That is, there is a lack of examples that indicate the difference between Beth definability and interpolation. In this thesis we present plenty of such examples. For example, as we already mentioned, all extensions of the basic interpretability logic IL have the Beth property. On the other hand, besides the systems IL and ILP, so far no extensions have been found with interpolation. Another example is presented by the guarded fragment and the packed fragment of first order logic. We will prove the Beth theorem for these two fragments and also for all of their finite variable fragments. However, it will be shown that none of these logics has interpolation, apart from the one- and two variable case. The conclusion that we draw from all this is that the interpolation property is much stronger than the definability property.

1.2 Organization of the dissertation

Below, we give a brief survey of the thesis. A more detailed outline of the individual chapters can be found on their respective title pages.

Chapter 2 The aim of the next chapter is to make the reader familiar with the main themes of this dissertation: definability and interpolation. The chapter is written in a rather informal manner with an emphasis on giving simple examples. We discuss the precise relationship between the relevant properties, summarize the state of the art, and provide ample references to the literature. This chapter also takes a look at the matter from an algebraic perspective.

Chapter 3 This chapter is of an abstract algebraic nature in which algebraic equivalents of several Beth definability properties are given. We also supply many applications of these characterizations. The chapter contains an introduction to the abstract algebraic framework(s) we are working in.

Chapter 4 and chapter 5 These chapters can be seen as case studies in which we extend known methods for proving interpolation and definability. The fourth chapter concerns interpretability logics (these are non-standard modal logics), the final chapter deals with guarded fragments of first order logic.

Appendix and index Here we supply a brief summary of the notions and terminology that we assume the reader to be familiar with. For the reader's convenience, we included an index and a list of symbols.

1.3 About the origins of the various chapters

Most of the results presented in this dissertation have been published previously. We give references below. However, it should be mentioned that all of this material has been rigorously edited for the present occasion.

Parts of chapter 3 are based on my master's thesis [Hoogland, 1996], written under the supervision of I. Németi. The sections 3.2–3.5 contain material of [Blok and Hoogland, 2001]. This paper has been written jointly with W. Blok. Section 3.6 is based on [Hoogland, 2000].

Chapter 4 is an extended version of [Hoogland and Marx, 2000], which was written with M. Marx. Section 4.5 reports on joint work with M. Marx and M. Otto and has appeared as the conference paper [Hoogland et al., 1999].

Chapter 5 has been published as [Areces et al., 2000b], written in collaboration with C. Areces and D. de Jongh.

Chapter 2

A general picture of definability and interpolation

Outline of the chapter

The goal of this chapter is to make the reader familiar with the two main themes of this dissertation: definability and interpolation.

In the first two sections we discuss definability properties, resp. interpolation properties. These include Beth's definability property, the weak Beth property, the projective Beth property, Craig's interpolation property (global and local version), uniform interpolation, Lyndon interpolation and Robinson's consistency property. These properties turn out to be closely connected to certain algebraic properties concerning amalgamation and the surjectiveness of (some specific) epimorphisms. In section 2.5 we sketch this algebraic background. Section 2.6 treats the connection between logic and algebra.

2.1 Preface

This thesis is a contribution to the investigations into definability and interpolation. It therefore seems appropriate to start with providing a general picture of these investigations, if only to put the present work into some perspective. But there are more reasons for this.

First of all, the literature on (Beth) definability and interpolation is extensive but of a fragmentary nature. Investigations tend to concentrate on particular logics, and results are scattered throughout the literature. A general picture is missing.

Moreover, properly speaking there is no such thing as *the* interpolation property or *the* Beth definability property. There is actually a cluster of notions around the concepts of interpolation and definability. This makes a statement like "For a large class of logics, the interpolation property is equivalent to the Beth property" highly ambiguous; under some readings it is true, under most others it is not.

The first aim of the present chapter is to clarify this situation by discussing the precise relation between the relevant properties. This also involves a motivation for the choice of these properties. A second goal is to get some working experience with these properties. In general, proofs of (and even counterexamples to) interpolation theorems and definability theorems are rather complicated. This causes an acquaintance with these properties to be theoretical at best. In this chapter simple examples are presented in abundance which may yield some intuitions.

In agreement with the above aims, we give ample references to the literature. However, it should be realized that this is by no means a complete list of references on this topic. Rather than giving a complete overview, in this chapter we try to give a representative picture of definability and interpolation that makes a good starting point for the remaining part of this thesis.

For a rather considerable part, the work in this thesis is of a (universal) algebraic nature. In accordance with this, also the present chapter contains an algebraic component. However, this need not scare off the reader who prefers a purely logical approach. Let him be reassured that for his purposes the first three sections suffice, while these do not mention any algebra.

Warning to the reader We chose to be rather informal in this chapter. What we want to get across is a general picture, not a bundle of technicalities. This may imply that we unjustly assume familiarity with certain notions and systems. Some of these are defined in the appendix, for others references are given below. But most importantly, we want to emphasize that for a general understanding of the chapter it is not necessary to be acquainted with all the systems we consider. The examples are only meant to illuminate the text: in case they obscure the matter, they may be safely skipped.

Another warning concerns our use of the word 'logic'. In this chapter we use the word 'logic' as an informal notion, even though it appears in theorems, etc.

Again, the reason is that we want to focus on the general ideas rather than on specifying the appropriate framework. However, in chapter 3 we *will* provide a clear framework in which concepts like 'logic' can, and will, be properly defined.

Some references Our modal terminology has been explained in the appendix to this thesis. There we also explicate what is understood by infinitary logics, intermediate logics and finite variable fragments. For a thorough introduction to modal logics, we refer to the forthcoming [Blackburn et al., 2001]. A good reference to infinitary logics is [Keisler, 1971], whereas [Hodkinson, 1993] contains a lively introduction to finite variable fragments. The prerequisites on abstract algebraic logic can be found in chapter 3. The necessary universal algebraic knowledge can be obtained from any textbook on this subject.

2.1.1. Convention In this chapter, we associate with any modal logic a *global consequence relation*. A definition can be found at page 183. However, the results we discuss for global modal logics also have implications for modal logics with a *local consequence relation*. The interested reader is referred to the appendix.

A second convention concerns first order languages. For simplicity, in this chapter these are purely relational languages. That is, *without* constants and function symbols. However, we note that the results of Beth and Craig for first order logic that are the main topic of discussion in this chapter do hold if constants and functions symbols are included in the language. ⊣

2.2 Definability

The problem we shall be concerned with is the following. Given a theory Γ, how to determine whether some primitive notion r is definable in terms of certain other notions? To make this question more concrete, let us for the moment discuss the situation for relational languages. Later we will consider propositional languages.

2.2.1 Predicate logics

Let $\mathcal{L}$ be a set of relation symbols and Γ a set of sentences in the language $\mathcal{L}$. Let $R \in \mathcal{L}$, and set $\mathcal{L}_0 = \mathcal{L} \setminus R$. A formula φ in the language $\mathcal{L}_0$ is called an *explicit definition* for R relative to Γ in the logic S if

$$\Gamma \models_S \forall \boldsymbol{v}(R\boldsymbol{v} \leftrightarrow \varphi(\boldsymbol{v})),$$

where $\boldsymbol{v}$ is a sequence of variables of the appropriate length. We emphasize that the symbol R may occur in Γ, but does not occur in φ.

As Padoa observed in 1900, in order to show that a relation symbol R does *not* have an explicit definition in the language $\mathcal{L}_0$ with respect to a theory Γ, it clearly suffices to find two interpretations I, I' which both satisfy Γ and which are such that $I(P) = I'(P)$, for all relation symbols $P \in \mathcal{L}_0$, but $I(R) \neq I'(R)$. In modern terminology, we say that in this case Γ does not *implicitly define* R.[1] As this notion plays a central role in this dissertation, let us give a precise formulation.

2.2.1. Definition [**Implicit definition**] Let S be a predicate logic, and let $\mathcal{L}_0$, $\mathcal{L}$, R, Γ be as above. Let R' be a relation symbol which is not in $\mathcal{L}$ and which is of the same arity as R. By Γ' we denote the result of renaming R to R' in Γ. We say that the theory Γ *implicitly defines* the relation symbol R in S if any $\mathcal{L}_0$-structure has at most one expansion to a model of Γ. This is often expressed by

$$\Gamma, \Gamma' \models_S \forall \boldsymbol{v}(R\boldsymbol{v} \leftrightarrow R'\boldsymbol{v}), \tag{2.1}$$

where again $\boldsymbol{v}$ is a sequence of variables of the appropriate length.[2] ⊣

2.2.2. Example [Hájek, 1977] contains a nice example of an implicit definition for first order logic on finite models. That is, the syntax of the logic S is the usual syntax of first order logic, but as models of S we only consider the *finite* first order structures.

Let $\mathcal{L} = \{<, R\}$, where R is a unary relation symbol and $<$ is binary. Let γ_1 be the $\mathcal{L}$-sentence expressing that $<$ is a linear order, and let γ_2 express that there exists a first element in this order that moreover has the property R. Finally, $\gamma_3(x, y)$ expresses that y is an immediate successor of x in this order. That is, $\gamma_3(x, y) = x < y \wedge \forall z(x < z \rightarrow (y < z \vee y = z))$. Define

$$\Gamma = \{\gamma_1, \gamma_2, \forall x \forall y(\gamma_3(x, y) \rightarrow (R(x) \leftrightarrow \neg R(y)))\}.$$

We will show that Γ implicitly defines R on finite models. To this end, consider a finite model $\mathcal{M}$ of Γ. As $<$ linearly orders the finite model $\mathcal{M}$, the elements of $\mathcal{M}$ can be written as $\{m_1, \ldots, m_k\}$, for some $k \in \omega$, with $m_1 < \cdots < m_k$. According to Γ, it is the case that $m_i \in I^{\mathcal{M}}(R)$ iff i is odd. Hence, the interpretation of R is fixed. This shows that Γ implicitly defines R on finite models.

Note that on infinite models for Γ, the interpretation of R need not be fixed. Take for example the non-negative real numbers, and consider the usual interpretation of $<$. Then any interpretation of R which contains the element 0 yields a model of Γ. ⊣

So far, we only noted that if a relation symbol R has an explicit definition with respect to a theory Γ, then certainly Γ implicitly defines R. What about the

[1] Note that the notion of an implicit definition is originally a *semantic* notion of definability.

[2] If S is a complete logic, then (2.1) is clearly equivalent to the syntactic condition that $\Gamma, \Gamma' \vdash_S \forall \boldsymbol{v}(R\boldsymbol{v} \leftrightarrow R'\boldsymbol{v})$. In some of the literature, this condition is taken to define the notion of an implicit definition. However, we point out that the concept of an implicit definition is in essence of a semantic nature.

converse? That is, if R does not have an explicit definition with respect to Γ, are we always able to exhibit two interpretations that both satisfy Γ, that agree on $\mathcal{L}_0$ but that interpret R differently? The answer depends on the choice of the underlying logic. Logics for which this is indeed the case are said to have the *Beth definability property.*

2.2.3. Definition [**Beth (definability) property**] A predicate logic S has the *Beth (definability) property* if for any R, Γ as in Definition 2.2.1, if Γ implicitly defines R in S, then there exists an explicit definition for R relative to Γ in S. ⊣

In a slogan, the Beth property can be phrased as 'implicit definability equals explicit definability'.[3] In 1953, E.W. Beth proved the striking theorem that first order logic has this property. According to [Craig, 1957, page 269] this result

> [...] showed that for all first-order systems a certain model-theoretic notion of definability coincides with a certain proof-theoretic notion.

It is worth the effort to compare this result with Gödel's completeness theorem. For, following Tarski, Beth distinguishes within the field of logic a *theory of deducibility* and a *theory of definability.* Both these theories have a syntactic and a semantic component. Implicit definability is a model-theoretic notion of definability, whereas explicit definability is a proof-theoretic notion of definability. Parallel to the question whether a certain conclusion can be derived from a certain set of premises, is the question whether a certain notion can be (syntactically) defined in terms of certain given notions. Beth puts it as follows, cf. [van Ulsen, 2000, page 116].

> A formal system will be *complete* from the standpoint of the theory of definability if, whenever within this system a notion a is not definable in terms of certain notions $t_1, \ldots, t_n$, this can be proved by means of Padoa's method.

That is, if this system has the Beth property.

2.2.4. Example An example of a logic which fails to have the Beth property, is first order logic on finite structures. To see this, we go back to Example 2.2.2. Recall that the theory Γ defined there implicitly defines the relation symbol R on finite models. We will show that R does not have an explicit definition relative to Γ on finite structures.

Reasoning by contradiction, suppose such an explicit definition φ does exist. Let n be the quantifier rank of φ, and let $\mathcal{M}$ be a model for Γ of size 2^{n+2} and $\mathcal{N}$ a

[3] In Beth's original article this terminology does not appear. For Beth, *definable* means explicitly definable. He doesn't adopt a name for implicit definability. Craig distinguishes 'semantic definability' as opposed to 'syntactic definability'. The notions of 'implicit' and 'explicit' definability are introduced in [Robinson, 1956].

model for Γ of size $2^{n+2}+1$. Note that such models exist. In particular, $\mathcal{M}, \mathcal{N}$ are linear orders. We recall from [Gurevich, 1984] that for any $k \in \omega$, any two linear orders of size $\geq 2^k$ satisfy the same first order formulas of rank $\leq k$ in the language $\{<\}$.

Consider the formula $\theta = \exists x(\forall y(x \neq y \rightarrow y < x) \wedge \varphi(x))$. Given our assumptions on φ, this formula would be true in a finite model for Γ if and only if the cardinality of this structure is an odd number. In particular, $\mathcal{N}$ satisfies θ but $\mathcal{M}$ does not. As θ is a formula of rank $\leq n+2$ in the language $\{<\}$, this contradicts Gurevich's theorem. We conclude that R does not have an explicit definition relative to Γ on finite structures. $\dashv$

2.2.5. Example A more difficult example of a predicate logic without the Beth property is L^k, the k-variable fragment of first order logic, for all $k \geq 2$. This result was first proven in [Sain, 1990] by algebraic means. The following example from [Hodkinson, 1993] gives a direct logical proof of the failure of the Beth property in L^k, for $k \geq 3$.

Fix a finite graph $\langle G, E\rangle$, and suppose that $<$ is an irreflexive linear order on G. Say, $G = \{g_1, \ldots, g_n\}$ and suppose G is ordered as $g_1 < g_2 < \cdots < g_n$. Let the formula γ_n express that $<$ is an n-element irreflexive linear order. Note that this can be done without using the identity symbol '=', and using 3 variables only. The formula $\gamma_{i,j}(x, y)$ expresses that x, y are the ith, respectively jth elements along $<$. Let Γ be the union of $\{\gamma_n\}$ together with the sets

- $\{\forall x \forall y(\gamma_{i,j}(x, y) \rightarrow E(x, y)) \ : \ G \models E(g_i, g_j), 1 \leq i, j \leq n\}$ and
- $\{\forall x \forall y(\gamma_{i,j}(x, y) \rightarrow \neg E(x, y)) \ : \ G \not\models E(g_i, g_j), 1 \leq i, j \leq n\}$.

Essentially, Γ describes the diagram of G. We claim that whenever G is a *rigid* graph (i.e., without non-trivial automorphisms), then Γ implicitly defines $<$. To see this, suppose $\langle M, E, <\rangle$ and $\langle M, E, \prec\rangle$ both satisfy Γ. Then $<$ linearly orders M as $m_1 < \cdots < m_n$, say. Moreover, for some permutation $\sigma : n \longrightarrow n$, the order $\prec$ linearly orders M as $m_{\sigma(1)} \prec \cdots \prec m_{\sigma(n)}$.

Note that this certainly restricts the interpretation of $<$, but in general does not determine it completely. For example, take $G = \{g_1, g_2, g_3, g_4\}$, $I^G(E) = \{\langle g_1, g_3\rangle, \langle g_3, g_1\rangle, \langle g_2, g_4\rangle, \langle g_4, g_2\rangle\}$, and suppose $<$, $\prec$ linearly order G as $g_1 < g_2 < g_3 < g_4$ and $g_4 \prec g_1 \prec g_2 \prec g_3$ respectively (see Figure 2.1). Then $\langle G, E, <\rangle$ and $\langle G, E, \prec\rangle$ both satisfy Γ, but $I^G(<) \neq I^G(\prec)$.

Returning to the proof of our claim, we define the map $f : G \longrightarrow G$ by $f(g_i) = g_{\sigma(i)}$, for all $i \leq n$. We leave it to the reader to verify that f is a graph automorphism. Now, suppose G is *rigid*. Then for all $i \leq n$, $i = \sigma(i)$. This implies that $I^{\mathcal{M}}(<) = I^{\mathcal{M}}(\prec)$. In other words, in this case Γ implicitly defines $<$.

Hodkinson goes on to show that if G satisfies the so-called k-extension axiom, then $<$ can not be explicitly defined with respect to Γ in L^k. Not even in the extended

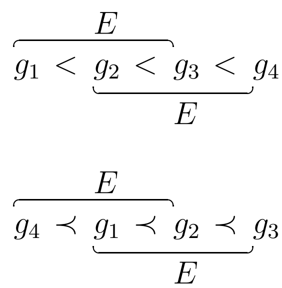


Figure 2.1: Two models for Γ.

first order language L^k_∞ which allows for infinite conjunctions and disjunctions. For, suppose $<$ has an explicit definition $\varphi_<$ in L^k_∞ relative to Γ. That is, $\Gamma \models \forall x \forall y (x < y \leftrightarrow \varphi_<(x, y))$, where $\varphi_<(x, y)$ is a formula in the language $\{E\}$. It is easy to verify that the graph reduct of any model of Γ is isomorphic to G. Therefore, if G satisfies the k-extension axiom, then so does any other model of Γ. We recall from [Hodkinson, 1993, Theorem 8.2] that any graph satisfying the k-extension axiom has quantifier elimination in L^k_∞. This implies that modulo Γ, the formula $\varphi_<$ is equivalent to a quantifier-free formula $\psi(x, y)$ in the language $\{E\}$. The upshot of this is twofold. First,

$$\Gamma \models \forall x \forall y (x < y \leftrightarrow \psi(x, y)). \tag{2.2}$$

Second, the formula $\psi(x, y)$ will be symmetric, because graphs are (i.e., if $\psi(x, y)$ is true in a model $\mathcal{M}$ under a valuation v, then so is $\psi(y, x)$). But $x < y$ is not symmetric. A contradiction with (2.2).

Summarizing, we see that all that is needed to refute the Beth property for L^k_∞ and L^k, for all $k \geq 3$, is to find a finite rigid graph which satisfies the k-extension axiom. In [Andréka et al., 1995] such a graph is constructed. ⊣

The Beth property also fails in a large number of quantificational modal logics. Most notably, it fails for first order $\mathsf{S5}$, cf. [Fine, 1979]. Fine also shows that this failure persists when the constant domain axiom-schema, also called *the Barcan formula*, $\forall x \Box \varphi = \Box \forall x \varphi$ is added to $\mathsf{S5}$. Even more, the Beth property fails in every system between K and $\mathsf{S5}$ if this axiom-schema is added. On the other hand, the quantificational modal logics K, T, $\mathsf{S4}$ have the Beth property, cf. [Gabbay, 1972]. Also intuitionistic predicate logic has the Beth property. This has been proven syntactically in [Schütte, 1962]. [Gabbay, 1971] gives a semantic proof of this fact using Kripke models. This proof also works for the minimal calculus, the positive calculus (i.e., without the connective $\neg$), and several extensions of intuitionistic predicate logic including those obtained by adding one or both of the following axiom schemas: $\forall x \neg\neg\varphi(x) \to \neg\neg\forall x \varphi(x)$ and $\neg\varphi \vee \neg\neg\varphi$. Actually, Gabbay proves the interpolation property for all of the above logics. As we shall see in the next section, this entails the Beth property. Another example of a predicate logic with the Beth property is the guarded fragment of first order logic which will be discussed in chapter 4.

2.2.2 Propositional logics

For a propositional language, the Beth property can be formulated as follows. Below, for a theory Γ and a sequence of propositional variables $\boldsymbol{p}$ we sometimes write $\Gamma(\boldsymbol{p})$ to denote that no formula in Γ contains any variables other than those in $\boldsymbol{p}$. By $\Gamma(\boldsymbol{p}, q)$ we denote the result of substituting the variable q for the variable r in every formula of the theory $\Gamma(\boldsymbol{p}, r)$.

2.2.6. Definition [**Beth property for a propositional logic**] Let S be a propositional logic and suppose the sequence of variables $\boldsymbol{p}$ does not contain the variables r and r'. S is said to have the *Beth (definability) property* if for any set of formulas $\Gamma(\boldsymbol{p}, r)$ the condition

$$\Gamma(\boldsymbol{p}, r), \Gamma(\boldsymbol{p}, r') \models_S r \leftrightarrow r', \tag{2.3}$$

implies that there exists some formula $\varphi_r(\boldsymbol{p})$ such that

$$\Gamma(\boldsymbol{p}, r) \models_S r \leftrightarrow \varphi_r(\boldsymbol{p}).$$

The formula $\varphi_r(\boldsymbol{p})$ is called an *explicit definition* for r in S relative to Γ. Condition (2.3) formally expresses that Γ *implicitly defines* r in terms of $\boldsymbol{p}$. ⊣

Note that the above definition is similar to the earlier definition of the Beth property for a predicate logic, where this time propositional variables play the role of relation symbols.

2.2.7. Example As a first example, let us consider classical propositional calculus (CPC). Consider the theory $\Gamma(p_1, p_2, r) = \{r \to p_1, r \to p_2, p_1 \to (p_2 \to r)\}$. This can be thought of as a logical formulation of the usual lattice-theoretic definition of the meet operation as the infimum of $\{p_1, p_2\}$. As the reader may verify, $\Gamma(p_1, p_2, r), \Gamma(p_1, p_2, r') \models_{\mathsf{CPC}} r \leftrightarrow r'$. That is, Γ implicitly defines r in terms of $\{p_1, p_2\}$ in CPC.

Clearly, r has an explicit definition φ_r relative to Γ in CPC. Simply take $\varphi_r = p_1 \wedge p_2$. However, in the implicational fragment of CPC such explicit definition can not be found, as is well-known. We conclude that this fragment $\mathsf{CPC}^{\to}$ does not have the Beth property. An algebraic proof of this fact is given by Proposition 3.5.6. ⊣

A nice example of a modal logic in which the Beth property fails is the following, due to Maksimova.

2.2.8. Example Von Wright's logic *WL* is the logic of the frame $\langle Z, succ\rangle$, where Z denotes the set of integers, and *succ* is the successor relation. *WL* contains two unary modalities $\Box$ ('always') and $\bigcirc$ ('tomorrow') whose meaning is defined, for any $z \in Z$, by

$$z \models \Box\varphi \text{ iff } z' \models \varphi, \text{ for any } z' \in Z,$$

$$z \models \bigcirc\varphi \text{ iff } succ(z) \models \varphi.$$

Let $\Gamma(p,r) = \{\Diamond p, \Box(p \to r), \Box(r \leftrightarrow \neg \bigcirc r)\}$. The set Γ expresses that p holds somewhere in the model, p only holds at points in time where r is true, and the truth of r changes every day. This implies that under every valuation that satisfies Γ there is a point z at which p is true, and r is true at precisely those points z' such that the distance between z and z' is even. This shows that the interpretation of r is determined by the interpretation of p in every model for Γ. That is, Γ implicitly defines r.

However, r is not explicitly definable with respect to Γ. To see this, we consider the model $\mathcal{M} = \langle Z, succ, v\rangle$, where the valuation v is defined by $v(p) = \{0\}$ and $v(r) = \{2n, -2n \ : \ n \in \omega\}$. Note that $\mathcal{M} \models \Gamma$. Moreover, by induction on the complexity of φ it can be shown that for all formulas $\varphi(p)$ and all $z > 0$,

$$\mathcal{M}, z \models \varphi \text{ iff } \mathcal{M}, succ(z) \models \varphi.$$

The only non-trivial case is for φ of the form $\bigcirc\psi$. Then we have the following equivalences. $\mathcal{M}, z \models \bigcirc\psi$ iff $\mathcal{M}, succ(z) \models \psi$ iff (by induction hypothesis) $\mathcal{M}, succ(succ(z)) \models \psi$ iff $\mathcal{M}, succ(z) \models \bigcirc\psi$.

Suppose φ_r explicitly defines r relative to Γ. As $\mathcal{M}, 2 \models r$, then $\mathcal{M}, 2 \models \varphi_r$. By the above claim then $\mathcal{M}, 1 \models \varphi_r$, whence $\mathcal{M}, 1 \models r$. Contradiction. We conclude that the Beth property fails for Von Wright's logic. $\dashv$

Notwithstanding the above two examples, a large number of propositional logics has the Beth property. The following observation is due to Kreisel.

2.2.9. Theorem [Kreisel, 1960] *Any intermediate logic*[4] *has the Beth property.*

Proof: Suppose $\Gamma(\boldsymbol{p}, r)$ implicitly defines r in the intermediate logic S. That is, for every variable r',

$$\Gamma(\boldsymbol{p}, r), \Gamma(\boldsymbol{p}, r') \models_S r \leftrightarrow r'. \tag{2.4}$$

[4]An *intermediate logic* is any logic between classical propositional calculus and intuitionistic propositional calculus. A proper definition can be found in the appendix.

We have to find an explicit definition for r relative to Γ. To this end, we first substitute the constant $\top$ for r' in (2.4), which yields that $\Gamma(\boldsymbol{p}, r), \Gamma(\boldsymbol{p}, \top) \models_S r$. By compactness we may assume Γ to be finite. From the deduction theorem, we obtain that

$$\Gamma(\boldsymbol{p}, r) \models_S \bigwedge_{\gamma\in\Gamma} \gamma(\boldsymbol{p}, \top) \to r. \tag{2.5}$$

On the other hand, any intermediate logic S' has a strong form of the replacement theorem which can be formulated as follows. Let φ be a formula, ψ one of its subformulas, and let φ' denote the result of replacing ψ in φ by the formula ψ'. Then $\models_{S'} (\psi \leftrightarrow \psi') \to (\varphi \leftrightarrow \varphi')$. In particular, in the logic S we have that $\models_S (r \leftrightarrow \top) \to [\bigwedge_{\gamma\in\Gamma} \gamma(\boldsymbol{p}, r) \leftrightarrow \bigwedge_{\gamma\in\Gamma} \gamma(\boldsymbol{p}, \top)]$. From this it follows that $\models_S \bigwedge_{\gamma\in\Gamma} \gamma(\boldsymbol{p}, r) \to (r \to \bigwedge_{\gamma\in\Gamma} \gamma(\boldsymbol{p}, \top))$. So if we apply the deduction theorem, we see that

$$\Gamma(\boldsymbol{p}, r) \models_S r \to \bigwedge_{\gamma\in\Gamma} \gamma(\boldsymbol{p}, \top). \tag{2.6}$$

(2.5) and (2.6) show that $\bigwedge_{\gamma\in\Gamma} \gamma(\boldsymbol{p}, \top)$ is an explicit definition of r with respect to Γ. ■

In the above proof it is essential that for all formulas φ, ψ it is the case that $\varphi \models_S \psi$ iff $\models_S \varphi \to \psi$. This is the reason why Theorem 2.2.9 can not be extended to modal logics. However, any normal extension of $\mathsf{K4}$ does have *a* deduction theorem, viz., $\varphi \models_{\mathsf{K4}} \psi$ iff $\models_{\mathsf{K4}} (\varphi \wedge \Box\varphi) \to \psi$. Maksimova exploits the ideas in Kreisel's proof to show that this deduction theorem suffices to obtain an analogue of Theorem 2.2.9, a remarkable result.

2.2.10. Theorem [Maksimova, 1993] *Any normal extension of the modal logic* $\mathsf{K4}$ *has the Beth property.*

The proof of this theorem is much more complicated than the proof of Theorem 2.2.9.

Theorem 2.2.10 can not be extended to *any* modal logic. For example, it can not be extended to all logics containing the axiom $\Box\Box p \to \Box\Box\Box p$. This axiom can be seen as a weak form of transitivity which states that, loosely speaking, if a point y can be reached from a point x in three R-transitions, then y can already be reached from x in two R-transitions. [Maksimova, 1993, Theorem 2] gives the following example of a modal logic that contains the axiom $\Box\Box p \to \Box\Box\Box p$ and that does not have the Beth property.

2.2.11. Example Let S be the logic of the frame $\mathcal{F}$ below.[5] Note that $\mathcal{F}$ satisfies the weak form of transitivity described above. Hence, S contains the axiom $\Box\Box p \to \Box\Box\Box p$.

[5]In the appendix (on page 184) is explained what we understand by a *logic of a frame*.

Figure 2.2: The frame $\mathcal{F}$.

Consider the formula $\gamma(r) = r \leftrightarrow \Box\neg r$. The only valuation that satisfies γ in $\mathcal{F}$ is the valuation that maps r to $\{b\}$. This implies that γ implicitly defines r in S.

Suppose φ_r is an explicit definition of r relative to γ. Then φ_r is a variable free formula such that $\gamma(r) \models_S r \leftrightarrow \varphi_r$. Let v be the same valuation as before. By induction on the complexity of φ it can be shown that for all variable free formulas φ we have that $\langle\mathcal{F}, v\rangle, a \models \varphi$ iff $\langle\mathcal{F}, v\rangle, b \models \varphi$. We then argue as follows. Recall that $\langle\mathcal{F}, v\rangle, b \models r$. As φ_r is an explicit definition of r with respect to γ and v satisfies γ, it follows that $\langle\mathcal{F}, v\rangle, b \models \varphi_r$. By the above claim, then $\langle\mathcal{F}, v\rangle, a \models \varphi_r$. Again, this entails that $\langle\mathcal{F}, v\rangle, a \models r$. A contradiction. We conclude that S does not have the Beth property. ⊣

Theorem 2.2.10 implies that many of the well-known modal logics have the Beth property including the systems S4, S5, GL and Grz. Other important modal systems, like K, T and B, are also known to have the Beth property. This follows from the interpolation property of these systems, as will be explained in section 2.3.

Well-known logics which fail to have the Beth property can be found e.g., among many-valued logics. Corollary 3.5.11 shows that for any $n \geq 3$ the Beth property fails in the n-valued Łukasiewicz logic $Ł_n$. The Beth property also fails in a large number of relevance logics. More specifically, [Urquhart, 1999] shows failure of the Beth property for any logic between the relevance logics T and R. This includes the entailment logic E. We will prove this algebraically for an even wider class of relevance logics which includes the basic relevance logic B (see Corollary 3.5.17).

2.2.3 Related definability properties

The following model-theoretic notion of definability is discussed in [Craig, 1957]. Let Γ be a set of sentences in the relational language $\{R, S_1, \ldots S_n\}$, and let $k \leq n$.

> We shall say that the set of models of Γ *defines* the value of R *in terms of* (the underlying domain and) the values of $S_1, \ldots, S_k$, if and only if any two models of Γ which agree in the underlying domain M and in the interpretation of $S_1, \ldots, S_k$, also agree in the interpretation of R. Then the set of models of Γ may be regarded as a *function* yielding for any domain M and interpretation of $S_1, \ldots S_k$ at most one R.

The notion of implicit definability that we considered so far is the special case where $k = n$. Craig shows that in first order logic the above model-theoretic notion of definability coincides with explicit definability. This is a stronger result than Beth's theorem. Nowadays, logics with a similar property are said to have the *projective Beth property*. Below we formulate this property for propositional logics. It is left to the reader to adapt this definition to predicate logics.

2.2.12. Definition [**Projective Beth property**] Consider disjoint sequences of variables $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{q'}$ that do not contain the variables r and r'. The propositional logic S is said to have the *projective Beth property* if for any set of formulas $\Gamma(\boldsymbol{p}, \boldsymbol{q}, r)$ the condition

$$\Gamma(\boldsymbol{p}, \boldsymbol{q}, r), \Gamma(\boldsymbol{p}, \boldsymbol{q'}, r') \models_S r \leftrightarrow r', \tag{2.7}$$

implies that there exists some formula $\varphi_r(\boldsymbol{p})$ such that

$$\Gamma(\boldsymbol{p}, \boldsymbol{q}, r) \models_S r \leftrightarrow \varphi_r(\boldsymbol{p}).$$

Note that the variables from $\boldsymbol{q}$ may occur in Γ, but do not occur in φ_r. ⊣

The Beth property is a special case of the projective Beth property, where $\boldsymbol{q}$ is the empty sequence. Results by Maksimova show that indeed the projective Beth property is a much stronger property. Recall that any extension of K4, hence in particular any extension of S5, has the Beth property. However, [Maksimova, 1999a, Theorem 2.6] implies that there are only four extensions of S5 with the projective Beth property. A proof of this fact can be found in Theorem 2.3.18 below. Also in the realm of intermediate logics the two properties are rather different. [Maksimova, 1999b] shows that there are only sixteen intermediate logics with the projective Beth property. For example, the intermediate logic Z_5 which is characterized by a 5-element linearly ordered Heyting algebra, does not have the projective Beth property, [Maksimova, 1999a, Theorem 4.2]. On the other hand, as we saw in Theorem 2.2.9, all intermediate logics have the Beth property.

The above quotation by W. Craig hints at another interesting deviation. It ends by saying that " [...] The set of models of Γ may be regarded as a *function* yielding for any domain M and interpretation of $S_1, \ldots, S_k$ *at most* one R." The question thus arises what happens if we insist that there exists *at least one* such R. This yields a stronger notion of an implicit definition, whence more chance of finding an explicit definition. That is, we obtain a weaker definability property. This property was first discussed in [Friedman, 1973]. Below, $Cons_{\boldsymbol{p}}(\Gamma) = \{\varphi(\boldsymbol{p}) \ : \ \Gamma \models \varphi(\boldsymbol{p})\}$.

2.2.13. Definition [**Weak Beth property**] The logic S is said to have the *weak Beth property* if relative to any set of sentences $\Gamma(\boldsymbol{p}, r)$ that implicitly defines r, there exists an explicit definition of r, as long as

$$\text{any model of } Cons_{\boldsymbol{p}}(\Gamma(\boldsymbol{p}, r)) \text{ expands to a model of } \Gamma. \tag{2.8}$$

This model is necessarily unique. ⊣

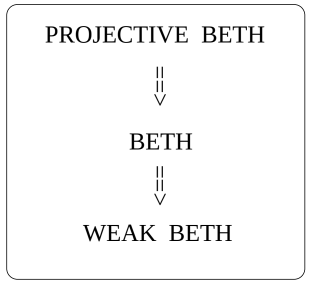


Figure 2.3: The connection between the various definability properties.

Note that the Beth property indeed implies the weak Beth property. Summarizing, we have the picture in Figure 2.3.

An implicit definition $\Gamma(\boldsymbol{p}, r)$ that satisfies condition (2.8) is called a *strong implicit definition*. For an example of a strong implicit definition, let us go back to Example 2.2.2. There it is shown that any finite linear order $\langle M, < \rangle$ (whence any model of $Cons_<(\Gamma)$) expands to a unique model $\langle M, <, R \rangle$ of Γ, by setting $m_i \in I^{\mathcal{M}}(R)$ iff i is odd. This means that Γ is a strong implicit definition of R. We recall from Example 2.2.4 that the relation symbol R does not have an explicit definition with respect to Γ. This allows us to conclude that first order logic on finite structures does not have the weak Beth property.

The implicit definition in Example 2.2.5 is *not* a strong implicit definition. To see this, note that in general the set $Cons_E(\Gamma)$ does not force its models to be of size n. However, any model of Γ has n elements. Therefore, not every model of $Cons_E(\Gamma)$ expands to a model of Γ. This leaves open the question whether the logic L^k has the weak Beth property. [Hodkinson, 1993] gives an ingenuous example, using structures called *multipedes*, showing that weak Beth fails in the k-variable fragments of first order logic, at least for $k \geq 5$.

When Friedman introduced the weak Beth property, he suggested that it was this definability property that really matters in the contexts of logics different from first order logic, cf. [Friedman, 1973]. This opinion is widely shared, cf. [Feferman, 1974] and [Makowsky and Shelah, 1979].[6] Interestingly, there are not many examples known of logics which show the difference between the Beth property and its weaker version. One such example is $L(Q)$. This is first order logic with identity and the additional quantifier $Q(x)$ with the interpretation "there are as many x as there are elements in the model." In [Friedman, 1973] it is proven

[6]In the footnote at page 10 we noted that sometimes in the literature a syntactic formulation of the notion of an implicit definition is given. Let us remark that the notion of a strong implicit definition can only be defined in a model-theoretic framework. Therefore, if one agrees with Friedman that the notion of a strong implicit definition is the proper formulation of the concept of definability one has in mind, then one is also obliged to agree that the concept of an implicit definition is essentially of a semantic nature. This supports the view we adopted in Definition 2.2.1.

that the Beth property fails for $L(Q)$, and it is asked whether this logic has the weak Beth property. [Mekler and Shelah, 1986] shows that it is consistent, assuming the consistency of ZF, that $L(Q)$ has the weak Beth property. It is an open question whether it is provable in ZFC that $L(Q)$ has the weak Beth property.

Another logic that should be mentioned in this respect is L^2, the 2-variable fragment of first order logic. As we mentioned in Example 2.2.5, the Beth property fails in L^2. It is a longstanding open question whether this logic has the weak Beth property, cf. [Sain, 1990], and [Hodkinson, 1993].

Summarizing, one can say that the weak Beth property properly formulates the definability property that we are interested in. However, for simplicity it is the Beth property, and not its weaker version, that is usually studied in the literature. This simplification is easily justified by the lack of examples of well-known logics that show the difference between these two properties.

2.3 Interpolation

This section is devoted to the Craig interpolation property ($\mathrm{CIP}^{\rightarrow}$). Section 2.3.1 contains some key references on this topic. The most important applications, which yield definability and preservation results, are discussed in section 2.3.2. In order to obtain the latter kind of results, a stronger interpolation property is introduced, the so-called Lyndon interpolation property. In section 2.3.3 we compare $\mathrm{CIP}^{\rightarrow}$ to an alternative, globally formulated, interpolation property. Section 2.3.4 is devoted to Robinson's consistency property and its relation to interpolation. We end this section with some additional motivation for the study of interpolation and a discussion on the different proof methods for interpolation.

2.3.1 Craig interpolation property

In 1957, W. Craig proved the following property for first order logic.

2.3.1. Definition [**Craig interpolation property** ($\mathrm{CIP}^{\rightarrow}$)] Let S be a logic which has the implication $\rightarrow$ among its logical connectives. S is said to have the *Craig interpolation property*, or $\mathrm{CIP}^{\rightarrow}$ for short,[7] if for any pair of S-formulas φ, ψ such that $\models_S \varphi \rightarrow \psi$, there exists an *interpolant* in S. That is, there exists an S-formula ϑ in the common language of φ, ψ such that $\models_S \varphi \rightarrow \vartheta$ and $\models_S \vartheta \rightarrow \psi$. ⊣

2.3.2. Remark In case the logic S does not contain constant formulas which denote *true* and *false*, the existence of an interpolant for $\models_S \varphi \rightarrow \psi$ is usually only required in case

$$\not\models \neg\varphi \text{ and } \not\models \psi. \tag{2.9}$$

[7] The choice for this terminology is motivated in Remark 2.3.17.

An example of such a logic is first order logic without identity. This logic does not have the property described in Definition 2.3.1. For example, there is no interpolant for $\models Px \rightarrow (Sx \leftrightarrow Sx)$. However, the interpolation theorem holds for first order logic without identity if we add condition (2.9). ⊣

Nowadays, there is an extensive literature on interpolation.[8] Even to such an extent that it is impossible to survey the whole field. In the following, we confine ourselves to some key references on this topic.

One of the first analogues of Craig's theorem was given by Schütte in 1962 for intuitionistic predicate logic. This result entails the interpolation theorem for classical first order logic. [Gabbay, 1971] shows interpolation for several extensions of intuitionistic predicate logic including the logic *CD* with constant domain and the extensions obtained by adding one or both of the following axiom schemas: $\forall x \neg\neg\varphi(x) \rightarrow \neg\neg\forall x \varphi(x)$ and $\neg\varphi \vee \neg\neg\varphi$.

The situation for intermediate logics has been thoroughly investigated by Maksimova. It turns out that in the whole continuum of intermediate logics, only eight have $\text{CIP}^{\rightarrow}$. These include CPC, intuitionistic propositional calculus (IPC), the inconsistent logic, Dummett's logic (i.e., the extension of IPC axiomatized by $x \rightarrow y \vee y \rightarrow x$), and the logic axiomatized by $\neg x \vee \neg\neg x$. The methods developed by Maksimova to obtain these results, can also be put to work to study fragments of intermediate logics. E.g., in [Maksimova, 1979] it is shown that the positive fragments of CPC, IPC and Dummett's logic are the only three consistent logics with $\text{CIP}^{\rightarrow}$ between the positive fragments of CPC and IPC. [Renardel de Lavalette, 1981] studies fragments of IPC, and concludes that there exists a continuum of such fragments without the interpolation property $\text{CIP}^{\rightarrow}$. However, most natural fragments of IPC have $\text{CIP}^{\rightarrow}$ as has been shown in [Renardel de Lavalette, 1989a, Hendriks, 2000]. In CPC, all fragments satisfy interpolation, as was proven by Ville, see [Kreisel and Krivine, 1967].

As a simple example, let us show that CPC has interpolation.

2.3.3. Theorem *Classical propositional calculus has* $\text{CIP}^{\rightarrow}$*.*

Proof: Suppose $\models_{\mathsf{CPC}} \varphi(\boldsymbol{p}, \boldsymbol{r}) \rightarrow \psi(\boldsymbol{p}, \boldsymbol{q})$, where $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r} = r_1, \ldots, r_n$ are disjoint sequences of variables. For any $X \subseteq \{1, \ldots, n\}$, let φ_X denote the result of substituting in φ all variables r_i by either $\top$ or $\bot$ according to whether or not $i \in X$. The formula $\bigvee_{X \subseteq \{1,\ldots,n\}} \varphi_X$ is an interpolant for φ, ψ. ■

Note that the choice of the interpolant in the above proof is a function of φ and the language $\{r_1, \ldots, r_n\}$, rather than of φ and ψ. In fact, we have shown the following stronger theorem. Below, for any formula φ, the *language of* φ is denoted by $\mathcal{L}_\varphi$. That is, $\mathcal{L}_\varphi$ is the set of propositional variables from which φ is built.

[8] For a motivation of the study of the interpolation property, we refer to section 2.3.2 and the discussion in section 2.3.5.

2.3.4. Theorem (Uniform interpolation theorem for CPC) *Let φ be a formula of* CPC *and $X \subseteq \mathcal{L}_\varphi$. Then there exists some formula ϑ_X of* CPC *such that*

1. $\mathcal{L}_{\vartheta_X} \subseteq X$,
2. $\models_{\mathsf{CPC}} \varphi \to \vartheta_X$, *and*
3. *for all formulas ψ such that $\mathcal{L}_\psi \cap \mathcal{L}_\varphi \subseteq X$ and $\models_{\mathsf{CPC}} \varphi \to \psi$, it is the case that $\models_{\mathsf{CPC}} \vartheta_X \to \psi$.*

A logic for which a theorem holds analogous to Theorem 2.3.4 is said to have the *uniform interpolation property.* This property turns out to be really stronger than CIP$^{\to}$.

2.3.5. Example [Henkin, 1963] contains the following example which shows that first order logic does not have the uniform interpolation property.

Let the first order sentence $\varphi(<)$ express that $<$ is an irreflexive, transitive order without endpoints. For any $n \in \omega$, let ψ_n be a formula in the language $\{=\}$ that expresses that there are at least n elements. Then $\models \varphi(<) \to \psi_n$, for any $n \in \omega$. Suppose there exists a uniform interpolant $\vartheta_=$ for $\varphi(<)$ and the set $\{=\}$. That is, there exists a formula $\vartheta_=$ which does not contain any relation symbols other than $=$, such that $\models \vartheta_= \to \psi_n$, for all $n \in \omega$. But it is well-known that if n exceeds the number of variables that occur in $\vartheta_=$, then $\not\models \vartheta_= \to \psi_n$. A contradiction. We conclude that first order logic does not have the uniform interpolation property. This shows that uniform interpolation is really stronger than CIP$^{\to}$. ⊣

Another example of a logic with CIP$^{\to}$ but without uniform interpolation is the modal logic S4, [Ghilardi and Zawadowski, 1995].

According to Theorem 2.3.4, an interpolant for $\models_{\mathsf{CPC}} \varphi \to \psi$ can be found from φ and some set $X \subseteq \mathcal{L}_\varphi$. In CPC, this is equivalent to the condition that an interpolant can be found from ψ and some set $Y \subseteq \mathcal{L}_\psi$. However, if a logic does not contain the axiom schema $\neg\neg x \to x$ then the two conditions are not interderivable. This is, for example, the case in intuitionistic propositional calculus (IPC). Nevertheless, it turns out that IPC has both versions of the uniform interpolation property, cf. [Pitts, 1992]. This surprising result needed a rather involved proof. [Visser, 1996] contains another proof of this fact and proves via the same method uniform interpolation for the basic modal logic K, provability logic GL and Grzegorczyck's logic Grz.

A logic without CIP$^{\to}$, whence certainly without the uniform interpolation property, is the infinitary logic L_∞, as was shown by [Malitz, 1971]. The following example is from [Barwise and van Benthem, 1999].

2.3.6. Example Let $<, \prec$ be binary relation symbols. The sentence $\varphi(<)$ says that the universe of the model $\mathcal{M}$ is well ordered by $<^M$ in order type ω. $\psi(\prec)$ says that the universe is well ordered by $\prec^M$ in order type ω_1. Obviously, the implication $\varphi(<) \to \neg\psi(\prec)$ is valid.

Suppose ϑ is an interpolant for $\varphi(<) \to \neg\psi(\prec)$ in L_∞. Then ϑ is an L_∞-formula in the language $\{=\}$ which is true in all models of size $\aleph_0$ but false in all models of size $\aleph_1$. However, [Malitz, 1971] shows that all infinite structures are indistinguishable in L_∞ if we use only $\{=\}$. A contradiction. Therefore, L_∞ does not have CIP$^\to$. $\dashv$

Note that the formulas φ, ψ in the above example can be written in L_{ω_2}. Hence, this example actually shows the failure of CIP$^\to$ in every infinitary logic between L_{ω_2} and L_∞. Moreover, we observe that φ, ψ use two variables only. This implies that CIP$^\to$ fails in the k-variable fragments of all these logics, for any $k \geq 2$.

The underlying idea in Example 2.3.6 is the following. There exist cardinalities s_1, s_2 with the following two properties. On the one hand it is possible to enforce models to be of size s_1, respectively s_2, by making use of additional predicate symbols. On the other hand, structures of size s_1 and s_2 are indistinguishable in the logic if we only use the identity symbol $=$. The same heuristics are followed in Example 2.3.7 below. A logic for which this idea does not work is L_{ω_1}. For, finite models of different size are distinguishable in L_{ω_1} by means of pure identity formulas, whereas structures of different infinite size can never be enforced. But not only does this idea not work in L_{ω_1}. It turns out that L_{ω_1} indeed has CIP$^\to$, as was shown in [Lopez-Escobar, 1965].

2.3.7. Example [Andréka et al., 1998, Theorem 3.5.1] We show that the finite variable fragments L^k of first order logic fail to have interpolation, for any $k \geq 2$. Let $k \geq 2$ and consider the unary predicates $P_1, \ldots, P_k$. The formula φ_k is the conjunction of the following four formulas,

1. $\forall x \bigvee_{1\leq i\leq k} P_i x$.
2. $\bigwedge_{1\leq i\leq k} \exists x P_i x$.
3. $\forall x \bigwedge_{1\leq i,j\leq k, i\neq j}(P_i x \to \neg P_j x)$.
4. $\forall x \forall y (\bigwedge_{1\leq i\leq k}((x \neq y \wedge P_i x) \to \neg P_i y))$.

The formula φ_k states a one-one correspondence between elements in the domain and the properties P_i. Hence, if φ_k is true in a model $\mathcal{M}$, then the domain of $\mathcal{M}$ is of size k. Similarly, define φ_{k+1} using different predicate symbols $R_1, \ldots, R_{k+1}$. Note that φ_k, φ_{k+1} use two variables only. In particular, they are L^k-formulas. Moreover, $\models \varphi_k \to \neg\varphi_{k+1}$.

Suppose L^k has interpolation. Then there exists a k-variable formula ϑ in the language $\{=\}$ such that $\models \varphi_k \to \vartheta$ and $\models \vartheta \to \neg\varphi_{k+1}$. However, it is well-known that k-variable formulas using $=$ only can not distinguish between domains with k and $k+1$ objects, respectively. $\dashv$

With only one variable at our disposal, no counting is possible. Therefore, no counterexample along the above lines is possible. But even more is the case: first order logic with just one variable has CIP$^{\rightarrow}$.

The results on CIP$^{\rightarrow}$ for infinitary logics and finite variable fragments have been summarized in the table below.

Logic	**Craig interpolation property**$^{\rightarrow}$
L^1	Yes
(1-variable fragment of f.o.l.)	[Pigozzi, 1972]
L^k, $1 < k < \omega$	No
(k-variable fragment of f.o.l.)	[Comer, 1969, Pigozzi, 1972]
First order logic L_ω	Yes
	[Craig, 1957]
Infinitary logic L_{ω_1}	Yes
	[Lopez-Escobar, 1965]
Infinitary logic L_α,	No
$\alpha > \omega_1$ or $\alpha = \infty$	[Malitz, 1971]

We note that according to Theorem 2.3.10 below, failure of CIP$^{\rightarrow}$ in L^k, for $k \geq 3$, also follows from Example 2.2.5 which shows the failure of the Beth property in these logics.

Interpolation has also been extensively studied for modal logics. CIP$^{\rightarrow}$ has been proven for many important propositional modal logics, including the systems K, $\mathsf{K4}$, T, $\mathsf{S4}$, [Gabbay, 1972], $\mathsf{S5}$, [Schumm, 1976], B (= K + $p \rightarrow \Box\Diamond p$), M (= K + $\Box\Diamond p \rightarrow \Diamond\Box p$), [Rautenberg, 1983], provability logic GL, [Smoryński, 1978] and Grzegorczyck's logic Grz, [Boolos, 1980]. For propositional dynamic logic, the answer is still unknown. [Maksimova, 1991] is a survey of results on interpolation in propositional modal logics. It shows e.g., that there exists only finitely many extensions of $\mathsf{S4}$ with interpolation. To be precise, no more than 37 have CIP$^{\rightarrow}$. Among the well-known $\mathsf{S4}$-extensions without interpolation, we mention $\mathsf{S4.3}$, i.e., the logic of linearly ordered frames, $\mathsf{GL.3}$ and $\mathsf{Grz.3}$. The situation is rather different for extensions of $\mathsf{K4}$, as there exists a continuum of normal extensions of $\mathsf{K4}$ with CIP$^{\rightarrow}$. To see this, we need the following notion which has been introduced by Maksimova. Let S be a normal modal logic. Let $\varphi(p)$ be a formula in at most the variable p. By $\varphi(\psi)$ we denote the result of substituting the formula ψ for p in $\varphi(p)$. The formula $\varphi(p)$ is called *S-conservative* if the following two conditions are satisfied:

1. $\varphi(\bot), \varphi(p), \varphi(q) \models_S \varphi(p \rightarrow q)$.
2. $\varphi(\bot), \varphi(p) \models_S \varphi(\Box p)$.

The essential property of an S-conservative formula $\varphi(p)$ is that

$$\varphi(\bot), \varphi(p_1), \ldots, \varphi(p_n) \models_S \varphi(\psi(p_1, \ldots, p_n)), \text{ for all formulas } \psi(p_1, \ldots, p_n). \tag{2.10}$$

As an example, we leave it to the reader to verify that the formula $\Box\Diamond p \to \Diamond\Box p$ is $\mathsf{S4}$-conservative. Evidently, all variable free formulas are S-conservative for any logic S. The proposition below generalizes Theorem 3 in [Rautenberg, 1983] concerning variable free formulas.

2.3.8. Proposition [Maksimova, 1991, Proposition 3] *Let S be a compact, normal modal logic, and let Ax be a set of S-conservative formulas. If S has* CIP$^{\to}$*, then $S + Ax$ has* CIP$^{\to}$*.*

Proof: Let S, Ax be as in the proposition. Assume $\models_{S+Ax} \varphi \to \psi$. That is,

$$Ax \models_S \varphi \to \psi. \tag{2.11}$$

We will give an interpolant. Without loss of generality, suppose $\mathcal{L}_{Ax} \subseteq \mathcal{L}_\varphi \cup \mathcal{L}_\psi$. Say $\mathcal{L}_\varphi = \{p_1, \ldots, p_n\}$ and $\mathcal{L}_\psi = \{q_1, \ldots, q_k\}$. As S is compact and conjunctive, we may assume Ax to be a single formula. By (2.10) and (2.11), $Ax(\bot), Ax(p_1), \ldots, Ax(p_n), Ax(q_1), \ldots, Ax(q_k) \models_S \varphi \to \psi$. By the local deduction property of S, there exists some $m \in \omega$ such that

$$\models_S (\bigwedge_{1\le i\le n} \boxdot^m Ax(p_i) \wedge \bigwedge_{1\le j\le k} \boxdot^m Ax(q_j) \wedge \boxdot^m Ax(\bot)) \to (\varphi \to \psi),$$

where $\boxdot^m Ax = Ax \wedge \Box Ax \wedge \cdots \wedge \Box^m Ax$. Then $\models_S (\bigwedge_{1\le i\le n} \boxdot^m Ax(p_i) \wedge \boxdot^m Ax(\bot) \wedge \varphi) \to (\bigwedge_{1\le j\le k} \boxdot^m Ax(q_j) \to \psi)$. Let ϑ be an interpolant for the above implication in S. One easily verifies that ϑ is also an interpolant for $\models_{S+Ax} \varphi \to \psi$. ■

Recall that the formula $\Box\Diamond p \to \Diamond\Box p$ is $\mathsf{S4}$-conservative. Moreover, we already mentioned that $\mathsf{S4}$ has CIP$^{\to}$. Hence, by Proposition 2.3.8 the system $\mathsf{S.4.1} = \mathsf{S4} + \Box\Diamond p \to \Diamond\Box p$ has CIP$^{\to}$. The proposition also implies that CIP$^{\to}$ holds for the deontic system $\mathsf{D} = \mathsf{K} + \Diamond\top$ and the system $\mathsf{D4}$. As Maksimova noted, $\mathsf{K4}$ has a continuum of extensions of the form $\mathsf{K4} + Ax$, with Ax variable free. In particular, all these axiom sets Ax are $\mathsf{K4}$-conservative. Therefore, as another corollary of Proposition 2.3.8 we obtain that there exists a continuum of normal extensions of $\mathsf{K4}$ with CIP$^{\to}$.

[Marx, 1995] gives a necessary condition for certain canonical modal logics to have interpolation. This is basically a reformulation of amalgamation in terms of *frames*. Typical examples of frame conditions that obstruct amalgamation are $\forall\exists$-formulas of the form "For every x, y with a certain property, there exists some z that is somehow related to them." Examples include the density axiom (i.e., $\forall x \forall y(Rxy \to \exists z(Rxz \wedge Rzy))$), the Church-Rosser property (i.e., $\forall x\forall y\forall z((Rxy \wedge$

$Rxz) \to \exists w(Ryw \wedge Rzw)))$, and the commutativity of the cylindrifications C_i from cylindric algebra theory (i.e., $\forall x \forall y \forall z((R_i xy \wedge R_j yz) \to \exists w(R_j xw \wedge R_i wz)))$. Marx concludes that interpolation fails in the corresponding modal logics axiomatized by respectively $\Diamond p \to \Diamond\Diamond p$, $\Diamond\Box p \to \Box\Diamond p$ and $C_i C_j p \to C_j C_i p$.[9]

As we learned in section 2.2, the Beth property fails in a large number of quantificational modal logics. As we will see in Theorem 2.3.10, this implies that $\mathrm{CIP}^{\to}$ also fails in these systems. However, first order K, T and S4 (all without the Barcan formula) have $\mathrm{CIP}^{\to}$, cf. [Gabbay, 1972].

2.3.2 Applications: definability and preservation

At the time of publication, the interpolation theorem for first order logic was seen as a 'fundamental lemma' which, in Craig's own words, " [...] seems a useful tool for further investigations. In particular, it may lend itself to questions of this kind: How is a certain model-theoretic property of a system reflected by theorems in the system?" Beth's theorem answers one such question, and indeed, this answer also follows from Craig's lemma. This was the foremost application of the interpolation lemma in [Craig, 1957].

Proof of Beth's theorem from the interpolation lemma for first order logic: Suppose Γ implicitly defines R in first order logic. That is, $\Gamma, \Gamma' \models \forall \boldsymbol{x}(R\boldsymbol{x} \leftrightarrow R'\boldsymbol{x})$, for all relation symbols R' of the same arity as R which do not occur in Γ. Here, Γ' denotes the result of substituting R' for R in Γ and $\models$ stands for the semantic consequence relation of first order logic. As first order logic is compact and conjunctive, we may assume Γ to be a single formula. By the deduction property, $\models (\Gamma \wedge R\boldsymbol{x}) \to (\Gamma' \to R'\boldsymbol{x})$. From Craig's lemma, we obtain an interpolant ϑ which does not contain the relation symbol R such that

$$\models (\Gamma \wedge R\boldsymbol{x}) \to \vartheta, \tag{2.12}$$

$$\models \vartheta \to (\Gamma' \to R'\boldsymbol{x}). \tag{2.13}$$

From (2.12) it follows that $\models \Gamma \to (R\boldsymbol{x} \to \vartheta)$ so, by the detachment property, $\Gamma \models R\boldsymbol{x} \to \vartheta$. Similarly, (2.13) implies that $\models \Gamma' \to (\vartheta \to R'\boldsymbol{x})$, whence $\Gamma' \models \vartheta \to R'\boldsymbol{x}$. Substituting R for R' in the latter derivation yields $\Gamma \models \vartheta \to R\boldsymbol{x}$. We conclude that $\Gamma \models R\boldsymbol{x} \leftrightarrow \vartheta$. In other words, ϑ is an explicit definition of R with respect to Γ. ■

An essential role in the above proof is played by the deduction theorem for first order logic. This property can be generalized as follows. Our generalization is very close to what is called a *local deduction theorem* in [Blok and Pigozzi, 1991].

[9] This result actually shows failure of the weaker interpolation property $\mathrm{CIP}^{\models}$, to be introduced below.

2.3.9. Definition [**Local deduction property**] A logic S is said to have the *local deduction property* if for each pair φ, ψ of formulas of S there exists a formula χ_φ in the same language as φ such that for all formulas ξ,

1. $\xi, \varphi \models \psi$ iff $\xi \models \chi_\varphi \to \psi$, and
2. $\varphi, \chi_\varphi \to \xi \models \xi$.

The formula χ_φ is called a *deduction term* for φ, ψ. ⊣

Note that the deduction property of first order logic is much stronger, as in that case the deduction term χ_φ can be uniformly chosen to be φ and does not depend on the pair φ, ψ. The situation is different for e.g., the basic modal logic K, where for every φ, ψ there exists some $n \in \omega$ such that $\varphi \wedge \Box\varphi \wedge \cdots \wedge \Box^n\varphi$ is a deduction term for φ, ψ.

We obtain the following general connection between $\mathrm{CIP}^{\to}$ and the Beth property.

2.3.10. Theorem *Let S be a compact, conjunctive logic with the local deduction property. If S has* $\mathrm{CIP}^{\to}$*, then S has the Beth property.*

Proof: We prove the theorem for a propositional logic S. The case for predicate logics is similar and is left to the reader. Suppose Γ implicitly defines r in S. That is, $\Gamma(\boldsymbol{p}, r), \Gamma(\boldsymbol{p}, r') \models_S r \leftrightarrow r'$. As S is compact and conjunctive, we may assume Γ to be a single formula. By the local deduction property, there exists some formula $\chi_{\Gamma'}(\boldsymbol{p}, r')$ such that $\Gamma(\boldsymbol{p}, r) \models_S \chi_{\Gamma'}(\boldsymbol{p}, r') \to (r \leftrightarrow r')$. Applying the local deduction theorem once again, we obtain a formula $\chi_\Gamma(\boldsymbol{p}, r)$ such that $\models_S \chi_\Gamma(\boldsymbol{p}, r) \to (\chi_{\Gamma'}(\boldsymbol{p}, r') \to (r \leftrightarrow r'))$. It follows that $\models_S (\chi_\Gamma(\boldsymbol{p}, r) \wedge r) \to (\chi_{\Gamma'}(\boldsymbol{p}, r') \to r')$. Since S had $\mathrm{CIP}^{\to}$, there exists an interpolant $\vartheta(\boldsymbol{p})$ for the above implication. Reasoning as in the proof of Beth's theorem from the interpolation lemma for first order logic, we see that ϑ is an explicit definition of r with respect to Γ. Details are left to the reader. ■

In particular, any intermediate logic with $\mathrm{CIP}^{\to}$ has the Beth property. Similarly, any compact, normal modal logic with $\mathrm{CIP}^{\to}$ has the Beth property. This has lead to the widespread belief that "interpolation implies definability." However, it is good to keep in mind that this is *not* a strict implication. That is, there exist logics with $\mathrm{CIP}^{\to}$ but without the Beth property. Simple examples include the implicational fragments of classical propositional logic and intuitionistic propositional logic. These logics have $\mathrm{CIP}^{\to}$ as has been shown in [Kreisel and Krivine, 1967] resp. [Renardel de Lavalette, 1989a]. However, they do not have the Beth property as we will see in Proposition 3.5.6. An example of a conjunctive logic with $\mathrm{CIP}^{\to}$ but without the Beth property is the following from [Andréka and Németi, 1996].

2.3.11. Example (**$\mathrm{CIP}^{\to} \not\Rightarrow$ Beth property**) The tense logic TL contains two unary modalities $\odot$ ('first time') and the earlier encountered $\bigcirc$ ('tomorrow').

TL is the axiomatic extension of the basic system K with the additional axioms: $\odot\neg\varphi \leftrightarrow \neg \odot \varphi$, $\bigcirc\neg\varphi \leftrightarrow \neg \bigcirc \varphi$, $\odot\varphi \leftrightarrow \odot \odot\varphi$ and $\odot\varphi \leftrightarrow \bigcirc \odot\varphi$. [Marx, 1995, Example 5.1.8] shows that TL has $\mathrm{CIP}^{\rightarrow}$. By Theorem 2.3.10, then TL also has the Beth property.

The logic TL^+ is obtained from TL by adding the following induction-like inference rule (IR): $\odot\varphi, \varphi \rightarrow \bigcirc \varphi/\varphi$. Note that TL and TL^+ have the same validities. As TL has $\mathrm{CIP}^{\rightarrow}$, this implies that TL^+ also has $\mathrm{CIP}^{\rightarrow}$. However, TL^+ does not have the local deduction property. Hence, the fact that TL^+ has $\mathrm{CIP}^{\rightarrow}$ need not imply that TL^+ has the Beth property. In fact, we claim that the Beth property fails for TL^+.

To prove this claim, consider the theory $\Gamma(r) = \{\odot r, \bigcirc\neg r\}$. Using the fact that the formulas $\bigcirc\varphi \rightarrow \bigcirc (\psi \rightarrow \varphi)$ and $\odot\varphi \rightarrow \odot (\psi \rightarrow \varphi)$ are valid in TL^+, one derives that $\Gamma(r), \Gamma(r') \vdash_{TL^+} \odot(r \leftrightarrow r')$ and $\Gamma(r), \Gamma(r') \vdash_{TL^+} \bigcirc(r \leftrightarrow r')$. The induction rule then yields $\Gamma(r), \Gamma(r') \vdash_{TL^+} r \leftrightarrow r'$. That is, Γ implicitly defines r in TL^+. On the other hand, we claim that r is not explicitly definable with respect to Γ in TL^+. To see this, we recall from [Andréka and Németi, 1996] that TL^+ is strongly complete with respect to the frame $\mathcal{N} = \langle \omega, succ \rangle$, where ω is the set of natural numbers, and for any $n \in \omega$,

$$
\begin{aligned}
n \models \odot\varphi &\text{ iff } 0 \models \varphi, \\
n \models \bigcirc\varphi &\text{ iff } succ(n) \models \varphi.
\end{aligned}
$$

Let the valuation v on $\mathcal{N}$ be defined by $v(r) = \{0\}$. Note that $\langle \mathcal{N}, v \rangle \models \Gamma$. Similar to Example 2.2.8 concerning von Wright's logic, one shows by induction on the complexity of φ that for every variable free formula φ and every $n \in \omega$ it is the case that $\langle \mathcal{N}, v \rangle, n \models \varphi$ iff $\langle \mathcal{N}, v \rangle, succ(n) \models \varphi$. Now we have all the ingredients to show that there is no explicit definition of r with respect to Γ in TL^+. For, suppose such definition φ_r does exist. Then φ_r is a variable free formula such that $\Gamma \models \varphi_r \leftrightarrow r$. As $\langle \mathcal{N}, v \rangle \models \Gamma$ and $\langle \mathcal{N}, v \rangle, 1 \models \neg r$, then $\langle \mathcal{N}, v \rangle, 1 \models \neg\varphi_r$. By the above observation this implies that $\langle \mathcal{N}, v \rangle, 0 \models \neg\varphi_r$, whence $\langle \mathcal{N}, v \rangle, 0 \models \neg r$. Contradiction. We conclude that TL^+ does not have the Beth property. $\dashv$

Another type of model-theoretic question that can be answered via the interpolation lemma concerns preservation theorems. The question is to characterize those formulas that are preserved under a given model-theoretic operation such as the operation of taking substructures, or homomorphic images. For proving such preservation theorems, the interpolant usually needs to have some additional properties. For example, [Malitz, 1969] contains the following result about the infinitary logic L_{ω_1}. If $\models_{\mathsf{L}_{\omega_1}} \varphi \rightarrow \psi$ and ψ is a universal formula in L_{ω_1}, then there exists a *universal* interpolant in L_{ω_1}. As a corollary, Malitz obtains a version of the Łos-Tarski theorem for L_{ω_1}. That is, an L_{ω_1}-sentence φ is preserved under submodels iff φ is equivalent to a universal L_{ω_1}-sentence. Also [Barwise and van Benthem, 1999] contains a number of results of this general character: an interpolation theorem with a preservation theorem as consequence.

The most important interpolation property of this kind has been formulated by R. Lyndon. [Lyndon, 1959a] studies “ [...] the formal structure of sentences whose validity is preserved under passage from an algebraic system to a homomorphic image of the system.” In this study, the notion of a positive formula plays a central role. For technical reasons, let, for the moment, formulas be built up from atomic formulas using conjunction, disjunction, negation and existential and universal quantification. A given occurrence of a relation symbol is said to be *positive* iff it occurs within the scope of an even number of negation signs. It is called *negative* otherwise. A formula φ is called *positive* if every occurrence of a relation symbol in φ is positive. The conclusion in [Lyndon, 1959b] is that “A sentence of the predicate calculus is preserved under homomorphism if and only if it is equivalent to a *positive* sentence.” The cornerstone in the proof of this result is the following interpolation lemma in which the occurrences of relation symbols are taken into account.

2.3.12. Theorem (Lyndon's interpolation theorem for first order logic) *Let φ, ψ be first order formulas that do not contain any function symbols. If $\models \varphi \to \psi$, then there exists a formula ϑ such that*

1. $\models \varphi \to \vartheta$ *and* $\models \vartheta \to \psi$, *and*
2. *every relation symbol (other than the identity) which occurs positively (resp. negatively) in ϑ occurs positively (resp. negatively) in both φ and ψ.*

For first order logic with function symbols (including constants), Lyndon interpolation does not hold. That is, a first order validity $\varphi \to \psi$ need not have an interpolant ϑ in which every function symbol which occurs positively (resp. negatively) in ϑ occurs positively (resp. negatively) in both φ and ψ. For example, consider the theorem

$$\models \exists x(x = c \wedge \neg P(x)) \to \neg P(c).$$

The constant c occurs only positively in the antecedent and does not occur positively in the consequent. Nevertheless, c must occur in any interpolant.

An analogue of Lyndon's theorem is known to hold for the propositional modal systems K, T, D, $\mathsf{S4}$, $\mathsf{S5}$, [Fitting, 1983], provability logic GL, [Sánchez Valencia,], and the first order systems K, $\mathsf{K4}$, $\mathsf{S4}$, [Maksimova, 1982]. These logics are said to have the *Lyndon interpolation property*. It seems to be an open question whether such an analogue exists for the logic Grz. [Maksimova, 1982] contains examples of modal logics with $\mathrm{CIP}^{\to}$ but without the Lyndon interpolation property. The simplest of them is the following.

Figure 2.4: The frame $\mathcal{F}$.

2.3.13. Example (**CIP$^{\rightarrow}$ $\not\Rightarrow$ Lyndon interpolation**)[Maksimova, 1991, Theorem 11] Let S be the logic of the frame $\mathcal{F}$ depicted in Figure 2.4. Using algebraic methods, Maksimova showed CIP$^{\rightarrow}$ for S. Now consider the formula

$$(\Diamond p \wedge \neg p \wedge \Box(\neg p \vee q)) \rightarrow (\neg q \vee \Box q),$$

which is evidently valid in S. Suppose ϑ is a Lyndon-interpolant for this formula in S. Then ϑ contains at most the variable q which moreover occurs only positively. Define the valuation v on $\mathcal{F}$ by $v(p) = v(q) = \{b\}$. Note that $\langle \mathcal{F}, v\rangle, a \models \Diamond p \wedge \neg p \wedge \Box(\neg p \vee q)$. Hence, $\langle \mathcal{F}, v\rangle, a \models \vartheta$. By induction on the complexity of the formula it can be seen that the only formula $\chi(q)$ in which q occurs only positively and which is true in a under the valuation v, is the constant formula $\top$. However, $\top$ clearly is not an interpolant, as $\langle \mathcal{F}, v\rangle, b \not\models \top \rightarrow (\neg q \vee \Box q)$. A contradiction. This shows that S does not have the Lyndon interpolation property. ⊣

An example of a proof of a Lyndon interpolation theorem can be found in section 4.7 where we prove (a restricted version of) the Lyndon interpolation property for the guarded fragment of first order logic.

2.3.3 A global interpolation property

In general, there are several possible formulations of the interpolation property. A well-known alternative is the following.

2.3.14. Definition [**$\models$-Craig interpolation property** (CIP$^{\models}$)] The logic S is said to have the $\models$-*Craig interpolation property*, or CIP$^{\models}$ for short, if for any pair of S-formulas φ, ψ such that $\varphi \models_S \psi$ there exists an interpolant in S. That is, there exists an S-formula ϑ in the common language of φ, ψ such that $\varphi \models_S \vartheta$ and $\vartheta \models_S \psi$. ⊣

One easily verifies that in the presence of a local deduction theorem as defined in 2.3.9, CIP$^{\rightarrow}$ implies CIP$^{\models}$. This is in particular the case for all normal modal logics, and all intermediate logics. The other implication does not hold, not even for systems with the deduction property. [Maksimova, 1980] gives an example of a normal extension of S4 with the deduction property and with CIP$^{\models}$ but without CIP$^{\rightarrow}$. Without any form of deduction theorem, there is no correlation between the two interpolation properties at all. This was to be expected. For, CIP$^{\rightarrow}$ depends only on the validities of the logic, whereas CIP$^{\models}$ depends on the consequence relation $\models$. To say that a logic fails to have a local deduction property precisely means that these two features are not strongly linked.

Simple examples of logics with $\mathrm{CIP}^{\models}$ but without $\mathrm{CIP}^{\rightarrow}$ can be constructed as follows.

2.3.15. Example ($\mathbf{CIP}^{\models} \not\Rightarrow \mathbf{CIP}^{\rightarrow}$) Let S be any logic without $\mathrm{CIP}^{\rightarrow}$. We define the extension S^+ by adding the following (infinite number of) rules to S:

$$\{\varphi/\bot \ : \not\models_S \varphi\}.$$

Note that S^+ does not have $\mathrm{CIP}^{\rightarrow}$, as S^+ has the same theorems as S. However, S^+ has $\mathrm{CIP}^{\models}$. For, suppose $\varphi \models_{S^+} \psi$. Either $\not\models_S \varphi$ and $\bot$ is an interpolant. Or $\models_S \varphi$, whence $\models_{S^+} \psi$, and $\top$ is an interpolant. $\dashv$

Next, an example from [Andréka and Németi, 1996] of a logic with $\mathrm{CIP}^{\rightarrow}$ but without $\mathrm{CIP}^{\models}$.

2.3.16. Example ($\mathbf{CIP}^{\rightarrow} \not\Rightarrow \mathbf{CIP}^{\models}$) The modal logic *ML* is the extension of the basic system K with the one extra rule $\Diamond\top \rightarrow \Diamond\Diamond\top/\Diamond p \rightarrow \Diamond\Diamond p$. Note that the logic *ML* has the same validities as K. Since this latter logic is known to have $\mathrm{CIP}^{\rightarrow}$ it follows that *ML* also has $\mathrm{CIP}^{\rightarrow}$.

To see that *ML* fails to have $\mathrm{CIP}^{\models}$, one first notes (as is done in [Marx, 1995, Theorem 5.6.6.1] that the extension S of K with the axiom $\Diamond p \rightarrow \Diamond\Diamond p$ fails to have this property. This yields two formulas φ, ψ such that $\varphi \models_S \psi$ without an interpolant in S. Then $(\Diamond\top \rightarrow \Diamond\Diamond\top) \wedge \varphi \models_{ML} \psi$. We claim that this pair does not have an interpolant. For suppose ϑ is such an interpolant. Note that ϑ is a formula in the common language of φ and ψ. Then $\varphi \models_S \varphi \wedge (\Diamond\top \rightarrow \Diamond\Diamond\top) \models_{ML} \vartheta \models_{ML} \psi$. As $\models_{ML} \subseteq \models_S$ it follows that ϑ is an interpolant for φ and ψ in S. A contradiction. $\dashv$

Another example of a modal logic with $\mathrm{CIP}^{\models}$ but without $\mathrm{CIP}^{\rightarrow}$ is the system $\mathsf{S4.3.2}$. This extension of $\mathsf{S4}$ is defined by the axiom $\Box p \vee \Box(\Box p \rightarrow (\Box q \vee \Box\neg\Box q))$. The proof can be found in [Maksimova, 1980] and is by no means trivial. To show that $\mathsf{S4.3.2}$ has $\mathrm{CIP}^{\models}$, Maksimova uses algebraic methods and proceeds as in Example 2.5.5 below. That $\mathsf{S4.3.2}$ does not have $\mathrm{CIP}^{\rightarrow}$ follows from the main theorem of [Maksimova, 1980] which gives a complete classification of all extensions of $\mathsf{S4}$ with $\mathrm{CIP}^{\rightarrow}$.

2.3.17. Remark We already noted that in the presence of a local deduction theorem $\mathrm{CIP}^{\rightarrow}$ implies $\mathrm{CIP}^{\models}$. In these contexts it is therefore reasonable to distinguish a *weak interpolation property* from a *strong interpolation property*, as has been done in the literature. However, from the general perspective taken in this chapter in which the strong version need not imply the weak one, such terminology seems inappropriate. Therefore we adhere to the more suggestive $\mathrm{CIP}^{\rightarrow}$ and $\mathrm{CIP}^{\models}$. Other names that appear in the literatures are *local interpolation property* and *global interpolation property.* $\dashv$

In the previous section we introduced the projective Beth property. To indicate the difference between this definability property and the usual Beth property we referred to results by Maksimova that imply that only four normal extensions of S5 have the projective Beth property, whereas they all have the Beth property. In fact, Maksimova came to this conclusion via earlier results in [Maksimova, 1980] according to which only four extensions of S5 have $\mathrm{CIP}^{\models}$, together with the following theorem which states that in extensions of S5 the projective Beth property implies $\mathrm{CIP}^{\models}$.

2.3.18. Theorem [Maksimova, 1999a, Theorem 2.6] *Let S be a normal extension of* S5 *with the projective Beth property. Then S has* $\mathrm{CIP}^{\models}$.

Proof: Let S be a normal extension of S5 with the projective Beth property. Assume that $\varphi(\boldsymbol{p},\boldsymbol{q_1}) \models_S \psi(\boldsymbol{p},\boldsymbol{q_2})$. We need to give an interpolant. From the deduction theorem for extensions of S5 it follows that $\models_S \Box\varphi(\boldsymbol{p},\boldsymbol{q_1}) \to \Box\psi(\boldsymbol{p},\boldsymbol{q_2})$. This implies that $\models_S [\Box(r \to \Box\varphi(\boldsymbol{p},\boldsymbol{q_1})) \wedge \Box(\Box\psi(\boldsymbol{p},\boldsymbol{q_2'}) \to r')] \to (r \to r')$, and $\models [\Box(r' \to \Box\varphi(\boldsymbol{p},\boldsymbol{q_1'})) \wedge \Box(\Box\psi(\boldsymbol{p},\boldsymbol{q_2}) \to r)] \to (r' \to r)$. Hence, $r \to \Box\varphi(\boldsymbol{p},\boldsymbol{q_1}) \wedge \Box\psi(\boldsymbol{p},\boldsymbol{q_2}) \to r, r' \to \Box\varphi(\boldsymbol{p},\boldsymbol{q_1}') \wedge \Box\psi(\boldsymbol{p},\boldsymbol{q_2}') \to r' \models_S r' \leftrightarrow r$. By the projective Beth property, there exists some formula $\chi(\boldsymbol{p})$ such that

$$[r \to \Box\varphi(\boldsymbol{p},\boldsymbol{q_1})] \wedge [\Box\psi(\boldsymbol{p},\boldsymbol{q_2}) \to r] \models_S \chi(\boldsymbol{p}) \leftrightarrow r. \tag{2.14}$$

Set $r = \top$ in (2.14). This yields $\Box\varphi(\boldsymbol{p},\boldsymbol{q_1}) \models_S \chi(\boldsymbol{p})$. As $\models_S$ is the *global* consequence relation (see the appendix), $\varphi(\boldsymbol{p},\boldsymbol{q_1}) \models_S \Box\varphi(\boldsymbol{p},\boldsymbol{q_1})$, and we see that $\varphi(\boldsymbol{p},\boldsymbol{q_1}) \models_S \chi(\boldsymbol{p})$.

For $r = \bot$ in (2.14) we obtain $\neg\Box\psi(\boldsymbol{p},\boldsymbol{q_2}) \models_S \neg\chi(\boldsymbol{p})$. By the deduction property, $\models_S \Box\neg\Box\psi(\boldsymbol{p},\boldsymbol{q_2}) \to \neg\chi(\boldsymbol{p})$, whence, $\models_S \chi(\boldsymbol{p}) \to \neg\Box\neg\Box\psi(\boldsymbol{p},\boldsymbol{q_2})$. Recall that S5 contains the axiom $\neg\Box\neg\Box p \to p$. Therefore, $\models_S \chi(\boldsymbol{p}) \to \psi(\boldsymbol{p},\boldsymbol{q_2})$ and certainly $\chi(\boldsymbol{p}) \models_S \psi(\boldsymbol{p},\boldsymbol{q_2})$. We conclude that $\chi(\boldsymbol{p})$ is an interpolant for $\varphi(\boldsymbol{p},\boldsymbol{q_1}), \psi(\boldsymbol{p},\boldsymbol{q_2})$. ■

In the remaining part of this subsection, we discuss the merits of the two interpolation properties $\mathrm{CIP}^{\to}$ and $\mathrm{CIP}^{\models}$.

Which of the two interpolation properties is more significant, depends on what is understood by a logic. If a logic is seen as a *set of theorems*, then $\mathrm{CIP}^{\to}$ is the more interesting notion as it concerns only the theorems. If, on the other hand, one takes a logic to be a *set of rules of inference*, it is only natural to define interpolation in terms of the consequence relation, i.e., as $\mathrm{CIP}^{\models}$. To illustrate the difference between these two views, consider the global and local version of a modal system S (see the appendix for a definition). These logics share the same theorems, but do not have the same inference rules. E.g., the rule of necessitation is valid in the global version of S, but not in the local one.

Logic as a set of inference rules is the prevailing point of view in for example the field of algebraic logic. Hence in this area, $\mathrm{CIP}^{\models}$ is emphasized. There is also some algebraic support for this position. For, as we will see in the next section, $\mathrm{CIP}^{\models}$ is closely related to the *amalgamation property*, a well-known property in the universal algebra literature. On the other hand, $\mathrm{CIP}^{\rightarrow}$ is not related to any such property (the *superamalgamation property* which corresponds to $\mathrm{CIP}^{\rightarrow}$ in a similar way has been especially designed for this purpose).

Another drawback of $\mathrm{CIP}^{\rightarrow}$ is that it can only be formulated for implicative logics, whereas $\mathrm{CIP}^{\models}$ has a general formulation. A generalization of $\mathrm{CIP}^{\rightarrow}$ which is suggested in [Czelakowski and Pigozzi, 1999] is Δ-interpolation, where $\Delta(x, y)$ is a set of formulas in the two variables x, y. E.g., $\{\rightarrow\}$-interpolation is the usual $\mathrm{CIP}^{\rightarrow}$.

Now that we mention Δ-interpolation, one may wonder whether it is the most natural choice for $\Delta(x, y)$ to be $x \rightarrow y$. For example, in an algebraic study of CPC and first order logic the formula $x \leftrightarrow y$ plays a central role. This explains the remark in [Czelakowski and Pigozzi, 1999] that " [...] the real significance of $\mathrm{CIP}^{\rightarrow}$ for e.g., CPC or first order logic lies in the fact that $\mathrm{CIP}^{\rightarrow}$ implies $\{\leftrightarrow\}$-interpolation for these logics." Something else to take into account is that in first order logic the formula $x \rightarrow y$ has a special role as a deduction term. The point we want to make is that the interest in $\mathrm{CIP}^{\rightarrow}$ for first order logic depends on quite some particularities of this logic. In general, it may be of less importance.

2.3.19. Remark We note in this connection that also the Beth property depends on the notion of consequence. Similar to the distinction between $\mathrm{CIP}^{\rightarrow}$ and $\mathrm{CIP}^{\models}$, we could have introduced the $\rightarrow$-Beth property and the $\models$-Beth property. We only considered the $\models$-Beth property. The reason for this is twofold. First, there is a good intuition that defining a primitive notion is something that happens *globally* instead of *locally*. Second, [Maksimova, 1992b] shows that for a huge class of logics, the so-called *regular* logics, the $\rightarrow$-Beth property is equivalent to the $\mathrm{CIP}^{\rightarrow}$. Hence, in some disguise, we *do* study the $\rightarrow$-Beth property. $\dashv$

2.3.4 Robinson's consistency property

Around the same time as Craig provided an alternative proof of Beth's theorem by way of his interpolation lemma, another alternative was given by A. Robinson. His proof in [Robinson, 1956] is based on the fact that first order logic has the following property.

2.3.20. Definition [**Robinson's (joint) consistency property**] Consider the languages $\mathcal{L}_1$, $\mathcal{L}_2$, and let $\mathcal{L} = \mathcal{L}_1 \cap \mathcal{L}_2$. Suppose T is a complete theory in $\mathcal{L}$ and $T_1 \supset T$, $T_2 \supset T$ are consistent theories in $\mathcal{L}_1$, $\mathcal{L}_2$, resp. Then $T_1 \cup T_2$ is consistent in the language $\mathcal{L}_1 \cup \mathcal{L}_2$. $\dashv$

We will not derive Beth's theorem from the joint consistency theorem for first order logic. Instead, it will be shown in Theorem 2.3.21 that the latter theorem is equivalent to Craig's result from which, as we already know, the Beth property of first order logic can be obtained.

Although Craig's result is equivalent to Robinson's theorem, this by no means entails that the two authors cooperated. It even took some time to realize the equivalence of their work. As Craig explains, years later, in a letter to Dauben, [van Ulsen, 2000, page 140]

> Our exchanges about his joint-consistency theorem and my interpolation theorem were not very extended. [...] It took both of us some time, I believe, to realize that the respective theorems that we used in the proof of Beth's results were of intrinsic interest. [...] The equivalence of Robinson's theorem and mine, for first order logic, which only then became an issue, also did not come out until some time after our papers had been published.

Also their proof methods were rather different. Craig's proof combines semantic and syntactic elements, whereas Robinson's proof is purely along model-theoretic lines. Especially his use of *diagrams* (i.e., complete descriptions in terms of (negated) atomic sentences) has since then often been copied.

Let us give sufficient conditions under which the interpolation property and the consistency property are mutually derivable.

2.3.21. Theorem (Consistency property $\Leftrightarrow$ CIP$^{\models}$) *Let S be a compact logic such that the set of S-formulas is closed under the Boolean operations. Then S has* CIP$^{\models}$ *iff S has Robinson's consistency property.*

In view of the important role of the consistency property in the algebraic treatment of interpolation (cf. section 2.6), we include a proof of Theorem 2.3.21.

Proof: Let S be a compact logic, closed under the Boolean operations.

First, assume S has CIP$^{\models}$. Let $\mathcal{L}, \mathcal{L}_1, \mathcal{L}_2, T, T_1, T_2$ be as in Definition 2.3.20 and suppose $T_1 \cup T_2$ is not consistent. We will derive a contradiction. We have that $T_1, T_2 \models \bot$. As S is compact and conjunctive, we may assume T_1, T_2 to be single S-formulas. Then $\neg T_2$ is also an S-formula, as S is closed under negation. Summarizing, we have S-formulas $T_1, \neg T_2$ such that $T_1 \models \neg T_2$. By CIP$^{\models}$, there exists some formula $\vartheta \in \mathcal{L}_1 \cap \mathcal{L}_2$ such that

$$T_1 \models \vartheta, \tag{2.15}$$
$$\vartheta \models \neg T_2. \tag{2.16}$$

Recall that T_1 is consistent. Therefore, (2.15) implies that $T_1 \cap T_2 \not\models \neg\vartheta$. Since $T_1 \cap T_2$ is a complete theory in $\mathcal{L}_1 \cap \mathcal{L}_2$, then $T_1 \cap T_2 \models \vartheta$. From (2.16) we conclude that T_2 is inconsistent. A contradiction.

For the other direction, assume S has Robinson's consistency property. Let φ be a sentence in the language $\mathcal{L}_\varphi$, ψ a sentence in the language $\mathcal{L}_\psi$, and let $\mathcal{L} = \mathcal{L}_\varphi \cap \mathcal{L}_\psi$.

The case where φ, ψ contain free variables requires some minor adaptations and is left to the reader. Suppose $\varphi \models \psi$. We need to find an interpolant. To this end, set $\Theta = \{\vartheta \in \mathcal{L} \,:\, \varphi \models \vartheta\}$. The aim is to show that

$$\Theta \models \psi. \tag{2.17}$$

For then, by compactness, ψ is implied by some finite $\Theta_0 \subseteq \Theta$, whence $\bigwedge_{\vartheta \in \Theta_0} \vartheta$ is the desired interpolant.

To show (2.17), consider a model $\mathcal{M}$ of Θ. Define $Th_{\mathcal{L},\mathcal{M}} = \{\chi \in \mathcal{L} \,:\, \mathcal{M} \models \chi\}$. Suppose $\mathcal{M} \not\models \psi$. Then $Th_{\mathcal{L},\mathcal{M}} \cup \{\neg\psi\}$ is consistent.

We claim that $Th_{\mathcal{L},\mathcal{M}} \cup \{\varphi\}$ is consistent as well. For, suppose it is not. Hence $Th_{\mathcal{L},\mathcal{M}}, \varphi \models \bot$. By the properties of S, then $\varphi \models \neg\vartheta_0$, for some $\vartheta_0 \in Th_{\mathcal{L},\mathcal{M}}$. Then $\neg\vartheta_0 \in \Theta$, whence $\mathcal{M} \models \neg\vartheta_0$. Quod non.

So far, we obtained two consistent theories $Th_{\mathcal{L},\mathcal{M}} \cup \{\neg\psi\}$ and $Th_{\mathcal{L},\mathcal{M}} \cup \{\varphi\}$. Note that their intersection equals $Th_{\mathcal{L},\mathcal{M}}$, which is a complete theory. From the Robinson's consistency property of S we derive the consistency of $Th_{\mathcal{L},\mathcal{M}} \cup \{\varphi\} \cup \{\neg\psi\}$. But $\varphi \models \psi$. A contradiction, which finishes the proof of (2.17). ■

If a given logic does not contain negation, the meaning of the Robinson property becomes inadequate since it may happen that even the set of all sentences has a model. [Ono, 1986] introduces and studies a variant of Robinson's property which is also applicable to logics which are not closed under negation.

As was noted before (on page 23), the infinitary logic L_{ω_1} has the interpolation property. However, the consistency property fails for this logic, cf. [Keisler, 1971]. In this case, the proof of Theorem 2.3.21 does not go through because the compactness theorem is no longer available. As one may verify, L_{ω_1} has a weaker form of consistency property in which the theories T_1, T_2 are assumed to be countable.

We remark that in the context of extended model theory, the Robinson consistency property is much stronger than $\mathrm{CIP}^{\models}$. In fact, as long as the number of symbols in a single sentence is finite, the consistency property yields compactness, [Barwise and Feferman, 1985, page 721]. Hence, such a logic has the Robinson consistency property just in case it satisfies both compactness and $\mathrm{CIP}^{\models}$.

2.3.5 Discussion

Additional motivation for the study of interpolation In one of the preceding subsections we motivated the interest in the interpolation property by referring to its role in proving definability and preservation theorems. Another reason which is usually put forward is the role of the interpolation property as an indication of the (non-)existence of a good proof system. We cite from [Barwise and Feferman, 1985, page 17],

> One can use the interpolation property as a yardstick for measuring whether there is a good proof theory. In case of $\mathcal{L}(Q_1)$[10] knowing that

[10] At present it does not matter what this logic looks like.

> interpolation fails shows that one is not going to have a good Gentzen style proof system for $\mathcal{L}(Q_1)$. What Feferman was after was a richer logic that had a better completeness theorem in this sense, and he was using the interpolation property as a model-theoretic test for such a better theorem.

However, one needs to be careful here. E.g., the modal systems S4.3 and G.3 fail to have interpolation despite having a simple tableau axiomatization, cf. [Rautenberg, 1983].

Other support to study the interpolation property comes from the field of software design. For it turns out that certain important structuring properties of specification languages can be conveniently formalized in terms of interpolation properties. In particular, the modularity property is such. This property is crucial from the standpoint of the formal approach to specification development. Roughly speaking, it guarantees that one may safely specify complex data types by combining the specifications of more primitive types. In fact, it guarantees that under certain conditions the union of two specifications (i.e., theories in the specification language) remains conservative, in a sense that we do not make precise. The point is that this property resembles Robinson's consistency property which guarantees that under certain conditions the union of two theories remains consistent. In view of Theorem 2.3.21 it is not too surprising that the proofs of the modularization theorem involve (some version of) interpolation, cf. [Veloso, 1996]. More information can be found in [Renardel de Lavalette, 1989b].

I would finally like to argue that for logicians the study of interpolation is a useful approach in order to gain a deeper understanding of logics. For, not only is the presence of an interpolation theorem a nice asset. The absence of such theorem is sometimes just as informative. This is best understood on the basis of some examples.

As a matter of fact, first order logic is not algebraizable. This doesn't mean that first order logic can't be studied from an algebraic perspective, but to do so requires some suitable reformulation of the logic. An important objective in algebraic logic is therefore to find "the right" algebraic version of first order logic. A good candidate of such an algebraic variant of first order logic is the system PR_ω. This system is studied in e.g. [Pigozzi, 1972]. There it is shown that $\text{CIP}^\rightarrow$ fails in PR_ω. This induces Pigozzi to conclude that "To get a deductive system that corresponds exactly to the standard first-order logic [...], PR_ω has to be augmented by certain axioms, of a different character than those considered so far." This example illustrates in which respect I regard the study of interpolation a *useful line of approach in the fine-tuning of the design of a formal system.*

The following citation from [Barwise and Feferman, 1985, page 17] supports this view. Here Barwise discusses the failure of interpolation in the logic $\mathcal{L}(Q_1)$. This logic is an expansion of first order logic with the additional quantifier "there exists at least uncountably many".

> The counterexamples that were found to the Craig and Beth theorems for $\mathcal{L}(Q_1)$ and related logics have repeatedly suggested additional concepts that were in the constellation of notions around countability but that were not definable in $\mathcal{L}(Q_1)$. That is, the counterexamples all suggested that we just did not yet have the right logic, rather than that there was an essential obstacle.

Another example where failure of interpolation gives new insight, is presented by the *guarded fragment*. This fragment of first order logic is introduced in [Andréka et al., 1998] where it is conceived as a generalization of modal logic. This claim is supported by the fact that the guarded fragment shares many of the nice features of the basic modal logic K. However, Theorem 4.4.5 shows that the guarded fragment does not have $\mathrm{CIP}^{\rightarrow}$. At first sight, this result seems to undermine the above claim. However, a closer analysis shows that on the contrary we actually underestimated the modal character of this logic. At the risk of being too technical, let us try to explain. A more careful exposition is to be found in chapter 4. Recall that via the standard translation from modal logic into first order logic, formulas of the form $\Box_a p$ translate to $\forall y(R_a xy \to Py)$. So, the relation symbol R_a corresponds to the modality $\Box_a$ and the relation symbol P corresponds to the propositional variable p. Then we have the following difference concerning the notion of 'common language' between multi-modal logics and first order logic. An interpolant for a multi-modal implication $\varphi \to \psi$ may contain any modality (just as it may contain any logical connective). However, an interpolant for the first order translation $\varphi^{tr} \to \psi^{tr}$ may contain only those relation symbols R_a that occur both in φ^{tr} and in ψ^{tr}. This is a much stronger condition. Phrased in modal terms, it states that an interpolant may contain only modalities that occur both in φ and in ψ. Indeed, this property easily fails in multi-modal logics. Just as it fails in the guarded fragment. However, the guarded fragment has a 'modal' interpolation property in which an interpolant for a guarded implication $\varphi \to \psi$ may contain any relation symbol R_a that corresponds to a modality. This analysis confirms the modal character of the guarded fragment: it does not have the stronger interpolation theorem which is indeed also too strong for most multi-modal logics but it does have the interpolation property that is usually studied in modal logic. This is an example where failure of interpolation forces us to take a closer look at the matter.

Interpolation, in as far as it entails the Beth property, is a completeness property in the theory of definition. Hence, failure of interpolation in a formal system S indicates some expressive inadequacies of S. As we have seen, quantificational modal logics present prominent examples of logics without $\mathrm{CIP}^{\rightarrow}$ and without the Beth property. [Areces et al., 2000a] takes up this fact to make a case for *hybrid logics*. For, it is argued, hybrid logic precisely adds the two features needed for an interpolation and definability theorem; not only do the hybrid extensions have these properties, they are somehow the minimal ones. Therefore, the failure of interpolation together with the analysis in [Areces et al., 2000a] sheds light upon the effects of the hybrid machinery.

Failure of interpolation is also interesting in another respect, as it leaves open several ways for reparation. In some cases this paves the way to interesting new directions. Let us see some examples. In [Fitting, 2000] the quantificational modal system S5 (a notorious example of a logic *without* interpolation) is extended with propositional quantifiers. The resulting system turns out to have interpolation. This revives an interest in propositional quantifiers. Anther example has to do with the many-valued logics of Łukasiewicz. [Krzystek and Zachorowski, 1977] shows that these logics do not have the interpolation property. Triggered by this result, [Malinowski and Michalczyk, 1981] defines in a natural way a consequence relation which is determined by the Łukasiewicz matrix, and an implication which corresponds to the original implication in this matrix. This calculus turns out to be much better behaved: not only does it have the interpolation property but it also has the deduction property with respect to this newly defined connective. Yet another example is presented in [Barwise and van Benthem, 1999]. The failure of $\mathrm{CIP}^{\rightarrow}$ in the infinitary logic L_∞ is the starting point for Barwise and van Benthem to investigate an alternative interpolation property for this logic. In analyzing the proof of this interpolation theorem, they develop a method which yields many interpolation and preservation results. This includes a version of the Łos-Tarski theorem for L_∞, and a modal invariance theorem for the same logic.

Being logicians, we would like to gain a deeper understanding of logics. Clearly, knowing that a logic has interpolation is important information. But, as the above examples illustrate, also a negative result on interpolation has its value: it often leads to new and interesting directions of research. Therefore, studying interpolation is a suitable approach to study logical systems.

Proof methods for interpolation To round off this section on interpolation, let us devote a few words on the different proof methods for interpolation that can be found in the literature. The two most common approaches are either of a model-theoretic or of a proof-theoretic nature. Some proofs (like the original proof of Craig) combine semantic and syntactic elements.

A proof of interpolation which uses a Gentzen-style calculus first appears in Schütte's article on intuitionistic predicate logic. This method has also been used for showing interpolation in fragments of IPC, [Renardel de Lavalette, 1989a], infinitary logics, [Lopez-Escobar, 1965], and modal logics, [Czermak, 1975]. The key tool in these proofs is a complete and cut-free sequent calculus, sometimes with some additional properties (as needed e.g., in [Pitts, 1992]). By induction on the structure of the proof, one then derives the existence of an interpolant. The use of tableau calculi is another option, taken e.g., in [Rautenberg, 1983].

A purely model-theoretic argument for interpolation in first order logic has been given in [Henkin, 1963]. This proof centers around the construction of some model having certain desirable properties. The construction of this model is very similar to the construction which is carried out in Henkin's famous completeness proof for first order logic, and has since then often been copied. This kind of proof is especially popular for modal logics, see e.g., [Gabbay, 1972, Smoryński, 1978, Boolos, 1980], but also e.g., for infinitary logics, cf. [Makkai, 1969].

We note that both in the semantic and in the syntactic approach an essential step towards interpolation is the completeness proof. In this respect, a third, algebraic approach might be advantageous as an algebraic completeness result is always available. This method is explained in some detail in section 2.6.

Finally, we balance some of the pros and cons of both methods against each other. For disproving interpolation, the semantic method is the more obvious approach. What about proving interpolation? [Gabbay, 1972] polemizes by stating that "The semantic method is illuminating, since it applies uniformly to many systems." However, in [Rautenberg, 1983] the use of tableau calculi also leads to a uniform proof for a similar number of systems. In favor of the syntactic approach, it has been held that such a proof yields specific information concerning the interpolant (e.g., concerning the occurrences of modalities), and also gives a procedure to construct the interpolant. Against that it can be said that also [Visser, 1996] actually constructs a (uniform) interpolant via a semantic proof that uses Kripke models and a notion of bisimulation that characterizes formulas of a certain bounded modal depth. And e.g., [Sánchez Valencia,] and [Hoogland, 2000] establish semantic proofs for Lyndon interpolation, a typical example where additional information on the interpolant is required. Indeed, the co-existence of such a large number of results obtained from either direction is itself an indication that the scale doesn't tip to either side.

2.4 Table of results

Below, one finds a table of results on the Beth property and the interpolation property $\mathrm{CIP}^{\rightarrow}$ for many of the well-known logics studied in the literature.

The results on CPC follow from the corresponding results on first order logic. Direct proofs have been given above (cf. Theorem 2.2.9 and Theorem 2.3.3.) Failure of interpolation in the many-valued logics of Łukasiewicz has been shown in [Krzystek and Zachorowski, 1977]. In the other cases where a reference is missing, the result follows from Theorem 2.3.10 which states that in many cases the Beth property is implied by $\mathrm{CIP}^{\rightarrow}$.

Logic	**Beth property**	**Craig interpolation**$^{\to}$
CPC	Yes	Yes
CPC$^{\to}$	No Example 2.2.7	Yes [Kreisel and Krivine, 1967]
First order logic	Yes [Beth, 1953]	Yes [Craig, 1957]
Intuitionistic predicate logic	Yes	Yes [Schütte, 1962]
S5 (propositional)	Yes	Yes [Schumm, 1976]
S4.3 (propositional)	Yes [Maksimova, 1993]	No [Maksimova, 1980]
GL (propositional)	Yes	Yes [Smoryński, 1978]
Grz (propositional)	Yes	Yes [Boolos, 1980]
K (first order)	Yes	Yes [Gabbay, 1972]
S4 (first order)	Yes	Yes [Gabbay, 1972]
S5 (first order)	No [Fine, 1979]	No
L^1 (1-variable fragment of f.o.l.)	Yes	Yes [Pigozzi, 1972]
L^k, $1 < k < \omega$ (k-variable fragment of f.o.l.)	No [Andréka et al., 1982]	No [Comer, 1969] [Pigozzi, 1972]
Guarded fragment	Yes Theorem 4.5.1	No Theorem 4.4.5
Infinitary logic L_{ω_1}	Yes	Yes [Lopez-Escobar, 1965]
Infinitary logic L_{α}, $\alpha > \omega_1$ or $\alpha = \infty$	No [Gregory, 1974]	No [Malitz, 1971]
Ł$_n$, $n > 2$ (Łukasiewicz n-valued logic)	No Corollary 3.5.11	No
Relevance logic R	No [Urquhart, 1999] Corollary 3.5.17	No
Entailment logic E	No [Urquhart, 1999] Corollary 3.5.17	No

2.5 Algebraic equivalents of interpolation and definability

The interpolation and definability results from the preceding sections turn out to be closely connected to algebraic theorems to the effect that certain classes of algebras have the amalgamation property or the property of having only surjective epimorphisms. Both properties were already studied in the universal algebraic literature independent of any logical considerations. In the present section we sketch this algebraic background. Subsequently, we discuss the connection between logic and algebra.

2.5.1 Amalgamation properties

The amalgamation property was first considered in [Schreier, 1927], where it was investigated for groups. In a general form, the amalgamation property was first formulated by Fraisse, cf. [Fraïssé, 1954], in connection with certain embedding properties. The strong amalgamation property is introduced in [Jónsson, 1956].

2.5.1. Definition [**(Strong) Amalgamation property ((S)AP)**] Let $\mathbb{K}$, $\mathbb{K}'$ be classes of similar algebras. For $\boldsymbol{A}, \boldsymbol{B_1}, \boldsymbol{B_2} \in \mathbb{K}$, and embeddings $e_1 : \boldsymbol{A} \rightarrowtail \boldsymbol{B_1}$ and $e_2 : \boldsymbol{A} \rightarrowtail \boldsymbol{B_2}$, the quintuple $\langle \boldsymbol{A}; e_1, \boldsymbol{B_1}; e_2, \boldsymbol{B_2} \rangle$ is called a $\mathbb{K}$-*amalgam*. We say that this amalgam can be *amalgamated* in $\mathbb{K}'$ if there exists some $\boldsymbol{C} \in \mathbb{K}'$ and embeddings $f_1 : \boldsymbol{B_1} \rightarrowtail \boldsymbol{C}$, $f_2 : \boldsymbol{B_2} \rightarrowtail \boldsymbol{C}$, such that $f_1 \circ e_1 = f_2 \circ e_2$. That is, if the diagram in Figure 2.5 commutes.

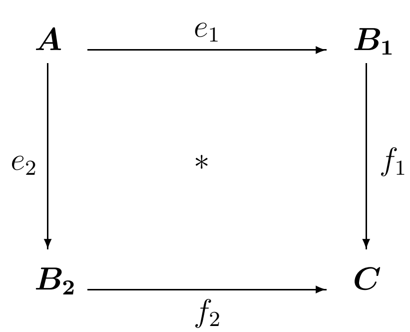


Figure 2.5: The amalgamation property.

This diagram can be *strongly amalgamated* if it can be amalgamated so that $\mathrm{range}(f_1) \cap \mathrm{range}(f_2) \subseteq \mathrm{range}(f_1 \circ e_1)$ (cf. Figure 2.6). A class of similar algebras $\mathbb{K}$ is said to have the *amalgamation property*, or AP for short, if every $\mathbb{K}$-amalgam can be amalgamated in $\mathbb{K}$. $\mathbb{K}$ has the *strong amalgamation property* (or SAP) if every $\mathbb{K}$-amalgam can be strongly amalgamated in $\mathbb{K}$. ⊣

Consider the diagram in Figure 2.5. Clearly, $\mathrm{range}(f_1 \circ e_1) \subseteq \mathrm{range}(f_1)$. Moreover, as $f_1 \circ e_1 = f_2 \circ e_2$, also $\mathrm{range}(f_1 \circ e_1) = \mathrm{range}(f_2 \circ e_2) \subseteq \mathrm{range}(f_2)$. That is,

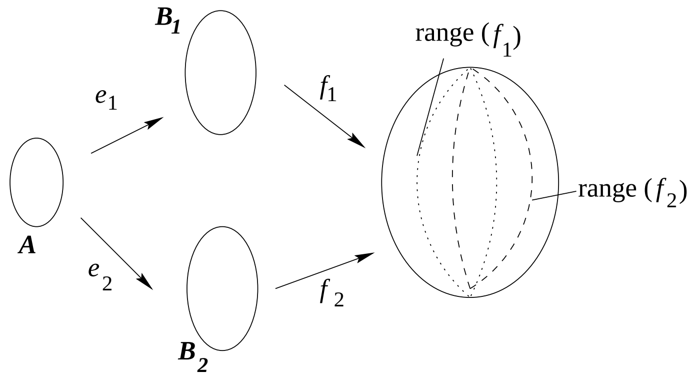


Figure 2.6: The strong amalgamation property.

range$(f_1 \circ e_1) \subseteq$ range$(f_1) \cap$ range(f_2). We conclude that in case the diagram is *strongly* amalgamated, then in fact range$(f_1 \circ e_1) =$ range$(f_1) \cap$ range(f_2).

2.5.2. Example [Jónsson, 1956] The class $\mathbb{L}$ of all lattices has the strong amalgamation property. To see this, consider an $\mathbb{L}$-amalgam $\langle \boldsymbol{A}; e_1, \boldsymbol{B_1}; e_2, \boldsymbol{B_2} \rangle$. We can assume that $\boldsymbol{A}$ is a sublattice of $\boldsymbol{B_1}$, $\boldsymbol{B_2}$ and $A = B_1 \cap B_2$.

We define a partial order on the set $P = B_1 \cup B_2$ as follows. For $i = 1, 2$ and $b, b' \in B_i$ we set $b \leq^P b'$ iff $b \leq^{B_i} b'$. For $b \in B_i, b' \in B_j$, $i \neq j$, let $b \leq^P b'$ iff there exists some $a \in A$ such that $b \leq^{B_i} a \leq^{B_j} b'$. One readily verifies that $\langle P, \leq^P \rangle$ is a poset. As usual, we can turn this poset into a partial lattice $\boldsymbol{P} = \langle P, \wedge^P, \vee^P \rangle$ by setting $p \wedge^P p' = \inf\{p, p'\}$ in case $\inf\{p, p'\}$ exists, and $p \wedge^P p'$ is undefined otherwise. Similar for $p \vee^P p'$. Note that $\boldsymbol{A}, \boldsymbol{B_1}, \boldsymbol{B_2}$ are sublattices of $\boldsymbol{P}$. The result now follows from the well-known theorem that every partial lattice can be embedded in a lattice. $\dashv$

A good starting point for a discussion on the amalgamation property is the following quotation from [Jónsson, 1965] in which Jónsson motivates his interest in the amalgamation property as follows.

> Model theory has [...] demonstrated that investigations in classes of structures of a specified kind, e.g., of equational classes of algebras, are worthwhile. It is only natural that one should there be primarily concerned with elementary classes or at least with L-classes for some language L, and that the axioms should be in some sense of a model-theoretic character. However, certain properties of a different nature play a role in these investigations, and it seems reasonable that in some cases it might be profitable to start with such properties as axioms characterizing the classes to be investigated. Of course, only time can tell which properties deserve to be studied in this manner, but there are some that have already shown up in so many different contexts that

> they seem worthy of special attention. This discussion centers around two such properties, the embedding property and the amalgamation property.

A similar motivation can be found in [Comer, 1969], where Comer refers to the amalgamation property as " [...] an extremely useful tool in model-theoretic investigations." As an example of a theorem in which the amalgamation property plays the axiomatic role that Jónsson refers to above, we mention [Jónsson, 1962]. This article generalizes many of the basic theorems concerning algebraic field extensions to classes which, first of all, are axiomatized by universal sentences, and second, have the amalgamation property.

Examples of classes of algebras with the amalgamation property include the classes of (abelian) groups, fields and (distributive) lattices. On the other hand, this property fails in the classes of semi-groups, rings and modular lattices. These examples suggested to Jónsson that "One does not see much hope of finding general results that assert that if an elementary class is characterized by axioms of such and such form, then this class has the amalgamation property." Nearly twenty years later the situation has not much improved. The chapter on amalgamation in the glossary [Kiss et al., 1983] starts by asserting that "This area is among the neglected fields of universal algebra. There seems to exist no general theory or result which would provide deeper information."

An exception is the following theorem which, under special circumstances, reduces the problem of whether or not a class of algebras $\mathbb{K}$ has the amalgamation property to the corresponding problem for a potentially more tractable subclass of $\mathbb{K}$. A closely related result is [Pigozzi, 1972, Theorem 1.2.5], which in fact turns out to be equivalent for a large number of classes. Below, by $Sir(\mathbb{K})$ we denote the class of subdirectly irreducible members of $\mathbb{K}$.

2.5.3. Theorem [Grätzer and Lakser, 1971, Theorem 3] *Let $\mathbb{K}$ be a variety satisfying the congruence extension property, such that $Sir(\mathbb{K})$ is closed under taking subalgebras. Then $\mathbb{K}$ satisfies the amalgamation property iff every $Sir(\mathbb{K})$-amalgam can be amalgamated in $\mathbb{K}$.*

In [Grätzer et al., 1973] this theorem is applied to show that the class of all pseudocomplemented distributive lattices has the amalgamation property. Another corollary is the following.

2.5.4. Corollary [Werner, 1978, page 27] *A discriminator variety has the amalgamation property iff the class of its simple, non-singleton members has the amalgamation property.*

Although the general theory on amalgamation is lean, there are " [..] strong theorems on concrete structures ", as put in [Kiss et al., 1983]. For example, the situation with respect to lattices has been completely described. It turns out that there are precisely three varieties of lattices that satisfy the amalgamation property: trivial lattices, distributive lattices and all lattices. Obtaining this result has not been easy at all. [Jónsson, 1956] establishes the amalgamation property for the class of all lattices, and [Pierce, 1968] does the same for the class of distributive lattices. In the meanwhile, cf. [Jónsson, 1965], it has been asked whether the class of modular lattices has the amalgamation property. When Grätzer, Jónsson and Lasker answer this question negatively in 1973, it is mentioned that " [...] this problem has been around for more than a decade." It took another decade to completely solve the question of amalgamation in varieties of lattices with the final result, mentioned above, in [Day and Jezek, 1984].

Another example of a strong theorem on concrete structures concerns Heyting algebras. [Maksimova, 1979] shows the existence of (exactly) eight varieties of Heyting algebras with the amalgamation property, among which the class of all Heyting algebras (HA). As an example, let us sketch a proof of this last statement. To this end, we use the representation of a Heyting algebra as an algebra of subsets of a partially ordered set.

2.5.5. Example [Day, 1972] For a partially ordered set Z, let $Up(Z)$ denote the set of upward closed subsets of Z. Let

$$P(Z) = \langle Up(Z), \cap, \cup, \rightarrow, \neg, 1\rangle,$$

where $1 = Z$, $X \rightarrow Y = \{z \ : \ \forall y \geq z(y \in X \Rightarrow y \in Y)\}$ and $\neg X = (X \rightarrow \emptyset) \cup (\emptyset \rightarrow X)$. One readily verifies that $P(Z)$ is a Heyting algebra. The set of all prime filters on a Heyting algebra $\boldsymbol{A}$ is denoted by $PF(\boldsymbol{A})$. It is well-known that the homomorphism $em : \boldsymbol{A} \longrightarrow P(PF(\boldsymbol{A}))$ which maps $a \in A$ to the set $\{F \in PF(\boldsymbol{A}) \ : \ a \in F\}$ is an embedding.

Consider the HA-amalgam $\langle \boldsymbol{A}; e_1, \boldsymbol{A_1}; e_2, \boldsymbol{A_2}\rangle$. Define

$$W = \{\langle F_1, F_2\rangle \in PF(\boldsymbol{A}_1) \times PF(\boldsymbol{A}_2) \ : \ \forall a \in A(e_1(a) \in F_1 \Leftrightarrow e_2(a) \in F_2)\}.$$

Let $i \in \{1, 2\}$. The map $f_i : \boldsymbol{A_i} \longrightarrow P(W)$ is defined by $f_i(b) = \{\langle F_1, F_2\rangle \in W \ : \ b \in F_i\}$.

2.5.6. Claim *f_i is a homomorphism.*

2.5.7. Claim *f_i is injective.*

Obviously $f_1 \circ e_1 = f_2 \circ e_2$. Hence, this completes the amalgam.

The only non-trivial step in the proof of Claim 2.5.6 is to show for any $b, b' \in A_i$ that $(f_i(b) \rightarrow f_i(b')) \subseteq f_i(b \rightarrow b')$. This step essentially involves the following lemma. For readers familiar with the notion of bisimulation, we note the similarity of this lemma with the zig-clause in the definition of a bisimulation.

2.5.8. Lemma *If $\langle F_1, F_2\rangle \in W$ and $F_1 \subseteq G_1$, for some $G_1 \in PF(\boldsymbol{A_1})$, then there exists some G_2 such that $\langle G_1, G_2\rangle \in W$ and $F_2 \subseteq G_2$.*

The proof of Claim 2.5.7 uses the *prime ideal theorem* and the following lemma whose proof consists of an (easy) application of that same prime ideal theorem.

2.5.9. Lemma *For all $F_1 \in PF(\boldsymbol{A_1})$ there exists some $F_2 \in PF(\boldsymbol{A_2})$ with $\langle F_1, F_2\rangle \in W$.*

Details can be found in [Maksimova, 1979, Lemma 6 and Lemma 8]. ⊣

The above examples are meant to illustrate that indeed a lot is known about amalgamation on concrete structures.

Finally, another amalgamation property that should be mentioned is the *superamalgamation property* introduced in 1979 by Maksimova. It differs in nature from the amalgamation properties considered so far as its study has a purely logical motivation: it is the algebraic counterpart of CIP$^{\rightarrow}$. This connection will be discussed in subsection 2.6.1. The superamalgamation property only applies to partially ordered algebras. These are structures of the form $\langle \boldsymbol{A}, \leq\rangle$, where $\leq$ is a partial order on the domain of the algebra $\boldsymbol{A}$. This distinguished partial order corresponds, in a sense we will not make precise, to the implication in the corresponding logic.

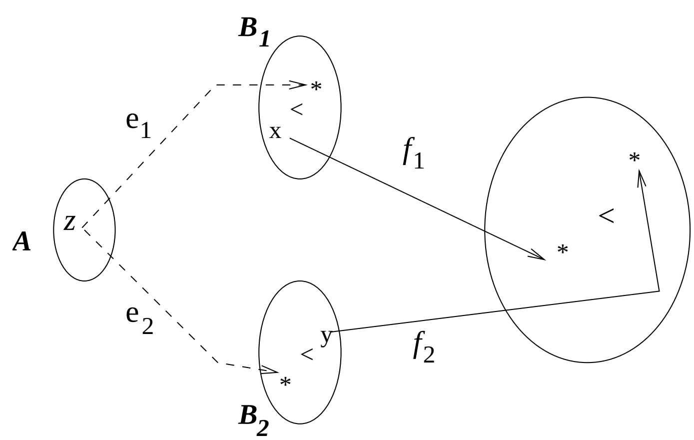


Figure 2.7: The superamalgamation property.

2.5.10. Definition [**Superamalgamation property (SUPAP)**] Let $\mathbb{K}$ be a class of partially ordered algebras. A $\mathbb{K}$-amalgam $\langle \boldsymbol{A}; e_1, \boldsymbol{B_1}; e_2, \boldsymbol{B_2}\rangle$ can be *superamalgamated* if it can be strongly amalgamated via embeddings $f_1 : \boldsymbol{B_1} \rightarrowtail \boldsymbol{C}$, $f_2 : \boldsymbol{B_2} \rightarrowtail \boldsymbol{C}$, such that for $i, j \in \{1, 2\}$,

$$\forall x \in B_i \forall y \in B_j [f_i(x) \leq f_j(y) \Rightarrow \exists z \in A(x \leq e_i(z) \text{ and } e_j(z) \leq y)].$$

See also Figure 2.7. $\mathbb{K}$ is said to have the *superamalgamation property*, or SUPAP for short, if every $\mathbb{K}$-amalgam can be superamalgamated in $\mathbb{K}$. ⊣

In summary, we have introduced three amalgamation properties: the amalgamation property AP, the strong amalgamation property SAP, and the superamalgamation property SUPAP. By definition, SUPAP is stronger than SAP which in its turn implies AP. In subsection 2.5.3 we study the reverse implications. It turns out that SAP is in fact equivalent to the union of AP together with the property of surjectiveness of epimorphisms which we are about to introduce.

2.5.2 Surjectiveness of epimorphisms

A property that is often studied in parallel with amalgamation properties is surjectiveness of epimorphisms. We will see in the next subsection that in the presence of this property the amalgamation properties AP and SAP are equivalent.

2.5.11. Definition [**Epimorphism**] Let $\mathbb{K}$ be a class of algebras, and $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{K}$. A homomorphism $h : \boldsymbol{A} \to \boldsymbol{B}$ is called a $\mathbb{K}$-*epimorphism*, or $\mathbb{K}$-*epi*, if it meets the following condition. For any $\boldsymbol{C} \in \mathbb{K}$ and any pair of homomorphisms $k, k' : \boldsymbol{B} \to \boldsymbol{C}$,

$$\text{if } k \circ h = k' \circ h, \text{ then } k = k'.$$

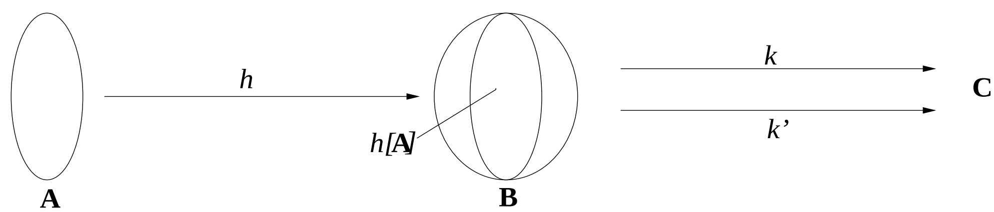


Figure 2.8: An epimorphism.

This is sometimes expressed by saying that h is 'cancelable on the right.' A concept closely related to that of a $\mathbb{K}$-epimorphism is that of a $\mathbb{K}$-*epic subalgebra*. The algebra $\boldsymbol{A} \subseteq \boldsymbol{B}$ is called *epic in* $\mathbb{K}$ (or, a $\mathbb{K}$-epic subalgebra) if the identity from $\boldsymbol{A}$ into $\boldsymbol{B}$ is a $\mathbb{K}$-epimorphism. ⊣

The notion of an epimorphism is the category-theoretic version of the notion of surjectiveness. However, as McLarty puts it in his introduction to category theory [McLarty, 1992], "It is not accurate to think of an epic arrow $h : \boldsymbol{A} \longrightarrow \boldsymbol{B}$ as 'onto', [...]. It is better to think of an epic as 'covering enough of $\boldsymbol{B}$' that any two different arrows out of $\boldsymbol{B}$ must disagree somewhere within the part covered by h."

Obvious examples of epimorphisms are surjections. For any class of algebras $\mathbb{K}$ the question thus arises whether surjections are the only $\mathbb{K}$-epimorphisms. Classes for which this is the case are henceforth said to have the property of *Surjectiveness of Epimorphisms*, or ES for short.

We note that if $\boldsymbol{A} \subsetneq \boldsymbol{B}$ is epic in $\mathbb{K}$, then $id : \boldsymbol{A} \longrightarrow \boldsymbol{B}$ is a non-surjective $\mathbb{K}$-epi. On the other hand, if $h : \boldsymbol{A} \to \boldsymbol{B}$ is a non-surjective $\mathbb{K}$-epimorphism, then $h(\boldsymbol{A})$ is a proper $\mathbb{K}$-epic subalgebra of $\boldsymbol{B}$ (provided $h(\boldsymbol{A}) \in \mathbb{K}$). This shows that if $\mathbb{K}$ is closed under subalgebras or closed under homomorphic images, then $\mathbb{K}$ has ES if and only if no algebra in $\mathbb{K}$ has a proper $\mathbb{K}$-epic subalgebra.

Examples of classes with ES are Boolean algebras, Heyting algebras, (abelian) groups (cf. e.g., [Burgess, 1965]), semilattices and the class of all lattices. As we shall see in section 3.5, classes of algebras with non-surjective epimorphisms include distributive lattices, De Morgan algebras, and Kleene algebras. ES also fails for modular lattices (see [Freese, 1979]), semigroups and fields.

As a simple example, we show that all epimorphisms of the class of abelian groups are surjective.

2.5.12. Example [Burgess, 1965, Theorem 12] We recall from group theory that if N is a normal subgroup of G, then the relation Θ_N defined by $x\Theta_N y \Leftrightarrow x \cdot y^{-1} \in N$ is a congruence relation on G. Moreover, if G is an abelian group then also G/Θ_N is abelian. This quotient algebra is usually denoted by G/N.

Let $f : G_1 \longrightarrow G_2$ be a non-surjective homomorphism between two abelian groups G_1, G_2. We will show that f is not an epimorphism in the class of abelian groups.

As f is non-surjective, $f(G_1)$ is a proper normal subgroup of G_2. Let $g, h : G_2 \longrightarrow G_2/f(G_1)$ be the canonical map and the constant one map respectively. That is, for all $x \in G_2$, $g(x) = x/f(G_1)$, and $h(x) = 1/f(G_1) = f(G_1)$.

Let $x \in f(G_1)$. Then $g(x) = f(G_1) = h(x)$. That is, $g \circ f = h \circ f$. On the other hand, $y \in g(y)$, for all $y \in G_2$. In particular, if $x_0 \in G_2 \setminus f(G_1)$, then $x_0 \in g(x_0)$ but $x_0 \notin f(G_1) = h(x_0)$. That is, $x_0 \in g(x_0) \setminus h(x_0)$, and therefore $g(x_0) \neq h(x_0)$. This shows that if f is non-surjective, then f is not an epimorphism in the class of abelian groups. We conclude that abelian groups have the property ES. $\dashv$

The next example shows that ES fails in the class of distributive lattices. For more examples on failure of ES, we refer to section 3.5.

2.5.13. Example Consider the four-element distributive lattice $\boldsymbol{M_2}$ depicted in Figure 2.9 and its three-element subalgebra $\boldsymbol{A}$ with domain $\{a, c, d\}$. We claim that $\boldsymbol{A}$ is an epic subalgebra of $\boldsymbol{M_2}$ in the category of distributive lattices. To see this, let g, h be homomorphisms from $\boldsymbol{M_2}$ into some distributive lattice $\boldsymbol{C}$ that agree on $\{a, c, d\}$. It needs to be shown that $gb = hb$. We compute as follows. $hb = hb \wedge hd = hb \wedge (ga \vee gb) = hb \wedge (ha \vee gb) \leq (hb \wedge ha) \vee (hb \wedge gb) = hc \vee (hb \wedge gb) \leq (hc \vee hb) \wedge (hc \vee gb) = hb \wedge gb \leq gb$. This implies that $hb \leq gb$. Analogously, one shows that $gb \leq hb$. We conclude that $gb = hb$. Therefore, $id : \boldsymbol{A} \longrightarrow \boldsymbol{M_2}$ is a non-surjective epimorphism in the class of distributive lattices. $\dashv$

The question whether or not a class of algebras has the property ES is in general not easy to answer. For example, the case of rings turned out to be all but

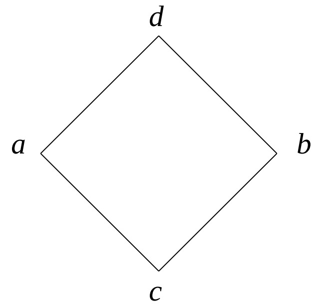


Figure 2.9: The distributive lattice $\boldsymbol{M_2}$.

apparent. In 1965, J. Isbell remarks that indeed, "The answer will be many volumes." In other important cases (e.g., for the class of orthomodular lattices) the answer is still not known. In a series of papers, Isbell investigates the surjectivity problem in a general setting. One of his conclusions in [Isbell, 1973] is that even very strong conditions on a variety V fail to imply that V has the property ES.

One of the difficulties in deciding whether a class has ES or not concerns the sensitivity of this property under either extending or restricting the class of algebras. To see this, consider two classes of algebras $\mathbb{K} \subset \mathbb{K}'$.

It is possible that all $\mathbb{K}$-epimorphisms are surjective, while there does exist a non-surjective $\mathbb{K}'$-epimorphism $h : \boldsymbol{A} \longrightarrow \boldsymbol{B}$, for $\{\boldsymbol{A}, \boldsymbol{B}\} \not\subseteq \mathbb{K}$. For example, take for $\mathbb{K}$ the class of 1-element algebras. Epis are surjective here, but need not be in larger classes. As another example, take for $\mathbb{K}$ the class CA_α of all cylindric algebras of some fixed dimension $\alpha \geq 2$. This class has non-surjective epimorphisms as is shown in [Daigneault, 1964]. However, the subclass of all its locally finite members has ES according to an unpublished result by Andréka, Comer and Németi.

For the other direction, we note that a $\mathbb{K}$-epi need not be a $\mathbb{K}'$-epi. For example, consider the four-element distributive lattice $\boldsymbol{M_2}$ and its three-element subalgebra $\boldsymbol{A}$ defined in Example 2.5.13. We saw that the identity $id : \boldsymbol{A} \longrightarrow \boldsymbol{M_2}$ is an epimorphism in the category of distributive lattices. That is, any two homomorphisms k, k' from $\boldsymbol{M_2}$ into some distributive lattice $\boldsymbol{C}$ are equal if they agree on $\boldsymbol{A}$. However, the fact that the algebra $\boldsymbol{C}$ is distributive is essential in this proof: the map $id : \boldsymbol{A} \longrightarrow \boldsymbol{M_2}$ is *not* an epimorphism in the category of all lattices. In fact, as we will see in the next subsection, the category of all lattices has the property ES. In this case, $\mathbb{K}'$ has ES while $\mathbb{K}$ does not.

Notwithstanding the fact that general results on ES are hard to come by, Comer showed that for discriminator varieties, ES carries over to the whole variety when true for the class of its simple members under the further assumption of the amalgamation property.

2.5.14. Theorem [Comer, 1969] *Let V be a discriminator variety such that the class of its simple members has AP and ES. Then V satisfies ES.*

Comer applies this theorem to show that the class CA_1 of all cylindric algebras of dimension 1 has ES. A somewhat more general, though less perspicuous theorem can be found in [Bruns and Harding, 1999]. Their aim is to show ES for certain varieties of orthomodular lattices, though the question of ES for the class of all orthomodular lattices remains open.

Theorem 2.5.14 can also be applied to the class of Boolean algebras (BA). Recall that the two-element Boolean algebra is the only subdirectly irreducible member of BA. Hence, as BA is a discriminator variety, it is the only simple member of BA. This directly implies that the class of simple Boolean algebras has AP and ES. From Theorem 2.5.14 we conclude that BA has ES. Below, we give a direct proof of this fact. This proof can be adapted to show ES for a large number of (canonical) varieties of Boolean algebras with operators.

2.5.15. Example Below, for a poset L and $l \in L$, the set $\{y \in L \ : \ l \leq y\}$ is denoted by $[l)$.

Consider the Boolean algebras $\boldsymbol{S} \subset \boldsymbol{L}$, and the element $x \in L \setminus S$. To see that $\boldsymbol{S}$ is not epic in BA, we define $W = \{\langle U_1, U_2\rangle \in \mathcal{U}f\boldsymbol{L} \times \mathcal{U}f\boldsymbol{L} \ : \ U_1 \cap S = U_2 \cap S\}$. Let $\boldsymbol{A}$ be the Boolean set algebra with the powerset of W as its domain. Finally, for $i = 1, 2$ define the map e_i from L to A by $e_i(l) = \{\langle U_1, U_2\rangle \in W \ : \ l \in U_i\}$.

From the properties of an ultrafilter it follows that e_i is a Boolean homomorphism from $\boldsymbol{L}$ to $\boldsymbol{A}$, for $i = 1, 2$. Moreover, e_i is an embedding. For, let l, l' be different elements in L. Without loss of generality, suppose $l \not\leq l'$. The set $[l) \cup \{-l'\}$ has the finite intersection property, whence it can be extended to an ultrafilter U on L. Then $\langle U, U\rangle$ in $e_1(l) \setminus e_1(l') \cap e_2(l) \setminus e_2(l')$. In particular, $e_1(l) \neq e_1(l')$ and $e_2(l) \neq e_2(l')$.

It follows directly from the construction that $(e_1)_{\restriction S} = (e_2)_{\restriction S}$. What remains to be shown is that $e_1(x) \neq e_2(x)$. To this end, first note that $(S \cap [x)) \cup (S \cap [-x))$ has the finite intersection property. For, assume it does not. In this case, there exist $s, t \in S$ such that $x \leq s$, $-x \leq t$ and $s \wedge t = 0$. Then $-x \leq t \leq -s$. In other words, $s \leq x$, whence $x = s \in S$. A contradiction. We conclude that there exists some $U \in \mathcal{U}f\boldsymbol{S}$ which extends $(S \cap [x)) \cup (S \cap [-x))$.

Assume that there exists some $u \in U$ such that $u \wedge x = 0$. That is, $x \leq -u$ whence $-u \in S \cap [-x) \subseteq U$, which is not the case. This shows that that $U \cup \{x\}$ has the finite intersection property. Hence we can extend $U \cup \{x\}$ to an ultrafilter U_1 on $\boldsymbol{L}$. Similarly, we extend $U \cup \{-x\}$ to some $U_2 \in \mathcal{U}f\boldsymbol{L}$. Note that $U_1 \cap S = U = U_2 \cap S$. This shows that $\langle U_1, U_2\rangle \in e_1(x) \setminus e_2(x)$. We conclude that $e_1(x) \neq e_2(x)$. $\dashv$

The careful reader may note the similarity between this proof and the proof of the amalgamation property for Heyting algebras in example 2.5.5. This similarity is no coincidence as the properties AP and ES are indeed closely related. In the next subsection we specify this relation.

2.5.3 On the relations between ES, AP and SAP

In general, the properties ES and AP are independent. For example, in section 3.5 we will see that the varieties of Stone algebras, Kleene algebras and De Morgan algebras all fail to have ES, whereas it is known (cf. e.g., [Kiss et al., 1983]) that all three of them have the amalgamation property. Another example is the class of distributive lattices for which it is well-known that ES fails but AP holds. Vice versa, the variety of rings satisfying $R^n = 0$ and the variety of semigroups with 0 satisfying $S^n = 0$ have the property ES but do not have AP, for $n \geq 4$ (cf. [Kiss et al., 1983, page 89]).

On the other hand, as we will see in this subsection, there are interesting connections between the properties ES and AP. The strong amalgamation property turns out to play an important role in these investigations. Most of the results presented here can be found in [Kiss et al., 1983], though mainly without proof. Corollary 2.5.24 was not explicitly stated there. Running ahead of things, we mention that Figure 2.11 contains a schematic overview of the main results of this subsection.

To capture the extent to which SAP is stronger than AP, we introduce the following property.

2.5.16. Definition [**Intersection property of amalgamation (IPA)**] A class of similar algebras $\mathbb{K}$ has the *intersection property of amalgamation (IPA)* if every $\mathbb{K}$-amalgam that can be amalgamated in $\mathbb{K}$, can be strongly amalgamated in $\mathbb{K}$. $\dashv$

2.5.17. Fact Let $\mathbb{K}$ be a class of similar algebras. $\mathbb{K}$ has SAP if and only if $\mathbb{K}$ has AP and $\mathbb{K}$ has IPA.

In [Isbell, 1966] the following terminology is introduced. It provides a clear perspective on the property IPA and its relationship to the property ES. Below, for algebras $\boldsymbol{S}, \boldsymbol{L}$, the notation $\boldsymbol{S} \subseteq \boldsymbol{L}$ (resp. $\boldsymbol{S} \subset \boldsymbol{L}$) expresses that $\boldsymbol{S}$ is a subalgebra (resp. proper subalgebra) of $\boldsymbol{L}$.

2.5.18. Definition [**Absolutely closed, saturated**] Let $\mathbb{K}$ be a class of similar algebras, $\boldsymbol{S}, \boldsymbol{L} \in \mathbb{K}$ and $\boldsymbol{S} \subseteq \boldsymbol{L}$. The *dominion* $\mathrm{Dom}_L(S)$ of $\boldsymbol{S}$ in $\boldsymbol{L}$ is the set of elements l of L with the property that for each pair of homomorphisms $f, g : \boldsymbol{L} \longrightarrow \boldsymbol{A}$ into some $\boldsymbol{A} \in \mathbb{K}$, it is the case that $f{\restriction_S} = g{\restriction_S}$ implies that $f(l) = g(l)$. An algebra $\boldsymbol{S}$ is called *saturated* if $\mathrm{Dom}_L(S) \subset L$, for each $\boldsymbol{L} \supset \boldsymbol{S}$. $\boldsymbol{S}$ is called *absolutely closed* if $\mathrm{Dom}_L(S) = S$, for each $\boldsymbol{L} \supseteq \boldsymbol{S}$. $\dashv$

To say that $\boldsymbol{S}$ is an epic subalgebra of $\boldsymbol{L}$ precisely means that $\mathrm{Dom}_L(S) = L$. In other words, $\boldsymbol{S}$ is a non-surjective $\mathbb{K}$-epic subalgebra iff $\boldsymbol{S}$ is not saturated. This shows the first item in the proposition below. The proofs of the other items are more involved.

2.5.19. Proposition *Let $\mathbb{K}$ be a class of similar algebras.*

1. *$\mathbb{K}$ has ES iff every $\boldsymbol{S} \in \mathbb{K}$ is saturated.*

2. *If $\mathbb{K}$ has IPA, then every $\boldsymbol{S} \in \mathbb{K}$ is absolutely closed.*

3. *If $\mathbb{K}$ is closed under subalgebras and direct products and every $\boldsymbol{S} \in \mathbb{K}$ is absolutely closed, then $\mathbb{K}$ has IPA.*

Proof: To show (2), consider a class of algebras $\mathbb{K}$ with IPA, some $\boldsymbol{S}, \boldsymbol{L} \in \mathbb{K}$ such that $\boldsymbol{S} \subset \boldsymbol{L}$, and some $l \in L \setminus S$. We show that $l \notin \mathrm{Dom}_L(S)$. Obviously, the amalgam $\langle \boldsymbol{S}; id, \boldsymbol{L}; id, \boldsymbol{L} \rangle$ can be amalgamated in $\mathbb{K}$. As $\mathbb{K}$ has IPA, it can be strongly amalgamated. That is, there exists some $\boldsymbol{A} \in \mathbb{K}$ and a pair of embeddings $f, f' : \boldsymbol{L} \rightarrowtail \boldsymbol{A}$ such that $f{\restriction_S} = f'{\restriction_S}$ and $\mathrm{range}(f) \cap \mathrm{range}(f') = f(S)$. As $l \notin S$, it follows that $f(l) \notin f(S) = \mathrm{range}(f) \cap \mathrm{range}(f')$. This implies that $f(l) \notin \mathrm{range}(f')$. In particular, $f(l) \neq f'(l)$. Hence, f, f' witness the fact that $l \notin \mathrm{Dom}_L(S)$. For (3), we consider a $\mathbb{K}$-amalgam $\langle \boldsymbol{A}; e_1, \boldsymbol{B}; e_2, \boldsymbol{C} \rangle$ that can be amalgamated in $\mathbb{K}$. That is, there exists some $\boldsymbol{D} \in \mathbb{K}$ and embeddings $f_1 : \boldsymbol{B} \rightarrowtail \boldsymbol{D}, f_2 : \boldsymbol{C} \rightarrowtail \boldsymbol{D}$, such that $f_1 \circ e_1 = f_2 \circ e_2$. It needs to be shown that the amalgam can be strongly amalgamated in $\mathbb{K}$. Write $X = \mathrm{Dom}_D(f_1(e_1(A)))$. As $f_1(e_1(\boldsymbol{A})) \in \mathrm{S}\mathbb{K} = \mathbb{K}$, we see that $f_1(e_1(\boldsymbol{A}))$ is absolutely closed. Hence, for every $d \in D \setminus X$, there exists some $\boldsymbol{E}_d \in \mathbb{K}$ and a pair of homomorphisms $g_d, g'_d : \boldsymbol{D} \longrightarrow \boldsymbol{E_d}$ such that $(g_d){\restriction_{f_1(e_1(A))}} = (g'_d){\restriction_{f_1(e_1(A))}}$ but $g_d(d) \neq g'_d(d)$. Note that $\prod_{d \in D\setminus X} \boldsymbol{E}_d \in \mathbb{K}$. Define the pair $g, g' : \boldsymbol{D} \longrightarrow \prod_{d \in D \setminus X} \boldsymbol{E}_d$ by $g(x) = \langle g_d(x) \; : \; d \in D \setminus X \rangle$ and $g'(x) = \langle g'_d(x) \; : \; d \in D \setminus X \rangle$. Then $g \circ f_1 : \boldsymbol{B} \longrightarrow \prod_{d \in D \setminus X} \boldsymbol{E}_d$ and $g' \circ f_2 : \boldsymbol{C} \longrightarrow \prod_{d \in D \setminus X} \boldsymbol{E}_d$ strongly amalgamate the amalgam $\langle \boldsymbol{A}; e_1, \boldsymbol{B}; e_2, \boldsymbol{C} \rangle$ in $\mathbb{K}$. Note that in case $D \setminus X = \emptyset$, then $D = \mathrm{Dom}_D(f_1(e_1(A))) = f_1(e_1(A))$. The last equality holds as $f_1(e_1(A))$ is absolutely closed. Hence, $D = \mathrm{range}(f_1 \circ e_1) = \mathrm{range}(f_2 \circ e_2)$ and we conclude that $\mathrm{range}(f_1) \cap \mathrm{range}(f_2) = \mathrm{range}(f_1 \circ e_1)$. Therefore, in this case the amalgam $\langle \boldsymbol{A}; e_1, \boldsymbol{B}; e_2, \boldsymbol{C} \rangle$ is in fact strongly amalgamated by $\boldsymbol{D}, f_1, f_2$. ■

Note that if an algebra is absolutely closed, then it is certainly saturated. Hence, from Proposition 2.5.19.1-2 we obtain the following.

2.5.20. Corollary (IPA ⇒ ES) *If a class of algebras $\mathbb{K}$ has IPA, then $\mathbb{K}$ has ES.*

Clearly, SAP is stronger than IPA (cf. Fact 2.5.17). Hence the above corollary implies that SAP is stronger than ES. We can apply this to the class $\mathbb{L}$ of all lattices. In Example 2.5.2 we saw that $\mathbb{L}$ has the strong amalgamation property. It follows that all $\mathbb{L}$-epimorphisms are surjective.

[Kiss et al., 1983] shows that, under certain conditions on $\mathbb{K}$, ES together with AP implies IPA. For us it is interesting to note that instead of ES, the weaker property ES_1 suffices for this implication. This property plays an important role in chapter 3 where it is shown to be the algebraic counterpart of the Beth property.

2.5.21. Definition [**Surjectiveness of almost onto epis** (ES_1)] A homomorphism $h : \boldsymbol{A} \longrightarrow \boldsymbol{B}$ is said to be *almost onto* if there exists some $b \in B$ such that B is generated by $h[A] \cup \{b\}$. A class of algebras $\mathbb{K}$ has the property ES_1 if every almost onto $\mathbb{K}$-epimorphism is surjective. $\dashv$

Note that ES is indeed a stronger property than ES_1, as we claimed.

2.5.22. Theorem *Let $\mathbb{K}$ be a class of algebras closed under subalgebras and direct products. If $\mathbb{K}$ has both AP and* ES_1*, then $\mathbb{K}$ has IPA.*

Proof: Let $\mathbb{K}$ be as in the theorem and let $\boldsymbol{S} \in \mathbb{K}$. By the third clause of Proposition 2.5.19 it suffices to show that S is absolutely closed. To this end, consider some $\boldsymbol{L} \supset \boldsymbol{S}$ and some $l \in L \backslash S$. We need to find a pair of homomorphisms f, g from $\boldsymbol{L}$ into some $\mathbb{K}$-algebra such that $f_{\restriction S} = g_{\restriction S}$ but $f(l) \neq g(l)$.

First, consider the subalgebra $\boldsymbol{L}' \subseteq \boldsymbol{L}$ which is generated by $S \cup \{l\}$. Note that $\boldsymbol{L}' \in \mathbb{K}$. As $\mathbb{K}$ has the property ES_1, there exists some $\boldsymbol{A} \in \mathbb{K}$ and a pair of homomorphisms $f, f' : \boldsymbol{L}' \longrightarrow \boldsymbol{A}$ such that $f_{\restriction S} = f'_{\restriction S}$ but $f(l) \neq f'(l)$. Without loss of generality, we may assume f, f' to be injective. (If the reader is not convinced, he may copy the argument below after replacing $\boldsymbol{A}$ by $\boldsymbol{L}' \times \boldsymbol{A}$, and f, f' by $g, g' : \boldsymbol{L}' \longrightarrow \boldsymbol{L}' \times \boldsymbol{A}$ which map any $x \in L'$ to $\langle x, f(x)\rangle$, resp. $\langle x, f'(x)\rangle$. We will not do this.)

We then consider the $\mathbb{K}$-amalgam $\langle \boldsymbol{L}'; f, \boldsymbol{A}; id, \boldsymbol{L}\rangle$. As $\mathbb{K}$ has the amalgamation property, there exists some $\boldsymbol{B} \in \mathbb{K}$ and a pair of embeddings $h : \boldsymbol{A} \rightarrowtail \boldsymbol{B}$ and $h' : \boldsymbol{L} \rightarrowtail \boldsymbol{B}$ such that $h \circ f = h' \circ id$. Next, consider the $\mathbb{K}$-amalgam $\langle \boldsymbol{L}'; h \circ f', \boldsymbol{B}; h' \circ id, \boldsymbol{B}\rangle$. Again by the amalgamation property, there exists some $\boldsymbol{C} \in \mathbb{K}$ and a pair of embeddings $j : \boldsymbol{B} \rightarrowtail \boldsymbol{C}$ and $j' : \boldsymbol{B} \rightarrowtail \boldsymbol{C}$ such that $j \circ h \circ f' = j' \circ h' \circ id$.

Then $j \circ h', j' \circ h' : \boldsymbol{L} \longrightarrow \boldsymbol{C}$ are embeddings. Moreover, $j(h'(s)) = j(h(f(s))) = j(h(f'(s))) = j'(h'(s))$, for $s \in S$. Finally, as $f(l) \neq f'(l)$, also $j(h'(l)) = j(h(f(l))) \neq j(h(f'(l))) = j'(h'(l))$. This shows that the pair $j \circ h', j' \circ h'$ has all the required properties. ■

The upshot of the above results is the following corollary which specifies the extent to which SAP is stronger than AP. This is the main result of the present subsection.

2.5.23. Corollary *Let $\mathbb{K}$ be a class of algebras closed under subalgebras and direct products. $\mathbb{K}$ has SAP iff $\mathbb{K}$ has both AP and ES.*

Proof: By definition, SAP implies AP. Moreover, by the remark after Corollary 2.5.20, SAP implies ES. The other direction follows from Theorem 2.5.22 and Fact 2.5.17. ■

From Theorem 2.5.22 and Corollary 2.5.20 we obtain sufficient conditions on a class of algebras $\mathbb{K}$ under which the properties ES and ES_1 are equivalent.

2.5.24. Corollary *Let $\mathbb{K}$ be a class of algebras closed under subalgebras and direct products with the amalgamation property. $\mathbb{K}$ has ES_1 iff $\mathbb{K}$ has ES.*

Let us apply the above results to varieties of Heyting algebras. By Theorem 2.2.9, all intermediate logics have the Beth property. From our algebraic characterization given later in chapter 3 it follows that all varieties of Heyting algebras have the property ES_1. We recall from [Maksimova, 1979] that there exists exactly eight varieties of Heyting algebras with the amalgamation property, among which the class of all Heyting algebras. From Corollary 2.5.24 it then follows that all of Maksimova's eight varieties have the property ES. In particular, the class of all Heyting algebras has the property ES. Moreover, by Corollary 2.5.23 we conclude that a variety of Heyting algebras has the amalgamation property if and only if it has the strong amalgamation property.

2.6 Logic versus algebra

Earlier, we motivated the transition from definability and interpolation to epimorphisms and amalgamation by stating that the latter constitute the algebraic counterparts of the former. It is time to amplify this statement.

For a large number of logics S, the consequence relation $\vdash_S$ is closely related to the algebraic consequence relation $\models_{Mod^*(S)}$ of some corresponding class of algebras Mod^*S. For example, CPC is in this way related to the class of Boolean algebras, and IPC to the class of Heyting algebras. In chapter 3 we introduce the framework of abstract algebraic logic in which the exact relation can be identified. For the moment, the important point is that this is not a mere completeness result. There is more to it. The logic S and the corresponding class of algebras Mod^*S are so much alike that meta-logical theorems about S can be obtained via studying Mod^*S. Here one should think of compactness theorems and deduction theorems, but also of theorems concerning interpolation and definability.

2.6.1 Interpolation and amalgamation

The concept of the connection between amalgamation and interpolation originated with Daigneault. [Daigneault, 1964] contains an algebraic version, in the context of polyadic algebras, of Craig's interpolation theorem. The first part of the proof is to establish the amalgamation property for polyadic algebras, from which Daigneault then derives his algebraic interpolation theorem. [Johnson, 1970] generalizes these results. Johnson mentions in his article that: "With these stronger results, Robinson's, Beth's and Craig's theorems follow for Keisler's logic,[11] though we shall defer this to a later paper." However, these announced results seem to have never been published. The work of Daigneault and Johnson is complemented

[11] By Keisler's logic, Johnson refers to an algebraic version of first order logic.

in [Comer, 1969] where sufficient conditions are given for failure of the amalgamation property in the classes of polyadic algebras under consideration. Comer acknowledges that "The amalgamation property is an extremely useful tool in the development of algebraic analogues to logical theorems," but in his article the amalgamation property is not used as such.

[Pigozzi, 1972] contains the first systematic use of the connection between amalgamation and interpolation. [Pigozzi, 1972, Theorem 1.2.8] entails that Craig's theorem for any standard system of first order logic is equivalent to the statement that a certain class of cylindric algebras (the locally finite cylindric algebras of arbitrary dimension, to be precise) has the amalgamation property. Thus a proof of one theorem automatically entails a proof of the other. Pigozzi calls it the main purpose of his paper " [...] to study in detail the mutual implications between these two kinds of results and using these implications to obtain new results of either kind."

The findings of [Pigozzi, 1972] were rather surprising, as can be gathered from the introductory remark

> Recent developments have brought to light unexpected connections between two kinds of results —

From then on, an algebraic study of the interpolation property has been increasingly popular. In [Czelakowski and Pigozzi, 1999], nearly twenty-five years later, Pigozzi writes

> The correlation between interpolation theorems of logic and certain properties of the class of models related to the amalgamation property is well-known.

This has foremost been the virtue of Maksimova. In a series of papers, Maksimova extends the link between the amalgamation property and the interpolation property to all intermediate and normal modal logics and uses it to transfer algebraic results into logical theorems and vice versa. In this context, the distinction between $\mathrm{CIP}^{\rightarrow}$ and $\mathrm{CIP}^{\models}$ shows up. For this reason, Maksimova introduces the superamalgamation property as the algebraic counterpart of $\mathrm{CIP}^{\rightarrow}$ for normal modal logics. [Madarász, 1998] generalizes this result to non-normal modal logics.

The most general result of this kind which encompasses both Maksimova's results and some of the work in [Pigozzi, 1972] can be found in [Czelakowski, 1982]. There it is shown that for a broad class of compact, propositional logics with the deduction property, $\mathrm{CIP}^{\models}$ is equivalent to the amalgamation property of the corresponding class of matrix models.

[Pigozzi, 1972] also initiated investigations on the connection between $\mathrm{CIP}^{\models}$ and amalgamation in a general, universal algebraic context. In such a general context,

$\mathrm{CIP}^{\models}$ ramifies into several interpolation properties with a corresponding ramification of the amalgamation property. For example, $\mathrm{CIP}^{\models}$ turns out to be somewhat weaker than the logical equivalent of the amalgamation property. Even stronger is *Maehara's interpolation property*, introduced in [Maehara and Takeuti, 1961], which is formulated as a modularity property and which is of special interest for the specification theory community. For a compact logic with the local deduction property, these three interpolation properties are equivalent. However, in general they are not. [Bacsich, 1975] and, much later, [Czelakowski and Pigozzi, 1999] can be seen as continuations of this aspect of [Pigozzi, 1972].

Let us sketch briefly how the link between $\mathrm{CIP}^{\models}$ and amalgamation in algebraic logic is established. To this end, we go back to [Daigneault, 1964]. As we mentioned before, this article contains an algebraic version of Craig's interpolation theorem. The proof is carried out for polyadic algebras, but is similar to the logical argument which shows that Robinson's consistency lemma implies Craig's theorem (cf. Theorem 2.3.21). In Daigneault's proof, the role of Robinson's lemma is played by the amalgamation theorem for polyadic algebras (see [Daigneault, 1964, Theorem 2.6]). So, actually there are two links involved. First, the well-known correspondence between Robinson's lemma and Craig's theorem, and second, a link between Robinson's lemma and the amalgamation theorem.

To shed some light on this second connection, let us recall Robinson's lemma. It can be stated as saying that whenever two first order models $\mathcal{M}_1, \mathcal{M}_2$ have elementarily equivalent reducts in the common part of their languages $\mathcal{L}_1$, resp. $\mathcal{L}_2$, there is a third model $\mathcal{M}_3$ over the union of the languages whose reduct in the language $\mathcal{L}_i$ is elementarily equivalent to $\mathcal{M}_i$, for $i = 1, 2$. The important observation is that in the above formulation the notion of 'elementary equivalence' can not be replaced by 'isomorphism'. In other words, it is not the *models* that can be amalgamated, but the *theories of the models* that can be amalgamated. Hence, Robinson's consistency lemma can be seen as a *theory amalgamation property*. Now we are one step away from understanding why in an algebraic logic context the amalgamation property corresponds to the interpolation property. What we need to realize is that, loosely speaking, a logic is 'algebraizable' if the Lindenbaum-Tarski process can be carried out. This process basically implies that models can be built directly from theories. Therefore, it is no surprise that for these logics there is a close correspondence between the amalgamation property for models (i.e., the usual amalgamation property), and the amalgamation property for theories (i.e., Robinson's lemma).

2.6.1. Remark An abstract algebraic theorem of the form "Logic S has property P iff the corresponding class of algebras has a certain algebraic property $alg(P)$" can only be expected to hold for logics which are, to a certain extent, algebraizable. This will be made precise in chapter 3. The point we want to make here is that the most general correspondence which is known to hold between the amalgamation property and $\mathrm{CIP}^{\models}$ also requires the logic to be in addition compact and to have a deduction theorem. The reason for this is apparent from the above analysis in which we argue that the link between AP and $\mathrm{CIP}^{\models}$ is actually divided in two parts: a link between AP and Robinson's consistency property and one between Robinson's property and $\mathrm{CIP}^{\models}$. In order to establish the first link, the logic needs to be algebraizable. However, the condition on compactness and deduction property is required for the second link, as we saw in Theorem 2.3.21. ⊣

As a final remark, we note that Lyndon's interpolation property and the uniform interpolation property have not been studied extensively via algebraic methods. Interesting open problems are to find algebraic equivalents (similar to amalgamation) of these two properties.

2.6.2 Definability and epimorphisms

The essence of the connection between definability and surjectiveness of epimorphisms is captured by the slogan

> Epimorphisms correspond to implicit definitions.
> Surjections correspond to explicit definitions.

In order to explicate these correspondences, let us go back to Example 2.2.7 about the failure of the Beth property in the implicational fragment of CPC. The algebraic counterpart of this logic is the variety generated by the $\rightarrow$-reduct of the two-element Boolean algebra. This algebra is denoted by $\mathbf{2}^{\rightarrow}$ and the resulting variety is sometimes called the class of *Tarski algebras*. Figure 2.10 depicts the Tarski algebra $\boldsymbol{B} = \mathbf{2}^{\rightarrow} \times \mathbf{2}^{\rightarrow}$ and its subalgebra $\boldsymbol{A}$ with domain $B \setminus \{d\}$.

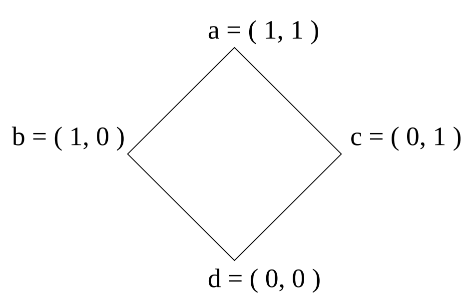


Figure 2.10: The Tarski algebra $\mathbf{2}^{\rightarrow} \times \mathbf{2}^{\rightarrow}$.

Below, we view algebraic terms as logical formulas and vice versa. Similarly, we take the elements in the algebra as propositional variables.

It is important to realize that any subuniverse is closed under the fundamental operations. Hence the fact that $d \notin A$ means that $d \neq t^{B}(a, b, c)$, for all terms $t(a, b, c)$. This can be seen as an algebraic way of saying that d is not explicitly definable in terms of $\{a, b, c\}$. This makes a reasonable case that explicit definability corresponds to surjectivity.

On the other hand, as we saw in Example 2.2.7, the theory of $\boldsymbol{B}$, i.e., the set $Th_B = \{\varphi(a,b,c,d) \;:\; \varphi^B = 1^B\}$ implicitly defines d in terms of A. (Actually, Example 2.2.7 shows that already a subset of Th_B implicitly defines d.) Therefore, any two interpretations of the variables from B in some model of Th_B that agree on the variables from A are equal. This looks very similar to the statement that any two homomorphisms from $\boldsymbol{B}$ into some Tarski algebra which agree on $\boldsymbol{A}$ are equal. This illustrates the similarity between implicit definability and epimorphisms.

A general connection between definability and surjectiveness of epimorphisms first appeared in [Németi, 1984]. One of the main results of this manuscript found its way into [Henkin et al., 1985], see Theorem 5.6.10 therein. There, an algebraic version of the Beth property is formulated and it is shown that under certain conditions a class of algebras $\mathbb{K}$ has this algebraic Beth property if and only if $\mathbb{K}$ has the property ES_1 (cf. Definition 2.5.21). To formulate the Beth property for algebras is quite natural if these algebras are taken as a semantics for a certain logic. Nevertheless, Theorem 5.6.10 is not a transfer result of the kind "Logic S has property P iff the corresponding class of algebras has property $alg(P)$", if only because a general (abstract algebraic) theory is missing which explicitly relates logics to algebras. In [Andréka et al., 1994] such a theory is developed, and a connection between surjectiveness of epis and (some infinite version of) the Beth property is formulated as Conjecture B6. A proof can be found in this thesis, Theorem 3.3.14. The most general result of this kind, is Theorem 3.3.8. It states that an equivalential logic S has the (infinite) Beth property if and only if its class of matrix models has the property ES.

In the meanwhile, the connection between definability and epimorphisms has been employed in [Sain, 1990] to give algebraic proofs for the Beth property in (variants of) first order logic. Other applications can be found in section 3.5 where non-surjective epimorphisms are presented that show the failure of the Beth property in the implicational fragments of CPC, IPC, in the many-valued logics of Łukasiewicz and in a large number of relevance logics.

[Maksimova, 1992b] gives an algebraic property that slightly differs from ES_1 and which is equivalent to the Beth property in varieties of Boolean algebras with operators. She also extensively investigates the Beth property in families of modal logics via the algebraic semantics of these logics, cf. [Maksimova, 1989] and [Maksimova, 1993]. However, this latter work does not explicitly refer to epimorphisms.

An algebraic analogue of the projective Beth property for modal and intermediate logics has been given in [Maksimova, 1999a] and [Maksimova, 1999b]. Theorem 3.4.3 generalizes these results to equivalential logics.

Theorem 3.6.12 is an algebraic characterization of the weak Beth property. This proves [Andréka et al., 1994, Conjecture B10]. Alternative characterizations have been found by Sain, Madarász and Németi, cf. [Sain, 1998].

2.6.3 A schematic overview

In the diagram below, a solid arrow indicates that the depicted relation holds in full generality. If the relation is subject to certain additional conditions, then the arrow is dashed and the conditions are stated in the text below. Absence of an arrow means that no connection, at any level of generality, is known to exist. Again, a counterexample can be found in the text below.

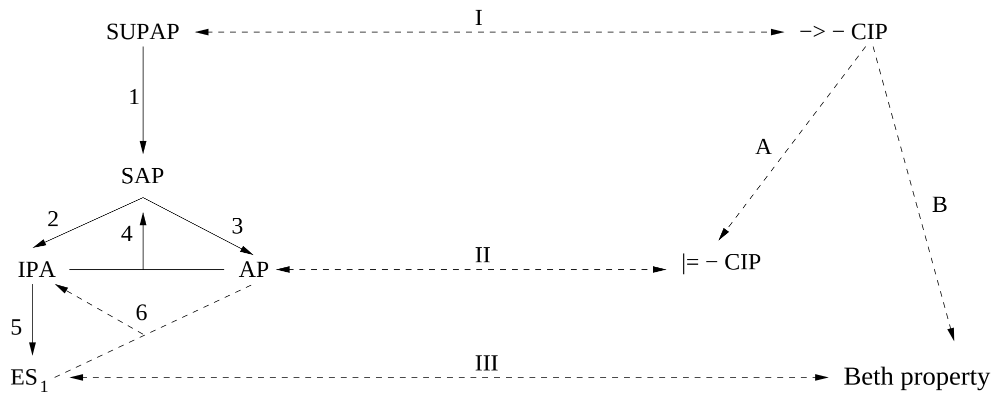


Figure 2.11: Relations between definability, interpolation, amalgamation and ES.

(1)–(4): Follow directly from the definitions of the notions involved.

(5): Corollary 2.5.20.

(6): According to Theorem 2.5.22, any class of algebras $\mathbb{K}$ which is closed under subalgebras and direct products has IPA, if $\mathbb{K}$ has both AP and ES_1.

$\mathrm{ES}_1 \not\Rightarrow$ IPA: The variety of rings satisfying $R^n = 0$ and the variety of semigroups satisfying $S^n = 0$ have ES_1 but do not have IPA for $n \geq 4$, as shown in [Kiss et al., 1983, page 89]. In particular, they do not have AP.

AP $\not\Rightarrow \mathrm{ES}_1$: In section 3.5 we will see that distributive lattices, Stone algebras, Kleene algebras and De Morgan algebras do not have ES_1. As is well-known, they do have AP (see, e.g., [Kiss et al., 1983]).

SAP $\not\Rightarrow$ SUPAP: An example of a Boolean algebra with operators with SAP but without SUPAP can be found in [Maksimova, 1992a, Theorem 1].

For any equivalential logic S, its class of matrix models is denoted by Mod^*S. If S is in fact algebraizable, then Mod^*S is a class of algebras. These concepts will all be explained in full detail in chapter 3.

(I). Let S be an algebraizable logic such that Mod^*S is a variety of normal Boolean algebras with operators or a variety of Boolean algebras with operators (not necessarily normal) with only unary modalities. Then S has $\mathrm{CIP}^{\rightarrow}$ if and only if Mod^*S has SUPAP. This is the main result in [Madarász, 1998] which extends earlier results by Maksimova.

(II). [Czelakowski, 1982] shows that an equivalential, compact and conjunctive logic S has $\mathrm{CIP}^{\models}$ if Mod^*S has the amalgamation property. The reverse implication holds if the logic also has a local deduction theorem. It is claimed in [Andréka and Németi, 1996, Theorem 0.7] that the local deduction property is essential in this direction. In fact, they claim that the tense logic TL^+ defined in Example 2.3.16 has $\mathrm{CIP}^{\models}$, while they show that AP fails in $\mathit{Mod}^*(TL)^+$.

(III). According to Theorem 3.3.8, an equivalential logic S has the Beth property if and only if Mod^*S has ES_1.

(A). For any logic with the local deduction property, $\mathrm{CIP}^{\rightarrow}$ implies $\mathrm{CIP}^{\models}$. A logic without the local deduction property, with $\mathrm{CIP}^{\rightarrow}$ but without $\mathrm{CIP}^{\models}$ is defined in Example 2.3.16.

(B). Any compact, conjunctive logic with the local deduction property and $\mathrm{CIP}^{\rightarrow}$ has the Beth property, cf. Theorem 2.3.10. Example 2.3.16 presents a compact, conjunctive logic without the local deduction property, with $\mathrm{CIP}^{\rightarrow}$ but without the Beth property.

Beth property $\not\Rightarrow$ $\mathrm{CIP}^{\models}$: According to [Maksimova, 1980], in the continuum of normal extensions of the modal logic $\mathsf{S4}$ there are only finitely many with $\mathrm{CIP}^{\models}$. On the other hand, any normal extension of $\mathsf{K4}$, hence in particular any normal extension of $\mathsf{S4}$, has the Beth property, cf. Theorem 2.2.10. Other examples can be found in [Maksimova, 1979] where it is shown that in the continuum of intermediate logics there are exactly eight logics with $\mathrm{CIP}^{\models}$, whereas according to Theorem 2.2.9 they all have the Beth property. We also mention that [Visser, 1998] shows failure of $\mathrm{CIP}^{\models}$ in all interpretability logics between IL and ILM_0. On the other hand, in [Areces et al., 2000b] it is proven that any interpretability logic has the Beth property.

$\mathrm{CIP}^{\models}$ $\not\Rightarrow$ Beth property: [Kreisel and Krivine, 1967] shows that any fragment of CPC has $\mathrm{CIP}^{\models}$. However, the implicational fragment of CPC does not have the Beth property, as we learned in Example 2.2.7. Another example is the modal logic defined in Example 2.2.11. From [Maksimova, 1992a, Theorem 2.b] it follows that this logic has $\mathrm{CIP}^{\models}$. But, as we have seen, it fails to have the Beth property.

Chapter 3

Algebraic equivalents of definability properties

Outline of the chapter

In this chapter we give algebraic equivalents of several Beth definability properties.

In the first section we recall the approach to abstract algebraic logic that can be found in e.g., [Blok and Pigozzi, 1992] and [Blok and Pigozzi, 2001]. This allows us to formulate and prove the main theorem of the chapter, Theorem 3.3.8, which gives an algebraic characterization of the Beth property for equivalential logics. We also explicitly formulate this theorem for the special case of algebraizable logics. Section 3.4 contains a characterization of the projective Beth property. In section 3.5 we supply many applications of our results and give simple examples of non-surjective epimorphisms. In the final two sections we switch to the model-theoretic framework. Section 3.6 contains the necessary preliminaries to give a characterization of the weak Beth property in this framework. This results in Theorem 3.6.12.

3.1 Introduction to abstract algebraic logic

Historically, certain classes of algebras have been studied out of a *logical* interest. E.g., when Boole started the investigations in what is nowadays known as classical propositional logic, he did so by studying a specific class of algebras, the so-called Boolean algebras. Similarly, de Morgan's study of reasoning with relations took an algebraic form. In other cases (e.g., intuitionistic logic, and modal logic) where the systems were originally defined and studied in a logical context, it turned out that algebraic methods and results (on Heyting algebras and modal algebras respectively) could provide useful information.

A natural question to ask is to which logics the methods from algebra can be fruitfully applied. Any answer to this question will first have to settle issues like: what do we mean by a logic? And, how far are we willing to stretch the methods from (universal) algebra? A systematic study of these issues has been undertaken in the field of abstract algebraic logic. Let us take a closer look at some of its findings.

The starting point in abstract algebraic logic is to canonically associate with any logic S a class of algebra-like structures Mod^*S with respect to which the logic is complete. However, it is not just a completeness result we are after. We want S and Mod^*S to be so much alike that as much information as possible about S can be obtained via studying Mod^*S. Here one should think of information like: does S have a deduction property, an interpolation property, is S compact, etc. This likeness of S and Mod^*S is determined by the degree in which the consequence relation of the logic S, i.e., $\vdash_S$, and the semantic consequence relation $\models_{Mod^*(S)}$ are interrelated.

For *algebraizable* logics the relationship between these consequence relations is very close, as they can be recaptured from each other. More precisely, there exists a translation tr_1 from equations to formulas and a translation tr_2 from formulas to equations such that $\models_{Mod^*(S)}$ is interpretable in $\vdash_S$ via tr_1, $\vdash_S$ is interpretable in $\models_{Mod^*(S)}$ via tr_2 and moreover, applying one translation after the other yields an expression which is equivalent to the original. In this case we replace Mod^*S by a quasivariety of algebras, and the associated consequence relation $\models_{Mod^*(S)}$ is the usual notion of equational consequence. As an example, let us consider classical propositional calculus (CPC). Let s, t be terms, and φ a formula. Treating terms as formulas, and vice versa, the translations $tr_1(t \approx s) = t \leftrightarrow s$, and $tr_2(\varphi) = \varphi \approx \top$ have the above properties, where $Mod^*(\mathsf{CPC})$ is the class of Boolean algebras.

Note that these translations make use of the biconditional $\leftrightarrow$ and the constant $\top$. Actually, this turns out to be a characteristic feature of algebraizable logics: roughly speaking, *any* algebraizable logic has two (sets of) connective(s) with the characteristic properties of the biconditional $\leftrightarrow$, respectively the constant $\top$, in CPC.[1] Examples of algebraizable logics include intuitionistic propositional calculus

[1] To suggest that in an algebraizable logic there is a set of connectives with the characteristic

(IPC), normal modal logics (provided the consequence relation is defined *globally*, see the appendix), the many-valued logics of Post and Łukasiewicz, relevance logic R, the implicational logic BCK, and many others.

For some logics the relations $\vdash_S$ and $\models_{Mod^*(S)}$ are not interderivable, as they are for algebraizable systems, but still one of the relations can be interpreted in the other. In other words, these systems only have one of the two connectives $\leftrightarrow$, $\top$. Logics with a $\leftrightarrow$-like connective are known as *equivalential* logics. These logics form the main focus of attention in this chapter. Note that the presence of this connective does not necessarily imply that $True$ is expressible within the logic. As an example, consider any modal logic S with the *local* consequence relation. That is, $\varphi \vdash_S \psi$ is read as: If the formula φ holds in a Kripke model $\mathcal{M}$ at a certain node w, then also ψ is true in $\mathcal{M}$ at node w. (This has been explained in more detail in the appendix.) In this case there is an appropriate translation from equations to formulas by setting e.g., $tr_1(t \approx s) = \{\Box^n(t \leftrightarrow s) : n \in \omega\}$, where for any modal formula φ, $\Box^0\varphi = \varphi$ and $\Box^{n+1}\varphi = \Box\Box^n\varphi$. However, a translation like tr_2 can not be found. (For a proof, see [Blok and Pigozzi, 2001, Example V.3.10].)

The equivalential logics are included in the class of *protoalgebraic* logics. In these logics a weak form of biconditional can be defined by means of *parameters*. The presence of these parameters weakens the link between a protoalgebraic logic S and its corresponding class Mod^*S, as compared to the situation for an equivalential logic, but this link is still tight enough to allow for algebraic methods in the study of these logics. Protoalgebraic logics have been thought of as the largest class of logics to which the methods of (universal) algebra apply. However, work of Font and Jansana provides a framework in which even non-protoalgebraic logics can be the object of an algebraic study (cf. [Font and Jansana, 1996]).

In this chapter we study equivalences between various Beth definability properties and surjectiveness of (certain kinds of) epimorphisms. A link between these two properties was first laid by Németi, cf. [Németi, 1984] and Theorem 5.6.10 in [Henkin et al., 1985]. The main theorem of this chapter is the most general result of this kind to date. It states that an equivalential logic S has the Beth property iff all epimorphisms of the class Mod^*S are surjective. Subsequently it will be shown that the proof of this characterization is generally applicable in proving equivalences, mutatis mutandis, between all kinds of definability properties and surjectiveness of various kinds of epimorphisms. This gives us for example equally general characterizations of the weak Beth property and projective Beth property.

In the literature, two general approaches to the algebraic study of logics can be found. The first approach has its roots in the theory of logical matrices and

properties of the constant $\top$ is a bit misleading. In the case of relevance logic, for example, the formula $\varphi \to \varphi$ cannot be taken as such. It would be more accurate to say that in an algebraizable logic S the filters of the matrices from Mod^*S are equationally definable. Since these filters can somehow be regarded as the set of 'true' states, this explains that in a sense the notion of truth is always definable in an algebraizable logic.

has been developed in a series of papers by Blok and Pigozzi. Also the work of Czelakowski, Font, Jansana and Herrmann, among others, has contributed to this development. The second approach has been initiated by Andréka, Németi and Sain, and shows a strong influence of abstract model theory. The main result of this chapter has been formulated in the framework of Blok and Pigozzi. The reason for this is twofold. First, this approach seems to be more common, see e.g., [Font et al., 2000]. Second, Blok and Pigozzi explicitly discuss how to apply algebraic methods to logics which are not necessarily algebraizable but, say, equivalential. This enables us to give a more general characterization that encompasses this larger class of logics. However, one of the key virtues of the second approach is " [...] its ability to deal explicitly with the usual (model-theoretic or possible-world) semantics of many logics.", [Font and Jansana, 1994, page 22]. This ability is necessary to formulate an algebraic equivalent of the weak Beth property which, as we have seen in subsection 2.2.3, can itself only be defined in a model-theoretic framework. Therefore, when it comes to characterizing the weak Beth property we switch to this different framework.

We assure the reader that all the notions introduced so far will be explained in full detail below. Moreover, the present chapter is completely self-contained. In particular, no prior knowledge of (abstract) algebraic logic is assumed. However, we do assume some familiarity with the basic notions of universal algebra that can be found in any textbook on this subject. A good reference is [Burris and Sankappanavar, 1981].

3.1.1. Notation Matrices are denoted by calligraphic letters $\mathcal{A}$, $\mathcal{B}$ etc., the underlying algebras by the corresponding boldface letters $\boldsymbol{A}, \boldsymbol{B}$, and their universes by A, B, etc. Elements of A^k are sometimes referred to by boldface letters, e.g., $\boldsymbol{a} = \langle a_1, \ldots, a_k \rangle$. If $h : \boldsymbol{A} \longrightarrow \boldsymbol{B}$ is a homomorphism and $\boldsymbol{a} \in A^k$, we write $h(\boldsymbol{a}) = \langle h(a_1), \ldots, h(a_k) \rangle$. For $F \subseteq A^k$, write $h(F) = \{h(\boldsymbol{f}) : \boldsymbol{f} \in F\}$. Let us also mention the following. The *diagonal relation* Δ_A on a set A is the relation $\{\langle a, a \rangle : a \in A\}$. For any map f and subset X of the domain of f, $f_{\lceil X}$ denotes the restriction of f to X, that is, $f_{\lceil X} = \{\langle x, f(x) \rangle : x \in X\}$. For maps $f : A \longrightarrow B$, $g : B \longrightarrow C$, the *composition* $g \circ f : A \longrightarrow C$ is defined for all $x \in A$ by $(g \circ f)(x) = g(f(x))$. For sets $X \subseteq Y$, the map $id : X \longrightarrow Y$ denotes the identity. We use the following operations on classes of similar algebras $\mathbb{K}$: $\mathrm{I}\mathbb{K} = \{\boldsymbol{A} : \boldsymbol{A}$ is isomorphic to some $\boldsymbol{B} \in \mathbb{K}\}$, $\mathrm{S}\mathbb{K} = \{\boldsymbol{A} : \boldsymbol{A}$ is a subalgebra of some $\boldsymbol{B} \in \mathbb{K}\}$, $\mathrm{P}\mathbb{K} = \mathrm{I}\{\prod_{j \in J} \boldsymbol{A}_j : J$ is a set, and $(\forall j \in J)\ \boldsymbol{A}_j \in \mathbb{K}\}$. Note that the class $\mathrm{P}\mathbb{K}$ is closed under taking isomorphisms. We finally point out that we apply set-theoretic notation to classes. For instance, $\bigcup C$ denotes the union of all members of C, also when C is a *proper* class. ⊣

3.2 Logic as a consequence relation

In this section we give an outline of the approach to abstract algebraic logic that is introduced by Blok and Pigozzi, see e.g., [Blok and Pigozzi, 1992] and [Blok and Pigozzi, 2001]. It also serves to fix notation and terminology.

3.2.1 k-deductive systems

We adopt the framework of k-deductive systems which has been developed in [Blok and Pigozzi, 1992] and [Blok and Pigozzi, 2001]. 1-deductive systems are the usual systems of sentential logic. The generalization to higher dimensions can be seen as a general framework which encompasses sentential logics and suitably modified versions of predicate logic, as well as equational logic.

Let us fix a proper class *Var* of sentential variables. It is essential here that we have sets of variables of arbitrary size at our disposal. By a *(sentential) language type* we understand a set $\mathcal{L} = \{\omega_i : i \in I\}$ of connectives of finite rank. Given a set $X \subset \mathit{Var}$, the *formulas of type* $\mathcal{L}$ *over* X are defined recursively in the usual way: every variable in X and every connective in $\mathcal{L}$ of rank 0 is a formula, and if ω is an n-ary connective in $\mathcal{L}$ and $\varphi_1, \dots, \varphi_n$ are formulas of type $\mathcal{L}$ over X, then so is $\omega(\varphi_1, \dots, \varphi_n)$. The set of formulas of type $\mathcal{L}$ over X is denoted by $Fm_{\mathcal{L}}(X)$. As usual, the set $Fm_{\mathcal{L}}(X)$ can be given the structure of an $\mathcal{L}$-algebra by taking the connectives in $\mathcal{L}$ as algebraic operations. The *algebra of* $\mathcal{L}$*-formulas over* X is denoted by $\boldsymbol{Fm}_{\mathcal{L}}(X)$. We omit the type if it is clear from the context. A *substitution* is a homomorphism $\sigma : \boldsymbol{Fm}(X) \longrightarrow \boldsymbol{Fm}(Y)$ between formula algebras.

Let $1 \leq k \leq \omega$. nd let $\mathcal{L}$ be a language type. By a k*-formula over* X we mean a k-tuple of ordinary formulas over X, i.e., a member of $(Fm(X))^k$. To emphasize that we are considering k-formulas rather than formulas, we denote k-formulas in bold Greek type. We write $Fm^k(X)$ for $(Fm(X))^k$. The union of $Fm^k(X)$, over all subsets X of *Var*, is denoted by Fm.

3.2.1. Definition [**k-deductive system**] Let $\mathcal{L}$ be a language type. A *k-deductive system* S (over $\mathcal{L}$) is a system of ordered pairs

$$\{\langle Fm_{\mathcal{L}}(X), Cn_{S(X)}\rangle \ : \ X \text{ is a subset of } \mathit{Var}\},$$

where $Cn_{S(X)} : \mathcal{P}(Fm^k_{\mathcal{L}}(X)) \longrightarrow \mathcal{P}(Fm^k_{\mathcal{L}}(X))$ is a map, satisfying the following four conditions for all sets $X, Y \subset \mathit{Var}$, and $\Gamma, \Delta \subseteq Fm^k_{\mathcal{L}}(X)$:

1. $\Gamma \subseteq Cn_{S(X)}(\Gamma)$.

2. $Cn_{S(X)}(\Gamma) = Cn_{S(X)}Cn_{S(X)}(\Gamma)$.

3. If $\Gamma \subseteq \Delta$, then $Cn_{S(X)}(\Gamma) \subseteq Cn_{S(X)}(\Delta)$.

4. $\sigma Cn_{S(X)}(\Gamma) \subseteq Cn_{S(Y)}\sigma(\Gamma)$, for every $\sigma : \boldsymbol{Fm}_{\mathcal{L}}(X) \longrightarrow \boldsymbol{Fm}_{\mathcal{L}}(Y)$.

The component $\langle Fm_{\mathcal{L}}(X), Cn_{S(X)}\rangle$ of S which is determined by the set X is denoted by $S(X)$. The map $Cn_{S(X)}$ is called a *consequence relation* on $Fm^k_{\mathcal{L}}(X)$. ⊣

By conditions 3.2.1.1–3, $Cn_{S(X)}$ is a closure operator on $Fm^k_{\mathcal{L}}(X)$. Condition 4 is called *substitution invariance* (also the name *structurality* appears in the literature). This condition in particular implies that for sets of variables $X \subseteq Y$, the component $S(Y)$ is *conservative* over $S(X)$. That is, for $\Gamma \subseteq Fm^k_{\mathcal{L}}(X)$, $Cn_{S(X)}(\Gamma) = Cn_{S(Y)}(\Gamma) \cap Fm^k_{\mathcal{L}}(X)$. To see this, let $\sigma : \boldsymbol{Fm}_{\mathcal{L}}(Y) \longrightarrow \boldsymbol{Fm}_{\mathcal{L}}(X)$ be a substitution that fixes the variables in X, and let $\boldsymbol{\varphi} \in Fm^k_{\mathcal{L}}(X)$. If $\boldsymbol{\varphi} \in Cn_{S(Y)}(\Gamma)$, then $\sigma(\boldsymbol{\varphi}) \in \sigma Cn_{S(Y)}(\Gamma)$, whence, by substitution invariance, $\boldsymbol{\varphi} = \sigma\boldsymbol{\varphi} \in Cn_{S(X)}\sigma(\Gamma) = Cn_{S(X)}(\Gamma)$. Vice versa, if $\boldsymbol{\varphi} \in Cn_{S(X)}(\Gamma)$, then by substitution invariance $\boldsymbol{\varphi} \in Cn_{S(Y)}(\Gamma)$, as the identity from $Fm^k_{\mathcal{L}}(X)$ to $Fm^k_{\mathcal{L}}(Y)$ is a substitution.

For $\Gamma \subseteq Fm^k_{\mathcal{L}}$, let $var(\Gamma)$ denote the set of sentential variables occurring in formulas in Γ. We write $Cn_S(\Gamma)$ for $Cn_{S(var(\Gamma))}(\Gamma)$.

All the familiar sentential logics can be formalized as 1-deductive systems. The most important example of a 2-deductive system is (quasi)-equational logic, where a 2-formula $\langle\varphi, \psi\rangle$ is to be interpreted as the equation $\varphi \approx \psi$. Other examples of 2-deductive systems naturally occur when we think of the 2-formula $\langle\varphi, \psi\rangle$ as $\varphi \leq \psi$. Also, essential use of k-deductive systems is made in the treatment of Gentzen systems; here $\langle\varphi, \psi\rangle$ stands for the sequent $\varphi \vdash \psi$. For these and other examples the reader is referred to [Blok and Pigozzi, 1992], [Blok and Pigozzi, 2001] or [Czelakowski and Pigozzi, 1999]. In this chapter, the terms 'deductive system' and 'logic' are used interchangeably.

Often, a k-deductive system is defined as a family of pairs $\langle Fm_{\mathcal{L}}(X), \vdash_{S(X)}\rangle$, where the relation $\vdash_{S(X)} \subseteq \mathcal{P}(Fm^k_{\mathcal{L}}(X)) \times Fm^k_{\mathcal{L}}(X)$ is given via a set of axioms and inference rules. However, this relation can naturally be defined in terms of the consequence relation as follows. Let $Cn_{S(X)}$ be a consequence relation on $Fm^k_{\mathcal{L}}(X)$. The consequence relation $\vdash_{S(X)} \subseteq \mathcal{P}(Fm^k_{\mathcal{L}}(X)) \times Fm^k_{\mathcal{L}}(X)$ is defined by the condition that, for all $\Gamma \cup \{\boldsymbol{\varphi}\} \subseteq Fm^k_{\mathcal{L}}(X)$,

$$\Gamma \vdash_{S(X)} \boldsymbol{\varphi} \text{ iff } \boldsymbol{\varphi} \in Cn_{S(X)}(\Gamma).$$

We write $\Gamma \vdash_S \boldsymbol{\varphi}$ if $\boldsymbol{\varphi} \in Cn_{S(X)}(\Gamma)$, for some set $X \subseteq Var$ that contains all variables occurring in $\{\boldsymbol{\varphi}\} \cup \Gamma$. By substitution invariance, this is equivalent to $\boldsymbol{\varphi} \in Cn_{S(X)}(\Gamma)$, for *every* such set X.

3.2.2. Convention Henceforth, $\mathcal{L}$ is a fixed but arbitrary language type and $1 \leq k \leq \omega$. Unless stated otherwise, every deductive system S is assumed to be of type $\mathcal{L}$. ⊣

3.2.2 Basics on matrices

In the next subsection we present a matrix semantics for k-deductive systems. Here we recall (from e.g., [Blok and Pigozzi, 1992]) some basic notions from universal algebra in the context of matrices.

An $\langle \mathcal{L}, k\rangle$-matrix, or simply a k-matrix, is an ordered pair $\mathcal{A} = \langle \boldsymbol{A}, F_A\rangle$, where $\boldsymbol{A}$ is an $\mathcal{L}$-algebra, and $F_A \subseteq A^k$. $\boldsymbol{A}$ is called the *underlying algebra* of $\mathcal{A}$, and F_A its *filter*.

Let $\mathcal{A} = \langle \boldsymbol{A}, F_A\rangle$, $\mathcal{B} = \langle \boldsymbol{B}, F_B\rangle$ be two k-matrices. $\mathcal{A}$ is a *submatrix* of $\mathcal{B}$ if $\boldsymbol{A}$ is a subalgebra of $\boldsymbol{B}$ and $F_A = F_B \cap A^k$. A *matrix homomorphism* h from $\mathcal{A}$ to $\mathcal{B}$ (notation: $h : \mathcal{A} \longrightarrow \mathcal{B}$) is an algebra homomorphism from $\boldsymbol{A}$ to $\boldsymbol{B}$ such that $h(F_A) \subseteq F_B$. If $h(F_A) = F_B$, then h is called a *strict* matrix homomorphism. A matrix homomorphism $h : \mathcal{A} \longrightarrow \mathcal{B}$ is *injective* (notation: $h : \mathcal{A} \rightarrowtail \mathcal{B}$) if h is injective as an algebra homomorphism and $h(F_A) = F_B \cap h(A)^k$; h is *surjective* (notation: $h : \mathcal{A} \twoheadrightarrow \mathcal{B}$) if h is surjective as an algebra homomorphism. A matrix *isomorphism* is a matrix homomorphism which is both injective and surjective. We will often refer to a matrix homomorphism simply as a homomorphism.

Given a matrix $\mathcal{A} = \langle \boldsymbol{A}, F_A\rangle$, and a congruence θ on $\boldsymbol{A}$, the quotient matrix $\mathcal{A}/\theta$ is the pair $\langle \boldsymbol{A}/\theta, F_A/\theta\rangle$, where $\boldsymbol{A}/\theta$ denotes the quotient algebra $\boldsymbol{A}$ factorized by θ, and $F_A/\theta = \{a/\theta : a \in F_A\}$. By the *kernel* of a matrix homomorphism $h : \mathcal{A} \longrightarrow \mathcal{B}$ (notation: $ker(h)$) we mean the relation kernel of h, i.e., the set $\{\langle a, a'\rangle \in A \times A \;:\; h(a) = h(a')\}$.

The following lemma is a matrix version of the well-known homomorphism lemma from universal algebra.

3.2.3. Lemma (Homomorphism lemma) *Let $\mathcal{A}, \mathcal{B}, \mathcal{C}$ be k-matrices, and $g : \mathcal{A} \twoheadrightarrow \mathcal{B}$ a strict, onto matrix homomorphism. For any matrix homomorphism $f : \mathcal{A} \longrightarrow \mathcal{C}$ with $ker(g) \subseteq ker(f)$ there exists a matrix homomorphism $h : \mathcal{B} \longrightarrow \mathcal{C}$ such that $f = h \circ g$.*

Proof: Consider matrices $\mathcal{A} = \langle \boldsymbol{A}, F_A\rangle$, $\mathcal{B} = \langle \boldsymbol{B}, F_B\rangle$, $\mathcal{C} = \langle \boldsymbol{C}, F_C\rangle$, a strict, onto homomorphism $g : \mathcal{A} \twoheadrightarrow \mathcal{B}$ and a homomorphism $f : \mathcal{A} \longrightarrow \mathcal{C}$. Define the map $h : \mathcal{B} \longrightarrow \mathcal{C}$ as follows. For any $b \in B$, set $h(b) = f(a)$, where $a \in A$ is such that $g(a) = b$. As g is onto, such a exists. Moreover, if $g(a) = g(a')$, then $\langle a, a'\rangle \in ker(g) \subseteq ker(f)$. That is, $f(a) = f(a')$, and we conclude that h is well-defined. The reader easily verifies that h is an algebra homomorphism. To see that h is a matrix homomorphism, choose some $\boldsymbol{b} \in F_B$. As g is strict, $\boldsymbol{b} = g(\boldsymbol{a})$, for some $\boldsymbol{a} \in F_A$. Recall that f is a matrix homomorphism, and therefore $h(\boldsymbol{b}) = f(\boldsymbol{a}) \in F_C$, as desired. ■

In Definition 2.5.11, we defined the notion of an algebra epimorphism. This notion can be extended to matrices in the following, obvious, way. Let $\mathbb{K}$ be a class of matrices, and $\mathcal{A}, \mathcal{B} \in \mathbb{K}$. A matrix homomorphism $h : \mathcal{A} \longrightarrow \mathcal{B}$ is a $\mathbb{K}$-*epimorphism* if for any $\mathcal{C} \in \mathbb{K}$ and any pair of matrix homomorphisms k, k' :

$\mathcal{B} \longrightarrow \mathcal{C}$ it is the case that $k \circ h = k' \circ h$ implies that $k = k'$. The matrix $\mathcal{A} \subseteq \mathcal{B}$ is called *epic in* $\mathbb{K}$ (or, a $\mathbb{K}$-epic submatrix) if the identity from $\mathcal{A}$ into $\mathcal{B}$ is a $\mathbb{K}$-epimorphism. Our notion of a matrix epimorphism reduces to the usual notion of an algebra epimorphism if we view the class of algebras $\mathbb{K}$ as the class of matrices $\{\langle \boldsymbol{A}, \Delta_A \rangle \ : \ \boldsymbol{A} \in \mathbb{K}\}$.

3.2.3 Matrix semantics

The study of matrix semantics originated with [Łukasiewicz and Tarski, 1930] and was systematically developed in [Łoś and Suszko, 1958] and [Czelakowski, 1981]. In [Blok and Pigozzi, 1992] a matrix semantics for k-deductive systems has been introduced. This will be recapitulated in the present subsection.

A k-matrix $\mathcal{A}$ defines for every set $X \subset Var$ a *semantic consequence relation* $\models_{\mathcal{A}(X)} \subseteq \mathcal{P}(Fm^k(X)) \times Fm^k(X)$ over $\mathcal{L}$ as follows. For any $\Gamma \cup \{\boldsymbol{\varphi}\} \subseteq Fm^k(X)$,

$$\Gamma \models_{\mathcal{A}(X)} \boldsymbol{\varphi} \overset{\text{def}}{\Leftrightarrow} (\forall h : \boldsymbol{Fm}(X) \longrightarrow \boldsymbol{A})[(h(\Gamma) \subseteq F_A) \Rightarrow (h(\boldsymbol{\varphi}) \in F_A)].$$

For a class of k-matrices $\mathbb{M}$, we write $\Gamma \models_{\mathbb{M}(X)} \boldsymbol{\varphi}$ iff $\Gamma \models_{\mathcal{A}(X)} \boldsymbol{\varphi}$, for all $\mathcal{A} \in \mathbb{M}$. Note that for sets $X, Y \subset Var$ and $\Gamma \cup \{\boldsymbol{\varphi}\} \subseteq Fm^k(X) \cap Fm^k(Y)$, we have $\Gamma \models_{\mathcal{A}(X)} \boldsymbol{\varphi}$ iff $\Gamma \models_{\mathcal{A}(Y)} \boldsymbol{\varphi}$. Therefore, we are allowed to omit the reference to the set of variables and simply write $\Gamma \models_{\mathcal{A}} \boldsymbol{\varphi}$.

For future use, we introduce the following consequence relation between sets of k-formulas and *pairs* of 1-formulas. Given a k-matrix $\mathcal{A}$, a set of variables $X \subset Var$, a set of k-formulas $\Gamma \subseteq Fm^k(X)$ and two 1-formulas $\varphi, \psi \in Fm(X)$, we define

$$\Gamma \models_{\mathcal{A}(X)} \langle \varphi, \psi \rangle \overset{\text{def}}{\Leftrightarrow} (\forall h : \boldsymbol{Fm}(X) \longrightarrow \boldsymbol{A})[(h(\Gamma) \subseteq F_A) \Rightarrow (h(\varphi) = h(\psi))].$$

This relation will be used in our formulation of the Beth property in Definition 3.3.1.

Let S be a k-deductive system. A k-matrix $\mathcal{A}$ is called an *S-matrix* (or a *matrix model for S*) if for any $\Gamma \cup \{\boldsymbol{\varphi}\} \subseteq Fm^k$,

$$\Gamma \vdash_S \boldsymbol{\varphi} \text{ implies } \Gamma \models_{\mathcal{A}} \boldsymbol{\varphi}.$$

For any $\mathcal{L}$-algebra $\boldsymbol{A}$, a set $F \subseteq A^k$ is called an *S-filter* on $\boldsymbol{A}$ if the k-matrix $\langle \boldsymbol{A}, F \rangle$ is an S-matrix. We write $Fi_S(\boldsymbol{A})$ for the set of all S-filters on $\boldsymbol{A}$. Given a k-deductive system S, an *S-theory* is, as usual, a set of k-formulas Γ that is closed under the consequence relation, i.e., $\Gamma = Cn_S(\Gamma)$. Note that $\Gamma = Cn_S(\Gamma)$ if and only if $\Gamma \in Fi_S(\boldsymbol{Fm}(var(\Gamma)))$. Hence, the S-theories are precisely the S-filters on formula algebras.

Note that the homomorphic pre-image of a filter, is again a filter. That is, given a homomorphism $h : \boldsymbol{A} \longrightarrow \boldsymbol{B}$ and a k-deductive system S, $h^{-1}(G) \in Fi_S(\boldsymbol{A})$ if $G \in Fi_S(\boldsymbol{B})$. However, in general, $F \in Fi_S(\boldsymbol{A})$ does not imply $h(F) \in Fi_S(\boldsymbol{B})$. We adopt the notation $h_S(F)$ for the least S-filter on $\boldsymbol{B}$ that contains $h(F)$.

By definition, any k-deductive system S is sound with respect to its class of matrix models. To see that it is also complete, assume $\Gamma \not\vdash_S \boldsymbol{\varphi}$, for some $\Gamma \cup \{\boldsymbol{\varphi}\} \subseteq Fm^k(X)$. Define $\mathcal{A} = \langle \boldsymbol{Fm}(X), Cn_{S(X)}(\Gamma)\rangle$. As $Cn_{S(X)}(\Gamma)$ is an S-theory, it is an S-filter on $\boldsymbol{Fm}(X)$, i.e., $\mathcal{A}$ is an S-matrix. Obviously, $\Gamma \not\models_{\mathcal{A}} \boldsymbol{\varphi}$. Thus every k-deductive system is complete with respect to its class of matrix models. However, this observation is not that significant since the underlying algebras of these matrices do not have any meaningful structure, whence the semantics is completely supplied by the filters. A much more interesting completeness theorem, from our point of view, is one which is more algebraic in nature in the sense that it correlates the theory of the class of underlying algebras with the logical properties of the deductive system. For such a completeness result, which is given by Theorem 3.2.6, we need to introduce some more notions.

Let $\langle \boldsymbol{A}, F_A\rangle$ be a k-matrix, and let Θ be a congruence on $\boldsymbol{A}$. Θ is said to be *compatible* with F_A if for any $\boldsymbol{a}, \boldsymbol{b} \in A^k$ such that $a_i \Theta b_i$, for all $1 \leq i \leq k$, it is the case that $\boldsymbol{a} \in F_A$ iff $\boldsymbol{b} \in F_A$. For any algebra $\boldsymbol{A}$, define the map $\Omega^{\boldsymbol{A}}$ from the powerset of A^k to the congruence lattice of $\boldsymbol{A}$ by setting, for any $X \subseteq A^k$,

$$\Omega^{\boldsymbol{A}}(X) = \bigvee\{\Theta \ : \ \Theta \text{ is a congruence on } \boldsymbol{A} \text{ compatible with } X\ \}.$$

Note that $\Omega^{\boldsymbol{A}}(X)$ is itself a congruence on $\boldsymbol{A}$ compatible with X, and thus is the largest congruence on $\boldsymbol{A}$ with this property. It is called *the Leibniz congruence of* $\boldsymbol{A}$ *over* X. As the next proposition shows, the Leibniz congruence relates those elements of a domain that have exactly the same properties, relative to the language. This is reminiscent of a criterion of equality proposed by Leibniz, and thereby justifies the congruence's name. Below, $\varphi^{\boldsymbol{A}}(a_1, \dots, a_n)$ denotes the interpretation of the formula $\varphi(x_1, \dots, x_n)$ in the algebra $\boldsymbol{A}$ when for all $i \leq n$ the variable x_i is interpreted as $a_i \in A$.

3.2.4. Proposition *Let $\langle \boldsymbol{A}, F_A\rangle$ be a k-matrix, $a, b \in A$. Then $\langle a, b\rangle \in \Omega^{\boldsymbol{A}}(F_A)$ iff for any formula $\varphi(x, y_1, \dots, y_n)$ and any $\boldsymbol{c} \in A^n$, $[\varphi^{\boldsymbol{A}}(a, \boldsymbol{c}) \in F_A \Leftrightarrow \varphi^{\boldsymbol{A}}(b, \boldsymbol{c}) \in F_A]$.*

Proof: The direction from left to right is immediate, as $\Omega^{\boldsymbol{A}}(F_A)$ is a congruence compatible with F_A. For the other direction, note that the set $\{\langle a, b\rangle \in A^2 : \boldsymbol{\varphi}^{\boldsymbol{A}}(a, \boldsymbol{c}) \in F_A \Leftrightarrow \boldsymbol{\varphi}^{\boldsymbol{A}}(b, \boldsymbol{c}) \in F_A$, for any formula $\varphi(x, y_1, \dots, y_n)$ and any $\boldsymbol{c} \in A^n\}$ is a congruence on $\boldsymbol{A}$ compatible with F_A. ■

A k-matrix $\langle \boldsymbol{A}, F_A\rangle$ with the property that $\Omega^{\boldsymbol{A}}(F_A) = \Delta_A$ is called *reduced*. Reduced matrices play a central role in the main theorem of this chapter. Let us therefore take a closer look at them. Given a k-matrix $\mathcal{A} = \langle \boldsymbol{A}, F\rangle$, we define $\mathcal{A}^* = \mathcal{A}/\Omega^{\boldsymbol{A}}(F)$. That is, $\mathcal{A}^* = \langle \boldsymbol{A}^*, F^*\rangle$, where $\boldsymbol{A}^* = \boldsymbol{A}/\Omega^{\boldsymbol{A}}(F)$ and $F^* = F/\Omega^{\boldsymbol{A}}(F)$. The reduced matrices are, up to isomorphism, precisely the ones of the form $\mathcal{A}^*$.

One readily verifies that for every $\Gamma \cup \{\varphi\} \subseteq Fm^k$, and all k-matrices $\mathcal{A}$,

$$\Gamma \models_{\mathcal{A}} \varphi \Leftrightarrow \Gamma \models_{\mathcal{A}^*} \varphi. \tag{3.1}$$

For any k-deductive system S, the class of reduced S-matrices plays an important role in the algebraic study of S. We therefore introduce the following notation.

3.2.5. Notation Let S be a k-deductive system. The class of reduced S-matrices is denoted by Mod^*S.

We observed above that every k-deductive system is complete with respect to its class of matrix models. Together with (3.1) this implies that every k-deductive system is complete with respect to its class of *reduced* matrix models. This theorem has first been proven as Theorem 5.6 in [Blok and Pigozzi, 1992].

3.2.6. Theorem (Reduced Completeness Theorem) *Let S be a k-deductive system, $\Gamma \cup \{\varphi\} \subseteq Fm^k$. Then*

$$\Gamma \vdash_S \varphi \Leftrightarrow \Gamma \models_{Mod^*S} \varphi.$$

In [Blok and Pigozzi, 2001] it is shown that this theorem reduces to the well-known algebraic completeness theorems for e.g., the systems of classical and intuitionistic propositional logic, the modal logic K, and the system BCK. Moreover, this class of reduced S-matrices is the algebraic counterpart of the system S in the sense that metalogical questions concerning S can be studied algebraically by investigating Mod^*S. For example, the main result in [Blok and Pigozzi, 2001] states that, under some mild conditions, a deductive system S has an abstract version of the deduction theorem if and only if Mod^*S has a well-known universal algebraic property, viz., that of having equationally definable principal relative congruences. Running ahead of things, we note that it is also with respect to reduced S-matrices that surjectiveness of epimorphisms is related to the Beth property.

3.2.4 Protoalgebraic and equivalential deductive systems

Protoalgebraic logics, introduced in [Blok and Pigozzi, 1986], can be thought of as k-deductive systems with a very weak implication.

3.2.7. Definition [**Protoalgebraic logic**] A k-deductive system S is called *protoalgebraic* if there exists a (possibly infinite) set of k-formulas $\Delta(p, q)$ in at most the two 1-variables p, q such that

$$\begin{array}{ll} \text{(R)} & \vdash_S \Delta(p,p) \\ \text{(MP)} & p, \Delta(p,q) \vdash_S q \end{array}$$

$\dashv$

Note that every expansion and every extension of a protoalgebraic k-deductive system is itself protoalgebraic.

In this chapter we will mostly study an important subclass of protoalgebraic systems, the so-called *equivalential* deductive systems. Equivalential 1-deductive systems were introduced in [Prucnal and Wroński, 1974] and have been extensively investigated in [Czelakowski, 1981].

3.2.8. Definition [**Set of equivalence formulas, Equivalential logic**] A (possibly infinite) set of k-formulas $\Delta(p, q)$ in at most the two 1-variables p, q is a *set of equivalence formulas* for a k-deductive system S if (R) and (MP) hold, and moreover for every n-ary function symbol $\omega \in \mathcal{L}$,

$$\text{(Con)}\ \Delta(p_1, q_1), \ldots, \Delta(p_n, q_n) \vdash_S \Delta(\omega(p_1, \ldots, p_n), \omega(q_1, \ldots, q_n)).$$

A k-deductive system S is called *equivalential* if S has a set of equivalence formulas. $\dashv$

By definition, every equivalential deductive system is protoalgebraic. The other inclusion does not hold. An example of a protoalgebraic system which is not equivalential is the weakest classical modal logic E. This 1-deductive system has all classical tautologies as axioms and is defined by the rules of modus ponens, substitution and the extensionality rule

$$\frac{\vdash_{\mathsf{E}} \varphi \leftrightarrow \psi}{\vdash_{\mathsf{E}} \Box\varphi \leftrightarrow \Box\psi}.$$

Since E is an expansion of CPC, the logic E is protoalgebraic. However, E is not equivalential, as has been shown in [Malinowski, 1989]. Note that in E the extensionality rule only applies to the set of E-theorems. The strengthening of E in which the rule $\varphi \leftrightarrow \psi / \Box\varphi \leftrightarrow \Box\psi$ is valid is indeed equivalential and has as a set of equivalence formulas the set $\{p \leftrightarrow q\}$.[2]

It can be shown that for a set of equivalence formulas $\Delta(p, q)$,

$$\Delta(p, q) \vdash_S \Delta(q, p),$$

and

$$\Delta(p, q), \Delta(q, r) \vdash_S \Delta(p, r).$$

This shows that a set of equivalence formulas $\Delta(p, q)$ collectively behaves like an equivalence connective.

[2]Readers familiar with modal logic may wonder whether the same holds for the local normal modal logic K_{loc}, since also in this system an important rule (in this case the necessitation rule) only applies to the theorems of the system. However, contrary to E, the logic K_{loc} is equivalential (see Example 3.2.9). Very briefly, the difference lies in the fact that for any modal formula $\varphi(p)$ there exists some $n \in \omega$ such that $\vdash_{\mathsf{K}_{loc}} \bigwedge_{0 \leq k \leq n} \Box^k (p \leftrightarrow q) \to (\varphi(p) \leftrightarrow \varphi(q))$, while this is not the case in E.

3.2.9. Example The formula $p \leftrightarrow q$ is an equivalence formula for CPC and for IPC. The set $\{p \to q, q \to p\}$ is a set of equivalence formulas for the implicational fragment of classical propositional logic, $\mathsf{CPC}^{\to}$. This set is also a set of equivalence formulas for the logics E of entailment, and R of relevance. These logics will be defined in section 3.5. For modal logics with a local consequence relation (see the appendix), the set $\{\Box^n(p \leftrightarrow q) : n \in \omega\}$ can be taken as such. Or, for extensions of $\mathsf{K4}$, simply $\{p \leftrightarrow q, \Box(p \leftrightarrow q)\}$. For modal logics with a global consequence relation, the formula $p \leftrightarrow q$ will do. Consequently, all these logics are equivalential. ⊣

The next theorem states that the Leibniz congruences are definable *within* a logic S if and only if S is equivalential.

3.2.10. Theorem [Blok and Pigozzi, 1992, Theorem 13.5] *A set of k-formulas $\Delta(p,q)$ is a set of equivalence formulas for a k-deductive system S iff for any $\mathcal{L}$-algebra $\boldsymbol{A}$, any $F \in Fi_S(\boldsymbol{A})$ and $a, b \in A$,*

$$\langle a, b\rangle \in \Omega^{\boldsymbol{A}}(F) \Leftrightarrow \Delta^{\boldsymbol{A}}(a,b) \subseteq F.$$

Proof: We follow a proof suggested by R. Jansana. The direction from right to left is straightforward. The other direction is divided in two parts. Let $\Delta(p,q)$ be a set of equivalence formulas for S. First, assume $\langle a, b\rangle \in \Omega^{\boldsymbol{A}}(F)$ and $\delta(p,q) \in \Delta(p,q)$. Then $(\delta^{\boldsymbol{A}}(a,a), \delta^{\boldsymbol{A}}(a,b)) \in \Omega^{\boldsymbol{A}}(F)$. By (R), $\delta^{\boldsymbol{A}}(a,a) \in F$. Therefore, $\delta^{\boldsymbol{A}}(a,b) \in F$, as desired. Second, assume $\Delta^{\boldsymbol{A}}(a,b) \subseteq F$. Then, using (Con), for any formula $\varphi(x, y_1, \ldots, y_n)$ and any $\boldsymbol{c} \in A^n$, $\Delta(\varphi^{\boldsymbol{A}}(a, \boldsymbol{c}), \varphi^{\boldsymbol{A}}(b, \boldsymbol{c})) \in F$. Hence, for any formula $\varphi(x, \boldsymbol{y})$ and any $\boldsymbol{c} \in A^n$ it is the case that $\varphi^{\boldsymbol{A}}(a, \boldsymbol{c}) \in F$ iff $\varphi^{\boldsymbol{A}}(b, \boldsymbol{c}) \in F$. By Proposition 3.2.4 this implies that $\langle a, b\rangle \in \Omega^{\boldsymbol{A}}(F)$. ■

The next theorem collects all the results about equivalential systems that we will need.

3.2.11. Theorem *For a k-deductive system S, the following are equivalent.*

1. *S is equivalential.*
2. *The class of reduced S-models is closed under submatrices and direct products.*
3. *For any $\mathcal{L}$-algebra $\boldsymbol{A}$, the Leibniz operator $\Omega^{\boldsymbol{A}}$ is monotonic and for any $F_A \in Fi_S(\boldsymbol{A})$ and subalgebra $\boldsymbol{B} \subseteq \boldsymbol{A}$,*

$$\Omega^{\boldsymbol{B}}(F_A \cap B^k) = \Omega^{\boldsymbol{A}}(F_A) \cap B^2.$$

4. *For any $\mathcal{L}$-algebra $\boldsymbol{A}$, the restriction of $\Omega^{\boldsymbol{A}}$ to $Fi_S(\boldsymbol{A})$ is monotonic and commutes with inverse homomorphisms. That is, for any $F \in Fi_S(\boldsymbol{A})$ and any homomorphism $h : \boldsymbol{A} \longrightarrow \boldsymbol{B}$ between $\mathcal{L}$-algebras, $\Omega^{\boldsymbol{A}} h^{-1} h_S(F) = h^{-1}\Omega^{\boldsymbol{B}} h_S(F)$, where $h_S(F)$ denotes the least S-filter on $\boldsymbol{B}$ that contains $h(F)$.*

Proof: For (1) ⇔ (2) see [Blok and Pigozzi, 1992, Theorem 13.12]. The equivalence of (1) and (3) is shown in [Blok and Pigozzi, 1992, Theorem 13.13], and (1) ⇔ (4) has been proven in [Herrmann, 1997, Theorem 4.5]. ■

As an easy corollary we obtain the following. Recall that $h_S(F)$ below denotes the least S-filter on $\boldsymbol{B}$ that contains $h(F)$.

3.2.12. Corollary *Let S be an equivalential k-deductive system, and $h : \boldsymbol{A} \longrightarrow \boldsymbol{B}$ a homomorphism between arbitrary $\boldsymbol{\mathcal{L}}$-algebras. For any $F \in Fi_S(\boldsymbol{A})$, $h\Omega^{\boldsymbol{A}}(F) \subseteq \Omega^{\boldsymbol{B}}(h_S(F))$.*

Proof: Let all data be as above. Since $F \subseteq h^{-1}h_S(F) \in Fi_S(\boldsymbol{A})$, it follows from the monotonicity of the Leibniz operator restricted to S-filters that $\Omega^{\boldsymbol{A}}(F) \subseteq \Omega^{\boldsymbol{A}}h^{-1}h_S(F) = h^{-1}\Omega^{\boldsymbol{B}}h_S(F)$. The latter equality holds as $\Omega^{\boldsymbol{A}}$ restricted to $Fi_S(\boldsymbol{A})$ commutes with inverse homomorphisms. Then $h\Omega^{\boldsymbol{A}}(F) \subseteq hh^{-1}\Omega^{\boldsymbol{B}}h_S(F) \subseteq \Omega^{\boldsymbol{B}}h_S(F)$. ■

For any k-deductive system S, let Alg^*S denote the class of algebra-reducts of matrices in Mod^*S. That is, Alg^*S consists of those algebras $\boldsymbol{A}$ for which there exists some filter $F \subseteq A^k$ such that $\langle \boldsymbol{A}, F\rangle \in Mod^*S$. Fix a set $X \subset Var$. The following theorem implies that for a protoalgebraic system S the variety generated by Alg^*S and the variety generated by the algebra $\boldsymbol{Fm}(X)/\Omega^{\boldsymbol{Fm}(X)}(Cn_S(\emptyset))$ are the same.

3.2.13. Theorem *Let S be a protoalgebraic k-deductive system, and let $X \subset Var$. For every $\varphi, \psi \in Fm(X)$,*

$$\boldsymbol{Fm}(X)/\Omega^{\boldsymbol{Fm}(X)}(Cn_S(\emptyset)) \models \varphi \approx \psi \textit{ iff } Alg^*S \models \varphi \approx \psi.$$

Proof: See Theorem 4.34 in [Jansana, 2000]. ■

The above theorem plays an important role in the proof of Proposition 3.5.14.

3.2.5 Translations and equivalent deductive systems

Deductive systems to which the methods from universal algebra most easily and most extensively apply, are known as *algebraizable*. In order to give a precise meaning to this notion, we need a concept of equivalence between deductive systems of arbitrary dimension. This we will recall from [Blok and Pigozzi, 2001]. Readers who are only interested in equivalential logics may safely skip this subsection.

Let $k, l \in \omega$, and let $\mathcal{L}$ be a language type. By a *(k,l)-translation in $\mathcal{L}$* we understand a (possibly infinite) set $\boldsymbol{\tau}$ of l-formulas of type $\mathcal{L}$ in a single k-variable.[3] That is, there exists some index set I such that for any k-variable $\boldsymbol{p}$,

$$\boldsymbol{\tau}(\boldsymbol{p}) = \{\langle \tau_1^i(\boldsymbol{p}), \ldots, \tau_l^i(\boldsymbol{p}\rangle \ : \ i \in I\}.$$

[3] Some authors require these translations to be finite. For our purposes this is not needed.

For $\boldsymbol{\varphi} \in Fm^k$, $\boldsymbol{\tau}(\boldsymbol{\varphi})$ denotes the set of l-formulas $\{\langle \tau_1^i(\boldsymbol{\varphi}), \ldots, \tau_l^i(\boldsymbol{\varphi})\rangle : i \in I\}$, where $\tau_1^i(\boldsymbol{\varphi})$ denotes the result of substituting in the formula τ_1^i the formula φ_1 for the variable p_1, the formula φ_2 for the variable p_2, etc. That is, $\tau_1^i(\boldsymbol{\varphi}) = \tau_1^i(\boldsymbol{p})[\varphi_1/p_1, \ldots, \varphi_k/p_k]$. For $\Gamma \subseteq Fm^k$, the set $\bigcup\{\boldsymbol{\tau}(\boldsymbol{\gamma}) : \boldsymbol{\gamma} \in \Gamma\}$ will be abbreviated to $\boldsymbol{\tau}(\Gamma)$.

It is useful to note that any (k, l)-translation $\boldsymbol{\tau}$ is *structural* in the sense that for any $\boldsymbol{\varphi} \in Fm^k$ and substitution σ, $\sigma(\boldsymbol{\tau}(\boldsymbol{\varphi})) = \boldsymbol{\tau}(\sigma(\boldsymbol{\varphi}))$.

Let S_1 be a k-deductive system and S_2 an l-deductive system over the same language type $\mathcal{L}$. A (k, l)-translation $\boldsymbol{\tau}$ in $\mathcal{L}$ is an *interpretation of* S_1 *in* S_2 if for all $\Gamma \cup \{\boldsymbol{\varphi}\} \subseteq Fm^k$,

$$\boldsymbol{\varphi} \in Cn_{S_1}(\Gamma) \Leftrightarrow \boldsymbol{\tau}(\boldsymbol{\varphi}) \in Cn_{S_2}(\boldsymbol{\tau}(\Gamma)).$$

We sometimes write $\tau : S_1 \longrightarrow S_2$ to denote that τ is an interpretation of S_1 into S_2.

As an example, consider the language type $\mathcal{L} = \{\rightarrow, \wedge, \vee, \neg, \bot, \top\}$. Let S_{BA} be the 2-deductive system over $\mathcal{L}$ given by the axiom schemas $\langle p, p\rangle$ and $\langle \varphi, \psi\rangle$, for every BA-axiom $\varphi \approx \psi$, together with the inference rules

$$\frac{\langle p, q\rangle}{\langle q, p\rangle}, \quad \frac{\langle p, q\rangle, \langle q, r\rangle}{\langle p, r\rangle} \quad \text{and} \quad \frac{\langle p_1, q_1\rangle, \ldots, \langle p_n, q_n\rangle}{\langle \omega(p_1, \ldots p_n), \omega(q_1, \ldots, q_n)\rangle},$$

for every n-ary connective $\omega \in \mathcal{L}$. The (1,2)-translation $\boldsymbol{\tau} : \mathsf{CPC} \longrightarrow S_{BA}$ defined by $\boldsymbol{\tau}(p) = \langle p, \top\rangle$ is an interpretation of CPC in S_{BA}. Conversely, the (2,1)-translation $\boldsymbol{\rho} : S_{BA} \longrightarrow \mathsf{CPC}$ that maps $\langle p, q\rangle$ to $p \leftrightarrow q$ is an interpretation of S_{BA} in CPC.

3.2.14. Definition [**Equivalent deductive systems**] Let S_1 be a k-deductive system and S_2 an l-deductive system over the same language type $\mathcal{L}$, where $l, k \in \omega$. S_1 and S_2 are *equivalent* if there exist interpretations $\boldsymbol{\tau} : S_1 \longrightarrow S_2$ and $\boldsymbol{\rho} : S_2 \longrightarrow S_1$ that are inverses to one another in the sense that for all $\boldsymbol{\varphi} \in Fm^k$, $\boldsymbol{\psi} \in Fm^l$,

1. $Cn_{S_1}(\boldsymbol{\varphi}) = Cn_{S_1}(\boldsymbol{\rho\tau}(\boldsymbol{\varphi}))$.
2. $Cn_{S_2}(\boldsymbol{\psi}) = Cn_{S_2}(\boldsymbol{\tau\rho}(\boldsymbol{\psi}))$.

In this case we say that S_1 and S_2 are *equivalent under the translations* $\boldsymbol{\tau} : S_1 \longrightarrow S_2$ and $\boldsymbol{\rho} : S_2 \longrightarrow S_1$. ⊣

Note that the above two conditions immediately imply that for arbitrary sets $\Gamma \subseteq Fm^k$ and $\Gamma' \subseteq Fm^l$ we have that $Cn_{S_1}(\Gamma) = Cn_{S_1}(\boldsymbol{\rho\tau}(\Gamma))$ and $Cn_{S_2}(\Gamma') = Cn_{S_2}(\boldsymbol{\tau\rho}(\Gamma'))$.

In [Blok and Pigozzi, 2001, Example V.1.3] it is verified that the deductive systems S_{BA} and CPC are equivalent under the translations $\boldsymbol{\tau}, \boldsymbol{\rho}$ given above.

In the next section we show that the Beth property is respected by this notion of equivalence, cf. Theorem 3.3.7. In order to establish this theorem, we need the following results from [Blok and Pigozzi, 2001]. Lemma 3.2.15 states that each translation induces a transformation on filters that commutes with homomorphisms. To see why this is the case, fix some $\mathcal{L}$-algebra $\boldsymbol{A}$. Recall that $\varphi^{\boldsymbol{A}}(a_1, \dots, a_k)$ denotes the interpretation of the formula $\varphi(x_1, \dots, x_k)$ in $\boldsymbol{A}$ when x_i is interpreted as $a_i \in A$, for $1 \leq i \leq k$. For a (k, l)-translation $\boldsymbol{\tau}$ in $\mathcal{L}$, and $\boldsymbol{a} \in A^k$, we define

$$\boldsymbol{\tau}^{\boldsymbol{A}}(\boldsymbol{a}) = \{\langle \tau_1^{i\boldsymbol{A}}(\boldsymbol{a}), \dots, \tau_l^{i\boldsymbol{A}}(\boldsymbol{a})\rangle : i \in I\}.$$

Let $X \subseteq A^k$, and let S be a k-deductive system. We write $\boldsymbol{\tau}^{\boldsymbol{A}}(X) = \bigcup\{\boldsymbol{\tau}^{\boldsymbol{A}}(\boldsymbol{a}) : \boldsymbol{a} \in X\}$. Moreover, the S-filter on $\boldsymbol{A}$ generated by X is denoted by $Cn_S^{\boldsymbol{A}}(X)$. This notion generalizes the consequence relation introduced in Definition 3.2.1. Finally, for an interpretation $\boldsymbol{\tau} : S \longrightarrow S_2$, we define the map $\boldsymbol{\tau}_{S_2} : A^k \longrightarrow Fi_{S_2}(\boldsymbol{A})$ by $\boldsymbol{\tau}_{S_2}(F) = Cn_{S_2}^{\boldsymbol{A}}(\boldsymbol{\tau}^{\boldsymbol{A}}(F))$. Recall that for any homomorphism $h : \boldsymbol{A} \longrightarrow \boldsymbol{B}$, any k-deductive system S and any $F \in Fi_S(\boldsymbol{A})$, by $h_S(F)$ we denote the least S-filter on $\boldsymbol{B}$ that contains $h(F)$.

3.2.15. Lemma *Let $h : \boldsymbol{A} \longrightarrow \boldsymbol{B}$ be a homomorphism between $\mathcal{L}$-algebras, and let $\boldsymbol{\tau} : S_1 \longrightarrow S_2$ be an interpretation of S_1 in S_2. Then $h_{S_2} \circ \boldsymbol{\tau}_{S_2} = \boldsymbol{\tau}_{S_2} \circ h_{S_1}$. In other words, the diagram in Figure 3.1 commutes.*

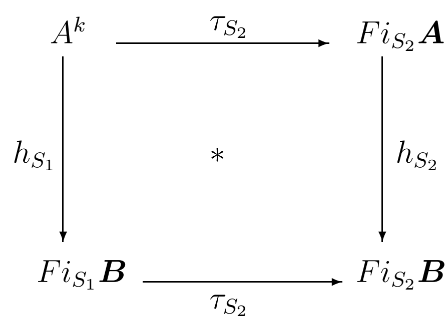


Figure 3.1: By Lemma 3.2.15 the diagram commutes.

Proof: See [Blok and Pigozzi, 2001, Lemma V.3.3]. ■

Let $\boldsymbol{A}$ be an algebra and $F \subseteq A^k$. Recall that $\Omega^{\boldsymbol{A}}(F)$ is the largest congruence on $\boldsymbol{A}$ compatible with F (see page 69). The lattice of congruences on $\boldsymbol{A}$ is denoted by $Co(\boldsymbol{A})$.

3.2.16. Theorem *Let the deductive systems S_1 and S_2 be equivalent under the translations $\boldsymbol{\tau} : S_1 \longrightarrow S_2$, and $\boldsymbol{\rho} : S_2 \longrightarrow S_1$. Let $\boldsymbol{A}$ be an $\mathcal{L}$-algebra. For any $F \in Fi_{S_1}(\boldsymbol{A})$, we have that $\Omega^{\boldsymbol{A}}(\boldsymbol{\tau}_{S_2}(F)) = \Omega^{\boldsymbol{A}}(F)$. That is, the diagram in Figure 3.2 commutes.*

Proof: See [Blok and Pigozzi, 2001, Theorem V.3.6]. ■

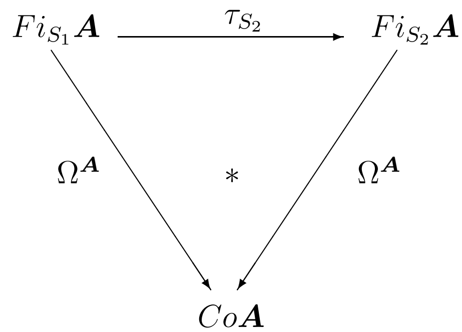


Figure 3.2: By Theorem 3.2.16 the diagram commutes.

3.3 An algebraic characterization of the Beth property

This section forms the heart of the present chapter in which we give an algebraic characterization of the Beth property for equivalential logics (cf. Theorem 3.3.8). In subsection 3.3.3 we explicitly formulate this result for a particularly well-behaved subclass of logics, viz., the so-called algebraizable logics. In the sequel, the proof of Theorem 3.3.8 will be used as a template to prove characterizations of several other definability properties.

3.3.1 The Beth property in abstract algebraic logic

In this subsection we aim to give a characterization of the Beth property which encompasses as many logics as possible. This implies that we will have to give a general formulation of the Beth property as the usual formulation (as in Definition 2.2.6) in terms of the connective $\leftrightarrow$ is not applicable to every logic.

We remind the reader that the consequence relation $\models_{\mathcal{A}(X)}$ has been introduced at page pagerefconsequence relation.

3.3.1. Definition [**Beth (definability) property**] Let S be a k-deductive system, and let P, R be disjoint sets of variables such that $Fm(P) \neq \emptyset$. A set of k-formulas $\Gamma \subseteq Fm^k(P \cup R)$ *implicitly defines* R *in terms of* P if for every $r \in R$ and every substitution σ such that $\sigma_{\restriction P} = id_{\restriction P}$, the following holds:

(I). For any $\mathcal{A} \in Mod^*S$, any set $X \subset Var$ that contains $P \cup R \cup \sigma R$,

$$\Gamma \cup \sigma\Gamma \models_{\mathcal{A}(X)} \langle r, \sigma(r)\rangle.$$

A formula $\varphi_r \in Fm(P)$ is called an *explicit definition* of $r \in R$ in terms of P with respect to Γ if

(II). For any $\mathcal{A} \in Mod^*S$, any set $X \subset Var$ that contains $P \cup \{r\}$,

$$\Gamma \models_{\mathcal{A}(X)} \langle r, \varphi_r\rangle.$$

S has the *Beth (definability) property* if for every Γ that implicitly defines R in terms of P, and every $r \in R$, there exists an explicit definition of r in terms of P with respect to Γ. ⊣

Let us see that Definition 3.3.1 yields the more familiar formulation of the Beth property in case the logic S is equivalential.

3.3.2. Proposition *Let S be an equivalential k-deductive system with a set of equivalence formulas $\Delta(x, y)$. Let P, R, Γ, σ, r be as in Definition 3.3.1. The following are equivalent:*

1. *(I) in Definition 3.3.1 holds.*
2. $\Gamma, \sigma\Gamma \vdash_S \Delta(r, \sigma r)$.

Proof: Assume (I). Let $\boldsymbol{A} = \boldsymbol{Fm}(P \cup R \cup \sigma R)/\Omega(Cn_S(\Gamma \cup \sigma\Gamma))$, $F_A = Cn_S(\Gamma \cup \sigma\Gamma)/\Omega(Cn_S(\Gamma \cup \sigma\Gamma))$, and $\mathcal{A} = \langle \boldsymbol{A}, F_A \rangle$. Let h be the natural map from $\boldsymbol{Fm}(P \cup R \cup \sigma R)$ to $\boldsymbol{A}$. Note that $\mathcal{A} \in Mod^*S$ and $h(\Gamma \cup \sigma\Gamma) \subseteq F_A$. Applying (I) gives that $h(r) = h(\sigma r)$, i.e., $\langle r, \sigma r \rangle \in \Omega^{\boldsymbol{Fm}(P \cup R \cup \sigma R)}(Cn_S(\Gamma \cup \sigma\Gamma))$. By Theorem 3.2.10, $\Delta(r, \sigma r) \subseteq Cn_S(\Gamma \cup \sigma\Gamma)$. In other words, $\Gamma, \sigma\Gamma \vdash_S \Delta(r, \sigma r)$.

For the other direction, assume $\Gamma, \sigma\Gamma \vdash_S \Delta(r, \sigma r)$. Consider some reduced matrix $\mathcal{A} = \langle \boldsymbol{A}, F \rangle$ and some set $X \subseteq Var$ as in (I). Let $h : \boldsymbol{Fm}(X) \longrightarrow \boldsymbol{A}$ be such that $h(\Gamma \cup \sigma\Gamma) \subseteq F$. As $\mathcal{A}$ is an S-matrix, then $\Delta^{\boldsymbol{A}}(h(r), h(\sigma r)) = h(\Delta(r, \sigma r)) \subseteq F$. Again by Theorem 3.2.10, this implies that $\langle h(r), h(\sigma r) \rangle \in \Omega^{\boldsymbol{A}}(F)$. Therefore $h(r) = h(\sigma r)$, as $\mathcal{A}$ is reduced. ■

3.3.3. Proposition *Let S, $\Delta(x, y)$, P, R, Γ, r be as in Proposition 3.3.2. Let $\varphi_r \in Fm(P)$. The following are equivalent:*

1. *φ_r is an explicit definition of r in terms of P with respect to Γ.*
2. $\Gamma \vdash_S \Delta(r, \varphi_r)$.

Proof: Left to the reader. ■

In the second expression in Proposition 3.3.2 resp. 3.3.3 the reader will recognize the notions of implicit resp. explicit definition in their usual formulation, such as in Definition 2.2.6.

The following auxiliary lemma follows easily from Proposition 3.3.3 and allows for a smooth presentation of the proof of Theorem 3.3.8. It states that an explicit definition may be chosen relative to some matrix $\mathcal{A}$ and homomorphism h.

3.3.4. Lemma *Let S be an equivalential k-deductive system and let P, R, Γ, r be as in Definition 3.3.1. Then r has an explicit definition in terms of P with respect to Γ if for any $\mathcal{A} \in Mod^*S$, any set $X \subset Var$ that contains $P \cup R$ and any map $h : \boldsymbol{Fm}(X) \longrightarrow \boldsymbol{A}$ such that $h(\Gamma) \subseteq F_A$ there exists some $\varphi_r \in Fm(P)$ such that $h(r) = h(\varphi_r)$.*

Proof: Assume the antecedent. Let $\boldsymbol{A} = \boldsymbol{Fm}(P \cup R)/\Omega(Cn_S\Gamma)$, $F_A = Cn_S\Gamma/\Omega(Cn_S\Gamma)$, and $\mathcal{A} = \langle \boldsymbol{A}, F_A \rangle$. Let h be the natural map from $\boldsymbol{Fm}(P \cup R)$ to $\boldsymbol{A}$. Note that $\mathcal{A} \in Mod^*S$ and $h(\Gamma) \subseteq F_A$. By assumption there exists some $\varphi_r \in Fm(P)$ such that $h(r) = h(\varphi_r)$. That is, $\langle r, \varphi_r \rangle \in \Omega^{\boldsymbol{Fm}(P\cup R)}(Cn_S\Gamma)$. By Theorem 3.2.10, $\Delta(r, \sigma\varphi_r) \subseteq Cn_S\Gamma$. In other words, $\Gamma \vdash_S \Delta(r, \varphi_r)$. By Proposition 3.3.3 this implies that φ_r is an explicit definition of r in terms of P with respect to Γ. ■

3.3.5. Remark Even though, as we have seen above, Definition 3.3.1 can be written in the usual format, it needs to be stressed that this definition is stronger than e.g., Definition 2.2.6. The difference is that Definition 3.3.1 is not restricted to the case where the set R is a singleton: R can be an arbitrary finite (or even an infinite) set. These three possibilities for the cardinality of R induce three Beth properties which we will, with some 'abus de language', baptize as the *singleton Beth property*, the *finite Beth property* and the *infinite Beth property*. The singleton version is the most studied of the three. This is why in chapter 2 we were interested in this property. However, for the purposes of the present chapter it makes sense to concentrate on the infinite Beth property: as we will shortly see, it allows for the most natural algebraic characterization from the proof of which characterizations for the other two Beth properties can easily be obtained. This has been done in Theorem 3.3.10. For completeness' sake let us anticipate Corollary 3.3.11 saying that in the present context we actually only need to distinguish two of the aforementioned properties; an equivalential logic has the singleton Beth property iff it has the finite Beth property. ⊣

3.3.6. Convention To prevent confusion, let us repeat that in this chapter by the Beth property we understand the infinite Beth property. ⊣

The next theorem states that the Beth property is preserved under equivalence of deductive systems, as defined in Definition 3.2.14.

3.3.7. Theorem *Let S_1, S_2 be equivalent deductive systems. Then*

$$S_1 \text{ has the Beth property} \Leftrightarrow S_2 \text{ has the Beth property.}$$

Proof: Let S_1 and S_2 be deductive systems which are equivalent under the translations $\boldsymbol{\tau} : S_1 \longrightarrow S_2$ and $\boldsymbol{\rho} : S_2 \longrightarrow S_1$. Suppose S_1 has the Beth property. To show that S_2 has this property too, consider disjoint sets of variables P, R, and let Γ implicitly define R in terms of P in S_2. We prove that every $r \in R$ has an explicit definition in terms of P with respect to Γ in S_2.

We first show that $\boldsymbol{\rho}\Gamma$ implicitly defines R in terms of P in S_1. Hereto, consider some $r \in R$ and a substitution σ such that $\sigma_{\restriction P} = id_{\restriction P}$. Let $\mathcal{A} = \langle \boldsymbol{A}, F_A \rangle \in Mod^*(S_1)$, X a set of sentential variables that contains $P \cup R \cup \sigma R$, and $h : \boldsymbol{Fm}(X) \longrightarrow \boldsymbol{A}$ such that $h(\boldsymbol{\rho}\Gamma \cup \sigma\boldsymbol{\rho}\Gamma) \subseteq F_A$. We need to show that $h(r) = h(\sigma r)$. Note the following.

$$\begin{aligned}
Cn_{S_2}(\Gamma \cup \sigma\Gamma) &= Cn_{S_2}(Cn_{S_2}\Gamma \cup Cn_{S_2}\sigma\Gamma),\\
&= Cn_{S_2}(Cn_{S_2}\boldsymbol{\tau\rho}\Gamma \cup Cn_{S_2}\boldsymbol{\tau\rho}\sigma\Gamma), \text{ by Definition 3.2.14},\\
&= Cn_{S_2}(\boldsymbol{\tau\rho}\Gamma \cup \boldsymbol{\tau\rho}\sigma\Gamma),\\
&= Cn_{S_2}\boldsymbol{\tau}(\boldsymbol{\rho}\Gamma \cup \boldsymbol{\rho}\sigma\Gamma),\\
&= Cn_{S_2}\boldsymbol{\tau}(\boldsymbol{\rho}\Gamma \cup \sigma\boldsymbol{\rho}\Gamma), \qquad \text{since } \boldsymbol{\rho} \text{ is structural.}
\end{aligned}$$

Recall that for any homomorphism $h : \boldsymbol{A} \longrightarrow \boldsymbol{B}$, any k-deductive system S and any $F \in Fi_S(\boldsymbol{A})$, by $h_S(F)$ we denote the least S-filter on $\boldsymbol{B}$ that contains $h(F)$. The map $\boldsymbol{\tau}_S$ has been defined just before Lemma 3.2.15. Then

$$\begin{aligned}
h(\Gamma \cup \sigma\Gamma) &\subseteq h_{S_2}Cn_{S_2}(\Gamma \cup \sigma\Gamma),\\
&= h_{S_2}Cn_{S_2}\boldsymbol{\tau}(\boldsymbol{\rho}\Gamma \cup \sigma\boldsymbol{\rho}\Gamma), \text{ by the above},\\
&= h_{S_2}\boldsymbol{\tau}_{S_2}(\boldsymbol{\rho}\Gamma \cup \sigma\boldsymbol{\rho}\Gamma), \text{ by definition of } \boldsymbol{\tau}_{S_2},\\
&= \boldsymbol{\tau}_{S_2}h_{S_1}(\boldsymbol{\rho}\Gamma \cup \sigma\boldsymbol{\rho}\Gamma), \text{ by Lemma 3.2.15},\\
&\subseteq \boldsymbol{\tau}_{S_2}F_A, \text{ as } h_{S_1}(\boldsymbol{\rho}\Gamma \cup \sigma\boldsymbol{\rho}\Gamma) \subseteq F_A.
\end{aligned}$$

By Theorem 3.2.16, $\Omega^{\boldsymbol{A}}(\boldsymbol{\tau}_{S_2}F_A) = \Omega^{\boldsymbol{A}}(F_A) = \Delta_A$. We conclude that the matrix $\langle \boldsymbol{A}, \boldsymbol{\tau}_{S_2}F_A\rangle \in Mod^*(S_2)$ and $h(\Gamma \cup \sigma\Gamma) \subseteq \boldsymbol{\tau}_{S_2}F_A$. As Γ implicitly defines R in terms of P in S_2, it follows that $h(r) = h(\sigma r)$. We have shown that $\boldsymbol{\rho}\Gamma$ implicitly defines R in terms of P in S_1.

Let $r \in R$. Since S_1 has the Beth property, r has an explicit definition $\varphi_r \in Fm(P)$ with respect to $\boldsymbol{\rho}\Gamma$ in S_1. We claim that φ_r is an explicit definition of r with respect to Γ in S_2. Let $\mathcal{B} = \langle \boldsymbol{B}, F_B\rangle \in Mod^*(S_2)$, $X \subset Var$ that contains $P \cup \{r\}$, and $h : \boldsymbol{Fm}(X) \longrightarrow \boldsymbol{B}$ such that $h(\Gamma) \subseteq F_B$. We show that $h(r) = h(\varphi_r)$. Again by Theorem 3.2.16, $\langle \boldsymbol{B}, \boldsymbol{\rho}_{S_1}F_B\rangle \in Mod^*(S_1)$. Using Lemma 3.2.15 we see that $h(\boldsymbol{\rho}\Gamma) \subseteq h_{S_1}(\boldsymbol{\rho}_{S_1}Cn_{S_1}\Gamma) = \boldsymbol{\rho}_{S_1}h_{S_2}Cn_{S_1}\Gamma \subseteq \boldsymbol{\rho}_{S_1}F_B$. As φ_r explicitly defines r with respect to $\boldsymbol{\rho}\Gamma$ in S_1, we conclude that $h(r) = h(\varphi_r)$. ■

3.3.2 Characterization of Beth property for equivalential logics

The following theorem is the main result of the present chapter. A history of this result can be found in subsection 2.6.2. Recall that the notion of a matrix epimorphism has been defined at page 68.

3.3.8. Theorem *Let S be an equivalential k-deductive system. S has the Beth property iff all Mod^*S-epimorphisms are surjective.*

In order to gain some intuition, consider a submatrix $\mathcal{A} \subseteq \mathcal{B}$ which is epic in some class of matrices $\mathbb{K}$. This means that any two homomorphisms from $\mathcal{B}$ into some $\mathbb{K}$-matrix which agree on A are equal (*). It is worth to compare this to the following situation where Γ implicitly defines R in terms of P in the logic S. That is, any two interpretations of the variables from $P \cup R$ in some S-model of Γ that agree on the variables in P are equal (**). This comparison suggests a close correspondence between $\mathbb{K}$-epimorphisms and implicit definitions in S, provided $\mathbb{K}$ and S are appropriately related. Note also that if $A = B$, then (*) is trivially fulfilled. Similarly, (**) is trivially fulfilled if every $r \in R$ has an explicit definition in terms of P with respect to Γ. In this case we might say that there are no proper implicit definitions in S. This leads to the idea that there are no proper implicit definitions in S (i.e., S has the Beth property) iff there are no proper epic submatrices in $\mathbb{K}$ (i.e., $\mathbb{K}$ has ES). Section 2.6.2 contains another informal exposition on the connection between the Beth property and surjectiveness of epimorphisms.

Proof: Let S be an equivalential k-deductive system.

Proof of BP $\Rightarrow$ ES: Assume that S has the Beth property. Let $f : \mathcal{A} \longrightarrow \mathcal{B} = \langle \boldsymbol{B}, F_B \rangle$ be a Mod^*S-epimorphism. That is, any two matrix homomorphisms k, k' from $\mathcal{B}$ into some $\mathcal{C} \in Mod^*S$ are equal as soon as $k \circ f = k' \circ f$. Our task is to show that f is surjective.

Consider a set $Z \subset Var$ with the same cardinality as B. Since Var contains sets of arbitrary size, such Z exists. Let $h : \boldsymbol{Fm}(Z) \twoheadrightarrow \boldsymbol{B}$ be a homomorphism which maps Z onto B. By Z_A (resp. $Z_{B\setminus A}$) we understand the sentential variables in Z which are mapped by h to $f(A)$ (resp. $B \setminus f(A)$). Note that $h^{-1}(F_B) \in Fi_S(\boldsymbol{Fm}(Z))$.

For brevity, write $F = h^{-1}(F_B)$. We claim that F implicitly defines $Z_{B\setminus A}$ in terms of Z_A.

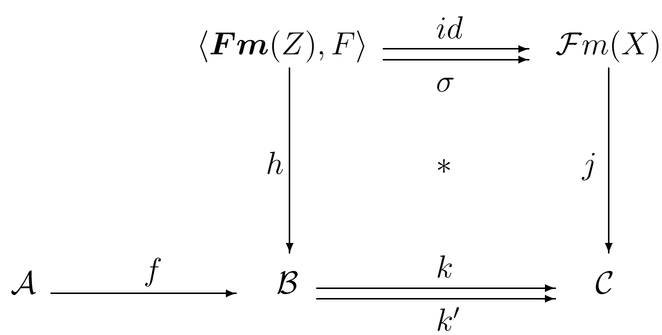
To see this, consider a substitution σ that fixes the variables in Z_A. Let $\mathcal{C} = \langle \boldsymbol{C}, F_C \rangle \in Mod^*S$, let X be a set of sentential variables that contains $Z \cup \sigma Z$ and let the map $j : \boldsymbol{Fm}(X) \longrightarrow \boldsymbol{C}$ be such that $j(F \cup \sigma F) \subseteq F_C$. We need to show that for all $z \in Z_{B\setminus A}$,

$$j(z) = j(\sigma z). \tag{3.2}$$

Without loss of generality, we may assume that $\Omega^{\boldsymbol{Fm}(X)}(Cn_S(F \cup \sigma F)) \subseteq ker(j)$. For, suppose we are able to prove (3.2) for any such j. That is, suppose (3.2) holds for any map $j : \boldsymbol{Fm}(X) \longrightarrow \boldsymbol{C}$ such that $j(F \cup \sigma F) \subseteq F_C$, and $\Omega^{\boldsymbol{Fm}(X)}(Cn_S(F \cup \sigma F)) \subseteq ker(j)$, where X is a arbitrary set of variables that contains $Z \cup \sigma Z$, and $\mathcal{C}$ is an arbitrary reduced S-matrix. We copy the proof of Proposition 3.3.2 (1) $\Rightarrow$ (2) as follows. Let $\boldsymbol{A} = \boldsymbol{Fm}(Z \cup \sigma Z)/\Omega(Cn_S(F \cup \sigma F))$, $F_A = Cn_S(F \cup \sigma F)/\Omega(Cn_S(F \cup \sigma F))$, and $\mathcal{A} = \langle \boldsymbol{A}, F_A \rangle$. Let n be the natural map from $\boldsymbol{Fm}(Z \cup \sigma Z)$ to $\boldsymbol{A}$. Note that $\mathcal{A} \in Mod^*S$ and $n(F \cup \sigma F) \subseteq F_A$. Moreover, $\Omega^{\boldsymbol{Fm}(X)}(Cn_S(F \cup \sigma F)) \subseteq ker(n)$. By supposition, (3.2) holds for n. That is, for every $z \in Z_{B\setminus A}$, $n(z) = n(\sigma z)$. In other words, $\langle z, \sigma z \rangle \in \Omega(Cn_S(F \cup \sigma F))$. By Theorem 3.2.10, this implies that $F, \sigma F \vdash_S \Delta(z, \sigma z)$. By Proposition 3.3.2 we conclude that F implicitly defines $Z_{B\setminus A}$ in terms of Z_A, as was to be shown.

We continue the proof of (3.2). Let $\mathcal{F}m(X) = \langle \boldsymbol{Fm}(X), Cn_S(F \cup \sigma F)\rangle$, and consider the matrix homomorphism $j : \mathcal{F}m(X) \longrightarrow \mathcal{C}$ and the matrix homomorphisms $\sigma, id : \langle \boldsymbol{Fm}(Z), F\rangle \longrightarrow \mathcal{F}m(X)$. As $F \subseteq Cn_{S(X)}(F \cup \sigma F) \cap Fm^k(Z)$, it follows from the monotonicity of $\Omega^{\boldsymbol{Fm}(Z)}$ restricted to S-filters, that $\Omega^{\boldsymbol{Fm}(Z)}(F) \subseteq \Omega^{\boldsymbol{Fm}(Z)}(Cn_{S(X)}(F \cup \sigma F) \cap Fm^k(Z))$. By the characterization in Theorem 3.2.11.3 the latter is included in $\Omega^{\boldsymbol{Fm}(X)}(Cn_{S(X)}(F \cup \sigma F))$, which in its turn is included in $\ker(j)$ by assumption. Since $\ker(h)$ is a congruence on $\boldsymbol{Fm}(Z)$ which is compatible with F, we have that $\ker(h) \subseteq \Omega^{\boldsymbol{Fm}(Z)}(F)$. We conclude that $\ker(h) \subseteq \ker(j)$. Moreover, $\sigma(\ker(h)) \subseteq \sigma\Omega^{\boldsymbol{Fm}(Z)}(F) \subseteq \Omega^{\boldsymbol{Fm}(\sigma Z)}\sigma_S F$. The last inclusion is due to Corollary 3.2.12. As $\sigma_S F \subseteq Cn_{S(X)}(F \cup \sigma F) \cap Fm^k(\sigma Z)$, it follows again from the monotonicity of $\Omega^{\boldsymbol{Fm}(\sigma Z)}$ restricted to S-filters and Theorem 3.2.11.3 that $\Omega^{\boldsymbol{Fm}(\sigma Z)}(\sigma_S F) \subseteq \Omega^{\boldsymbol{Fm}(X)}(Cn_{S(X)}(F \cup \sigma F)) \subseteq \ker(j)$. We deduce that $\ker(h) \subseteq \ker(j \circ \sigma)$.

Note that $h : \langle \boldsymbol{Fm}(Z), F\rangle \longrightarrow \mathcal{B}$ is a strict, onto matrix homomorphism. From the Homomorphism Lemma 3.2.3 we obtain matrix homomorphisms $k, k' : \mathcal{B} \longrightarrow \mathcal{C}$ such that $k \circ h = j \circ id$ and $k' \circ h = j \circ \sigma$.



We show that $k \circ f = k' \circ f$. Let $a \in A$. Then $f(a) = h(z_a)$, for some $z_a \in Z_A$. Recall that σ fixes Z_A. This gives $k(f(a)) = k(h(z_a)) = j(z_a) = j(\sigma z_a) = k'(h(z_a)) = k'(f(a))$. That is, $k \circ f = k' \circ f$. As f is a Mod^*S-epimorphism, it follows that $k = k'$.

To finish the proof of (3.2), let $z \in Z_{B \setminus A}$. Then $j(z) = k(h(z)) = k'(h(z)) = j(\sigma z)$.

This shows that F implicitly defines $Z_{B \setminus A}$ in terms of Z_A. Since S has the Beth property, it follows that every $z \in Z_{B \setminus A}$ has some explicit definition in terms of Z_A with respect to F.

We are now in a position to prove that f is surjective. Let $b \in B$. We claim that b lies in the range of f. By construction, $b = h(z)$, for some $z \in Z$. If $z \in Z_A$, we are done. Therefore, suppose $z \in Z_{B \setminus A}$. As we deduced above, in this case z has some explicit definition $\varphi_z \in Fm(Z_A)$ with respect to F. As $\mathcal{B} \in Mod^*S$ and $h : \boldsymbol{Fm}(Z) \longrightarrow \boldsymbol{B}$ is such that $h(F) \subseteq F_B$, this implies that $h(z) = h(\varphi_z)$. Summarizing, we see that $b = h(z) = h(\varphi_z) \in f(A)$.

Proof of ES $\Rightarrow$ BP: Assume that all Mod^*S-epimorphisms are surjective. Let P, R be disjoint sets of variables such that $Fm(P) \neq \emptyset$, and let $\Gamma \subseteq Fm^k(P \cup R)$ implicitly define R in terms of P. We have to show that every $r \in R$ has an explicit definition in terms of P with respect to Γ.

Let $\mathcal{A} \in Mod^*S$, let X be a set of sentential variables that contains $P \cup R$ and let the map $h : \boldsymbol{Fm}(X) \longrightarrow \boldsymbol{A}$ be such that $h(\Gamma) \subseteq F_A$. By Lemma 3.3.4 it suffices to show for every $r \in R$ there exists some $\varphi_r \in Fm(P)$ such that

$$h(r) = h(\varphi_r). \tag{3.3}$$

To this end, consider the two matrices $\mathcal{H}_P = \langle h(Fm(P)), F_A \cap h(Fm(P)^k)\rangle$ and $\mathcal{H}_{P,R} = \langle h(Fm(P \cup R)), F_A \cap h(Fm(P \cup R)^k)\rangle$. Note that both matrices are submatrices of $\mathcal{A}$, whence $\mathcal{H}_P, \mathcal{H}_{P,R} \in \mathrm{S}Mod^*S = Mod^*S$. This last equality holds as S is equivalential, using the characterization in Theorem 3.2.11.2. If $\mathcal{H}_P = \mathcal{H}_{P,R}$, then clearly (3.3) holds. By our assumption it suffices to show that $\mathcal{H}_P$ is an epic submatrix of $\mathcal{H}_{P,R}$ in Mod^*S. This will be established in the remaining part of the proof.

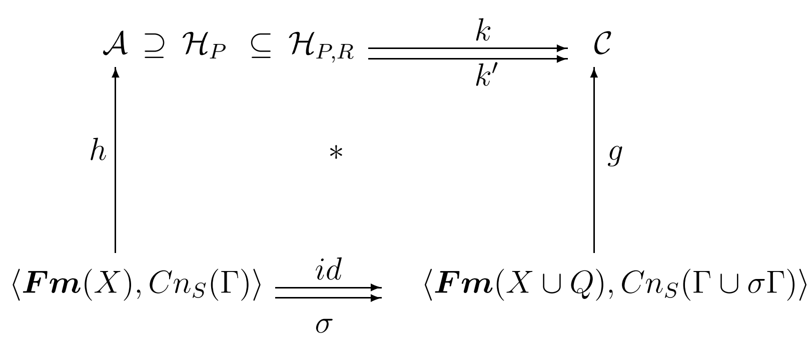
Consider some $\mathcal{C} = \langle \boldsymbol{C}, F_C\rangle \in Mod^*S$ together with a pair of matrix homomorphisms $k, k' : \mathcal{H}_{P,R} \longrightarrow \mathcal{C}$ such that $k_{\restriction h(Fm(P))} = k'_{\restriction h(Fm(P))}$. Our task is to show that $k = k'$.

Let Q be a set of variables disjoint from X, of the same size as R. Let $i : Q \rightarrowtail\!\!\!\!\rightarrow R$ be a bijection, and let σ be a substitution which interchanges q and $i(q)$, for every $q \in Q$, and which fixes all other variables. That is,

$$\begin{array}{rcll} \sigma(q) & = & i(q), & \text{if } q \in Q,\\ \sigma(r) & = & i^{-1}(r), & \text{if } r \in R,\\ \sigma(x) & = & x, & \text{otherwise.} \end{array}$$

In particular, σ fixes the variables in P. Define the homomorphism $g : \boldsymbol{Fm}(X \cup Q) \longrightarrow \boldsymbol{C}$ by

$$\begin{array}{rcll} g(x) & = & k(h(x)), & \text{for } x \in X,\\ g(q) & = & k'(h(i(q))), & \text{for } q \in Q. \end{array}$$



Let $\varphi \in Fm(P \cup R)$. The reader easily verifies that $g(\varphi) = k(h(\varphi))$ and $g(\sigma\varphi) = k'(h(\varphi))$. This implies that $g(\Gamma) = k(h(\Gamma)) \subseteq F_C$, as k is a matrix homomorphism.

Moreover, $g(\sigma\Gamma) = k'(h(\Gamma)) \subseteq F_C$, as k' is a matrix homomorphism. Whence $g(\Gamma \cup \sigma\Gamma) \subseteq F_C$.

Recall that Γ implicitly defines R in terms of P. By applying 3.3.1 (I) to $\mathcal{C}$, $X \cup Q$ and g we obtain for every $r \in R$ that $g(r) = g(\sigma r)$. As σ fixes the variables in P, this implies that for every $\varphi \in Fm(P \cup R)$, $g(\varphi) = g(\sigma\varphi)$.

This finishes the preparations for showing that $k = k'$. For, let $x \in h(Fm(P \cup R))$. Then $x = h(\varphi_x)$, for some $\varphi_x \in Fm(P \cup R)$. Therefore $k(x) = k(h(\varphi_x)) = g(\varphi_x) = g(\sigma\varphi_x) = k'(h(\varphi_x)) = k'(x)$. We conclude that $k = k'$.

This completes the proof of Theorem 3.3.8. ■

Recall that the Beth property we considered above is the infinite version as defined in Remark 3.3.5. However, the above proof also sheds light on the algebraic correspondents of the singleton and finite Beth property. Hereto, consider the following algebraic property that has already been introduced in chapter 2.

3.3.9. Definition [**Surjectiveness of almost onto epis (**ES_1**)**] A matrix homomorphism $h : \mathcal{A} \longrightarrow \mathcal{B}$ is said to be *almost onto* if there exists some $b \in B$ such that the set B is generated by $h(A) \cup \{b\}$. A class of matrices $\mathbb{K}$ has the property ES_1 if every almost onto $\mathbb{K}$-epimorphism is surjective. ⊣

The next theorem generalizes Németi's result in [Henkin et al., 1985, Theorem 5.6.10].

3.3.10. Theorem *Let S be an equivalential k-deductive system. S has the singleton Beth property iff Mod^*S has the property* ES_1.

Proof: Let S be an equivalential k-deductive system. We briefly sketch how Theorem 3.3.10 follows as a special case from the proof of Theorem 3.3.8.

Proof of BP $\Rightarrow$ ES: Assume that S has the singleton Beth property. Let $f : \mathcal{A} \longrightarrow \mathcal{B} = \langle \boldsymbol{B}, F_B \rangle$ be a Mod^*S-epimorphism and let $b_0 \in B$ be such that B is generated by $f(A) \cup \{b_0\}$. It suffices to show that b_0 lies in the range of f.

Consider a set $Z_A \subset Var$ with the same cardinality as A, and let $z_0 \in Var \setminus A$. Let $h : \boldsymbol{Fm}(Z_A \cup \{z_0\}) \twoheadrightarrow \boldsymbol{B}$ be a homomorphism which maps Z_A onto A and which maps z_0 to b_0.

Similar to the proof of Theorem 3.3.8 it can be shown that $h^{-1}(F_B)$ implicitly defines z_0 in terms of Z_A. By the singleton Beth property then z_0 has some explicit definition in terms of Z_A. As before, this implies that b_0 lies in the range of f.

Proof of ES $\Rightarrow$ BP: The proof for this direction goes through in its original form as the epimorphism $id : \mathcal{H}_P \longrightarrow \mathcal{H}_{P,R}$, around which this proof is centered, is clearly almost-onto in case R is a singleton. Details are left to the reader. ■

Following the above heuristics, it turns out that the finite Beth property corresponds to surjectiveness of those epimorphisms $f : \mathcal{A} \longrightarrow \mathcal{B}$ such that B is generated by the range of f together with some finite subset. By an easy induction argument it can be seen that this last property is equivalent to surjectiveness of almost-onto epimorphisms. Hence, from the proof of Theorem 3.3.8 we obtain the following corollary.

3.3.11. Corollary *Let S be an equivalential k-deductive system. S has the singleton Beth property iff S has the finite Beth property.*

As the reader may verify, the above corollary is not easily obtained via a direct, purely logical proof. In other words, this corollary is a nice example where the bridge between logic and algebra is really useful.

3.3.3 Characterization of Beth property for algebraizable logics

A *quasi-identity* is a formula of the form $(p_1 \approx q_1 \wedge \cdots \wedge p_n \approx q_n) \Rightarrow p \approx q$, or an identity. A class of algebras that is axiomatized by quasi-identities is called a *quasivariety*. It is easy to verify that every quasivariety is closed under taking subalgebras and direct products.

With each quasivariety $\mathbb{K}$ we associate a 2-deductive system $S_{\mathbb{K}}$ given by the axiom schemas $\langle p, p\rangle$ and $\langle \varphi, \psi\rangle$, for every $\mathbb{K}$-axiom $\varphi \approx \psi$, together with the inference rules

$$\frac{\langle p, q\rangle}{\langle q, p\rangle}, \quad \frac{\langle p, q\rangle, \langle q, r\rangle}{\langle p, r\rangle} \quad \text{and} \quad \frac{\langle p_1, q_1\rangle, \ldots, \langle p_n, q_n\rangle}{\langle \omega(p_1, \ldots p_n), \omega(q_1, \ldots, q_n)\rangle},$$

for every n-ary connective $\omega \in \mathcal{L}$, and the rule

$$\frac{\langle \varphi_1, \psi_1\rangle, \ldots, \langle \varphi_k, \psi_k\rangle}{\langle \varphi, \psi\rangle},$$

for every $\mathbb{K}$-axiom $(\varphi_1 \approx \psi_1 \wedge \cdots \wedge \varphi_k \approx \psi_k) \Rightarrow \varphi \approx \psi$. It can be shown that the reduced $S_{\mathbb{K}}$-matrices are precisely of the form $\langle \boldsymbol{A}, \Delta_A\rangle$, for some $\boldsymbol{A} \in \mathbb{K}$. Deductive systems of this kind are called *algebraic*. An example of an algebraic deductive system is S_{BA}, defined in subsection 3.2.5. The reduced matrices of S_{BA} are of the form $\langle \boldsymbol{A}, \Delta_A\rangle$, where $\boldsymbol{A}$ is a Boolean algebra.

Let $\mathbb{K}$ be a quasivariety and $S_{\mathbb{K}}$ the associated algebraic deductive system. We claim that $\mathbb{K}$ has ES iff $Mod^*(S_{\mathbb{K}})$ has ES. To see this, note that any algebra homomorphism $h : \boldsymbol{A} \longrightarrow \boldsymbol{B}$ can be seen as a matrix homomorphism $h : \langle \boldsymbol{A}, \Delta_A\rangle \longrightarrow \langle \boldsymbol{B}, \Delta_B\rangle$. One easily verifies that if $h : \boldsymbol{A} \longrightarrow \boldsymbol{B}$ is a $\mathbb{K}$-epimorphism, then $h : \langle \boldsymbol{A}, \Delta_A\rangle \longrightarrow \langle \boldsymbol{B}, \Delta_B\rangle$ is a $Mod^*(S_{\mathbb{K}})$-epimorphism. Note also that if $j : \mathcal{A} \longrightarrow \mathcal{B}$ is a $Mod^*(S_{\mathbb{K}})$-epimorphism, then the homomorphism $j : \boldsymbol{A} \longrightarrow \boldsymbol{B}$

between the underlying algebras is a $\mathbb{K}$-epimorphism. These two observations together imply that $\mathbb{K}$ has ES iff $Mod^*(S_{\mathbb{K}})$ has ES.

A k-deductive system S is called *algebraizable* if S is equivalent to some algebraic system $S_{\mathbb{K}}$.[4] In this case, the quasivariety $\mathbb{K}$ is called the *equivalent quasivariety* of S. Recall that in subsection 3.2.5 it has been shown that S_{BA} and CPC are equivalent. This implies that CPC is algebraizable with the class of Boolean algebras as its equivalent quasivariety (in this case even variety). The notion of "algebraizable logic" was introduced and studied in [Blok and Pigozzi, 1989].

In the introduction it was mentioned that a characteristic feature of algebraizable logics is the presence of two (sets of) connective(s) with the characteristic properties of the biconditional $\leftrightarrow$, respectively the constant $\top$, in CPC. The following theorem specifies this statement. Below, $\varphi \dashv\vdash_S \psi$ denotes that $\varphi \vdash_S \psi$ and $\psi \vdash_S \varphi$.

3.3.12. Theorem *A k-deductive system S is* algebraizable *if it has a finite set of equivalence formulas $\Delta_1(p,q), \dots, \Delta_n(p,q)$ and a finite set of defining equations $\delta_1(\boldsymbol{x}) \approx \varepsilon_1(\boldsymbol{x}), \dots, \delta_m(\boldsymbol{x}) \approx \varepsilon_m(\boldsymbol{x})$ that satisfy*

$$\{\Delta_i(\delta_j(\boldsymbol{x}), \varepsilon_j(\boldsymbol{x})) \ : \ i \leq n, j \leq m\} \dashv\vdash_S \boldsymbol{x}. \tag{3.4}$$

Proof: See [Blok and Pigozzi, 1992, Theorem 13.15]. ■

Note that if condition (3.4) holds in a k-deductive system, then it continues to hold in every extension over the same language. Hence, any extension of an algebraizable deductive system is itself algebraizable.

By Theorem 3.3.12, every algebraizable k-deductive system is equivalential. The other inclusion does not hold. Examples of equivalential deductive systems that are not algebraizable include all normal modal logics with a local semantic consequence relation (see the appendix). In Example 3.2.9 a set of equivalence formulas for these logics has been given. A proof of the fact that local normal modal logics are not algebraizable can be found in [Blok and Pigozzi, 2001, Theorem V.3.10].

3.3.13. Example The formula $p \leftrightarrow q$ and the equation $x \approx \top$ form a set of equivalence formulas and a defining equation for CPC, for IPC and also for any modal logic with a global semantic consequence relation. By Theorem 5.8 in [Blok and Pigozzi, 1989], the set $\{p \to q, q \to p\}$ together with the equation $x \wedge (x \to x) \approx x \to x$ can be taken as a set of congruence formulas and a defining

[4] Some authors require the translations under which S and $S_{\mathbb{K}}$ are equivalent to be *finite*. This requirement can be easily justified as almost all examples of algebraizable logics, if not all, are algebraizable in this stricter sense. Also, for some purposes the finiteness of the translations is essential. This is for example the case in the algebraic characterization of the deduction property that is given in [Blok and Pigozzi, 2001, Theorem VI.1.3]. It is unclear whether a generalization of this theorem exists to the non-finite case. However, for the present purposes the finiteness of the translations is not needed, and we therefore do not require this.

equation for the relevance logics R and RM. The same set of formulas and the equation $x \approx x \to x$ can be taken as such for the implicational logic BCK (cf. [Blok and Pigozzi, 1989, Theorem 5.10]), and the n-valued Łukasiewicz logics. By Theorem 3.3.12 this implies that all of the aforementioned logics are algebraizable.

As is well-known, the equivalent quasivariety of IPC is the class of Heyting algebras. In [Blok and Pigozzi, 1989, Theorem 5.11] it is shown that the class of BCK-algebras, which was introduced by Iséki in 1966 as a class of algebras related to the calculus BCK, is indeed the equivalent quasivariety of this logic. More on the algebraic semantics for relevance logics and many-valued logics can be found in section 3.5, page 95 and further. ⊣

For algebraizable logics, the characterization of the Beth property takes the form of Theorem 3.3.14. This theorem was first proven for $k = 1$ in [Hoogland, 2000, Theorem 1].

3.3.14. Theorem *Let S be an algebraizable k-deductive system, with equivalent quasivariety $\mathbb{K}$. S has the Beth property iff all $\mathbb{K}$-epimorphisms are surjective.*

Proof: Let S be equivalent to $S_{\mathbb{K}}$, for some quasivariety $\mathbb{K}$. By Theorem 3.3.7, S has the Beth property iff $S_{\mathbb{K}}$ has this property. By Theorem 3.3.8 this is the case if and only if $Mod^*(S_{\mathbb{K}})$ has ES. By the observation above, this is equivalent to the fact that $\mathbb{K}$ has ES. ■

3.4 Characterizing the projective Beth property

In subsection 2.2.3 we defined the projective Beth property for propositional logics. In a general abstract algebraic context this property can be formulated as follows.

3.4.1. Definition [**Projective Beth property**] Let S be a k-deductive system, and let P, Q, R be disjoint sets of variables such that $Fm(P) \neq \emptyset$. A set of k-formulas $\Gamma \subseteq Fm^k(P \cup Q \cup R)$ is a *projective implicit definition of R in terms of P via Q* if for every $r \in R$ and every substitution σ that fixes the variables in P the following holds:

For any $\mathcal{A} \in Mod^*S$, any $X \subset Var$ that contains $P \cup Q \cup R \cup \sigma Q \cup \sigma R$,

$$\Gamma, \sigma\Gamma \models_{\mathcal{A}(X)} \langle r, \sigma(r) \rangle.$$

S has the *projective Beth property* if for every projective implicit definition Γ of R in terms of P via Q, and every $r \in R$, there exists an explicit definition $\varphi_r \in Fm(P)$ of r in terms of P with respect to Γ, in the sense of Definition 3.3.1. ⊣

Obviously, the Beth property is an instance of the projective Beth property by taking $Q = \emptyset$. [Maksimova, 1999b, Theorem 3.1] and [Maksimova, 1999a, Theorem 3.6] give an algebraic equivalent of the projective Beth property in the context of intermediate logics and normal modal logics. This algebraic equivalent is a slight variant of surjectiveness of what I propose to call *projective epimorphisms.*

3.4.2. Definition [**Projective epimorphism**] Let $\mathbb{K}$ be a class of matrices, and $\mathcal{A}, \mathcal{B} \in \mathbb{K}$. A matrix homomorphism $h : \mathcal{A} \longrightarrow \mathcal{B}$ is a *projective* $\mathbb{K}$*-epimorphism* if h is either surjective, or there exists some $x \in B \backslash h(A)$ such that for any $\mathcal{C} \in \mathbb{K}$ and any pair of matrix homomorphisms $k, k' : \mathcal{B} \longrightarrow \mathcal{C}$ it is the case that

$$k \circ h = k' \circ h \text{ implies that } k(x) = k'(x). \tag{3.5}$$

In this case we call x a *witness* (of the fact that f is a projective epimorphism). ⊣

Note that for an epimorphism f, the implication in (3.5) holds *for all* $x \in B \backslash f(A)$. Hence, epimorphisms are examples of projective epimorphisms. Now we have the following generalization of Maksimova's aforementioned results.

3.4.3. Theorem (Characterization of projective Beth property) *Let S be an equivalential k-deductive system. S has the projective Beth property iff all projective epimorphisms of Mod^*S are surjective.*

Proof: Let S be an equivalential k-deductive system. We will follow the outline of the proof of Theorem 3.3.8, mutatis mutandis.

First, assume that S has the projective Beth property. Let $f : \mathcal{A} \longrightarrow \mathcal{B} = \langle \boldsymbol{B}, F_B \rangle$ be a projective Mod^*S-epimorphism. Suppose f is not surjective. By Definition 3.4.2 this implies the existence of some $b \in B \backslash f(A)$ for which (3.5) holds. We will derive a contradiction.

As before, consider a set $Z \subset Var$ with the same cardinality as B. Let $h : \boldsymbol{Fm}(Z) \twoheadrightarrow \boldsymbol{B}$ be an algebra homomorphism which maps Z onto B. By Z_A (resp. $Z_{B \backslash A}$) we understand the sentential variables in Z which are mapped by h to $f(A)$ (resp. $B \backslash f(A)$). We also distinguish $z_b \in Z$ as the variable that is mapped by h to b.

We claim that $h^{-1}(F_B)$ is a projective implicit definition of z_b in terms of Z_A via $Z_{B \backslash A}$. As the reader may verify, this can be shown in a similar way as in the proof of Theorem 3.3.8.

Since S has the projective Beth property, it follows that z_b has some explicit definition φ_{z_b} in terms of Z_A with respect to $h^{-1}(F_B)$. As in the proof of Theorem 3.3.8, this implies that $b = h(\varphi_{z_b}) \in f(A)$. But by assumption $b \notin f(A)$. Contradiction.

For the other direction, assume that all projective Mod^*S-epimorphisms are surjective. Let P, Q, R be disjoint sets of variables such that $Fm(P) \neq \emptyset$, and let

$\Gamma \subseteq Fm^k(P \cup Q \cup R)$ be a projective implicit definition of R in terms of P via Q. We have to show that every $r \in R$ has an explicit definition in terms of P with respect to Γ.

Let $\mathcal{A} = \langle \boldsymbol{A}, F_A \rangle \in Mod^*S$, let X be a set of sentential variables that contains $P \cup Q \cup R$ and let the map $h : \boldsymbol{Fm}(X) \longrightarrow \boldsymbol{A}$ be such that $h(\Gamma) \subseteq F_A$. Let $r_0 \in R$. Again by Lemma 3.3.4 it suffices to find some $\varphi_{r_0} \in Fm(P)$ such that $h(r_0) = h(\varphi_{r_0})$.

To this end, consider the two matrices $\mathcal{H}_P = \langle h(\boldsymbol{Fm}(P)), F_A \cap h(Fm(P)^k) \rangle \subseteq \mathcal{H}_{P,Q,R} = \langle h(\boldsymbol{Fm}(P \cup Q \cup R)), F_A \cap h(Fm(P \cup Q \cup R)^k) \rangle$. Our aim is to show that $\mathcal{H}_P = \mathcal{H}_{P,Q,R}$. By our assumption it suffices to prove that $id : \mathcal{H}_P \longrightarrow \mathcal{H}_{P,Q,R}$ is a projective epimorphism in Mod^*S. In fact, we claim that $h(r)$ witnesses this fact, for every $r \in R$.

Consider some $\mathcal{C} = \langle \boldsymbol{C}, F_C \rangle \in Mod^*S$ together with a pair of matrix homomorphisms $k, k' : \mathcal{H}_{P,Q,R} \longrightarrow \mathcal{C}$ such that $k_{\restriction h(Fm(P))} = k'_{\restriction h(Fm(P))}$. Our task is to show that $k(h(r)) = k'(h(r))$, for every $r \in R$.

Let Y be a set of variables disjoint from X, of the same size as $Q \cup R$. Let $i : Y \rightarrowtail\!\!\!\!\rightarrow Q \cup R$ be a bijection, and let σ be a substitution which interchanges y and $i(y)$, for every $y \in Y$, and which fixes all other variables. Define the homomorphism $g : \boldsymbol{Fm}(X \cup Y) \longrightarrow \boldsymbol{C}$ by

$$\begin{aligned} g(x) &= k(h(x)), && \text{for } x \in X, \\ g(y) &= k'(h(i(y))), && \text{for } y \in Y. \end{aligned}$$

As before, for all $\varphi \in Fm(P \cup Q \cup R)$ it is the case that $g(\varphi) = k(h(\varphi))$ and $g(\sigma\varphi) = k'(h(\varphi))$. Again, this implies that $g(\Gamma) \cup g(\sigma\Gamma) \subseteq F_C$. Since Γ is a projective implicit definition of R in terms of P, we obtain for every $r \in R$ that $g(r) = g(\sigma r)$. In particular $k(h(r_0)) = g(r_0) = g(\sigma r_0) = k'(h(r_0))$, as required.

This completes the proof of Theorem 3.4.3. ■

Let *singleton projective Beth* and *infinite projective Beth* denote the obvious variants with respect to the cardinality of R. A closer inspection of the above proof reveals the following implications: S has the singleton projective Beth property $\Rightarrow$ all projective Mod^*S-epis are surjective $\Rightarrow$ S has the infinite projective Beth property. Hence, the singleton projective Beth property implies the infinite projective Beth property. Since the other implication trivially holds, we obtain the following equivalence.

3.4.4. Corollary *Let S be an equivalential k-deductive system. Then S has the infinite projective Beth property iff S has the singleton projective Beth property.*

Like Corollary 3.3.11, the above corollary is another example where algebraic methods lead to non-trivial logical results.

3.5 Examples and applications

This section is divided in two parts. In the first part we formulate a general fact which specifies sufficient conditions on a class of algebras $\mathbb{K}$ to have non-surjective $\mathbb{K}$-epimorphisms. As a corollary we obtain, among others, the failure of ES in the classes of Stone algebras, double Stone algebras and De Morgan algebras. This answers open questions in [Kiss et al., 1983]. The second part of this section contains some applications of the characterization in Theorem 3.3.8. This yields a uniform proof of failure of ES in the class of Tarski algebras and in the class of Hilbert algebras. We also apply Theorem 3.3.8 in the other direction, which leads to a purely algebraic proof of failure of the Beth property in a large number of relevance logics and many-valued logics.

Throughout this section, we refer to the distributive lattices $\boldsymbol{M_2}$, $\mathbf{0} \oplus \boldsymbol{M_2} \oplus \mathbf{1}$ and $\boldsymbol{C_3}$. These are described in Figure 3.3. The terminology is standard in lattice theory.

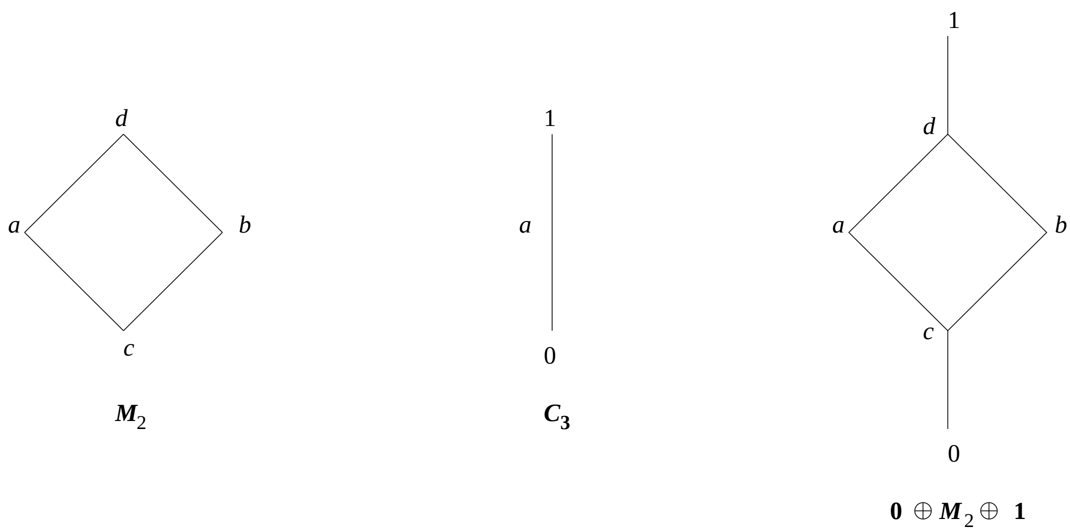


Figure 3.3: Examples of distributive lattices.

3.5.1 Sufficient conditions for failure of ES

Let $\boldsymbol{A}$ be the subalgebra of $\boldsymbol{M_2}$ with universe $\{a, c, d\}$. Example 2.5.13 showed that $\boldsymbol{A}$ is an epic subalgebra of $\boldsymbol{M_2}$ in the class of distributive lattices. This implies that the sublattice of $\mathbf{0} \oplus \boldsymbol{M_2} \oplus \mathbf{1}$ with universe $\{0, a, c, d, 1\}$ is also a proper epic subalgebra in the class of distributive lattices. We will come back to these examples. First we observe the following fact. Below, for any $\mathcal{L}$-algebra $\boldsymbol{A}$ and $\mathcal{L}' \subseteq \mathcal{L}$, we denote by $\boldsymbol{A}'$ the $\mathcal{L}'$-reduct of $\boldsymbol{A}$.

3.5.1. Fact Let $\mathbb{K}$ be a class of algebras over a language $\mathcal{L}$, and let $\mathcal{L}' \subseteq \mathcal{L}$. $\mathbb{K}'$ denotes the class of $\mathcal{L}'$-reducts of algebras in $\mathbb{K}$. Suppose $\mathbb{K}$ contains algebras $\boldsymbol{A}$, $\boldsymbol{B}$ such that

1. $\boldsymbol{A}$ is a proper subalgebra of $\boldsymbol{B}$,
2. $\boldsymbol{A}'$ is an epic subalgebra of $\boldsymbol{B}'$ in the class $\mathbb{K}'$.

Then ES fails in any quasivariety $\mathbb{K}''$ of $\mathcal{L}''$-reducts of algebras in $\mathbb{K}$, provided $\mathcal{L}' \subseteq \mathcal{L}'' \subseteq \mathcal{L}$ and $\boldsymbol{B}'' \in \mathbb{K}''$. In fact, $\boldsymbol{A}''$ is a proper epic subalgebra of $\boldsymbol{B}''$ in the class $\mathbb{K}''$. ⊣

In the following two paragraphs we apply Fact 3.5.1 together with the above examples on failure of ES in the class of distributive lattices in order to obtain results on failure of ES in some classes of algebras with a distributive lattice reduct. These results answer open questions in [Kiss et al., 1983].

Stone algebras An algebra $\boldsymbol{L} = \langle L, \wedge, \vee, *, 0, 1\rangle$ is a *pseudocomplemented lattice* if $\langle L, \wedge, \vee, 0, 1\rangle$ is a bounded distributive lattice and for every $x, y \in L$,

$$x \wedge y \approx 0 \Leftrightarrow y \leq x^*.$$

If, moreover, for every $x \in L$,

$$x^* \vee x^{**} \approx 1,$$

then $\boldsymbol{L}$ is known as a *Stone algebra*. A *dual pseudocomplemented lattice* $\boldsymbol{L'} = \langle L, \wedge, \vee, +, 0, 1\rangle$ satisfies for all $x, y \in L$, $x \vee y \approx 1 \Leftrightarrow x^+ \leq y$. If, moreover, for every $x \in L$, $x^+ \wedge x^{++} \approx 0$, then $\boldsymbol{L'}$ is a *dual Stone algebra*. A structure $\langle L, \wedge, \vee, *, +, 0, 1\rangle$ is called a *double Stone algebra* if $\langle L, \wedge, \vee, *, 0, 1\rangle$ is a Stone algebra and $\langle L, \wedge, \vee, +, 0, 1\rangle$ is a dual Stone algebra. A *double pseudocomplemented lattice* is defined analogously.

The study of pseudocomplemented lattices started with [Glivenko, 1929]. The notion of a Stone algebra first appeared in [Grätzer and Schmidt, 1957]. An accessible exposé on pseudocomplemented lattices (which includes a treatment of Stone algebras) can be found in [Balbes and Dwinger, 1974, Chapter VIII]

3.5.2. Proposition *The following classes do not have the property ES: Stone algebras, dual Stone algebras, double Stone algebras, pseudocomplemented lattices, dual pseudocomplemented lattices, double pseudocomplemented lattices.*

Proof: Let $\mathcal{L}' = \{\wedge, \vee\}, \mathcal{L} = \mathcal{L}' \cup \{+, *\}, \mathcal{L}'' = \mathcal{L}' \cup \{*\}$ and $\mathcal{L}''' = \mathcal{L}' \cup \{+\}$. Let $\mathbb{K}$ be the class of double Stone algebras. Consider the algebra $\boldsymbol{B} = \langle \mathbf{0} \oplus \boldsymbol{M_2} \oplus \mathbf{1}, *, +\rangle$, where the negation operation $*$ is defined by $0^* = 1$, and $1^* = a^* = b^* = c^* = d^* = 0$ and the dual negation operation $+$ is defined by

$1^+ = 0$, and $0^+ = a^+ = b^+ = c^* = d^+ = 1$. It is straightforward to check that $\boldsymbol{B}$ is a double Stone algebra, hence in particular a double pseudocomplemented lattice. Moreover, $\boldsymbol{B}''$ is a Stone algebra (hence, a pseudocomplemented lattice) and $\boldsymbol{B}'''$ is a dual Stone algebra (therefore, a dual pseudocomplemented lattice). Note that the set $B \setminus \{b\}$ is the universe of a subalgebra $\boldsymbol{A} \subset \boldsymbol{B}$. Finally, as we observed at the beginning of this subsection, $\boldsymbol{A}'$ is an epic subalgebra of $\boldsymbol{B}'$ in the class $\mathbb{K}'$. From Fact 3.5.1 it now directly follows that ES fails in all of the classes mentioned in the proposition. ■

De Morgan algebras An algebra $\langle L, \wedge, \vee, N\rangle$ is called a *De Morgan lattice* if $\langle L, \wedge, \vee\rangle$ is a distributive lattice and N is an involution of L which satisfies the De Morgan laws, i.e., for all $x, y \in L$,

- $NNx \approx x$.
- $N(x \wedge y) \approx Nx \vee Ny$.
- $N(x \vee y) \approx Nx \wedge Ny$.

De Morgan lattices were introduced in [Moisil, 1935]. In [Kalman, 1958] they are studied under the name of *distributive i-lattices*. A bounded De Morgan lattice $\langle L, \wedge, \vee, N, 0, 1\rangle$ is called a *De Morgan algebra*. De Morgan algebras are also known, e.g., in [Rasiowa, 1974], as *quasi-Boolean algebras*. A good introduction to the theory of De Morgan algebras is [Balbes and Dwinger, 1974, Chapter XI].

3.5.3. Proposition *ES fails in the class of De Morgan lattices and in the class of De Morgan algebras.*

Proof: Let $\mathcal{L}' = \{\wedge, \vee\}, \mathcal{L} = \mathcal{L}' \cup \{N, 0, 1\}$ and $\mathcal{L}'' = \mathcal{L}' \cup \{N\}$. Let $\mathbb{K}$ be the class of De Morgan algebras. Consider the algebra $\boldsymbol{B} = \langle \boldsymbol{M_2}, N, 0, 1\rangle$, where $0^B = c$, $1^B = d$ and the involutive operation N is defined by $Nd = c, Nc = d, Na = a$ and $Nb = b$. It is easy to verify that $\boldsymbol{B}$ is a De Morgan algebra. Note that $\boldsymbol{B}''$ is a De Morgan lattice. Moreover, the set $\{d, a, c\}$ is the universe of a subalgebra $\boldsymbol{A} \subset \boldsymbol{B}$. By Example 2.5.13, $\boldsymbol{A}'$ is an epic subalgebra of $\boldsymbol{B}'$ in the class $\mathbb{K}'$. From Fact 3.5.1 we conclude that ES fails in the variety of De Morgan algebras and De Morgan lattices. ■

The following example concerns a subclass of De Morgan algebras. It is illustrative to compare it to the previous example as this time, the result does not follow from Fact 3.5.1. Instead, it relies on the fact that the algebra into which we map satisfies the extra Kleene axiom.

Kleene algebras A *Kleene algebra* $\langle L, \wedge, \vee, N, 0, 1\rangle$ is a De Morgan algebra where for all $x, y \in L$,

$$x \wedge Nx \leq y \vee Ny.$$

Kleene algebras were studied in [Brignole and Monteiro, 1967]. They are called *normal i-lattices* in [Kalman, 1958].

Note that the De Morgan algebra $\boldsymbol{B}$ in the previous example is not a Kleene algebra.

3.5.4. Proposition *ES fails in the class of Kleene algebras.*

Proof: Let $\boldsymbol{B} = \langle \boldsymbol{C_3}, N\rangle$, where the involution N is defined by $N1 = 0, N0 = 1, Na = a$. One simply checks that $\boldsymbol{B}$ is a Kleene algebra. Moreover, the set $\{1, 0\}$ is the universe of a subalgebra $\boldsymbol{A} \subset \boldsymbol{B}$. We claim that $\boldsymbol{A}$ is an epic subalgebra of $\boldsymbol{B}$ in the class of Kleene algebras. For, let g, h be homomorphisms from $\boldsymbol{B}$ into some Kleene algebra. Then, $g(a) = g(Na) = N(ga)$. Similarly, $h(a) = N(ha)$. Therefore, by using the Kleene law, $g(a) = g(a) \wedge N(ga) \leq h(a) \vee N(ha) = h(a)$. Analogously, one shows that $h(a) \leq g(a)$, whence $h(a) = g(a)$. ■

The table in [Kiss et al., 1983] includes results on Stone algebras, double Stone algebras, De Morgan algebras and Kleene algebras, and it is indicated that all of these classes have the amalgamation property. However, the question of whether any of these classes has the property ES remained open. Our results above complete the table. Another interesting aspect of these examples is that they affirm the difference between the properties AP and ES.

3.5.2 Applications of the characterization theorem

In this subsection we present some applications of the characterization in Theorem 3.3.8. A nice example concerns the relevance logic B as it provides an example of an application of Theorem 3.3.8 to a non-algebraizable logic.

Tarski algebras, Hilbert algebras and $\rightarrow$-fragments of intermediate logics Let $\mathbf{2}^{\rightarrow}$ denote the $\rightarrow$-reduct of the two-element Boolean algebra. That is, $\mathbf{2}^{\rightarrow} = \langle\{0, 1\}, \rightarrow\rangle$, where $0 \rightarrow 0 = 0 \rightarrow 1 = 1 \rightarrow 1 = 1$ and $1 \rightarrow 0 = 0$. By the variety of *Tarski Algebras* (TA) we understand the variety generated by $\mathbf{2}^{\rightarrow}$. Rasiowa uses the term *implication algebras* for these structures, cf. [Rasiowa, 1974, page 30]. TA is known to have the congruence extension property. Moreover, as $\mathbf{2}^{\rightarrow}$ is the only subdirectly irreducible Tarski algebra, it is immediate that the class of subdirectly irreducible Tarski algebras has ES and the amalgamation property. We recall from chapter 2 the following theorem by Grätzer, Jonsson and Lakser.

3.5.5. Theorem [Grätzer et al., 1973, Theorem 3] *Let $\mathbb{V}$ be a variety with the congruence extension property. If the class of subdirectly irreducible members of $\mathbb{V}$ is closed under subalgebras and has the amalgamation property, then $\mathbb{V}$ has the amalgamation property.*

From this theorem we conclude that the variety of Tarski algebras has the amalgamation property.

A *Brouwerian semilattice* (BSL) is an algebra $\boldsymbol{L} = \langle L, \wedge, \rightarrow, 1\rangle$ where $\langle L, \wedge, 1\rangle$ is a meet semilattice with 1 and $\rightarrow$ is a relative pseudo-complementation operator. That is, for all $x, y \in L$,

$$x \rightarrow y = \max\{z \in L : z \wedge x \leq y\}.$$

For readers familiar with the notion of *residuation* we mention that this is equivalent to saying that $\rightarrow$ is a residual of $\wedge$. I.e., for all $x, y, z \in L$, $z \leq x \rightarrow y$ if and only if $z \wedge x \leq y$. Subalgebras of $\rightarrow$-reducts of Brouwerian semilattices are called *Hilbert algebras* (HI). It is well-known that HI is a variety, cf. [Cornish, 1982]. In [Pałasińska, 1988, Theorem 1] it is shown that this variety has the amalgamation property.

Note that any Tarski algebra is a Hilbert algebra. For, the two-element Boolean algebra is clearly a Brouwerian semilattice. Therefore, the $\rightarrow$-reduct of this algebra (i.e., the Tarski algebra $\mathbf{2}^{\rightarrow}$) is a Hilbert algebra. Since the variety of Tarski algebras is generated by the algebra $\mathbf{2}^{\rightarrow}$, we conclude that TA is included in the variety HI.

TA is the equivalent quasivariety semantics for the implicational fragment of classical propositional calculus, $\mathsf{CPC}^{\rightarrow}$, as defined in subsection 3.3.3. The variety of Hilbert algebras forms the equivalent quasivariety semantics for the implicational fragment of intuitionistic propositional calculus, $\mathsf{IPC}^{\rightarrow}$. For more information, see [Blok and Pigozzi, 1989, Section 5.2]. From the aforementioned results on amalgamation for TA and HI and the characterization of the interpolation property in [Czelakowski, 1982] it follows that the systems $\mathsf{CPC}^{\rightarrow}$ and $\mathsf{IPC}^{\rightarrow}$ have the interpolation property $\mathrm{CIP}^{\models}$. For $\mathsf{IPC}^{\rightarrow}$ this has directly been proven in [Renardel de Lavalette, 1989a].

The notion of an *intermediate logic* can be found in the appendix. We just recall that any intermediate logic S is such that $\models_{\mathsf{IPC}} \subseteq \models_S \subseteq \models_{\mathsf{CPC}}$.

3.5.6. Proposition *Let $S^{\rightarrow}$ be the implicational fragment of some intermediate logic S. Then $S^{\rightarrow}$ does not have the Beth property.*

Proof: The proof is an elaboration of Example 2.2.7. Consider an intermediate logic S, and let Γ be the usual definition of conjunction in terms of the implication. That is, $\Gamma = \{r \rightarrow p_1, r \rightarrow p_2, p_1 \rightarrow (p_2 \rightarrow r)\}$. For any variable r', let $\Gamma(r')$ denote the result of substituting r' for r in Γ. As the reader may verify, $\Gamma, \Gamma(r') \models_{\mathsf{IPC}}$

$\{r \to r', r' \to r\}$, for every variable r'. Therefore, $\Gamma, \Gamma(r') \models_{S^{\to}} \{r \to r', r' \to r\}$. In other words, Γ implicitly defines r in $S^{\to}$. However, in Example 2.2.7 we noted that even in $\mathsf{CPC}^{\to}$ (hence, certainly in $S^{\to}$), an explicit definition of r with respect to Γ can not be found. To prove this statement, suppose r has an explicit definition φ_r with respect to Γ in $\mathsf{CPC}^{\to}$. Consider the Tarski algebra $\boldsymbol{B} = \mathbf{2}^{\to} \times \mathbf{2}^{\to}$, and let $\boldsymbol{A}$ denote the subalgebra of $\boldsymbol{B}$ with domain $B \setminus \{0\}$. See also Figure 3.4.

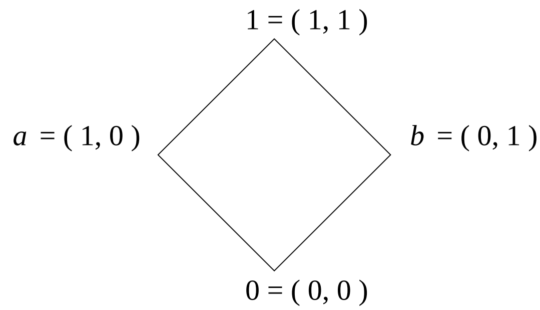


Figure 3.4: The Tarski algebra $\mathbf{2}^{\to} \times \mathbf{2}^{\to}$.

Consider the homomorphism $h : \boldsymbol{Fm}(\{p_1, p_2, r\}) \longrightarrow \boldsymbol{B}$ defined by $h(p_1) = a, h(p_2) = b$ and $h(r) = 0$. Note that $h(Fm(p_1, p_2)) \subseteq A$. Moreover, $h(\Gamma) = 1$. As $\Gamma \models_{TA} \{r \to \varphi_r, \varphi_r \to r\}$, then $h(r) = h(\varphi_r)$. But $h(\varphi_r) \in h(Fm(p_1, p_2)) \subseteq A$, and $h(r) = 0 \notin A$. Contradiction. We conclude that an explicit definition of r with respect to Γ does not exist in $S^{\to}$. ■

In particular, Proposition 3.5.6 implies that $\mathsf{CPC}^{\to}$ and $\mathsf{IPC}^{\to}$ do not have the Beth property.

With any quasivariety $\mathbb{K}$ of $\to$-reducts of Brouwerian semilattices that contains the algebra $\mathbf{2}^{\to}$ one may associate an intermediate logic S such that $\mathbb{K}$ is the equivalent quasivariety semantics of the $\to$-fragment of S. By virtue of Theorem 3.3.14 we obtain the following corollary.

3.5.7. Corollary *Let $\mathbb{K}$ be a quasivariety of $\to$-reducts of Brouwerian semilattices that contains the algebra $\mathbf{2}^{\to}$. Then $\mathbb{K}$ does not have the property ES.*

It follows that the variety of Tarski algebras and the variety of Hilbert algebras do not have the property ES. The meaning of this fact is twofold. First, it shows once more the different character of the properties ES and AP. Second, it implies that a theorem analogous to Theorem 3.5.5 but formulated for ES instead of AP requires some extra conditions.

Below, we present two Tarski algebras $\boldsymbol{A}, \boldsymbol{B}$ such that $\boldsymbol{A}$ is a proper epic subalgebra of $\boldsymbol{B}$ in the variety of Hilbert algebras. This directly shows that ES fails in the class of Tarski algebras and in the class of Hilbert algebras.

3.5.8. Example Note that the following identity is satisfied in any Brouwerian semilattice,

$$x \to (y \to z) \approx (x \wedge y) \to z. \tag{3.6}$$

Let $\boldsymbol{A}, \boldsymbol{B}$ denote the Tarski algebras defined in the proof of Proposition 3.5.6. That is, $\boldsymbol{B} = \mathbf{2}^{\to} \times \mathbf{2}^{\to}$, and $\boldsymbol{A}$ is the subalgebra of $\boldsymbol{B}$ with domain $B \setminus \{0\}$. We claim that $\boldsymbol{A}$ is an epic subalgebra of $\boldsymbol{B}$ in the variety of Hilbert algebras. To see this, consider some homomorphism f from $\boldsymbol{B}$ into some Hilbert algebra $\boldsymbol{C}$. Then $\boldsymbol{C}$ is a subalgebra of the $\to$-reduct of some Brouwerian semilattice $\boldsymbol{D}$. Now, $f(a) \to (f(b) \to f(0)) = f(a) \to f(a) = f(1) = 1^D$. Hence, by (3.6), $(f(a) \wedge^D f(b)) \to f(0) = 1^D$. Therefore, $(f(a) \wedge^D f(b)) \leq f(0)$. On the other hand, since $f(0) \to f(a) = f(1)$, we have that $f(0) \leq f(a)$. Similarly, $f(0) \leq f(b)$. Therefore, $f(0) \leq (f(a) \wedge^D f(b))$. Summarizing, we see that $f(0) = f(a) \wedge^D f(b)$. Hence, $f(0)$ is uniquely determined by $f(a)$ and $f(b)$. We conclude that $\boldsymbol{A} \subset \boldsymbol{B}$ is epic in the class of Hilbert algebras. ⊣

For more on the failure of ES in related classes of BCK-algebras, the reader is referred to [Pałasińska, 1988].

MV-algebras and many-valued logics For any $n \geq 3$, we define the n-element MV-algebra $\boldsymbol{L_n} = \langle \{0, \ldots, n-1\}, 0, n-1, +_{n-1}, \dot{-}, \wedge, \vee \rangle$, where $\wedge, \vee$ are the usual lattice operations, and for all $m, m' \in \{0, \ldots, n-1\}$,

- $m +_{n-1} m' = \min(m + m', n-1)$.
- $m \dot{-} m' = \max(m - m', 0)$.

MV-algebras were introduced and studied in [Chang, 1958]. For a class of algebras $\mathbb{K}$, let $Q(\mathbb{K})$ denote the quasivariety generated by $\mathbb{K}$.

3.5.9. Proposition $Q(\{\boldsymbol{L_n} : n \in \omega\})$ *does not have ES.*

Proof: Note that the following quasi-identity holds in $\boldsymbol{L_n}$, for any $n \in \omega$,

$$(x \dot{-} y \approx y \ \& \ x \dot{-} z \approx z) \to y \approx z. \tag{3.7}$$

Intuitively, this quasi-identity states that if $y = x/2$ and $z = x/2$, then $y = z$. Let $\boldsymbol{A}$ be the subalgebra of $\boldsymbol{L_3}$ with domain $\{0, 2\}$. We show that $\boldsymbol{A}$ is an epic subalgebra of $\boldsymbol{L_3}$ in the class $Q(\{\boldsymbol{L_n} : n \in \omega\})$. To this end, consider some $\boldsymbol{B} \in Q(\{\boldsymbol{L_n} : n \in \omega\})$, and $k, k' : \boldsymbol{A} \longrightarrow \boldsymbol{B}$ such that $k_{\restriction\{0,2\}} = k'_{\restriction\{0,2\}}$. Clearly $k(2) \dot{-} k(1) = k(1)$. Moreover, $k(2) \dot{-} k'(1) = k'(2) \dot{-} k'(1) = k'(1)$. Since $\boldsymbol{B}$ satisfies (3.7), it follows that $k(1) = k'(1)$. ■

By essentially the same argument it can be shown that,

3.5.10. Proposition *$Q(\boldsymbol{L_n})$ does not have ES, for any $n \geq 3$.*

Proof: Let $n \in \omega$. In the same spirit as before, the quasi-identity below expresses that if $y = x/n-1$ and $z = x/n-1$, then $y = z$.

$$(x \underbrace{\dot{-}y\dot{-}\ldots\dot{-}y}_{n-2 \text{ times}} \approx y \;\&\; x \underbrace{\dot{-}z\dot{-}\ldots\dot{-}z}_{n-2 \text{ times}} \approx z) \rightarrow y \approx z. \tag{3.8}$$

Obviously, (3.8) is valid in $\boldsymbol{L_n}$. We claim that the subalgebra of $\boldsymbol{L_n}$ with domain $\{0, n-1\}$ is epic in the class $Q(\boldsymbol{L_n})$. For, consider two homomorphisms k, k' from $\boldsymbol{L_n}$ into some algebra in $Q(\boldsymbol{L_n})$ such that $k(n-1) = k'(n-1)$. Reasoning as in the proof of Proposition 3.5.9, this time using the quasi-identity (3.8), we infer that $k(1) = k'(1)$. Since the element 1 generates $\boldsymbol{L_n}$, it follows that $k = k'$. We conclude that for $n \geq 3$, the subalgebra of $\boldsymbol{L_n}$ with domain $\{0, n-1\}$ is an example of a proper subalgebra which is epic in the class $Q(\boldsymbol{L_n})$. ■

MV-algebras form the algebraic counterpart of the many-valued logics introduced by Łukasiewicz and Tarski in 1930. An excellent reference to these, and other many-valued logics, is [Urquhart, 1984]. More precisely, the quasivariety of all MV-algebras (which happens to be a variety) is the equivalent quasivariety of the infinite-valued system $Ł_\infty$. For any $n \geq 3$, the quasivariety generated by the algebra $\boldsymbol{L_n}$ (which also turns out to be a variety) is the equivalent quasivariety of the n-valued Łukasiewicz logic $Ł_n$. This example will be discussed in [Blok and Pigozzi, 2001]. From the characterization in Theorem 3.3.8 and the above propositions we obtain the following corollary.

3.5.11. Corollary *For any $n \geq 3$, the many-valued logics $Ł_n$ do not have the Beth property. Moreover, the system $Ł_\infty$ does not have the Beth property either.*

3.5.12. Example For the reader familiar with Łukasiewicz's three-valued logic $Ł_3$ we mention the following. Define

$$\Gamma = (r \rightarrow p) \leftrightarrow r,$$

where $p \rightarrow q$ stands for $(2\dot{-}p) +_2 q$ and $p \leftrightarrow q$ is an abbreviation of $(p \rightarrow q) \wedge (q \rightarrow p)$. Using the algebraic semantics, it is straightforward to check that Γ implicitly defines r in the system $Ł_3$. However, an explicit definition of r with respect to Γ in $Ł_3$ can not be found. For, suppose such explicit definition $\varphi_r \in Fm(p)$ does exist. Consider the homomorphism $h : \boldsymbol{Fm}(\{p, r\}) \longrightarrow \boldsymbol{L_3}$ defined by $h(p) = 0$ and $h(r) = 1$. Note that $h(\varphi_r) \in \{0, 2\}$. Moreover, $h(\Gamma) = 2$. As $\Gamma \models r \leftrightarrow \varphi_r$, then $h(r \leftrightarrow \varphi_r) = 2$, i.e., $h(r) = h(\varphi_r)$. But $h(\varphi_r) \in \{0, 2\}$, and $h(r) = 1$. Contradiction. We conclude that an explicit definition of r with respect to Γ does not exist. ⊣

Relevance algebras and relevance logics The example below is based on the example used in [Urquhart, 1999] to show that the Beth property fails in many relevance logics. We use it to falsify ES in a large number of relevance algebras. By our characterization in Theorem 3.3.8, this comes down to the same thing. However, we choose to include this example as in our presentation the algebraic methodology is nicely separated from the logical core. Moreover, our example includes the basic relevance logic B, which is not covered by Urquharts theorem. This provides an example of an application of Theorem 3.3.8 to a non-algebraizable logic.

The basic relevance logic B is defined in [Routley et al., 1982] by the following axiom schemes:

$$\begin{array}{lc} A1 & A \to A \\ A2 & A \wedge B \to A \\ A3 & A \wedge B \to B \\ A4 & ((A \to B) \wedge (A \to C)) \to (A \to (B \wedge C)) \\ A5 & A \to A \vee B \\ A6 & B \to A \vee B \\ A7 & ((A \to C) \wedge (B \to C)) \to ((A \vee B) \to C \\ A8 & A \wedge (B \vee C) \to ((A \wedge B) \vee (A \wedge C)) \\ A9 & \sim\sim A \to A \end{array}$$

The rules of B are as follows:

$$\begin{array}{lc} R1 & A \to B, A/B \\ R2 & A, B/A \wedge B \\ R3 & A \to B, C \to D/(B \to C) \to (A \to D) \\ R4 & A \to\ \sim B/B \to\ \sim A \end{array}$$

The following derived rule will prove to be useful.

3.5.13. Proposition *In the system* B*, from* $A \to B$ *and* $B \to C$ *one can derive* $A \to C$*.*

Proof: We reason in the system B. From $A \to B$ and $B \to C$ we derive, via R3, $(B \to B) \to (A \to C)$. By A1, $B \to B$ is an axiom. Hence, by modus ponens, we infer $A \to C$. ■

The well-known relevance logics E and R are extensions of B. Definitions can be found in [Anderson and Belnap, 1975]. In [Blok and Pigozzi, 1989] it is shown that R is algebraizable, but E is not. By Theorem 3.3.12 this implies that B is not algebraizable either. On the other hand, the following proposition states that B, and hence E, is equivalential.

3.5.14. Proposition *The basic relevance logic* B *is equivalential with equivalence system* $\{p \to q, q \to p\}$.

Proof: By A1 and R1 it follows directly that B is protoalgebraic. It remains to be shown that B satisfies the rule (Con) from Definition 3.2.8. We will verify that

$$q \to p \vdash_{\mathsf{B}} {\sim}p \to {\sim}q. \tag{3.9}$$

We have the following derivation in B.

(1)	$\vdash_{\mathsf{B}} {\sim}p \to {\sim}p$	*A*1
(2)	${\sim}p \to {\sim}p \vdash_{\mathsf{B}} p \to {\sim}{\sim}p$	*R*4
(3)	$q \to p, p \to {\sim}{\sim}p \vdash_{\mathsf{B}} q \to {\sim}{\sim}p$	3.5.13
(4)	$q \to {\sim}{\sim}p \vdash_{\mathsf{B}} {\sim}p \to {\sim}q$	*R*4

From (1)–(4) it is easy to derive that $q \to p \vdash_{\mathsf{B}} {\sim}p \to {\sim}q$. This proves (3.9). The other cases, for $\wedge, \vee$ and $\to$, are straightforward and are therefore left to the reader. ■

The next proposition gives some information about the reduced B-matrices.

3.5.15. Proposition *Every algebra-reduct of a reduced* B*-matrix is a distributive lattice with additional operations* $\to, N$.

Recall from page 73 that the class of algebra-reducts of reduced B-matrices is denoted by $Alg^*(\mathsf{B})$.

Proof: We first observe the following generality. Fix a set $X \subset Var$, and define $\Theta = \{\langle\varphi,\psi\rangle \in Fm^2(X) \ : \vdash_{\mathsf{B}} \varphi \to \psi \text{ and } \vdash_{\mathsf{B}} \psi \to \varphi\}$. It can be shown that Θ is a congruence on $\boldsymbol{Fm_B}(X)$ which is compatible with $Cn_B(\emptyset)$. Hence, $\Theta \subseteq \Omega^{\boldsymbol{Fm}(X)}(Cn_B(\emptyset))$. Suppose that $\vdash_{\mathsf{B}} \varphi \to \psi$. Then $\langle\varphi \wedge \psi, \varphi\rangle \in \Theta \subseteq \Omega^{\boldsymbol{Fm}(X)}(Cn_B(\emptyset))$. Therefore, the identity $\varphi \wedge \psi \approx \varphi$ is satisfied in the algebra $\boldsymbol{Fm}(X)/\Omega^{\boldsymbol{Fm}(X)}(Cn_B(\emptyset))$. By Theorem 3.2.13 this implies that $\varphi \wedge \psi \approx \varphi$ is satisfied in the class $Alg^*\mathsf{B}$. That is, $\varphi \leq \psi$ is valid in $Alg^*\mathsf{B}$. Therefore, in order to show that all the distributive lattice-identities are valid in $Alg^*\mathsf{B}$, it suffices to show that $\vdash_{\mathsf{B}} \varphi \to \psi$ and $\vdash_{\mathsf{B}} \psi \to \varphi$, for every such identity $\varphi \approx \psi$. For most identities this is straightforward, and these are left to the reader. We only verify the non-trivial direction of the distributive law. That is, we will show that

$$\vdash_{\mathsf{B}} ((x \wedge y) \vee (x \wedge z)) \rightarrow (x \wedge (y \vee z)). \tag{3.10}$$

We reason as follows.

(1)	$\vdash_{\mathsf{B}} (x \wedge y) \rightarrow y$	$A3$
(2)	$\vdash_{\mathsf{B}} y \rightarrow (y \vee z)$	$A5$
(3)	$(x \wedge y) \rightarrow y, y \rightarrow (y \vee z) \vdash_{\mathsf{B}} (x \wedge y) \rightarrow (y \vee z)$	3.5.13
(4)	$\vdash_{\mathsf{B}} (x \wedge y) \rightarrow (y \vee z)$	$(1), (2), (3)$
(5)	$\vdash_{\mathsf{B}} (x \wedge y) \rightarrow x$	$A2$
(6)	$\vdash_{\mathsf{B}} ((x \wedge y) \rightarrow x) \wedge ((x \wedge y) \rightarrow (y \vee z))$	$R2, (4), (5)$

By axiom A4,

$$\vdash_{\mathsf{B}} [((x \wedge y) \rightarrow x) \wedge ((x \wedge y) \rightarrow (y \vee z))] \rightarrow [(x \wedge y) \rightarrow (x \wedge (y \vee z))]. \tag{3.11}$$

From this and (6) we obtain via R1 that

$$\vdash_{\mathsf{B}} (x \wedge y) \rightarrow (x \wedge (y \vee z)). \tag{3.12}$$

Similarly, one shows that $\vdash_{\mathsf{B}} (x \wedge z) \rightarrow (x \wedge (y \vee z))$. From this theorem together with (3.11), (3.12), R1 and R2 we easily obtain (3.10). ■

The reduced R-matrices have been investigated in [Font and Rodríguez, 1990]. It turns out that the algebra-reducts of reduced R-matrices are structures of the form $\langle L, \wedge, \vee, N, \rightarrow \rangle$, where $\langle L, \wedge, \vee, N \rangle$ is a De Morgan lattice which satisfies for all $x, y \in L$,

- $x \rightarrow (y \rightarrow z) \leq y \rightarrow (x \rightarrow z)$.
- $x \leq ((x \rightarrow y) \wedge z) \rightarrow y$.
- $x \rightarrow Ny \leq y \rightarrow Nx$.
- $x \rightarrow Nx \leq Nx$.
- $((x \rightarrow x) \wedge (y \rightarrow y)) \rightarrow z \leq z$.

In [Font and Rodríguez, 1990] the above structures are called *relevance algebras*. By Theorem 18 in that same paper a matrix $\langle \boldsymbol{A}, F_A \rangle$ is a reduced matrix for R if and only if $\boldsymbol{A}$ is a relevance algebra and F_A is the lattice filter generated by the set $\{x \rightarrow x \ : \ x \in A\}$.

Consider the algebra $\boldsymbol{C} = \langle \mathbf{0} \oplus \boldsymbol{M_2} \oplus \mathbf{1}, N, \rightarrow \rangle$, where the implication and negation operations are given in the table below. The implication is defined such that $x \rightarrow y = 0$ if $x \not\leq y$, and moreover $x \rightarrow 1 = 1$, $x \rightarrow d = {\sim}x$, $0 \rightarrow y = 1$, $c \rightarrow y = y$, $a \rightarrow a = a$ and $b \rightarrow b = b$. In [Thistlewaite et al., 1988], $\boldsymbol{C}$ is called *the Crystal lattice* and it is shown that $\boldsymbol{C}$ is a relevance algebra.

->	**1**	**d**	**a**	**b**	**c**	**0**	**N**
1	1	0	0	0	0	0	0
d	1	c	0	0	0	0	c
a	1	a	a	0	0	0	a
b	1	b	0	b	0	0	b
c	1	d	a	b	c	0	d
0	1	1	1	1	1	1	1

Figure 3.5: The Crystal lattice.

Let $F_C = \{a, b, c, d, 1\}$. Note that F_C is the smallest lattice filter on $\boldsymbol{C}$ containing the set $\{x \rightarrow x \ : \ x \in 0 \oplus M_2 \oplus 1\}$. By the aforementioned theorem in [Font and Rodríguez, 1990], this implies that $\langle \boldsymbol{C}, F_C \rangle$ is a reduced R-matrix.

3.5.16. Theorem *Let $\mathbb{M}$ be a class of reduced B-matrices that is closed under submatrices and that contains the Crystal matrix $\langle \boldsymbol{C}, F_C \rangle$. Then ES fails in $\mathbb{M}$.*

Proof: Let $\mathcal{L}' = \{\wedge, \vee\}$ and set $\mathcal{L} = \mathcal{L}' \cup \{\sim, \rightarrow\}$. Let $\mathbb{K}$ be the class of algebra-reducts of matrices in $\mathbb{M}$. That is, $\mathbb{K} = \{\boldsymbol{A} \ : \ \text{there exists some } F_A \subseteq A \text{ such that } \langle \boldsymbol{A}, F_A \rangle \in \mathbb{M}\}$. Finally, let $\mathbb{K}'$ be the class of $\mathcal{L}'$-reducts of algebras in $\mathbb{K}$. Note that the Crystal lattice $\boldsymbol{C}$ has a subalgebra $\boldsymbol{C}^-$ with universe $\{0, a, c, d, 1\}$. Moreover, the $\mathcal{L}'$-reduct of $\boldsymbol{C}^-$ (i.e., the distributive lattice with universe $\{0, a, c, d, 1\}$) is an epic subalgebra of $\boldsymbol{C}'$ (i.e., the distributive lattice $\mathbf{0} \oplus \boldsymbol{M_2} \oplus \mathbf{1}$) in the class $\mathbb{K}'$: this follows from the fact that the distributive lattice with universe $\{0, a, c, d, 1\}$ is an epic subalgebra of $\mathbf{0} \oplus \boldsymbol{M_2} \oplus \mathbf{1}$ in the class of distributive lattices (according to the example at page 89) together with the fact that any algebra in $\mathbb{K}'$ is a distributive lattice (by Proposition 3.5.15). From Fact 3.5.1 it follows that ES fails in $\mathbb{K}$. To be precise, it follows that $\boldsymbol{C}^-$ is an epic subalgebra of $\boldsymbol{C}$ in the class $\mathbb{K}$.

Note that the matrix $\mathcal{C}^- = \langle \boldsymbol{C}^-, F_C \backslash \{b\} \rangle$ is a submatrix of the Crystal matrix $\mathcal{C} = \langle \boldsymbol{C}, F_C \rangle$. We claim that $\mathcal{C}^-$ is an epic submatrix of $\mathcal{C}$ in the class $\mathbb{M}$. To see this, consider some $\mathcal{D} \in \mathbb{M}$ and two matrix homomorphisms $f, f' : \mathcal{C} \longrightarrow \mathcal{D}$ such that $f_{\restriction C^-} = f'_{\restriction C^-}$. It needs to be shown that $f = f'$. This becomes apparent if we view f, f' as algebra homomorphisms from $\boldsymbol{C}$ to $\boldsymbol{D}$. Recall that we have just shown that $\boldsymbol{C}^-$ is an epic subalgebra of $\boldsymbol{C}$ in the class $\mathbb{K}$. Since $\boldsymbol{D} \in \mathbb{K}$, it follows that $f = f'$. We conclude that $\mathcal{C}^-$ is a proper epic submatrix of $\mathcal{C}$ in the class $\mathbb{M}$. Therefore, $\mathbb{M}$ does not have the property ES. ■

Let S be an extension of B. Note that S is equivalential, since B is equivalential. Moreover, Mod^*S is a class of reduced matrix models for B that (by Theorem 3.2.11) is closed under submatrices. Above, we showed that the Crystal matrix $\mathcal{C}$ is a reduced R-matrix. Hence, if R happens to be an extension of S, then $\mathcal{C} \in Mod^*S$. In this case, Mod^*S fulfills all the conditions in Theorem 3.5.16 and we conclude that ES fails in Mod^*S. From Theorem 3.3.8 we obtain the following corollary.

3.5.17. Corollary *Let S be a relevance logic between* B *and* R. *Then S does not have the Beth property.*

3.6 The weak Beth property in algebraic logic

In chapter 2 we discussed the weak Beth property. It is our final objective in this chapter to give an algebraic characterization of this property. However, there is one extra difficulty involved. As we have seen in subsection 2.2.3, the weak Beth property is an intrinsically *model-theoretic* notion. It can only be defined by referring explicitly to the models of the logic. Therefore, before we can even start looking for an algebraic equivalent of this property we need to make sure that our abstract algebraic framework is capable of dealing with the explicit semantic features of logics. Unfortunately, the framework we worked in so far does not accomplish this.[5] Therefore, we will introduce a different approach to abstract algebraic logic which treats logics as being equipped with a specific set of models. This will be established in the next subsection. In subsection 3.6.2 we reap the fruits of this approach and give a characterization of the weak Beth property. This gives a solution to Problem 14 in [Sain, 1990].

3.6.1 The model-theoretic framework

In this subsection we sketch the model-theoretic framework that has been introduced and developed by the so-called Budapest school in algebraic logic. Relevant references include [Andréka and Németi, 1994], [Andréka et al., 1994] and [Andréka et al., 1995].

[5] For this reason, this framework is often referred to as the 'syntactic' approach, as to distinguish it from the semantic approach we are about to introduce. However, there is nothing especially syntactic to the former approach. Recall that in this framework a logic is seen as pair consisting of a language and a consequence relation; in which way this consequence relation is defined is besides the point. That is, it may be syntactically specified as an inference system but it may also be an algebraic consequence relation or, for that matter, a semantic consequence relation.

As before, let us fix a proper class *Var* of sentential variables. Recall that the notion of a formula of type t over some set of variables X has been defined at page 65. Given a set $X \subset Var$, a *logic* $\mathcal{L}(X)$ is seen as an ordered four-tuple consisting of

1. A set $Fm_{\mathcal{L}}(X)$ of *formulas* of some type $Cn(\mathcal{L})$ over the set X.
2. A class $M^X_{\mathcal{L}}$ of *models*. In general this will be a *proper* class.
3. A family $\{mng^X_{\mathcal{M}} : \mathcal{M} \in M^X_{\mathcal{L}}\}$ of functions such that each $mng^X_{\mathcal{M}}$ has domain $Fm_{\mathcal{L}}(X)$. These functions are called *meaning functions*.
4. A relation $\models^X_{\mathcal{L}}$ (called the *validity relation*) between $M^X_{\mathcal{L}}$ and $Fm_{\mathcal{L}}(X)$. This relation is linked to the meaning functions in such a way that for all formulas $\varphi, \psi \in Fm_{\mathcal{L}}(X)$ and every model $\mathcal{M} \in M^X_{\mathcal{L}}$,

$$\text{If } \mathcal{M} \models^X_{\mathcal{L}} \varphi \text{ and } mng^X_{\mathcal{M}}(\varphi) = mng^X_{\mathcal{M}}(\psi), \text{ then } \mathcal{M} \models^X_{\mathcal{L}} \psi.$$

We often omit the subscript $\mathcal{L}$ and superscriptions like X, Y, etc. if it causes no confusion.

From the validity relation $\models_{\mathcal{L}}$, we define the *semantic consequence relation* associated with $\mathcal{L}$ as follows. For every $\Gamma \cup \{\varphi\} \subseteq Fm_{\mathcal{L}}(X)$,

$$\Gamma \models_{\mathcal{L}} \varphi \text{ iff } (\forall \mathcal{M} \in M_{\mathcal{L}})[\text{ If } \mathcal{M} \models_{\mathcal{L}} \gamma, \text{ for all } \gamma \in \Gamma, \text{ then } \mathcal{M} \models_{\mathcal{L}} \varphi].$$

As usual, the symbol $\models_{\mathcal{L}}$ denotes both the validity relation and the semantic consequence relation.

We note that the above set-up is so general it does not assume any further connection, besides the one mentioned in 4, between the meaning functions and the validity relation. Stronger connections will be imposed in Definition 3.6.4 below.

As before, we will be interested in the family $\{\mathcal{L}(X) : X \text{ is a set of variables}\}$ rather than in one particular logic $\mathcal{L}(X)$. In the present terminology such family is called a *general* logic. Recall that $\boldsymbol{Fm}_{\mathcal{L}}(X)$ denotes the formula algebra of type $Cn(\mathcal{L})$ generated by the set of variables X.

3.6.1. Definition [General logic] Let $Cn(\mathbf{L})$ be a language type. A *general logic* is a class

$$\mathbf{L} = \{\mathcal{L}(X) : X \text{ is a subset of } Var\},$$

where for every set $X \subset Var$ the quadruple $\mathcal{L}(X) = \langle Fm(X), M^X, mng^X, \models^X \rangle$ is a logic in the above sense of type $Cn(\mathbf{L})$. Moreover, for any sets of variables X, Y, the following conditions hold:

1. If there is a bijection $f : X \rightarrowtail\!\!\!\!\twoheadrightarrow Y$ then the logic $\mathcal{L}(Y)$ is an "isomorphic copy" of $\mathcal{L}(X)$. That is, there exists a bijection $f^M : M^X \rightarrowtail\!\!\!\!\twoheadrightarrow M^Y$ such that for all $\varphi \in Fm(X)$ and all $\mathcal{M} \in M^X$,

$$mng^X_{\mathcal{M}}(\varphi) = mng^Y_{f^M(\mathcal{M})}(f^F(\varphi)), \text{ and}$$

$$\mathcal{M} \models^X \varphi \Leftrightarrow f^M(\mathcal{M}) \models^Y f^F(\varphi).$$

   Here f^F is the unique isomorphism from $\boldsymbol{Fm}_{\mathcal{L}}(X)$ onto $\boldsymbol{Fm}_{\mathcal{L}}(Y)$ extending f.

2. If $X \subseteq Y$, then $\{mng^X_{\mathcal{M}} : \mathcal{M} \in M^X\} = \{(mng^Y_{\mathcal{M}})\restriction_{Fm(X)} : \mathcal{M} \in M^Y\}$. $\dashv$

Definition 3.6.1 can be interpreted as saying that only the cardinality of the set of atomic formulas really matters.

3.6.2. Example Classical propositional calculus (CPC) can be defined as the family $\{\langle Fm_{\mathsf{CPC}}(X), M^X_{\mathsf{CPC}}, mng^X_{\mathsf{CPC}}, \models^X_{\mathsf{CPC}}\rangle \;:\; X \text{ is a subset of } Var\}$, where for every set of variables X the following conditions are satisfied:

1. The set $Fm_{\mathsf{CPC}}(X)$ of formulas of type $\{\wedge, \neg\}$ over the set X is defined by induction as usual.

2. M^X_{CPC} is the set of X-valuations. That is, M^X_{CPC} is the set of maps from X to $\{0, 1\}$.

3. For each valuation $v \in M^X_{\mathsf{CPC}}$, the meaning function mng^X_v is the unique homomorphism from $\boldsymbol{Fm}_{\mathsf{CPC}}(X)$ to the two-element Boolean algebra **2** which extends v. (Usually, this meaning function is also denoted by v, but in this example we use the notation mng^X_v to be perfectly clear.)

4. For all $v \in M^X_{\mathsf{CPC}}$ and all $\varphi \in Fm_{\mathsf{CPC}}(X)$, $v \models^X_{\mathsf{CPC}} \varphi \overset{\text{def}}{\Leftrightarrow} mng^X_v(\varphi) = 1$.

This shows that CPC is an example of a general logic. $\dashv$

Modal logics are other examples of general logics.

3.6.3. Example The system K4 is the logic of all transitive frames. It can be defined as the family $\{\langle Fm_{\mathsf{K4}}(X), M^X_{\mathsf{K4}}, mng^X_{\mathsf{K4}}, \models^X_{\mathsf{K4}}\rangle \;:\; X \text{ is a subset of } Var\}$, where for every set $X \subset Var$ the following conditions are satisfied:

1. The set $Fm_{\mathsf{K4}}(X)$ of formulas of type $\{\wedge, \neg, \Diamond\}$ over the set X is defined by induction as usual. In particular, if $\varphi \in Fm_{\mathsf{K4}}(X)$, then $\Diamond\varphi \in Fm_{\mathsf{K4}}(X)$.

2. M^X_{K4} is the class of pairs $\langle\langle W, R\rangle, v\rangle$, where

- W is a non-empty set.
- R is a binary, transitive relation on W.
- v is a map from X to subsets of W.

3. The meaning function $mng^X_{\mathcal{M}}$ of a model $\mathcal{M} = \langle\langle W, R\rangle, v\rangle \in M^X_{\mathsf{K4}}$ is the unique homomorphism from the formula algebra $\boldsymbol{Fm}_{\mathsf{K4}}(X)$ to the complex algebra $\langle \mathcal{P}(W), \cap, \setminus, m_R\rangle$ which extends v. Here, $\mathcal{P}(W)$ denotes the powerset of W, and for all $X \subseteq W$, $m_R(X) = \{w \in W \ : \ \exists w' \in W(wRw' \text{ and } w' \in X)\}$. In other words, the meaning function $mng^X_{\mathcal{M}}$ maps a formula φ to the set of points in W in which φ is true.

4. For all $\mathcal{M} = \langle\langle W, R\rangle, v\rangle \in M^X_{\mathsf{K4}}$ and all $\varphi \in Fm_{\mathsf{K4}}(X)$, $\mathcal{M} \models^X_{\mathsf{K4}} \varphi \overset{\text{def}}{\Leftrightarrow} mng^X_{\mathcal{M}}(\varphi) = W$.

Similarly every class of frames defines a general logic. ⊣

As we already mentioned, Definition 3.6.1 is very general. Its purpose is to give a very broad description of what can possibly be conceived as a logic. But, looking at the bulk of logics studied in the literature, one can't help noticing several similarities that are much more far-reaching than those described in Definition 3.6.1. Some of these similarities are listed in the following definition. In particular, the validity relation is more closely tied up to the meaning functions.

3.6.4. Definition [**Semantically algebraizable-, structural logic**] A general logic **L** is called *semantically algebraizable* if the following conditions are met:

1. For every set $X \subset \mathit{Var}$ and every $\mathcal{M} \in M^X$, the function $mng^X_{\mathcal{M}}$ is a homomorphism from $\boldsymbol{Fm}(X)$ into some algebra.

2. For every set $X \subset \mathit{Var}$, every $\mathcal{M} \in M^X$ and every homomorphism $h : \boldsymbol{Fm}(X) \longrightarrow mng^X_{\mathcal{M}}(\boldsymbol{Fm}(X))$ there exists some $\mathcal{N} \in M^X$ such that $h = mng^X_{\mathcal{N}}$.

3. There is a (possibly infinite) set of formulas $\Delta(p, q)$ in two variables such that for every set $X \subset \mathit{Var}$, every $\mathcal{M} \in M^X$ and every $\varphi, \psi \in Fm(X)$,
$$(mng^X_{\mathcal{M}}(\varphi) = mng^X_{\mathcal{M}}(\psi)) \Leftrightarrow \mathcal{M} \models^X \Delta(\varphi, \psi).$$

4. There is a (possibly infinite) set of pairs of formulas $\{\langle \varepsilon_i(p), \delta_i(p)\rangle \ : \ i \in I\}$ in one variable such that for every set $X \subset \mathit{Var}$, every $\mathcal{M} \in M^X$ and every $\varphi \in Fm(X)$,
$$\mathcal{M} \models^X \varphi \Leftrightarrow \mathcal{M} \models^X \{\Delta(\varepsilon_i(\varphi), \delta_i(\varphi)) \ : \ i \in I\}.$$

General logics satisfying the first two conditions are called *structural*. ⊣

3.6.5. Example Set $\Delta(\varphi, \psi) = \varphi \leftrightarrow \psi$, $\delta(\varphi) = \varphi$ and $\varepsilon(\varphi) = \varphi \leftrightarrow \varphi$, where $\varphi \leftrightarrow \psi$ is an abbreviation of $\neg(\varphi \wedge \neg\psi) \wedge \neg(\psi \wedge \neg\varphi)$. The reader easily verifies that with this choice of $\Delta, \delta, \varepsilon$, the logics CPC and IPC are algebraizable. Similar for every normal modal logic with a global consequence relation, an example of which is the logic $\mathsf{K4}$ defined in Example 3.6.3. Examples of logics that are algebraizable but where the above choice of $\Delta, \delta, \varepsilon$ does not work, are the n-valued Łukasiewicz logics. In this case, one may take $\Delta(\varphi, \psi) = \{\varphi \to \psi, \psi \to \varphi\}$, $\delta(\varphi) = \varphi$ and $\varepsilon(\varphi) = \varphi \to \varphi$. ⊣

Let us motivate Definition 3.6.4. Concerning the first condition we bear in mind that in general, a homomorphism is a map that is compositional with respect to the operations. Therefore, the first condition of Definition 3.6.4, according to which the meaning functions are homomorphisms, is one way of formulating the principle of compositionality of meaning. An equivalent way of putting the second condition in 3.6.4 is to say that the class of meaning functions is closed under substitutions. Formally, for every set of variables X, every model $\mathcal{M} \in M^X$ and every substitution $\sigma : \boldsymbol{Fm}(X) \longrightarrow \boldsymbol{Fm}(X)$ there exists some $\mathcal{N} \in M^X$ such that

$$mng^X_{\mathcal{N}} = mng^X_{\mathcal{M}} \circ \sigma. \tag{3.13}$$

Items 3 and 4 in Definition 3.6.4 require the existence of derived connectives that intuitively play the same role as the biconditional $\leftrightarrow$ and the constant $\top$ do in classical propositional logic.

This is an appropriate moment to compare the notion of a semantically algebraizable logic from Definition 3.6.4 to the notion of an algebraizable logic from the previous sections. For a detailed comparison between the two approaches, we refer to [Font and Jansana, 1994]. In this paper, Font and Jansana define a substitution property that is slightly stronger than condition 2 in Definition 3.6.4 (or, equivalently, than the substitution property described above), and which ensures that the model $\mathcal{N}$ in (3.13) also satisfies, for all $\varphi \in Fm(X)$, $\mathcal{N} \models \varphi \Leftrightarrow \mathcal{M} \models \sigma\varphi$. Theorem 2.2 in [Font and Jansana, 1994] states that the semantic consequence relations that are associated with structural general logics with this extended substitution property are precisely what we called *consequence relations* in the previous framework, cf. Definition 3.2.1.

Recall from Theorem 3.3.12 that a k-deductive system S is algebraizable if there exist (sets of) S-formulas $\Delta(p, q), \delta(p), \varepsilon(p)$ with the characteristic properties of the biconditional $\leftrightarrow$ and the constant $\top$ in CPC. This requirement looks very similar to conditions 3 and 4 in Definition 3.6.4. According to Corollary 2.10 in [Font and Jansana, 1994], these requirements are indeed similar. That is, the semantic consequence relations that are associated with semantically algebraizable general logics are precisely the *algebraizable logics* of the previous framework.

We now turn to algebras. With any structural logic $\mathcal{L}(X)$, two classes of algebras are associated. First, the *algebraic counterpart of the semantic component of* $\mathcal{L}(X)$,

$$\mathsf{Alg}_m(\mathcal{L}(X)) = \{mng^X_{\mathcal{M}}(\boldsymbol{Fm}(X)) : \mathcal{M} \in M^X\}.$$

The class $\mathsf{Alg}_m(\mathcal{L}(X))$ is very dependent on the presentation of the models of $\mathcal{L}(X)$. In particular, $\mathsf{Alg}_m(\mathcal{L}(X))$ is not even closed under isomorphisms. Note that this class is missing in the previous approach, for the simple reason that an k-deductive system does not have a semantic component. To define an analogue of Mod^*S, consider for any $\mathbb{K} \subseteq M^X$ the equivalence relation $\sim_{\mathbb{K}}$ on the formula algebra $\boldsymbol{Fm}(X)$ defined by

$$\varphi \sim_{\mathbb{K}} \psi \Leftrightarrow (\forall \mathcal{M} \in \mathbb{K})\, mng^X_{\mathcal{M}}(\varphi) = mng^X_{\mathcal{M}}(\psi). \tag{3.14}$$

Note that $\{\sim_{\mathbb{K}} : \mathbb{K} \subseteq M^X\}$ is a set (i.e., *not* a proper class), since $Fm(X)$ is a set. By condition 1 in 3.6.4, the relation $\sim_{\mathbb{K}}$ is a congruence relation. Let $\boldsymbol{Fm}(X)/{\sim_{\mathbb{K}}}$ denote the the quotient algebra of $\boldsymbol{Fm}(X)$ factorized by $\sim_{\mathbb{K}}$. Recall that I denotes the operation of taking isomorphic copies. Then *the algebraic counterpart of* $\mathcal{L}(X)$ is

$$\mathsf{Alg}_{\models}(\mathcal{L}(X)) = \mathrm{I}\ \{\boldsymbol{Fm}(X)/{\sim_{\mathbb{K}}} : \mathbb{K} \subseteq M^X\}.$$

The reader familiar with abstract algebraic logic may recognize the well-known Lindenbaum-Tarski construction.

With any structural general logic $\mathbf{L}$, we then associate the classes

$$\mathsf{Alg}_{\models}(\mathbf{L}) = \bigcup\{\mathsf{Alg}_{\models}(\mathcal{L}(X)) : X \text{ is a set of variables}\}, \tag{3.15}$$
$$\mathsf{Alg}_m(\mathbf{L}) = \bigcup\{\mathsf{Alg}_m(\mathcal{L}(X)) : X \text{ is a set of variables }\}. \tag{3.16}$$

For examples we refer to Example 3.6.6 below. For the moment, let us indicate the difference between the classes $\mathsf{Alg}_m(\mathbf{L})$ and $\mathsf{Alg}_{\models}(\mathbf{L})$. As we remarked above, the first class depends very much on the presentation of the models of $\mathbf{L}$. We claim that the class $\mathsf{Alg}_{\models}(\mathbf{L})$, on the other hand, only depends on the theories of the general logic $\mathbf{L}$, provided that $\mathbf{L}$ is semantically algebraizable. To see this, fix some set $X \subset Var$ and consider two classes $\mathbb{K}, \mathbb{K}' \subseteq M^X$ with the same theory. That is, we assume that a formula $\varphi \in Fm(X)$ is true in all $\mathbb{K}$-models if and only if φ is true in all $\mathbb{K}'$-models. We show that $\sim_{\mathbb{K}} = \sim_{\mathbb{K}'}$. To this end, consider formulas $\varphi, \psi \in Fm(X)$ such that $\varphi \sim_{\mathbb{K}} \psi$. Then, by (3.14), $(\forall \mathcal{M} \in \mathbb{K})\ mng^X_{\mathcal{M}}(\varphi) = mng^X_{\mathcal{M}}(\psi)$. Since $\mathbf{L}$ is semantically algebraizable this is equivalent to $(\forall \mathcal{M} \in \mathbb{K})\ \mathcal{M} \models^X \Delta(\varphi, \psi)$, (cf. item 3 in 3.6.4). That is, $\Delta(\varphi, \psi)$ is part of the theory of $\mathbb{K}$. But $\mathbb{K}$ and $\mathbb{K}'$ have the same theory. Hence, $\Delta(\varphi, \psi)$ is also part of the theory of $\mathbb{K}'$. Turning the above argumentation around we obtain that $\varphi \sim_{\mathbb{K}'} \psi$. This shows that for any two classes $\mathbb{K}, \mathbb{K}' \subseteq M^X$ with the same theory, the relations $\sim_{\mathbb{K}}$ and $\sim_{\mathbb{K}'}$ are equal. We conclude that the class $\mathsf{Alg}_{\models}(\mathbf{L})$ only depends on the theories of $\mathbf{L}$.

3.6.6. Example According to Example 3.6.5, the logics CPC and K4 are algebraizable. This means that the classes $\mathsf{Alg}_{\models}(\mathsf{CPC})$, $\mathsf{Alg}_m(\mathsf{CPC})$, $\mathsf{Alg}_{\models}(\mathsf{K4})$ and $\mathsf{Alg}_m(\mathsf{K4})$ are well-defined. Let us take a closer look at them. In what follows, we assume the reader is familiar with the Lindenbaum-Tarski construction. In short, by a Lindenbaum-Tarski algebra of the logic S we understand a quotient algebra of the form $\boldsymbol{Fm}_S(X)/\sim_T$, where $\sim_T$ denotes the relation of provable equivalence modulo the theory T. A Lindenbaum-Tarski algebra of CPC is simply called a Lindenbaum-Tarski algebra.

By (3.15), $\mathsf{Alg}_{\models}(\mathsf{CPC})$ is the class of (isomorphic copies of) Lindenbaum-Tarski algebras. A well-known theorem in algebraic logic states that every Boolean algebra is isomorphic to a Lindenbaum-Tarski algebra. Therefore, $\mathsf{Alg}_{\models}(\mathsf{CPC})$ is in fact the class of Boolean algebras.

On the other hand, $\mathsf{Alg}_m(\mathsf{CPC}) = \{\mathbf{2}\}$, where $\mathbf{2}$ denotes the two-element Boolean algebra.

Applying (3.15), $\mathsf{Alg}_{\models}(\mathsf{K4})$ is the class of (isomorphic copies of) Lindenbaum-Tarski algebras of the modal logic K4. Similar to the fact that the class of (isomorphic copies of) Lindenbaum-Tarski algebras is in fact the class of Boolean algebras, one shows that the class of (isomorphic copies of) Lindenbaum-Tarski algebras for the modal logic K4 is in fact the class of *transitive* Boolean algebras with operators. (These are Boolean algebras with operators of the form $\langle A, +, -, f_\Diamond \rangle$ that satisfy the condition $f_\Diamond f_\Diamond x \leq f_\Diamond x$.) Hence, it turns out that $\mathsf{Alg}_{\models}(\mathsf{K4})$ is the class of transitive Boolean algebras with operators.

We call a pair $\langle W, R \rangle$ a K4-*frame* if R is a transitive relation on W. By (3.16), $\mathsf{Alg}_m(\mathsf{K4}) = \mathrm{S}\{\langle \mathcal{P}(W), \cap, \setminus, m_R \rangle \,;\, \langle W, R \rangle \text{ is an K4-frame}\}$, where $\mathcal{P}(W)$ denotes the powerset of W. This class is known as the class of *complex algebras* of frames in K4. ⊣

Let us compare the algebraic counterpart $\mathsf{Alg}_{\models}(\mathbf{L})$ of a semantically algebraizable logic $\mathbf{L}$ with the equivalent quasivariety semantics as defined in subsection 3.3.3. To this end we associate, with any general logic $\mathbf{L}$ a 1-deductive system $S_{\mathbf{L}} = \{\langle Fm_{S_{\mathbf{L}}}(X), \vdash_{S_{\mathbf{L}}(X)} \rangle \;:\; X \text{ is a subset of } Var\}$, where for every X, $Fm_{S_{\mathbf{L}}}(X) = Fm_{\mathcal{L}}(X)$ and $\vdash_{S_{\mathbf{L}}(X)}$ is the semantic consequence relation $\models^X_{\mathcal{L}}$. By Theorem 3.7 in [Font and Jansana, 1994], if $\mathbf{L}$ is semantically algebraizable, then $S_{\mathbf{L}}$ is algebraizable. Moreover, Theorem 3.11 in that same paper says that in this case $\mathsf{Alg}_{\models}(\mathbf{L})$ is the largest equivalent algebraic semantics for $S_{\mathbf{L}}$ in the sense of Blok and Pigozzi.

Recall from Example 3.6.6 that $\mathsf{Alg}_{\models}(\mathsf{CPC}) = BA$ and $\mathsf{Alg}_m(\mathsf{CPC}) = \{\mathbf{2}\}$. By the well-known Stone theorem, SP $\mathbf{2} = BA$, where S (resp. P) denotes the operation of taking subalgebras (resp. isomorphic copies of direct products). Summarizing, we obtain the following relation between $\mathsf{Alg}_{\models}(\mathsf{CPC})$ and $\mathsf{Alg}_m(\mathsf{CPC})$: $\mathsf{Alg}_{\models}(\mathsf{CPC}) = BA = \mathrm{SP}\ \mathbf{2} = \mathrm{SP}\ \mathsf{Alg}_m(\mathsf{CPC})$. The following important result generalizes this observation.

3.6.7. Theorem [Andréka et al., 1994, Thm 3.2.17] *For a structural general logic* **L**,

$$\mathsf{Alg}_{\models}(\mathbf{L}) = \mathrm{SP}\ \mathsf{Alg}_m(\mathbf{L}).$$

For future reference, we mention the following result.

3.6.8. Lemma [Andréka et al., 1994, Claim 3.2.17.1] *Let* **L** *be a structural general logic,* $\boldsymbol{A} \in \mathsf{Alg}_m(\mathbf{L})$ *and* $h : \boldsymbol{Fm}(X) \longrightarrow \boldsymbol{A}$ *a homomorphism, for some set* $X \subset Var$. *Then there exists some* $\mathcal{N} \in M^X$ *such that* $mng^X_{\mathcal{N}} = h$.

Note that this lemma extends clause 2 in Definition 3.6.4, according to which such a model $\mathcal{N}$ exists in case $\boldsymbol{A} \in \mathsf{Alg}_m(\mathcal{L}(X))$.

Proof: Let **L**, $\boldsymbol{A}$, h, X be as in the lemma. Then $\boldsymbol{A} = mng^Y_{\mathcal{M}}(Fm(Y))$, for some set $Y \subset Var$ and some $\mathcal{M} \in M^Y$. By Definition 3.6.1, we may assume that either $X \subseteq Y$ or $Y \subseteq X$.

Assume $Y \subseteq X$. By item 2 in 3.6.1, $mng^Y_{\mathcal{M}}(Fm(Y)) = mng^X_{\mathcal{M}'}(Fm(Y))$, for some $\mathcal{M}' \in M^X$. Hence, h maps into $mng^X_{\mathcal{M}'}(Fm(Y)) \subseteq mng^X_{\mathcal{M}'}(Fm(X))$. The result follows from condition 2 in Definition 3.6.4.

Assume $X \subseteq Y$. Consider any homomorphism $h_{ex} : \boldsymbol{Fm}(Y) \longrightarrow \boldsymbol{A}$ that extends h. By clause 2 in Definition 3.6.4, $h_{ex} = mng^Y_{\mathcal{M}}$, for some $\mathcal{M} \in M^Y$. By condition 2 in Definition 3.6.1, $(mng^Y_{\mathcal{M}})_{\restriction Fm(X)} = mng^X_{\mathcal{N}}$, for some $\mathcal{N} \in M^X$. Summarizing, we have $h = (h_{ex})_{\restriction Fm(X)} = (mng^Y_{\mathcal{M}})_{\restriction Fm(X)} = mng^X_{\mathcal{N}}$. ■

3.6.2 A characterization of the weak Beth property

The weak Beth property has been discussed in subsection 2.2.3. In the present framework, this property can be formulated as follows. Recall that for any set of variables P and any set of formulas Γ, $Cons_P(\Gamma) = \{\varphi \in Fm(P) : \Gamma \models \varphi\}$ and $Mod^P(\Gamma) = \{\mathcal{M} \in M^P : \mathcal{M} \models \gamma,$ for all $\gamma \in \Gamma\}$.

3.6.9. Definition [Strong implicit definition, weak Beth property] Let **L** be a semantically algebraizable general logic, and let P, R be disjoint sets of variables such that $Fm(P) \neq \emptyset$. A set $\Gamma \subseteq Fm(P \cup R)$ *implicitly defines* R *in terms of* P if for every $r \in R$ and every substitution σ that fixes the variables from P,

$$\Gamma, \sigma\Gamma \models \Delta(r, \sigma r).$$

If Γ moreover has the property that

$$(\forall \mathcal{M} \in Mod^P(Cons_P(\Gamma)))(\exists \mathcal{N} \in Mod^{P\cup R}(\Gamma))\ (mng_{\mathcal{N}})_{\restriction Fm(P)} = mng_{\mathcal{M}}, \quad (3.17)$$

then Γ is a *strong implicit definition* of R in terms of P. A formula φ_r is an *explicit definition* of $r \in R$ with respect to Γ if

$$\Gamma \models \Delta(r, \varphi_r).$$

A general logic $\mathbf{L}$ has the *weak Beth property* if for every strong implicit definition Γ of R in terms of P, and every $r \in R$, there exists an explicit definition $\varphi_r \in Fm(P)$ of r with respect to Γ. ⊣

Note that the weak Beth property requires only explicit definability of some special implicit definitions. Therefore, one expects a counterstep on the algebraic side to consist in demanding surjectivity of a suitable subclass of epimorphisms. A possible subclass is the following.

3.6.10. Definition [$\mathbb{K}_0$**-extensible**] Let $\mathbb{K}_0 \subseteq \mathbb{K}$ be two classes of similar algebras. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{K}$ and $h : \boldsymbol{A} \longrightarrow \boldsymbol{B}$ be a homomorphism. h is said to be $\mathbb{K}_0$-*extensible* iff for every algebra $\boldsymbol{C} \in \mathbb{K}_0$ and every *surjection* $f : \boldsymbol{A} \twoheadrightarrow \boldsymbol{C}$ there exists some $\boldsymbol{D} \in \mathbb{K}_0$ and $g : \boldsymbol{B} \longrightarrow \boldsymbol{D}$ such that $\boldsymbol{C} \subseteq \boldsymbol{D}$ and $g \circ h = f$. ⊣

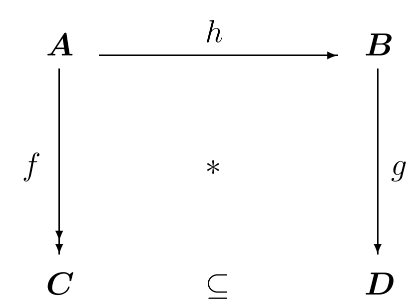


Figure 3.6: A $\mathbb{K}_0$-extensible homomorphism.

By means of the notion of $\mathbb{K}_0$-extensibility we are able to define an algebraic property that is to the weak Beth property as ES is to the Beth property. This property does not require the surjectivity of all $\mathsf{Alg}_{\models}(\mathbf{L})$-epimorphisms, but only of those $\mathsf{Alg}_{\models}(\mathbf{L})$-epimorphisms that are $\mathsf{Alg}_m(\mathbf{L})$-extensible.

3.6.11. Example Obviously, if all $\mathsf{Alg}_{\models}(\mathbf{L})$-epimorphisms are surjective, then so are all $\mathsf{Alg}_m(\mathbf{L})$-extensible $\mathsf{Alg}_{\models}(\mathbf{L})$-epimorphisms. This at once generates many examples of logics $\mathbf{L}$ such that all $\mathsf{Alg}_m(\mathbf{L})$-extensible $\mathsf{Alg}_{\models}(\mathbf{L})$-epimorphisms are surjective, viz., all logics $\mathbf{L}$ such that $\mathsf{Alg}_{\models}(\mathbf{L})$ has property ES.

Let us give an example of an $\mathsf{Alg}_m(\mathbf{L})$-extensible $\mathsf{Alg}_{\models}(\mathbf{L})$-epimorphism that is not surjective. To this end, consider the implicational fragment of CPC, denoted by $\mathsf{CPC}^{\rightarrow}$. This fragment can be defined as a general logic in a similar way as we defined CPC in Example 3.6.2. In particular, the meaning functions of $\mathsf{CPC}^{\rightarrow}$ map into $\mathbf{2}^{\rightarrow}$, that is, the $\rightarrow$-reduct of the two-element Boolean algebra. Hence, $\mathsf{Alg}_m(\mathsf{CPC}^{\rightarrow}) = \mathrm{S}\{\mathbf{2}^{\rightarrow}\}$. By Theorem 3.6.7, then $\mathsf{Alg}_{\models}(\mathsf{CPC}^{\rightarrow}) = \mathrm{SPS}\{\mathbf{2}^{\rightarrow}\} = \mathrm{SP}\{\mathbf{2}^{\rightarrow}\}$. This last equality holds as $\mathrm{PS} \subseteq \mathrm{SP}$.

Consider the algebra $\boldsymbol{B} = \mathbf{2}^{\rightarrow} \times \mathbf{2}^{\rightarrow}$ and its subalgebra $\boldsymbol{A}$ with domain $B \setminus \{0\}$. From Example 3.5.8 we learn that the identity $id : \boldsymbol{A} \longrightarrow \boldsymbol{B}$ is an epimorphism in the class $\mathrm{SP}\{\mathbf{2}^{\rightarrow}\}$. We claim that $id : \boldsymbol{A} \longrightarrow \boldsymbol{B}$ is $\mathrm{S}\{\mathbf{2}^{\rightarrow}\}$-extensible. For, consider some $\boldsymbol{C} \in \mathrm{S}\{\mathbf{2}^{\rightarrow}\}$ and some surjection $f : \boldsymbol{A} \twoheadrightarrow \boldsymbol{C}$. We will extend f to a homomorphism $f_{ext} : \boldsymbol{B} \longrightarrow \mathbf{2}^{\rightarrow}$. To this end, distinguish the following two cases.

Case 1. $\boldsymbol{C} = \mathbf{2}^{\rightarrow}$. In this case, set $f_{ext}(0) = 0$.

Case 2. $\boldsymbol{C}$ is the subalgebra of $\mathbf{2}^{\rightarrow}$ with domain $\{1\}$. This time, set $f_{ext}(0) = 1$.

In both cases, $f_{ext} : \boldsymbol{B} \longrightarrow \mathbf{2}^{\rightarrow}$ is a homomorphism which completes the diagram in Figure 3.6. We conclude that $id : \boldsymbol{A} \longrightarrow \boldsymbol{B}$ is an $\mathsf{Alg}_m(\mathsf{CPC}^{\rightarrow})$-extensible epimorphism in the class $\mathsf{Alg}_{\models}(\mathsf{CPC}^{\rightarrow})$. $\dashv$

For readers familiar with [Andréka et al., 1994] we point out that in comparison with that paper, we made in Definition 3.6.10 the additional requirement for f to be *onto*.

Now we have the following equivalence.

3.6.12. Theorem (Algebraic characterization of the weak Beth property) *Let* $\mathbf{L}$ *be a semantically algebraizable general logic. Then* $\mathbf{L}$ *has the weak Beth property iff all the* $\mathsf{Alg}_m(\mathbf{L})$*-extensible epimorphisms of* $\mathsf{Alg}_{\models}(\mathbf{L})$ *are surjective.*

Proof: Let $\mathbf{L}$ be a semantically algebraizable general logic.

WeakBP $\Rightarrow$ ExtES: Assume that $\mathbf{L}$ has the weak Beth property. Let $f : \boldsymbol{A} \longrightarrow \boldsymbol{B}$ be an $\mathsf{Alg}_m(\mathbf{L})$-extensible $\mathsf{Alg}_{\models}(\mathbf{L})$-epimorphism. Our task is to show that f is surjective.

Consider a set $Z \subset Var$ of the same cardinality as B. Let $h : \boldsymbol{Fm}(Z) \twoheadrightarrow \boldsymbol{B}$ be a homomorphism which maps Z onto B. As in the proof of Theorem 3.3.8, by Z_A (resp. $Z_{B\setminus A}$) we understand the sentential variables in Z which are mapped by h to $f(A)$ (resp. $B\setminus f(A)$).

Define the set of formulas

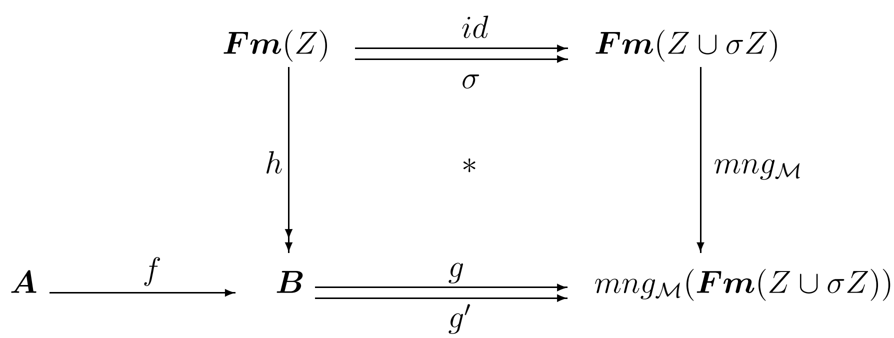
$$\Gamma = \bigcup\{\Delta(\varphi, \psi) : \ \varphi, \psi \in Fm(Z) \text{ and } h(\varphi) = h(\psi)\}.$$

We claim that Γ implicitly defines $Z_{B\setminus A}$ in terms of Z_A in a strong sense.

First we show that Γ is an implicit definition of $Z_{B\setminus A}$ in terms of Z_A. Hereto, consider a substitution σ that fixes the variables in Z_A, and some $\mathcal{M} \in Mod^{Z\cup\sigma Z}(\Gamma \cup \sigma\Gamma)$. It needs to be shown that for every $z \in Z_{B\setminus A}$,

$$mng_{\mathcal{M}}(z) = mng_{\mathcal{M}}(\sigma z). \tag{3.18}$$

Consider the homomorphism $mng_{\mathcal{M}} : \boldsymbol{Fm}(Z \cup \sigma Z) \longrightarrow mng_{\mathcal{M}}(\boldsymbol{Fm}(Z \cup \sigma Z))$.



Note that $\ker(h) \subseteq \ker(mng_{\mathcal{M}})$. For, suppose $h(\varphi) = h(\psi)$. Then $\Delta(\varphi, \psi) \subseteq \Gamma$. As $\mathcal{M}$ is a model of Γ, by clause 3 of Definition 3.6.4 this implies that $mng_{\mathcal{M}}(\varphi) = mng_{\mathcal{M}}(\psi)$. Hence, $\ker(h) \subseteq \ker(mng_{\mathcal{M}})$. By the well-known homomorphism lemma from universal algebra (a version for *matrices* can be found as Lemma 3.2.3), there exists some homomorphism $g : \boldsymbol{B} \longrightarrow mng_{\mathcal{M}}(\boldsymbol{Fm}(Z \cup \sigma Z))$ such that $g \circ h = (mng_{\mathcal{M}})_{\restriction Fm(Z)}$. Similarly, using the fact that $\ker(h) \subseteq \ker(mng_{\mathcal{M}} \circ \sigma)$, we derive the existence of some homomorphism $g' : \boldsymbol{B} \longrightarrow mng_{\mathcal{M}}(\boldsymbol{Fm}(Z \cup \sigma Z))$ such that $g' \circ h = mng_{\mathcal{M}} \circ \sigma$.

Take an arbitrary element a of A. By the surjectivity of h there is an element $z_a \in Z$ such that $f(a) = h(z_a)$. By definition of Z_A, $z_a \in Z_A$. Then $g(f(a)) = g(h(z_a)) = mng_{\mathcal{M}}(z_a) = mng_{\mathcal{M}}(\sigma z_a)$ (as σ fixes the variables in Z_A) $= g'(h(z_a)) = g'(f(a))$. In other words, $g \circ f = g' \circ f$. Since f is an $\mathsf{Alg}_{\models}(\mathbf{L})$-epimorphism, it follows that $g = g'$. This implies for every $z \in Z_{B\setminus A}$ that $mng_{\mathcal{M}}(z) = g(h(z)) = g'(h(z)) = mng_{\mathcal{M}}(\sigma z)$. This finishes the proof of (3.18), and shows that Γ implicitly defines $Z_{B\setminus A}$ in terms of Z_A.

Now we claim that Γ is a *strong* implicit definition. To prove this, we consider some $\mathcal{M} \in Mod^{Z_A}(Cons_{Z_A}(\Gamma))$. We need to find some $\mathcal{M}' \in Mod^{Z}(\Gamma)$ such that $(mng_{\mathcal{M}'})_{\restriction Fm(Z_A)} = mng_{\mathcal{M}}$.

Define for all $a \in A$ the map $j : \boldsymbol{A} \longrightarrow mng_{\mathcal{M}}(\boldsymbol{Fm}(Z_A))$ by

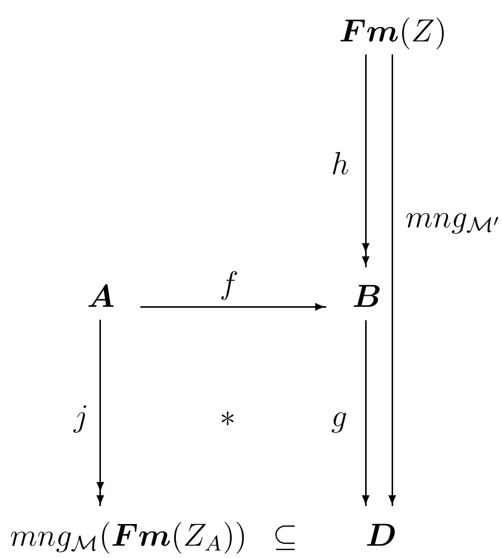
$$j(a) = mng_{\mathcal{M}}(z_a),$$

where $z_a \in Z_A$ is such that $h(z_a) = f(a)$. Since h maps Z_A onto $f(A)$ such a z_a exists. Let us verify that j is a homomorphism. To this end, consider some n-ary connective $\omega \in Cn(\mathbf{L})$, and $a_1, \ldots, a_n \in A$. We show that

$$j(\omega(a_1, \ldots, a_n)) = \omega(j(a_1), \ldots, j(a_n)). \tag{3.19}$$

Let $z_\omega, z_1, \ldots, z_n \in Z_A$ be such that $h(z_\omega) = f(\omega(a_1, \ldots, a_n))$ and $h(z_{a_i}) = f(a_i)$, for $1 \leq i \leq n$. Then $h(z_\omega) = f(\omega(a_1, \ldots, a_n)) = \omega(f(a_1), \ldots, f(a_n)) = \omega(h(z_{a_1}), \ldots, h(z_{a_n})) = h(\omega(z_{a_1}, \ldots, z_{a_n}))$. Therefore, Γ contains the set of formulas $\Delta(z_\omega, \omega(z_{a_1}, \ldots, z_{a_n}))$. As $\mathcal{M} \in Mod^{Z_A}(Cons_{Z_A}(\Gamma))$, then $mng_{\mathcal{M}} z_\omega = mng_{\mathcal{M}} \omega(z_{a_1}, \ldots, z_{a_n})$. Summarizing, we have $j(\omega(a_1, \ldots, a_n)) = mng_{\mathcal{M}} z_\omega = mng_{\mathcal{M}} \omega(z_{a_1}, \ldots, z_{a_n}) = \omega(mng_{\mathcal{M}} z_{a_1}, \ldots, mng_{\mathcal{M}} z_{a_n}) = \omega(j(a_1), \ldots, j(a_n))$. We have derived (3.19).

Moreover, j is well-defined. For, suppose $h(z_a) = h(z_{a'}) = f(a)$, for some $z_a, z_{a'} \in Z_A$. Then $\Delta(z_a, z_{a'}) \subseteq \Gamma$. As $\mathcal{M} \in Mod^{Z_A}(Cons_{Z_A}(\Gamma))$, it follows that $mng_{\mathcal{M}}(z_a) = mng_{\mathcal{M}}(z_{a'})$. Hence, $j(a)$ is well-defined.



Finally, we claim that j is onto. To see this, consider some $\varphi \in Fm(Z_A)$. We show that $mng_{\mathcal{M}}(\varphi)$ lies in the range of j. First note that, as h maps Z_A onto A, there exists some $z_\varphi \in Z_A$ such that $h(z_\varphi) = h(\varphi)$. Then $\Delta(z_\varphi, \varphi) \subseteq \Gamma$. Since $\mathcal{M} \in Mod^{Z_A}(Cons_{Z_A}(\Gamma))$, we infer that

$$mng_{\mathcal{M}}(z_\varphi) = mng_{\mathcal{M}}(\varphi) \tag{3.20}$$

Second, observe that $h(z_\varphi) \in f(A)$. Therefore, there exists some $a \in A$ such that $h(z_\varphi) = f(a)$. Then $j(a) = mng_{\mathcal{M}}(z_\varphi) = mng_{\mathcal{M}}(\varphi)$ (by (3.20). This shows that $mng_{\mathcal{M}}(\varphi)$ lies in the range of j. We conclude that j is a well-defined, onto homomorphism.

Since f is $\mathsf{Alg}_m(\mathbf{L})$-extensible, there exists some $\boldsymbol{D} \in \mathsf{Alg}_m(\mathbf{L})$ and a homomorphism $g : \boldsymbol{B} \longrightarrow \boldsymbol{D}$ such that $mng_{\mathcal{M}}(\boldsymbol{Fm}(Z_A)) \subseteq \boldsymbol{D}$ and $g \circ f = j$. By Lemma 3.6.8, there exists a model $\mathcal{M}' \in M^Z$ such that $g \circ h = mng_{\mathcal{M}'}$. We claim that $\mathcal{M}'$ has the desired properties, i.e., $\mathcal{M}'$ is a model of Γ such that $(mng_{\mathcal{M}'})_{\lceil Fm(Z_A)} = mng_{\mathcal{M}}$. First, let $z \in Z_A$. Then $h(z) = f(a_z)$, for some $a_z \in A$. Hence, $mng_{\mathcal{M}}(z) = j(a_z)$ (by definition of j) $= g(f(a_z))$ (as the diagram commutes) $= g(h(z)) = mng_{\mathcal{M}'}(z)$. This shows that $mng_{\mathcal{M}} = (mng_{\mathcal{M}'})_{\lceil Fm(Z_A)}$. Second, let $\Delta(\varphi, \psi) \subseteq \Gamma$. Then $h(\varphi) = h(\psi)$, whence $mng_{\mathcal{M}'}(\varphi) = g(h(\varphi)) = g(h(\psi)) = mng_{\mathcal{M}'}(\psi)$. Therefore, by 3.6.4.3, $\mathcal{M}'$ is a model of Γ.

We conclude that Γ is a *strong* implicit definition of $Z_{B\setminus A}$ in terms of Z. As $\mathbf{L}$ has the weak Beth property, it follows that every $z \in Z_{B\setminus A}$ has some explicit definition in terms of Z_A with respect to Γ.

We are now in a position to prove that f is surjective. Let $b \in B$. We will show that b lies in the range of f. By construction, $b = h(z)$, for some $z \in Z$. If $z \in Z_A$, we are done. Therefore, suppose $z \in Z_{B\setminus A}$. As we deduced above, in this case z has some explicit definition $\varphi_z \in Fm(Z_A)$ with respect to Γ. We claim that $b = h(\varphi_z)$, thereby showing that b lies in the range of f. For, suppose $b \neq h(\varphi_z)$. That is, $h(z) \neq h(\varphi_z)$. We derive a contradiction. By Theorem 3.6.7, $\boldsymbol{B} \in \mathrm{SP}\,\mathsf{Alg}_m(\mathbf{L})$. Hence, there exists some $\boldsymbol{B}_i \in \mathsf{Alg}_m(\mathbf{L})$ and a homomorphism $\pi_i : \boldsymbol{B} \longrightarrow \boldsymbol{B}_i$ such that $\pi_i(h(z)) \neq \pi_i(h(\varphi_z))$. By Lemma 3.6.8, there exists some model $\mathcal{N} \in M^Z$ such that $mng_{\mathcal{N}} = \pi_i \circ h$. Then $mng_{\mathcal{N}}(\varphi_z) \neq mng_{\mathcal{N}}(z)$. As φ_z is an explicit definition of z with respect to Γ, it follows that $\mathcal{N} \notin Mod^Z(\Gamma)$. On the other hand, consider some $\Delta(\varphi, \psi) \subseteq \Gamma$. Then $h(\varphi) = h(\psi)$. Hence, $mng_{\mathcal{N}}(\varphi) = \pi_i(h(\varphi)) = \pi_i(h(\psi)) = mng_{\mathcal{N}}(\psi)$. Therefore, $\mathcal{N} \in Mod^Z(\Gamma)$. Contradiction. We conclude that f is surjective.

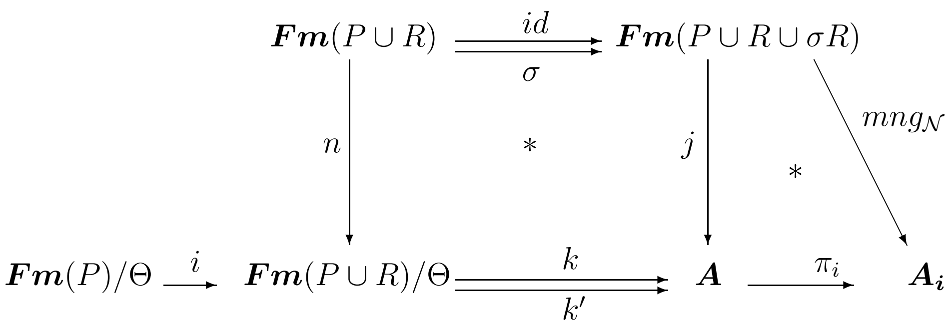
ExtES $\Rightarrow$ WeakBP: Assume that all $\mathsf{Alg}_m(\mathbf{L})$-extensible epimorphisms of $\mathsf{Alg}_{\models}(\mathbf{L})$ are surjective. Let P, R be disjoint sets of variables such that $Fm(P) \neq \emptyset$, and let $\Gamma \subseteq Fm(P \cup R)$ be a strong implicit definition of R in terms of P. We have to show that every $r \in R$ has an explicit definition in terms of P with respect to Γ.

Let $\Theta = \sim_{Mod(\Gamma)}$ be the congruence relation defined in (3.14). That is, for all $\varphi, \psi \in Fm(P \cup R)$,

$$\varphi \Theta \psi \Leftrightarrow (\forall \mathcal{M} \in Mod^{P\cup R}(\Gamma))\; mng_{\mathcal{M}}(\varphi) = mng_{\mathcal{M}}(\psi).$$

We show that the natural embedding $i : \boldsymbol{Fm}(P)/\Theta \rightarrowtail \boldsymbol{Fm}(P \cup R)/\Theta$ is an $\mathsf{Alg}_m(\mathbf{L})$-extensible epimorphism of $\mathsf{Alg}_{\models}(\mathbf{L})$. By assumption, this means that i is onto. This clearly implies that every $r \in R$ has an explicit definition in terms of P with respect to Γ. We first show that $i : \boldsymbol{Fm}(P)/\Theta \rightarrowtail \boldsymbol{Fm}(P \cup R)/\Theta$ is an $\mathsf{Alg}_{\models}(\mathbf{L})$-epimorphism. Second we show that it is $\mathsf{Alg}_m(\mathbf{L})$-extensible.

To see that $i : \boldsymbol{Fm}(P)/\Theta \rightarrowtail \boldsymbol{Fm}(P \cup R)/\Theta$ is an $\mathsf{Alg}_{\models}(\mathbf{L})$-epimorphism, consider some $\boldsymbol{A} \in \mathsf{Alg}_{\models}(\mathbf{L})$ and a pair of homomorphisms $k, k' : \boldsymbol{Fm}(P \cup R)/\Theta \longrightarrow \boldsymbol{A}$ such that for all $\varphi \in Fm(P)$, $k(\varphi/\Theta) = k'(\varphi/\Theta)$. We have to show that $k = k'$.



Let Q be a set of variables disjoint from $P \cup R$, of the same size as R. Let $i : Q \rightarrowtail\!\!\!\!\twoheadrightarrow R$ be a bijection, and let σ be a substitution which interchanges q and $i(q)$, for every $q \in Q$, and which fixes the variables in P. That is,

$$\begin{array}{rcll} \sigma(q) & = & i(q), & \text{if } q \in Q, \\ \sigma(r) & = & i^{-1}(r), & \text{if } r \in R, \\ \sigma(p) & = & p, & \text{if } p \in P. \end{array}$$

Let $n : \boldsymbol{Fm}(P \cup R) \longrightarrow \boldsymbol{Fm}(P \cup R)/\Theta$ be the natural mapping. Define the homomorphism $j : \boldsymbol{Fm}(P \cup R \cup \sigma R) \longrightarrow \boldsymbol{A}$ by

$$\begin{array}{rcll} j(p) & = & k(n(p)), & \text{for } p \in P, \\ j(r) & = & k(n(r)), & \text{for } r \in R, \\ j(\sigma r) & = & k'(n(r)), & \text{for } r \in R. \end{array}$$

Let $\varphi \in Fm(P \cup R)$. Then

$$j(\varphi) = k(n(\varphi)), \tag{3.21}$$

and, as the reader easily verifies,

$$j(\sigma\varphi) = k'(n(\varphi)). \tag{3.22}$$

Assume that $k \neq k'$. Then there exists some $r_0 \in R$ such that $k(n(r_0)) \neq k'(n(r_0))$. By Theorem 3.6.7, $\boldsymbol{A} \in \mathrm{SP}\ \mathsf{Alg}_m(\mathbf{L})$. Hence, there is some $\boldsymbol{A}_i \in \mathsf{Alg}_m(\mathbf{L})$ and a homomorphism $\pi_i : \boldsymbol{A} \longrightarrow \boldsymbol{A}_i$ such that

$$\pi_i(k(n(r_0))) \neq \pi_i(k'(n(r_0))). \tag{3.23}$$

By Lemma 3.6.8 there exists some model $\mathcal{N} \in M^{P \cup R \cup \sigma R}$ such that

$$\pi_i \circ j = mng_{\mathcal{N}}. \tag{3.24}$$

We claim that $\mathcal{N}$ is a model of Γ. For, let $\gamma \in \Gamma$. By items 3 and 4 in Definition 3.6.4, for all $\mathcal{M} \in Mod(\Gamma)$ it is the case that $mng_{\mathcal{M}}(\varepsilon\gamma) = mng_{\mathcal{M}}(\delta\gamma)$.[6] Hence, by definition of Θ,

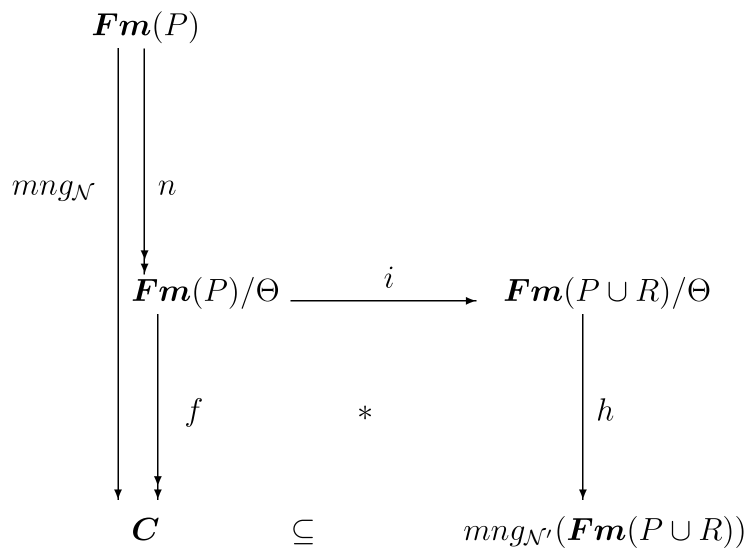
$$n(\varepsilon\gamma) = n(\delta\gamma). \tag{3.25}$$

Then $mng_{\mathcal{N}}(\varepsilon\gamma) = \pi_i(j(\varepsilon\gamma)) = \pi_i(k(n(\varepsilon\gamma))) = \pi_i(k(n(\delta\gamma))) = mng_{\mathcal{N}}(\delta\gamma)$. By the clauses 3 and 4 in Definition 3.6.4, this implies that $\mathcal{N}$ is a model of γ. Similarly, using the facts that $j \circ \sigma = k' \circ n$ and for every $\varphi \in Fm(P \cup R)$, $\varepsilon(\sigma(\varphi)) = \sigma(\varepsilon(\varphi))$, one shows that $\mathcal{N}$ is a model of $\sigma\Gamma$.

Recall that Γ implicitly defines R in terms of P. As $\mathcal{N}$ is a model of both Γ and $\sigma\Gamma$, this implies that for all $r \in R$, $mng_{\mathcal{N}}(r) = mng_{\mathcal{N}}(\sigma r)$. In particular, $mng_{\mathcal{N}}(r_0) = mng_{\mathcal{N}}(\sigma r_0)$. Then $\pi_i(k(n(r_0))) = \pi_i(j(r_0)) = mng_{\mathcal{N}}(r_0) = mng_{\mathcal{N}}(\sigma r_0) = \pi_i(k'(n(r_0)))$. A contradiction with (3.23). We conclude that $k = k'$.

This shows that $i : \boldsymbol{Fm}(P)/\Theta \rightarrowtail \boldsymbol{Fm}(P \cup R)/\Theta$ is an $\mathsf{Alg}_{\models}(\mathbf{L})$-epimorphism. It remains to be shown that it is $\mathsf{Alg}_m(\mathbf{L})$-extensible.

Hereto, consider an algebra $\boldsymbol{C} \in \mathsf{Alg}_m(\mathbf{L})$, and a surjection $f : \boldsymbol{Fm}(P)/\Theta \twoheadrightarrow \boldsymbol{C}$. Let $n : \boldsymbol{Fm}(P) \longrightarrow \boldsymbol{Fm}(P)/\Theta$ be the natural map. By Lemma 3.6.8, there exists some $\mathcal{N} \in M^P$ such that $f \circ n = mng_{\mathcal{N}}$. Note that $\boldsymbol{C} = mng_{\mathcal{N}}(\boldsymbol{Fm}(P))$, as f is onto.



We claim that $\mathcal{N} \in Mod^P(Cons_P(\Gamma))$. To see this, let $\varphi \in Cons_P(\Gamma)$. By clauses 3 and 4 in Definition 3.6.4, it suffices to show that

$$mng_{\mathcal{N}}(\varepsilon\varphi) = mng_{\mathcal{N}}(\delta\varphi). \tag{3.26}$$

[6]In fact, by item 4 in Definition 3.6.4 there exists a (possibly infinite) set of pairs of formulas $\{\langle\varepsilon_i, \delta_i\rangle \ : \ i \in I\}$ which plays this role. For simplicity, here we assume this set to be a single pair. The reader may easily check the general case.

We reason as follows. Since $\Gamma \models \varphi$, we have by condition 4 in 3.6.4 that for all $\mathcal{M} \in Mod(\Gamma)$, $mng_{\mathcal{M}}(\varepsilon\varphi) = mng_{\mathcal{M}}(\delta\varphi)$. Therefore, $n(\varepsilon\varphi) = n(\delta\varphi)$. Hence, $mng_{\mathcal{N}}(\varepsilon\varphi) = f(n(\varepsilon\varphi)) = f(n(\delta\varphi)) = mng_{\mathcal{N}}(\delta\varphi)$. This proves (3.26).

Now that we know that $\mathcal{N} \in Mod^P(Cons_P(\Gamma))$, it follows from the fact that Γ is a strong implicit definition of R that there exists some $\mathcal{N}' \in Mod^{P\cup R}(\Gamma)$ such that $(mng_{\mathcal{N}'})\restriction_{Fm(P)} = mng_{\mathcal{N}}$. Define $h : \boldsymbol{Fm}(P \cup R)/\Theta \longrightarrow mng_{\mathcal{N}'}(\boldsymbol{Fm}(P \cup R))$, for every $\varphi \in Fm(P \cup R)$, by

$$h(\varphi/\Theta) = mng_{\mathcal{N}'}(\varphi).$$

Note that if $\varphi\Theta\psi$, then $mng_{\mathcal{N}'}(\varphi) = mng_{\mathcal{N}'}(\psi)$, since $\mathcal{N}' \in Mod^{P\cup R}(\Gamma)$. Therefore, h is well-defined. We claim that $h \circ i = f$. For, let $\varphi \in Fm(P)$. Then $f(\varphi/\Theta) = mng_{\mathcal{N}}(\varphi) = mng_{\mathcal{N}'}(\varphi) = h(\varphi/\Theta)$. As $\boldsymbol{C} = mng_{\mathcal{N}}(\boldsymbol{Fm}(P)) \subseteq mng_{\mathcal{N}'}(\boldsymbol{Fm}(P \cup R))$, this completes the diagram. We conclude that the homomorphism $i : \boldsymbol{Fm}(P)/\Theta \rightarrowtail \boldsymbol{Fm}(P \cup R)/\Theta$ is $\mathsf{Alg}_m(\mathbf{L})$-extensible, as was to be shown. ■

3.6.13. Remark Other algebraic characterizations of the weak Beth property have been given by Madarász, Németi and Sain in [Sain, 1998]. These characterizations are of a category-theoretic nature and refer to certain connections between $\mathsf{Alg}_{\models}(\mathbf{L})$, $\mathsf{Alg}_m(\mathbf{L})$ and the class of all $\subseteq$-maximal elements of $\mathsf{Alg}_m(\mathbf{L})$. ⊣

Results in [Hoogland, 1996] show that Theorem 3.6.12 is the most general characterization of its kind possible in the present framework. More precisely, we can prove the following.

3.6.14. Theorem *There exists a structural general logic* $\mathbf{L}$ *satisfying condition 3 in 3.6.4 such that* $\mathbf{L}$ *does* not *have the weak Beth property, but all the epimorphisms of* $\mathsf{Alg}_{\models}(\mathbf{L})$ *are surjective.*

As the reader may verify, in our proof of the direction Weak BP⇒ Ext ES we did not use the fact that $\mathbf{L}$ satisfies condition 4 in 3.6.4. In other words, this direction holds already for structural general logics that satisfy condition 3.6.4.3. Again, this is the most general statement that can be made. That is, condition 3 in 3.6.4 is indeed needed.

3.6.15. Theorem *There exists a structural general logic* $\mathbf{L}$ *such that* $\mathbf{L}$ *has the Beth property, but* not *all the* $\mathsf{Alg}_m(\mathbf{L})$*-extensible epimorphisms of* $\mathsf{Alg}_{\models}(\mathbf{L})$ *are surjective.*

The above two theorems were obtained in cooperation with J. Madarász.

Chapter 4

Guarded fragments

Outline of the chapter

In this chapter we investigate interpolation and definability in the guarded fragment of first order logic and in some related fragments.

In the introduction we motivate the interest in the issue of interpolation and definability for the guarded fragment (GF) and for related fragments like the packed fragment (PF). In section 4.2 these fragments are formally defined and semantically characterized in terms of suitable notions of bisimulation. Readers familiar with the notion of a guarded bisimulation are advised to start reading in section 4.3, where counterexamples to interpolation in GF and PF are presented. The reason for the failure of interpolation in these fragments is discussed in section 4.4, giving rise to an alternative interpolation theorem for GF and PF. This alternative, it will be argued, is a natural property to investigate in a guarded context. Moreover, it is quite strong. Strong enough at least to entail the Beth definability theorem, as we will see in section 4.5. Even better, we will see that every guarded or packed finite variable fragment has the Beth property. In section 4.6, we investigate interpolation for these fragments. The main result of this section is that the guarded two variable fragment has interpolation. Finally, we show that the alternative interpolation theorem for GF that we proved in section 4.4 allows for a refinement similar to Lyndon's refinement of Craig's theorem for first order logic.

4.1 Introducing some guarded fragments

The guarded fragment and the packed fragment As is well-known, the basic modal logic K can be seen as a fragment of first order logic via the translation t which maps a proposition letter p to the atom Px, which commutes with the Boolean connectives, and which maps formulas of the form $\Diamond\varphi$ to $\exists y(Rxy \wedge \varphi^t(y))$ and $\Box\varphi$ to $\forall y(Rxy \rightarrow \varphi^t(y))$. The image of K under this translation is referred to as the *modal fragment*. This fragment turns out to behave excellently. It shares several nice model-theoretic properties with full first order logic (e.g., interpolation, Beth definability, and the Łos-Tarski property), and has in addition good algorithmic qualities: it is decidable and every satisfiable modal formula has a finite model and a tree model (in other words, the modal fragment has the finite model property and the tree model property). Moreover, the decidability of this fragment is robust in the sense that various expansions remain decidable. For example, adding features like counting quantifiers or fixed points to the modal fragment does not affect decidability.

In [Andréka et al., 1998] it is argued that the distinguishing characteristic of the modal fragment is its restriction on quantifier patterns. This brings Andréka, van Benthem and Németi to investigate the question to what extent we can loosen these quantifier restrictions while retaining the attractive modal behavior. By answering this question they hope to obtain a fine-structure of first order logic with the nice properties of modal logic but with a greater expressive power. The outcome is the *guarded fragment* (GF) which allows for quantifications of the form $\exists \boldsymbol{y}(R\boldsymbol{xy} \wedge \varphi(\boldsymbol{x}, \boldsymbol{y}))$ and $\forall \boldsymbol{y}(R\boldsymbol{xy} \rightarrow \varphi(\boldsymbol{x}, \boldsymbol{y}))$, where $\boldsymbol{x}, \boldsymbol{y}$ are finite sequences of variables and φ is a guarded formula with free variables among $\boldsymbol{x}, \boldsymbol{y}$ which *all* must appear in the atomic formula $R\boldsymbol{xy}$ (see Definition 4.2.3 for details).

Clearly, the guarded fragment extends the modal fragment. Moreover, as the following series of results shows, it also has many of the hoped-for nice properties. First of all, it is decidable, as is shown in [Andréka et al., 1998]. [Grädel, 1999] improves on this result by classifying the satisfiability problem for GF to be complete for deterministic double exponential time. Satisfiability for finite variable guarded fragments is even in $\mathsf{Exptime}$, in fact $\mathsf{Exptime}$-complete. [Grädel, 1999] also establishes the finite model property for GF, using a technique which involves the so-called Herwig Theorem. This technique has first been employed by Andréka, Hodkinson and Németi to show the finite model property for similar fragments, [Andréka et al., 1999]. What is more, GF has a certain tree model property, [Andréka et al., 1998]. Since the tree model property of the modal fragment can be seen as the main reason behind the robustness of the decidability of that fragment (cf. [Vardi, 1998]), this gives hope as to the robustness of GF. And indeed, adding least and greatest fixed points to GF yields a decidable expansion [Grädel and Walukiewicz, 1999]. Interpolation and definability were mentioned as other yardsticks by which to measure the modal behavior of GF. They are the object of investigation of the present chapter.

We also study a natural expansion of GF, called the *packed fragment* (PF). Roughly speaking, the packed fragment allows for quantifications of the form $\exists \boldsymbol{y}(\varphi \wedge \psi)$, where φ is a conjunction of atoms such that for every pair of free variables in ψ there exists some conjunct in which both variables occur. An example of a packed formula which is not guarded is the formula $\forall xyz((Exy \wedge Eyz \wedge Ezx) \rightarrow Dxyz)$. A precise definition of the packed fragment can be found in Definition 4.2.5. For readers familiar with [van Benthem, 1997] we note that PF bears close correspondence to the *loosely guarded fragment* introduced and studied there (cf. Remark 4.2.7). It turns out that PF shares many of the nice features of GF. For instance, [van Benthem, 1997] shows that PF is decidable. More specifically, just as for GF, the satisfiability problem for PF is complete for 2Exptime, while for finite variable packed fragments this problem is Exptime-complete, [Grädel, 1999]. Moreover, PF has a tree-model property, [Grädel, 1999]. The question as to whether PF has the finite model property has recently been solved affirmatively in [Hodkinson, 2000]. Interpolation and definability for PF are investigated in this chapter. As we will see, also with regard to these properties the two fragments exhibit a quite similar behavior.

Interpolation and definability Notwithstanding the fact that the guarded fragment and the packed fragment have been established as particularly well-behaved fragments of first order logic in many respects, in this chapter we show that the interpolation theorem fails for GF and PF. That is, we will exhibit two guarded (resp. packed) formulas φ, ψ such that $\varphi \rightarrow \psi$ is a validity, without a guarded (resp. packed) interpolant. The reason for this failure of interpolation becomes apparent if we take seriously the idea of these fragments being generalizations of the modal fragment: from this point of view, a natural interpolation theorem for these fragments is not so much a restriction of the interpolation theorem for first order logic, as a generalization of the interpolation theorem for modal logic. However, in multi-modal logics an interpolant is usually subject to slightly different conditions than in first order logic: it is confined to proposition letters in the common language but may contain non-shared modalities. In this chapter we generalize this condition to guarded and packed formulas, and show that GF and PF do have a generalized version of the modal interpolation property. Even better, it is possible to refine this theorem by taking into account the positive and negative occurrences of relation symbols. Moreover, this alternative interpolation property turns out to be quite strong. At least strong enough to entail the Beth definability property for GF and PF.

The upshot of these results is twofold. From our perspective it is interesting to note that the guarded and the packed fragment present natural examples of logics with the Beth property but without the interpolation property. This shows once again the difference between these two properties, a recurrent theme in this dissertation. Second, the fact that a generalized version of the interpolation theorem for multi-modal logics holds in the guarded fragment and the packed fragment affirms the modal character of these fragments. The search for this affirmation was the main motive for Andréka, van Benthem and Németi to be interested in the issue of interpolation for these fragments. So, one way to view our results is as providing the correct notion of interpolation for guarded fragments.

Guarded finite variable fragments In first order logic, restricting the number of variables causes failure of interpolation and definability. As shown in Example 2.2.5, for any $n \geq 2$, the n-variable fragment of first order logic fails to have the Beth property. In fact, this counterexample only uses binary relation symbols. With regard to interpolation the situation is even worse. As we have seen in Example 2.3.7, for any $n \in \omega$ there exist first order formulas φ_n, ψ_n in just two variables and unary relation symbols such that $\varphi_n \rightarrow \psi_n$ is valid but a first order interpolant for φ_n, ψ_n in n variables can not be found.

In this chapter, it will be shown that contrary to the situation for the full finite variable fragments of first order logic, any guarded and any packed finite variable fragment does have the Beth property. For interpolation such a result is not possible, for the simple reason that GF and PF themselves do not have this property. However, the positive results on interpolation for GF and PF mentioned above do carry over to all their finite variable fragments. Moreover, we will completely chart the behavior of interpolation with regard to guarded and packed finite variable fragments. Note that such a chart is in general not straightforward, as interpolation is a very sensitive property that is in general not preserved under either expanding or restricting the logic under consideration. For example, as we already saw, first order logic has interpolation whereas its two-variable fragment does not; then again, the basic modal logic K, in its turn a fragment of the two variable fragment, does have interpolation. However, for the guarded and packed fragments considered in this chapter we indicate a sharp boundary between failure and success of interpolation. More precisely, denoting the restriction on the arity of the relation symbols by k, and the restriction on the number of variables by n, we show that GF^k_n has interpolation if and only if $min(k, n) \leq 2$. Furthermore, PF^k_n has interpolation if and only if $n \leq 2$ or $k \leq 1$. In particular, we obtain the interpolation property for the guarded two variable fragment and the packed two variable fragment. This forms a striking contrast with the situation in the full 2-variable fragment of first order logic that we described above. The aforementioned result also assures the existence of a guarded interpolant in the first order sense for guarded formulas φ, ψ such that $\varphi \rightarrow \psi$ is a validity, provided that φ, ψ contain at most binary predicates. It is worthwhile to note that a similar result for packed formulas turns out to fail. This difference will be accounted for (cf. page 143).

The above results lead us to the conclusion that guarded and packed finite variable fragments are much better behaved with respect to interpolation and definability than the full finite variable fragments of first order logic. In particular, the guarded two variable fragment and the packed two variable fragment behave very nicely.

4.2 Preliminaries

In this section we formally define the guarded fragment and the packed fragment and we give semantic characterizations in terms of suitable notions of bisimulations. The guarded fragment has been introduced in [Andréka et al., 1998]. The notion of the packed fragment first appeared in [Marx, 1999b] and bears close correspondence to the loosely guarded fragment (also known as *pairwise guarded fragment*) introduced in [van Benthem, 1997].

4.2.1. Convention By a *language* $\mathcal{L}$ we understand in this chapter a relational first order language without function- or constant symbols. Besides variables, and the parentheses), (, we consider as *logical symbols* the connectives $\neg$, $\wedge$, $\vee$, $\rightarrow$, the quantifiers $\exists$, $\forall$ and the identity symbol =. $\dashv$

4.2.2. Notation Models are denoted by calligraphic letters like $\mathcal{M}$, $\mathcal{N}$, and their respective universes by M, N, etc. Finite tuples are sometimes referred to by boldface letters, like e.g., $\boldsymbol{m} = m_1, \dots m_k$. The interpretation of an k-ary predicate R in the model $\mathcal{M}$ (notation: $I^{\mathcal{M}}(R) \subseteq M^k$) is defined as usual. For any formula φ, we write $\varphi(v_1, \dots, v_n)$ to make explicit that the free variables in φ (notation: $free(\varphi)$) are among $\{v_1, \dots, v_n\}$. For a formula $\varphi(v_1, \dots, v_n)$, a model $\mathcal{M}$, and $m_1, \dots, m_n \in M$, we write $\mathcal{M} \models \varphi[m_1, \dots, m_n]$ iff each assignment which maps v_i to m_i satisfies φ in $\mathcal{M}$. As usual, if Σ is a set of formulas, and ψ a formula, then $\Sigma \models \psi$ denotes the *consequence* relation. That is, $\Sigma \models \psi$ iff any assignment into a model $\mathcal{M}$ which satisfies all formulas in Σ also satisfies ψ. In particular, $\varphi \models \psi$ is the same as to say that $\varphi \rightarrow \psi$ is valid, i.e., $\models \varphi \rightarrow \psi$. $\dashv$

Let $\boldsymbol{v} = v_1, \dots, v_k$. Below, we use *relativized quantifications* of the form $\exists \boldsymbol{v}(\gamma, \psi)$, which denote the first order formula $\exists v_1 \cdots \exists v_k(\gamma \wedge \psi)$. We call γ the *guard* of $\exists \boldsymbol{v}(\gamma, \psi)$.

4.2.3. Definition [Guarded Fragment] Let $\mathcal{L}$ be a language. As usual, the *atomic $\mathcal{L}$-formulas* (or, $\mathcal{L}$-atoms) are of the form:

1. $v_1 = v_2$, for variables v_1, v_2, not necessarily distinct.
2. $Pv_1 \cdots v_k$, for k-ary $P \in \mathcal{L}$ and variables $v_1, \dots, v_k$, not necessarily distinct.

The *guarded fragment* of first order logic (GF) is defined by induction as follows.

1. Any atomic formula is in GF.
2. If $\varphi, \psi \in$ GF, then $\varphi \wedge \psi$ and $\neg\varphi$ are in GF.
3. Let $\boldsymbol{v}$ be a finite, non-empty sequence of variables, $\psi \in$ GF and G an atomic formula such that $free(\psi) \subseteq free(G)$. Then $\exists \boldsymbol{v}(G, \psi) \in$ GF. $\dashv$

The formulas $\varphi \vee \psi$, $\varphi \to \psi$ and $\varphi \leftrightarrow \psi$ denote the usual abbreviations. Note that as a dual of guarded existential quantification we also get guarded universal quantification, of the form $\forall \boldsymbol{v}(G \to \psi)$.

A typical example of a formula in GF is $\forall v_1 v_2(Rv_1v_2 \to Rv_2v_1)$, expressing the symmetry of the relation R. On the other hand, the translation of the tense logical formula $\mathsf{Until}(\varphi, \psi)$, i.e., the formula $\exists v_2(v_1 < v_2 \wedge \psi(v_2) \wedge \forall v_3[(v_1 < v_3 \wedge v_3 < v_2) \to \varphi(v_3)])$, is not in GF as the guard in the subformula expressing betweenness (i.e., the formula $\forall v_3[(v_1 < v_3 \wedge v_3 < v_2) \to \varphi(v_3)])$ is not atomic.

4.2.4. Remark For readers familiar with [Andréka et al., 1998] we note that contrary to that paper, Definition 4.2.3 allows for identity atoms as guards. Since this issue does not affect decidability nor interpolation, we decided to concentrate on this slightly more general fragment. This also places us in line with [Grädel, 1999]. ⊣

Let γ be a formula with $free(\gamma) = \{v_1, \dots, v_k\}$. We say that γ *packs the set of variables* $\{v_1, \dots, v_k\}$ if γ is a conjunction of formulas of the form $v_i = v_j$, or $R(v_1, \dots, v_l)$, or $\exists \boldsymbol{v} R(v_1, \dots, v_l)$, such that every pair $v_i, v_j \in \{v_1, \dots, v_k\}$ is guarded by γ. That is, for every pair v_i, v_j there exists a conjunct in γ in which v_i and v_j both occur free. A relativized quantification $\exists \boldsymbol{v}(\gamma, \psi)$ is called *packed* if γ packs $free(\gamma)$ and $free(\psi) \subseteq free(\gamma)$. Again, we call γ the *guard* of $\exists \boldsymbol{v}(\gamma, \psi)$.

4.2.5. Definition [Packed Fragment] The *packed fragment* of first order logic (PF) is defined similarly to GF, by relaxing clause 3 as follows:

3'. If $\psi \in$ PF, then any packed relativized quantification $\exists \boldsymbol{v}(\gamma, \psi)$ is in PF. ⊣

Obviously, GF $\subseteq$ PF. The translation of $\mathsf{Until}(\varphi, \psi)$ is not packed itself, as the guard $(v_1 < v_3 \wedge v_3 < v_2)$ does not pack its own free variables $\{v_1, v_2, v_3\}$. However, the translation is equivalent to the packed $\exists v_2(v_1 < v_2 \wedge \psi(v_2) \wedge \forall v_3[(v_1 < v_2 \wedge v_1 < v_3 \wedge v_3 < v_2) \to \varphi(v_3)])$.

4.2.6. Notation Fix a countable set of variables $\{v_i : i \in \omega\}$, and let $k, n \in \omega$. By GF$_n$ we understand the fragment of GF that consists of formulas whose variables (free or bound) are among $v_1, \dots, v_n$. GF$_n^k$ denotes the collection of formulas in GF$_n$ which are built up from at most k-ary relation symbols. Similar notation will be used for fragments of PF. ⊣

4.2.7. Remark For readers familiar with [van Benthem, 1997], we point out the differences between the packed fragment and the so-called *loosely guarded fragment* (LGF) introduced and studied there. We recall from opus cit. that a relativized quantification $\exists \boldsymbol{v}(\gamma, \psi)$ is *loosely guarded* if γ is a conjunction of relational atoms such that $free(\gamma) = free(\psi) \cup \boldsymbol{v}$ and moreover every pair $\langle v, u \rangle$ such that $v \in \boldsymbol{v}$ and $u \in free(\gamma)$ is guarded by γ.

On the one hand, PF is more liberal. For, contrary to the case for LGF, formulas of the form $\exists \boldsymbol{v} R(v_1, \ldots, v_k)$ and identities may occur in the guard of a packed quantifier. This is a real expansion. For example, the packed formula $\exists v_2 v_3[(\exists v_4 R v_1 v_2 v_4 \wedge S v_2 v_3 \wedge S v_1 v_3) \wedge \neg R v_1 v_2 v_3]$ has no loosely guarded equivalent.

On the other hand, LGF is more liberal as in a loosely guarded quantifier $\exists \boldsymbol{v}(\gamma, \psi)$ a pair of variables from $free(\gamma) \setminus \boldsymbol{v}$ does not need to be guarded by γ. For this reason, the translation of $\mathsf{Until}(\varphi, \psi)$ is already in LGF.

For sentences though, PF is an expansion of LGF. For, let χ be a loosely guarded sentence, and let $\chi' = \exists \boldsymbol{v}(\gamma, \psi)$ be a subformula of χ which is not packed. Then some $u, u' \in free(\gamma) \setminus \boldsymbol{v}$ are not guarded by γ. As χ is a sentence, at some stage u, u' will be quantified over. More precisely, χ has a subformula of the form $\exists \boldsymbol{v}'(\gamma', \psi')$, such that χ' is a subformula of ψ' and $\boldsymbol{v}' \cap \{u, u'\} \neq \emptyset$ and $\{u, u'\} \subseteq free(\psi')$. By definition of LGF, there exists a conjunct G in γ' which guards u, u'. Replace χ' in χ by $\exists \boldsymbol{v}(\gamma \wedge G, \psi)$. The thus obtained sentence is equivalent to χ, and in this replacement u, u' are guarded. Repeating this procedure will eventually yield a packed equivalent of χ.

Viewed as an expressive fragment of first order logic which is still decidable, the choice between PF and LGF seems to be irrelevant. By introducing new predicates, one easily defines simple translations from PF to LGF and vice versa, such that a formula is satisfiable iff its translation is. We choose to study PF. The reason for this twofold. First, as we just saw, for sentences PF is a real expansion of LGF. Second, the definition of PF seems to be more straightforward. This is reflected in the fact that PF allows for a semantic characterization which naturally extends the characterization of GF known from [Andréka et al., 1998] (cf. Definition 4.2.9). $\dashv$

Similar to modal logics, the guarded fragment can be semantically analyzed via a suitable notion of bisimulation. This has been done in [Andréka et al., 1998]. Here we will recapitulate as much of these results as needed for our purposes. Moreover, we will introduce the notion of a *packed bisimulation*.

4.2.8. Definition [**live set, packed set**] Let Z be a finite subset of a model $\mathcal{M}$. Z is called *live* in $\mathcal{M}$ if Z is either a singleton, or there exists a relation R and a set X such that $Z \subseteq X$ and the elements of X are R-related, in any order or multiplicity. E.g., if $\langle m_2, m_1, m_1, m_3 \rangle \in I^{\mathcal{M}}(R)$, then $\{m_2, m_3\}$ is live in $\mathcal{M}$. In this case we say that $\{m_2, m_3\}$ is *R-live* (in $\mathcal{M}$). For any language $\mathcal{L}$, $Z \subseteq^l_{\mathcal{L}} \mathcal{M}$ denotes that Z is $\mathcal{L}$-live in $\mathcal{M}$. That is, Z is R-live in $\mathcal{M}$, for some $R \in \mathcal{L}$. Z is called *$\mathcal{L}$-packed* in $\mathcal{M}$ if any pair $m, m' \in Z$ is $\mathcal{L}$-live in $\mathcal{M}$. Notation: $Z \subseteq^p_{\mathcal{L}} \mathcal{M}$. $\dashv$

Below, by a finite *partial $\mathcal{L}$-isomorphism* f we mean a finite one-to-one partial map between two models $\mathcal{M}, \mathcal{N}$ which preserves $\mathcal{L}$-relations in both ways. That is, for every k-ary $R \in \mathcal{L}$ and any $m_1, \ldots, m_k \in M$, $[\langle m_1, \ldots, m_k \rangle \in I^{\mathcal{M}}(R)$ iff $\langle f(m_1), \ldots, f(m_k) \rangle \in I^{\mathcal{N}}(R)]$. By the *image* of a map $f : X \longrightarrow Y$ we understand the set $\{f(x) : x \in X\}$, and we refer to X as the *domain* of f.

4.2.9. Definition [Guarded bisimulation, packed bisimulation] A non-empty set F of finite partial $\mathcal{L}$-isomorphisms between two models $\mathcal{M}$ and $\mathcal{N}$ is called a *guarded* $\mathcal{L}$*-bisimulation* between $\mathcal{M}$ and $\mathcal{N}$ if for any $f : X \longrightarrow Y \in F$ the following holds:

1. For any $Z \subseteq^l_{\mathcal{L}} \mathcal{M}$ there is a $g \in F$ with domain Z such that g and f agree on the intersection of their domains. (The *zig*-condition)

2. For any $W \subseteq^l_{\mathcal{L}} \mathcal{N}$ there is a $g \in F$ with image W such that g^{-1} and f^{-1} agree on the intersection of their domains. (The *zag*-condition)

Stipulating the zig- and zag-condition for $\mathcal{L}$-packed subsets yields the notion of a *packed* $\mathcal{L}$*-bisimulation* ⊣

4.2.10. Notation We write $\mathcal{M}, m_1 \cdots m_k \sim_F \mathcal{N}, n_1 \cdots n_k$ to denote that F is a guarded bisimulation between $\mathcal{M}$ and $\mathcal{N}$ with some $f \in F$ such that $f(m_i) = n_i$, for $i \leq k$. ⊣

Guarded bisimulations are defined in such a way as to preserve guarded formulas. That is, for a guarded $\mathcal{L}$-formula $\varphi(v_1, \dots, v_k)$, a guarded $\mathcal{L}$-bisimulation F between models $\mathcal{M}$, $\mathcal{N}$, $f \in F$, and $m_1, \dots, m_k \in dom(f)$ it is the case that $\mathcal{M} \models \varphi[m_1, \dots, m_k] \;\Leftrightarrow\; \mathcal{N} \models \varphi[f(m_1), \dots, f(m_k)]$. This can be shown by a straightforward induction on the complexity of φ. For atomic formulas this follows directly from the definition according to which f is a one-to-one map which preserves $\mathcal{L}$-relations in both ways. The zig- and zag-conditions precisely take care of the induction step for existential quantification. Indeed, preservation under guarded bisimulations is the characteristic feature of GF, in the sense of the following Characterization Theorem from [Andréka et al., 1998]: up to logical equivalence, GF precisely consists of those first order formulas that are preserved under guarded bisimulations. Similarly, one shows that preservation under packed bisimulations is the characteristic feature of PF. This result is closely related to the characterization in terms of δ-bisimulations in [Marx, 1999b, Fact 3.11].

In the definition of a guarded bisimulation in [Andréka et al., 1998], the role of live sets in Definition 4.2.9 is taken over by what Andréka, van Benthem and Németi call *guarded* sets. The main difference is that singletons need not be guarded, whereas they are always live. This semantically reflects the fact that we allow for identity atoms as guards. Mutatis mutandis, all arguments in [Andréka et al., 1998] and in particular the characterization theorem also apply to guarded formulas and guarded bisimulations as defined here.

In the final part of this section we recall from [Marx, 1997] a reduction of the satisfiability problem of PF to a particularly nice subfragment of the guarded fragment. This fragment is the first order counterpart of the modal logic of relations *MLR* from [Venema and Marx, 1999]. We use the same reduction in section 4.4 to obtain our positive results on interpolation. The reduction is given as follows.

Let $\chi \in \mathrm{PF}_k^k$, and let C be a k-ary relation symbol not occurring in χ. χ^C denotes the result of replacing each packed quantification in χ of the form $\exists \boldsymbol{v}(\gamma, \psi)$, by $\exists \boldsymbol{uv}(Cv_1 \cdots v_k, \gamma \wedge \psi)$, where $\exists \boldsymbol{u}$ binds exactly the variables $\{v_1, \ldots, v_k\} \setminus free(\gamma \wedge \psi)$. Note that $\chi^C \in \mathrm{GF}_k^k$. Moreover, $Cv_1 \cdots v_k$ is the only guard in χ^C and the relation symbol C solely occurs at guarding position. We call formulas in GF_k^k with this property *MLR_k^C–formulas.*

Define the set $\mathrm{LC} = \{\forall v_1 \cdots v_k(Cv_1 \cdots v_k \to Cv_{\sigma(1)} \cdots v_{\sigma(k)}) \mid \sigma : k \longrightarrow k\}$. As shown in [Venema and Marx, 1999], LC axiomatizes the class of models on which C is interpreted as a *local cube.*

4.2.11. Theorem [Marx, 1997] *Let* $\varphi \in \mathrm{PF}_k^k$. *Then*

$$\models \varphi \ \ \textit{iff} \ \ \mathrm{LC}, Cv_1 \cdots v_k \models \varphi^C.$$

Proof: First, let $\mathcal{M} = (D, I) \not\models \varphi$. Expand $\mathcal{M}$ to a model $\mathcal{M}^+$ by setting $I^+(C) = {}^kD$. Obviously $\mathcal{M}^+ \models LC \wedge \forall \boldsymbol{v} Cv_1 \cdots v_k$, but $\mathcal{M}^+ \not\models \varphi^C$.

For the other direction, we use the fact that guarded formulas can be satisfied in unraveled models or, using a different terminology, tree models. This is established in several papers, e.g., [Andréka et al., 1998], [Marx, 1997] and [Grädel, 1999]. We use Grädel's formulation. Assume that $LC, Cv_1 \cdots v_k \not\models \varphi^C$. Thus the GF_k^k formula $LC \wedge Cv_1 \cdots v_k \wedge \neg\varphi^C$ is satisfiable. By [Grädel, 1999, Lemma 3.6] and the construction in the proof of Theorem 3.7 in [Grädel, 1999] this formula is satisfiable in a tree model $\mathcal{M}$ in which every packed set is C–live. We claim that in this model $\mathcal{M}$, for every assignment $s \in I(C)$ and for every subformula ψ of φ, we have

$$\mathcal{M} \models \psi[s] \text{ if and only if } \mathcal{M} \models \psi^C[s].$$

Obviously from this the theorem follows. The only non–trivial case of the inductive proof is the following. Let $\boldsymbol{v} = v_1, \ldots, v_n$, $\boldsymbol{u} = u_1, \ldots, u_m$, and $\boldsymbol{s} = s_1, \ldots, s_m$. Let $\mathcal{M} \models \exists \boldsymbol{v}(\varphi(\boldsymbol{u}, \boldsymbol{v}), \psi(\boldsymbol{u}, \boldsymbol{v}))[\boldsymbol{s}]$. Then $\mathcal{M} \models \varphi(\boldsymbol{u}, \boldsymbol{v}) \wedge \psi(\boldsymbol{u}, \boldsymbol{v})[\boldsymbol{s}, \boldsymbol{t}]$, for some $\boldsymbol{t} = t_1, \ldots, t_n$. As $\varphi(\boldsymbol{u}, \boldsymbol{v})$ packs all its free variables, the set $\{s_1, \ldots, s_m, t_1, \ldots, t_n\}$ is packed. Thus, by the assumption on the model $\mathcal{M}$, it is C–live. Hence, as $\mathcal{M} \models LC$, every k–tuple with elements from this set is in $I(C)$. Thus the inductive hypothesis applies and $\mathcal{M} \models Cv_1 \cdots v_k \wedge \varphi(\boldsymbol{u}, \boldsymbol{v}) \wedge \psi^C(\boldsymbol{u}, \boldsymbol{v})[\boldsymbol{s}, \boldsymbol{t}, r_1, \ldots, r_l]$, for appropriately chosen $r_1, \ldots, r_l$. Thus, $\mathcal{M} \models \exists \boldsymbol{vv'}(Cv_1 \cdots v_k \wedge \varphi(\boldsymbol{u}, \boldsymbol{v}) \wedge \psi^C(\boldsymbol{u}, \boldsymbol{v}))[\boldsymbol{s}]$, which is the C–translation of the formula $\exists \boldsymbol{v}(\varphi(\boldsymbol{u}, \boldsymbol{v}), \psi(\boldsymbol{u}, \boldsymbol{v}))$. ■

We will apply Theorem 4.2.11 in section 4.4 in order to derive results on interpolation for GF and PF from an interpolation theorem for MLR_k^C-formulas.

4.3 Failure of interpolation in GF and PF

In this section, we present examples showing that interpolation fails both in the guarded fragment and in the packed fragment. More precisely, we define guarded sentences φ, ψ in three variables using at most ternary relations such that $\models \varphi \to \psi$, which lack a guarded interpolant in any number of variables. Next, a similar example will be given for packed sentences built up from three variables and binary relations. We first recall the definition of the interpolation property.

4.3.1. Definition [**Interpolation in fragments of first order logic**] A fragment $\mathcal{F}$ of first order logic is said to have the *interpolation property* if for any pair of formulas $\varphi, \psi \in \mathcal{F}$ such that $\models \varphi \to \psi$ there exists an interpolant in $\mathcal{F}$. That is, there exists a formula $\vartheta \in \mathcal{F}$ which is built up from relation symbols which occur both in φ and ψ such that $\models \varphi \to \vartheta$ and $\models \vartheta \to \psi$. $\dashv$

Note that in the present context the interpolation properties $\mathrm{CIP}^{\to}$ and $\mathrm{CIP}^{\models}$ that we distinguished in chapter 2 (cf. Definition 2.3.1 and Definition 2.3.14) are equivalent due to the deduction theorem of first order logic.

4.3.2. Theorem (Failure of interpolation in guarded fragment) *There exist sentences $\varphi, \psi \in \mathrm{GF}^3_3$ such that $\models \varphi \to \psi$, without a guarded interpolant in any number of variables.*

Proof: For $i = 1, 2$, define the languages $\mathcal{L}_i$ and the GF^3_3-theories T_i as follows. $\mathcal{L}_1$ consists of one ternary predicate Δ, and the binary predicates R, B, G, P, W. Intuitively, Δxyz holds for nodes x, y and z which together form a triangle, and Rxy (resp. B, G, P, W) holds whenever the edge between the nodes x and y is painted red (resp. blue, green, purple and white). $\mathcal{L}_2$ does not contain the predicate Δ, but instead contains two extra binary predicates R_1, R_2. Theory T_1 is defined by

$$\exists xyz[\Delta xyz \wedge Rxy \wedge Bxz \wedge Gzy \wedge \exists z(\Delta xyz \wedge Pxz \wedge Wzy)]. \tag{4.1}$$

T_1 demands the existence of a situation as e.g., occurs in model $\mathcal{M}_1$ in Figure 4.1. Theory T_2 is the conjunction of the sentences

$$\forall xy[R_1xy \to Rxy], \quad \forall xy[R_2xy \to Rxy], \quad \forall xy[Rxy \to (R_1xy \vee R_2xy)]. \tag{4.2}$$

$$\forall xy[R_1xy \to \forall y(Pxy \to \forall x[Wyx \to \forall y(Ryx \to \neg R_1yx)])]. \tag{4.3}$$

$$\forall xy[R_2xy \to \forall y(Bxy \to \forall x[Gyx \to \forall y(Ryx \to \neg R_2yx)])]. \tag{4.4}$$

The formulas numbered (4.2) can be read as saying that the color red has precisely two shades R_1 and R_2. By (4.3) and (4.4) the situations in Figure 4.2 are prohibited.

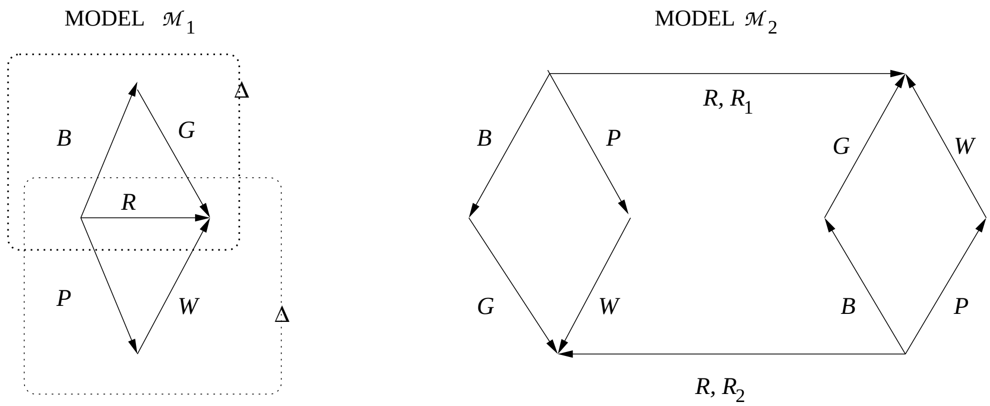


Figure 4.1: Guarded $\mathcal{L}_1 \cap \mathcal{L}_2$-bisimilar models.

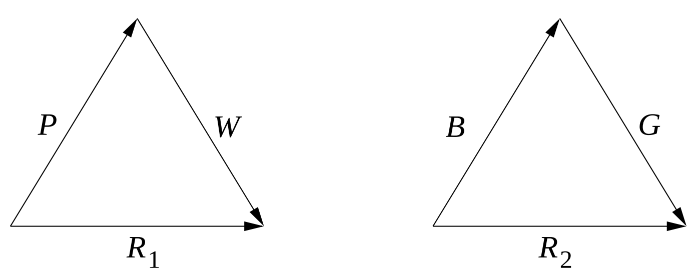


Figure 4.2: (4.3) and (4.4) prohibit these situations.

From Figures 4.1 and 4.2 it is clear that $\models T_1 \rightarrow \neg T_2$. However, T_1 and $\neg T_2$ do not have a guarded interpolant. For, consider the models $\mathcal{M}_1$, $\mathcal{M}_2$ in Figure 4.1. Obviously, $\mathcal{M}_i$ is a model for T_i. Let F be the set of finite partial maps from $\mathcal{M}_1$ to $\mathcal{M}_2$ which map an $\mathcal{L}_1 \cap \mathcal{L}_2$-live subset of $\mathcal{M}_1$ to a subset of $\mathcal{M}_2$ which is live by the same predicates. The reader easily checks that F is a guarded $\mathcal{L}_1 \cap \mathcal{L}_2$-bisimulation. Let us suppose that there does exist some guarded interpolant ϑ for $T_1, \neg T_2$. As $\mathcal{M}_1 \models T_1$, then $\mathcal{M}_1 \models \vartheta$. Since guarded $\mathcal{L}$-formulas are preserved by guarded $\mathcal{L}$-bisimulation, we obtain that $\mathcal{M}_2 \models \vartheta$. Hence, $\mathcal{M}_2 \models \neg T_2$. Contradiction. ■

4.3.3. Corollary GF, GF_3 *and* GF^3 *do not have the interpolation property.*

It is not hard to check that the packed formula $\exists xyz[Rxy \wedge Bxz \wedge Gzy \wedge \exists z(Pxz \wedge Wzy)]$ is an interpolant for the theories T_1 and $\neg T_2$ in the above example. This leaves open the question of interpolation for PF. However, along the same lines as above an example can be given showing that also PF fails to have interpolation.

4.3.4. Theorem (Failure of interpolation in the packed fragment) *There exist sentences* $\varphi, \psi \in \text{PF}_3^2$ *such that* $\models \varphi \rightarrow \psi$, *without a packed interpolant in any number of variables.*

Proof: For $i = 1, 2$, define the languages $\mathcal{L}_i$ and the PF_3^2 theories T_i as follows. $\mathcal{L}_1$ contains the binary predicates B, W, G, P, O, Y, R and J. $\mathcal{L}_2$ does not contain J, but has two extra binary predicates R_1, R_2. T_1 is the sentence

$$\exists xyz[Bxy \wedge Gyz \wedge Jxz \wedge \exists y(Rxy \wedge Wzy \wedge \exists z(Pxz \wedge Jzy \wedge \exists x(Ozx \wedge Yxy)))].$$

Model $\mathcal{M}_1$ in Figure 4.4 is a typical model of T_1. Theory T_2 is the conjunction of (4.2) together with

$$\forall xy[R_1xy \rightarrow \forall y(Bxy \rightarrow \forall x[Gyx \rightarrow \forall y(Wxy \rightarrow \forall x[Rxy \rightarrow \neg R_1xy])])],$$

and

$$\forall xy[R_2xy \rightarrow \forall y(Pxy \rightarrow \forall x[Oyx \rightarrow \forall y(Yxy \rightarrow \forall x[Rxy \rightarrow \neg R_2xy])])].$$

This theory T_2 prohibits the situations in Figure 4.3 below.

The figures clearly show that $\models T_1 \rightarrow \neg T_2$. We will imitate the model-theoretic argument in the proof of Theorem 4.3.2 to see that T_1 and $\neg T_2$ do not have a packed interpolant. This time consider the models $\mathcal{M}_1$, $\mathcal{M}_2$ in Figure 4.4. Again, $\mathcal{M}_i$ is a model for T_i. One readily verifies that the models $\mathcal{M}_1$ and $\mathcal{M}_2$ are packed $\mathcal{L}_1 \cap \mathcal{L}_2$-bisimilar. The crucial observation is that all $\mathcal{L}_1 \cap \mathcal{L}_2$-packed sets in $\mathcal{M}_1$ and $\mathcal{M}_2$ are $\mathcal{L}_1 \cap \mathcal{L}_2$-live. Therefore, the set F of finite partial maps from $\mathcal{M}_1$ to $\mathcal{M}_2$ which map an $\mathcal{L}_1 \cap \mathcal{L}_2$-live subset of $\mathcal{M}_1$ to a subset of $\mathcal{M}_2$ which is live by the same predicates, forms a *packed* $\mathcal{L}_1 \cap \mathcal{L}_2$-bisimulation between $\mathcal{M}_1$ and $\mathcal{M}_2$. Since packed $\mathcal{L}$-sentences are invariant under packed $\mathcal{L}$-bisimulation, the same argument as in the proof of the previous theorem leads to the desired contradiction. ■

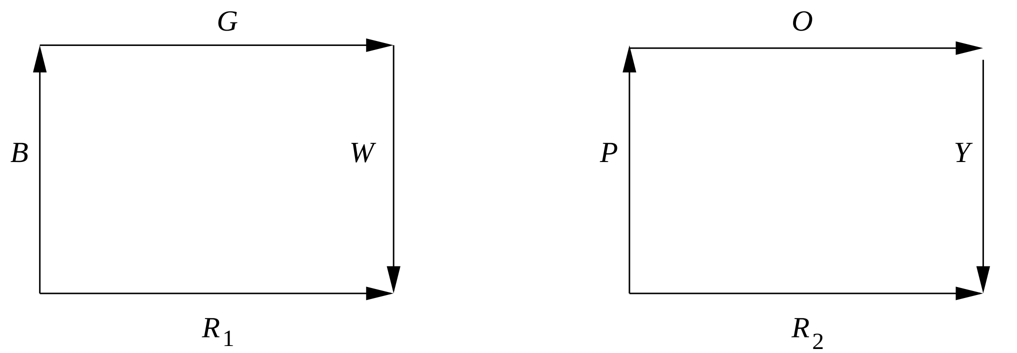


Figure 4.3: Theory T_2 prohibits these situations.

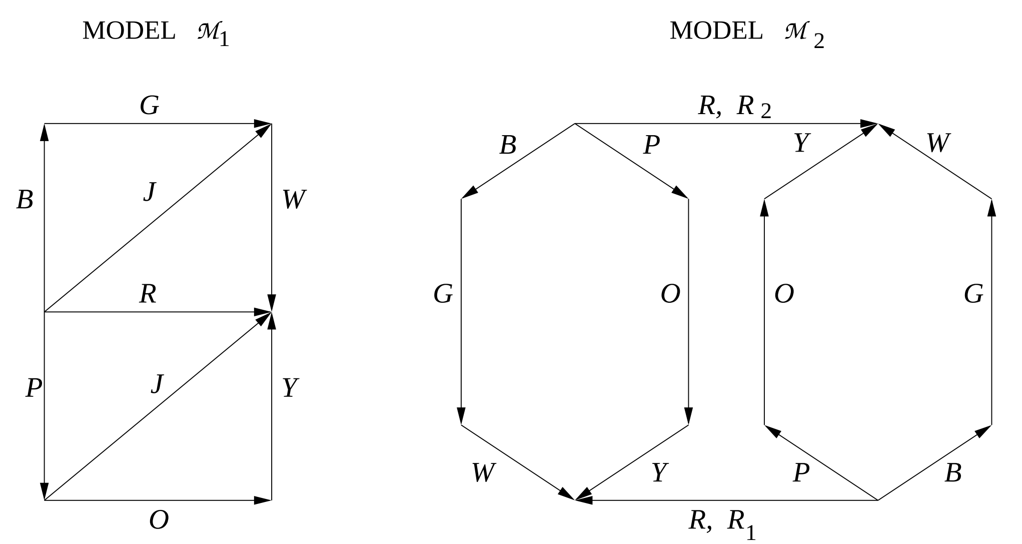


Figure 4.4: Packed $\mathcal{L}_1 \cap \mathcal{L}_2$-bisimilar models.

4.3.5. Corollary PF, PF_3 *and* PF^2 *do not have the interpolation property.*

Recapitulating the results in this section, we see that interpolation fails in GF and PF if we have at least 3 variables at our disposal. We also note that in the example showing the failure of interpolation in GF we made use of a ternary relation, whereas in the case of PF we could do with binary relations only. In section 4.6 it will be shown that these are the smallest counterexamples possible.

4.3.6. Remark According to Convention 4.2.1, we consider purely relational languages. Without deviating too much from this, let us just observe that treating constants as in [Grädel, 1999] and allowing e.g., guarded formulas of the form $T_1 = \forall x(Px \rightarrow (x = c))$, will yield failure of interpolation even in GF_1^1. For, consider $T_2 = \forall x(P_1x \rightarrow (Px \wedge \neg P_2x)) \wedge \forall x(P_2x \rightarrow Px) \wedge \exists x P_1x \wedge \exists x P_2x$. Then $\models T_1 \rightarrow \neg T_2$, and similar to the previous examples one verifies the absence of any guarded interpolant. ⊣

4.4 Restoring interpolation

To see why the interpolation property of first order logic fails in restriction to the guarded fragment, it is useful to compare it to the interpolation property studied in modal logic. In modal logic, the interpolant is usually confined to proposition letters in the common language but may contain non-shared modalities. Also in first order logic this kind of interpolation is studied, when we view $\exists x$ as a modality. This is e.g., the case in cylindric algebra theory, cf. [Henkin et al., 1985]. The point is that strengthening the requirement on the common language to include common modalities results in a much stronger interpolation property. [van Benthem, 1999] shows this stronger property for the basic multi-modal logic K_n. [Marx, 1999a] generalizes this result to Sahlqvist axiomatizable multi-modal logics whose axioms correspond to universal Horn frame conditions which do not specify any interaction between the different accessibility relations (e.g., bi-modal $\mathsf{S5}$). However, when there is interaction, the stronger interpolation property is easily lost as the following example from [van Benthem, 1999] shows. Consider the multi-modal logic defined by the axiom $\Diamond_1 p \rightarrow \Diamond_2 p$. This logic does not have the stronger interpolation property. For, in this logic $\Diamond_1 True \rightarrow \Diamond_2 True$ is a theorem whereas the only formulas in the common language (in the strong sense) are $True$ and $False$, which are obviously not interpolants. However, this logic does have the usual interpolation property. Summarizing, we conclude that the requirement on the common language to include common modalities yields a much stronger interpolation property.

Thinking of guarded formulas as translations of modal formulas, we see that Definition 2.3.1 formulates exactly this stronger version of interpolation, where 'common language' means the set of common relation symbols which includes both the relations which are translated proposition letters and the relations that are obtained in translating the modalities. This suggests to consider an alternative interpolation property for the guarded fragment that more closely resembles the one that is usually studied in modal logic. This will be done in the present section.

We first give an interpolation theorem for a particularly nice subfragment of GF that we already encountered, the so-called MLR^C_k-formulas. From this theorem we then derive, in a uniform way, a 'modal' interpolation theorem for the guarded fragment and the packed fragment. A word of warning is called for, since this section is of a highly technical nature. We therefore advice the reader to concentrate on the basic idea, viz. the analogy with modal logic. Other technical concepts that are introduced (although quite naturally and with good reason) need not bother the reader at a first reading.

In order to formulate a generalized version of the modal interpolation property for the guarded fragment, we will first have to distinguish occurrences of relation symbols as guards from other occurrences.

4.4.1. Notation For any guarded or packed formula φ we understand by $\mathcal{L}_{G(\varphi)}$ the set of relations that occur in the guard of some quantification in φ. Similarly, $\mathcal{L}_{\bar{G}(\varphi)}$ denotes the set of relations that occur in φ at a non-guard position. We call a formula φ *uniformly guarded* in a formula χ, if χ is the only guard occurring in φ. ⊣

The sets $\mathcal{L}_{G(\varphi)}$ and $\mathcal{L}_{\bar{G}(\varphi)}$ are guarded counterparts of the sets of modalities and proposition letters, respectively, occurring in a modal formula. Note however that whereas there is a syntactic distinction between modalities and proposition letters, the sets $\mathcal{L}_{G(\varphi)}$ and $\mathcal{L}_{\bar{G}(\varphi)}$ are not necessarily disjoint. For example, in $\varphi = \exists x(Px \wedge \forall y(Sxy \to Py))$, the relation P occurs both as a guard and as a non-guard.

The main lemma of this section is an interpolation theorem for MLR^C_k-formulas. These were introduced at page 125 as those formulas in GF^k_k which are uniformly guarded in $Cv_1 \cdots v_k$, while the relation symbol C only occurs at guarding position. Recall that $\mathrm{LC} = \{\forall v_1 \cdots v_k(Cv_1 \cdots v_k \to Cv_{\sigma(1)} \cdots v_{\sigma(k)}) \mid \sigma : k \longrightarrow k\}$.

4.4.2. Lemma (Interpolation for MLR^C_k formulas) *Let $k \in \omega$, and let $\varphi, \psi \in MLR^C_k$ such that $LC, Cv_1 \cdots v_k \models \varphi \to \psi$. Then there exists an interpolant $\vartheta \in MLR^C_k$ such that*

1. *$LC, Cv_1 \cdots v_k \models \varphi \to \vartheta$ and $LC, Cv_1 \cdots v_k \models \vartheta \to \psi$,*
2. *$\mathcal{L}_{\bar{G}(\vartheta)} \subseteq \mathcal{L}_{\bar{G}(\varphi)} \cap \mathcal{L}_{\bar{G}(\psi)}$, and*
3. *$free(\theta) \subseteq free(\varphi) \cup free(\psi)$.*

We prove Lemma 4.4.2 at the end of this section. First we continue our story and show that the guarded fragment indeed has a 'modal' interpolation property. The main result is Theorem 4.4.5 which states that under some mild assumptions (that e.g., all sentences satisfy) every $\varphi, \psi \in \mathrm{GF}^k_k$ such that $\varphi \to \psi$ is a validity, have an interpolant ϑ in GF^k_k in which only

- relations from the common language occur at non-guard positions, and
- guards in either φ or ψ occur at guard positions.

So, the guards in the interpolant need not be in the common language but they do occur as a guard in either φ or ψ. Note also that the non-guards in the interpolant are only required to occur in φ and ψ: not necessarily as non-guards. It is possible to strengthen this result by showing ϑ to be built up from non-guards in the common language which moreover occur both in φ and in ψ as non-guards. However, this takes one extra variable. We come back to this in Remark 4.4.6. At this point, all we want to convey is the analogy with the interpolation property for modal logics: an interpolant may use non-shared modalities, but uses shared proposition letters only.

It is our aim to specify and prove the above exposition. To this end, we first observe that for any relation symbol R and $n \in \omega$, the set of R-live n-tuples can be defined by a guarded formula. More precisely, there exists a formula $\mathsf{L}ive_R(v_1, \dots, v_n) \in \mathrm{GF}$ such that for any model $\mathcal{M}$ and $m_1, \dots, m_n \in M$:

$$\mathcal{M} \models \mathsf{L}ive_R(v_1, \dots, v_n)[m_1, \dots, m_n] \text{ iff the set } \{m_1, \dots, m_n\} \text{ is } R\text{-live in } \mathcal{M}.$$

Details on how to obtain this formula can be found in the appendix following this chapter. For any *finite* language $\mathcal{L}$ we further obtain a formula $\mathsf{L}ive_{\mathcal{L}}(v_1, \dots, v_n)$ defining the set of $\mathcal{L}$-live n-tuples. If $\mathcal{L}$ contains at most k-ary relations and $n \leq k$, then $\mathsf{L}ive_{\mathcal{L}}(v_1, \dots, v_n) \in \mathrm{GF}^k_k$. Moreover, as the following proposition states, we may use the formula $\mathsf{L}ive_{\mathcal{L}}$ as a guard while staying within the guarded fragment, modulo logical equivalence.

4.4.3. Proposition *Let $\mathcal{L}$ be a language which contains at most k-ary relation symbols. Any first order formula φ in k variables which is uniformly guarded in $\mathsf{L}ive_{\mathcal{L}}(v_1, \dots, v_k)$ is equivalent to some $\varphi' \in \mathrm{GF}^k_k$ such that $\mathcal{L}_{G(\varphi')} = \mathcal{L}$ and $\mathcal{L}_{\bar{G}(\varphi')} = \mathcal{L}_{\bar{G}(\varphi)}$.*

Proof: See the appendix at the end of this chapter. ■

The formula $\mathsf{L}ive_{\mathcal{L}}$ and Proposition 4.4.3 will also play an important role in section 4.5. Theorem 4.4.5 shows the existence of an interpolant for validities $\varphi \to \psi$ under the conditions that $free(\psi) \subseteq free(\varphi)$ and $\varphi, \neg\psi$ having the following property.

4.4.4. Definition [**Self-guarded formula**] A formula $\varphi(v_1,\dots,v_n)$ is called *self-guarded* if $\models \varphi \to \mathsf{Live}_{\mathcal{L}_{G(\varphi)}}(v_1,\dots,v_n)$. ⊣

Recall that singletons are live sets. Therefore, any formula in at most one free variable is self-guarded. In particular, sentences are self-guarded. Restricting attention to self-guarded formulas brings the syntax of GF closer to the semantic intuition as captured by the notion of a guarded bisimulation. For, consider a guarded bisimulation F. According to Definition 4.2.9, the domain of a partial isomorphism $f \in F$ need not be live. When we consider self-guarded formulas only, we can change the definition of a bisimulation such that the domain of every $f \in F$ *is* live while the characterization theorem still goes through. For this reason, the proof of the interpolation theorem is much simplified by only considering self-guarded formulas. Note that for decidability this matter does not show up, as a guarded formula $\varphi(\boldsymbol{x})$ is satisfiable iff the self-guarded formula $\exists \boldsymbol{x}(R\boldsymbol{x} \wedge \varphi(\boldsymbol{x}))$ is satisfiable, where R is some hitherto unused predicate symbol. Obviously, changing the language is no option when investigating interpolation.

We are now in a position to precisely formulate the modal interpolation property of the guarded fragment. As before, by $\mathcal{L}_\varphi$ (read: *the language of* φ) we denote the set of relation symbols occurring in φ.

4.4.5. Theorem (GF^k_k **has modal interpolation**) *Let $k \in \omega$. Let $\varphi, \psi \in \mathrm{GF}^k_k$ be such that $free(\psi) \subseteq free(\varphi)$ and $\varphi, \neg\psi$ are self-guarded. If $\models \varphi \to \psi$, then there exists an interpolant $\vartheta \in \mathrm{GF}^k_k$ satisfying the following conditions:*

1. $\models \varphi \to \vartheta$ *and* $\models \vartheta \to \psi$.
2. $\mathcal{L}_{G(\vartheta)} \subseteq \mathcal{L}_{G(\varphi)} \cup \mathcal{L}_{G(\psi)}$.
3. $\mathcal{L}_{\bar{G}(\vartheta)} \subseteq \mathcal{L}_\varphi \cap \mathcal{L}_\psi$.
4. *ϑ is equivalent to a formula χ with $\mathcal{L}_{\bar{G}(\chi)} \subseteq \mathcal{L}_{\bar{G}(\vartheta)}$ and which is uniformly guarded in $\mathsf{Live}_{\mathcal{L}_{G(\varphi)} \cup \mathcal{L}_{G(\psi)}}(v_1,\dots,v_k)$.*

The condition that $free(\psi) \subseteq free(\varphi)$ exempts us from the obligation to give an interpolant for implications like $\models \varphi(u) \to \psi(u,v)$. Rightfully so, as this implication is equivalent to $\models \varphi(u) \to \forall v(\psi(u,v))$ which happens to lie outside the guarded context.

Proof: Let k, φ, ψ be as in the theorem. By simply renaming the variables we may assume $free(\varphi) = \{v_1,\dots,v_l\}$, for some $l \leq k$. Moreover, we assume $free(\psi) = free(\varphi)$. For, from the fact that $\models \varphi \to \psi$ we infer that $\models \underbrace{\exists \boldsymbol{v}(\mathsf{Live}_{\mathcal{L}_{G(\varphi)}}(v_1,\dots,v_l) \wedge \varphi)}_{\varphi'} \to \psi$, where $\exists \boldsymbol{v}$ binds all variables in $free(\varphi) \setminus free(\psi)$. Then $free(\varphi') = free(\psi)$, and an interpolant for φ', ψ is also an interpolant for φ, ψ.

Since $\models \varphi \to \psi$, it follows from Theorem 4.2.11 that $\mathrm{LC}, Cv_1 \cdots v_k \models \varphi^C \to \psi^C$. Let $\vartheta \in MLR_k^C$ be the interpolant obtained by Lemma 4.4.2.

Let us write $\mathcal{L}$ for $\mathcal{L}_{G(\varphi)} \cup \mathcal{L}_{G(\psi)}$, and let $\vartheta^L = \vartheta[\mathsf{L}ive_{\mathcal{L}}(v_1, \ldots, v_k)/Cv_1 \cdots v_k]$. (That is, we replace the predicate $Cv_1 \cdots v_k$ everywhere in ϑ by the formula $\mathsf{L}ive_{\mathcal{L}}(v_1, \ldots, v_k)$.) Note that $free(\vartheta^L) \subseteq \{v_1, \ldots, v_l\}$.

By Proposition 4.4.3, ϑ^L has an equivalent $\vartheta' \in \mathrm{GF}_k^k$ such that $\mathcal{L}_{G(\vartheta')} = \mathcal{L}$ and $\mathcal{L}_{\bar{G}(\vartheta')} = \mathcal{L}_{\bar{G}(\vartheta^L)} = \mathcal{L}_\varphi \cap \mathcal{L}_\psi$. Hence, if we show that $\models \varphi \to \vartheta^L$ and $\models \vartheta^L \to \psi$, then ϑ' turns out to be an interpolant for φ, ψ which satisfies the conditions in Theorem 4.4.5.

We first show that $\models \varphi \to \vartheta^L$. To see this, consider an $\mathcal{L} \cup \mathcal{L}_\varphi$-model $\mathcal{M}$ and $\boldsymbol{m} \in M^l$ such that $\mathcal{M} \models \varphi[\boldsymbol{m}]$. Since φ is self-guarded, $\boldsymbol{m} \in I^{\mathcal{M}}(\mathsf{L}ive_{\mathcal{L}}(v_1, \ldots, v_l))$. We then expand the model $\mathcal{M}$ to the model $(\mathcal{M}, C)$ by setting $I^{(\mathcal{M},C)}(C) = I^{\mathcal{M}}(\mathsf{L}ive_{\mathcal{L}}(v_1, \ldots, v_k))$. Then

$$(\mathcal{M}, C) \models LC. \tag{4.5}$$

Extend the $\mathcal{L}$-live l-tuple $\boldsymbol{m}$ to an k-tuple $\boldsymbol{mn}$ by repeating some of its elements. Note that $\boldsymbol{mn} \in I^{\mathcal{M}}(\mathsf{L}ive_{\mathcal{L}}(v_1, \ldots, v_k))$, hence,

$$\boldsymbol{mn} \in I^{(\mathcal{M},C)}(C). \tag{4.6}$$

By induction on the complexity of χ it can be shown that for all $\chi \in \mathrm{GF}_k^k$ such that $\mathcal{L}_\chi \subseteq \mathcal{L} \cup \mathcal{L}_\varphi$ and $\mathcal{L}_{G(\chi)} \subseteq \mathcal{L}$, and all $\boldsymbol{t} \in M^k$, $\mathcal{M} \models \chi[\boldsymbol{t}] \Leftrightarrow (\mathcal{M}, C) \models \chi^C[\boldsymbol{t}]$. In particular,

$$(\mathcal{M}, C) \models \varphi^C[\bar{m}\boldsymbol{n}]. \tag{4.7}$$

As ϑ is the MLR_k^C interpolant for $LC, Cv_1, \ldots, v_k \models \varphi^C \to \psi^C$, it follows from (4.5), (4.6) and (4.7) that $(\mathcal{M}, C) \models \vartheta[\boldsymbol{mn}]$.

Recall that ϑ^L is obtained from ϑ by replacing the predicate $Cv_1 \cdots v_k$ by the formula $\mathsf{L}ive_{\mathcal{L}}(v_1, \ldots, v_k))$. As $I^{(\mathcal{M},C)}(C) = I^{(\mathcal{M},C)}(\mathsf{L}ive_{\mathcal{L}}(v_1, \ldots, v_k))$, we see that $(\mathcal{M}, C) \models \vartheta^L[\boldsymbol{mn}]$. Since $free(\theta^L)$ are among $\{v_1, \ldots, v_l\}$, then ϑ^L is already satisfied in $(\mathcal{M}, C)$ by the l-tuple $\bar{m}$. As C does not occur in ϑ^L, also $\mathcal{M} \models \vartheta^L[\boldsymbol{m}]$. This establishes that $\models \varphi \to \vartheta^L$.

To show that $\models \vartheta^L \to \psi$, one proceeds by proving that $\models \neg\psi \to \neg\vartheta^L$. This requires a similar argument as above, using the fact that $\neg\psi$ is self-guarded. This completes the proof of Theorem 4.4.5. ■

4.4.6. Remark Let us compare the interpolation theorem formulated in Theorem 4.4.5 with the interpolation property usually studied in modal logics. In the latter, the interpolant is built up from proposition letters that occur both in φ and ψ (as proposition letters). This is not completely similar to condition 3 in Theorem 4.4.5 according to which an interpolant ϑ for the validity $\models \varphi \to \psi$ is allowed to contain relation symbols from $(\mathcal{L}_\varphi \cap \mathcal{L}_\psi) \setminus (\mathcal{L}_{\bar{G}(\varphi)} \cap \mathcal{L}_{\bar{G}(\psi)})$ at non-guard positions. That is, the non-guards from which the interpolant is built, do not necessarily occur both in φ and ψ *as non-guards*. This allowance is necessary, as is witnessed by the next example.

Let $\varphi = \exists v(Pv, Rv)$ and $\psi = \exists v(Rv, Pv)$. Obviously, $\models \varphi \to \psi$, and both φ and ψ are interpolants satisfying the conditions in Theorem 4.4.5. However, an interpolant $\vartheta \in \mathrm{GF}_1$ such that $\mathcal{L}_{\bar{G}(\vartheta)} \subseteq \mathcal{L}_{\bar{G}(\varphi)} \cap \mathcal{L}_{\bar{G}(\psi)} = \emptyset$ can not be found.

In order to obtain a truly modal interpolation property, we should adopt another truly modal feature, namely the syntactic distinction between modalities and proposition letters. Such a distinction between state and action predicates in the guarded fragment has been advocated by van Benthem, [van Benthem, 1998]. We note that condition 3 in Theorem 4.4.5 can also be strengthened to $\mathcal{L}_{\bar{G}(\vartheta)} \subseteq \mathcal{L}_{\bar{G}(\varphi)} \cap \mathcal{L}_{\bar{G}(\psi)}$ if the interpolant ϑ is allowed to contain one extra variable. Hereto, simply replace every non-guard occurrence of a relation $R\boldsymbol{v} \in (\mathcal{L}_\varphi \cap \mathcal{L}_\psi) \setminus (\mathcal{L}_{\bar{G}(\varphi)} \cap \mathcal{L}_{\bar{G}(\psi)})$ in ϑ by $\exists v_{k+1}(R\boldsymbol{v}, \mathit{True})$. In the example above, this gives e.g, the formula $\vartheta = \exists v(Pv, \exists u(Rv, v = v))$ as an interpolant in GF_2 for φ, ψ. $\dashv$

A result similar to Theorem 4.4.5 holds for the packed fragment. Actually, one of the advantages of proving Theorem 4.4.5 via Lemma 4.4.2 is that it provides a uniform method which is also applicable to the packed fragment. We just need a packed formula which can play the role of $\mathsf{L}ive_\mathcal{L}$. Hereto we define, for any finite language $\mathcal{L}$ and $n \in \omega$,

$$\mathsf{P}ack_\mathcal{L}(v_1, \dots, v_n) \stackrel{\text{def}}{=} \bigwedge_{i,j \le n} \mathsf{L}ive_\mathcal{L}(v_i, v_j).$$

Then for all $\mathcal{L}$-models $\mathcal{M}$ and $m_1, \dots, m_n \in M$:

$$\mathcal{M} \models \mathsf{P}ack_\mathcal{L}(v_1, \dots, v_n)[m_1, \dots, m_n] \text{ iff } \{m_1, \dots, m_n\} \text{ is } \mathcal{L}\text{-packed in } \mathcal{M}.$$

Moreover, the packed equivalent of Proposition 4.4.3 holds.

4.4.7. Proposition *Let $\mathcal{L}$ be a language which contains at most k-ary relation symbols. Any first order formula φ in k variables which is uniformly guarded in $\mathsf{P}ack_\mathcal{L}(v_1, \dots, v_k)$ is equivalent to some $\varphi' \in \mathrm{PF}_k^k$ such that $\mathcal{L}_{G(\varphi')} = \mathcal{L}$ and $\mathcal{L}_{\bar{G}(\varphi')} = \mathcal{L}_{\bar{G}(\varphi)}$.*

Proof: See the appendix following this chapter. ■

In view of the remarks concerning Theorem 4.4.5, we restrict ourselves in the interpolation theorem for PF to validities $\varphi \to \psi$ such that $free(\psi) \subseteq free(\varphi)$ and $\varphi, \neg\psi$ have the following property.

4.4.8. Definition [**Self-packed formula**] A formula $\varphi(v_1, \ldots, v_n)$ is called *self-packed* if $\models \varphi \to \mathsf{P}ack_{\mathcal{L}_{G(\varphi)}}(v_1, \ldots, v_n)$. ⊣

4.4.9. Theorem (**PF_k^k has modal interpolation**) *Let $k \in \omega$. Let $\varphi, \psi \in \mathrm{PF}_k^k$ be such that $free(\psi) \subseteq free(\varphi)$ and $\varphi, \neg\psi$ are self-packed. If $\models \varphi \to \psi$, then there exists an interpolant $\vartheta \in \mathrm{PF}_k^k$ satisfying the following conditions:*

1. *$\models \varphi \to \vartheta$ and $\models \vartheta \to \psi$.*
2. *$\mathcal{L}_{G(\vartheta)} \subseteq \mathcal{L}_{G(\varphi)} \cup \mathcal{L}_{G(\psi)}$.*
3. *$\mathcal{L}_{\bar{G}(\vartheta)} \subseteq \mathcal{L}_{\varphi} \cap \mathcal{L}_{\psi}$.*
4. *ϑ is equivalent to a formula χ with $\mathcal{L}_{\bar{G}(\chi)} \subseteq \mathcal{L}_{\bar{G}(\vartheta)}$ and which is uniformly guarded in $\mathsf{P}ack_{\mathcal{L}_{G(\varphi)} \cup \mathcal{L}_{G(\psi)}}(v_1, \ldots, v_k)$.*

Proof: Similar to the proof of Theorem 4.4.5. This time, $\mathsf{P}ack_{\mathcal{L}}$ and Proposition 4.4.7 play the role of $\mathsf{L}ive_{\mathcal{L}}$ and Proposition 4.4.3. ■

As observed in Remark 4.4.6, condition 3 in Theorem 4.4.9 may be strengthened to $\mathcal{L}_{\bar{G}(\vartheta)} \subseteq \mathcal{L}_{\bar{G}(\varphi)} \cap \mathcal{L}_{\bar{G}(\psi)}$, if we allow for one extra variable in the interpolant.

The rest of this section is devoted to the proof of the main lemma, i.e., 4.4.2. To this end, we need the following preliminaries which will continue to play a role in future sections.

4.4.10. Definition [**Guarded $\mathcal{L}_1/\mathcal{L}_2$-Bisimulation**] Consider languages $\mathcal{L}_1$, $\mathcal{L}_2$. A *guarded $\mathcal{L}_1/\mathcal{L}_2$-bisimulation* between models $\mathcal{M}$ and $\mathcal{N}$ is defined as a non-empty set of finite partial $\mathcal{L}_2$-isomorphisms between $\mathcal{M}, \mathcal{N}$ with zig- and zag- condition (cf. Definition 4.2.9) stipulated for $\mathcal{L}_1$-live sets only. ⊣

In the above definition we did not specify any relationship between the languages $\mathcal{L}_1$ and $\mathcal{L}_2$. However, if we want to syntactically characterize the guarded formulas that are preserved under this type of bisimulation, we have to add the requirement that $\mathcal{L}_1 \subseteq \mathcal{L}_2$.

4.4.11. Lemma (Preservation lemma) *Let $\mathcal{L}_1 \subseteq \mathcal{L}_2$, and let $\varphi \in \mathrm{GF}$ be such that $\mathcal{L}_{G(\varphi)} \subseteq \mathcal{L}_1$ and $\mathcal{L}_{\bar{G}(\varphi)} \subseteq \mathcal{L}_2$. Then φ is preserved under guarded $\mathcal{L}_1/\mathcal{L}_2$-bisimulations.*

Proof: By a straightforward induction on the complexity of φ. ■

4.4.12. Notation Let $\mathcal{F}$ be a fragment of first order logic, and $\mathcal{L}$ a language. For models $\mathcal{M}, \mathcal{N}$, and $\boldsymbol{m} \in M^k, \boldsymbol{n} \in N^k$, write $\mathcal{M}, \boldsymbol{m} \equiv_{\mathcal{L}}^{\mathcal{F}} \mathcal{N}, \boldsymbol{n}$ to denote that $\mathcal{M} \models \theta[\boldsymbol{m}]$ iff $\mathcal{N} \models \theta[\boldsymbol{n}]$, for all $\theta \in \mathcal{F}$ in the language $\mathcal{L}$. Similarly, $\mathcal{M}, \boldsymbol{m} \equiv_{\mathcal{L}_1/\mathcal{L}_2}^{\mathcal{F}} \mathcal{N}, \boldsymbol{n}$ denotes equivalence with respect to formulas ϑ such that $\mathcal{L}_{G(\vartheta)} \subseteq \mathcal{L}_1$ and $\mathcal{L}_{\bar{G}(\vartheta)} \subseteq \mathcal{L}_2$. ⊣

Recall that MLR_k^C-formulas are uniformly guarded in the k–ary predicate C. Besides that, C only occurs at guarded position in such formulas. Therefore, by Lemma 4.4.11, any MLR_k^C-formula φ with non-guard relations in $\mathcal{L}$ is preserved under $\{C\}/\mathcal{L} \cup \{C\}$ bisimulations.

In the proof of the Characterization Theorem for GF in [Andréka et al., 1998] (cf. Theorem 4.2.2 therein) it is shown that the relation of guarded $\mathcal{L}$-equivalence between ω-saturated structures induces a guarded $\mathcal{L}$-bisimulation. Similarly, the following lemma can be shown.

4.4.13. Lemma *Let $\mathcal{M}$,$\mathcal{N}$ be ω-saturated models. Let F be the set of partial maps f from $\mathcal{M}$ to $\mathcal{N}$ such that for any $\boldsymbol{x} \in dom(f)$, $\mathcal{M}, \boldsymbol{x} \equiv_{C/\mathcal{L}}^{\mathrm{GF}_n} \mathcal{N}, f(\boldsymbol{x})$ and $\boldsymbol{x} \in I^{\mathcal{M}}(C)$ iff $f(\boldsymbol{x}) \in I^{\mathcal{N}}(C)$. If F is non-empty, then F is a guarded $\{C\}/\mathcal{L} \cup \{C\}$-bisimulation between $\mathcal{M}$ and $\mathcal{N}$.*

Proof: Standard. ■

Finally, we turn to the proof of Lemma 4.4.2.

Proof of Lemma 4.4.2: We will show 'amalgamation via bisimulation' in the same spirit as e.g., the proof of interpolation for the basic modal logic K in [Andréka et al., 1998, Theorem 2.5]. Its main construction is a deviation of an amalgamation method introduced in [Day, 1972] in the context of Heyting algebras (cf. Example 2.5.5) and applied e.g., in [Németi, 1985] to cylindric relativized set algebras.

Let k, φ, ψ be as in the lemma, and let $\boldsymbol{v} = v_1 \cdots v_k$. Set

$$\Theta = \{\vartheta \in MLR_k^C \ : \ \mathcal{L}_{\bar{G}(\vartheta)} \subseteq \mathcal{L}_{\bar{G}(\varphi)} \cap \mathcal{L}_{\bar{G}(\psi)} \text{ and } LC, C\boldsymbol{v} \models \varphi \to \vartheta\}.$$

Our aim is to show that

$$LC, C\boldsymbol{v}, \Theta \models \psi, \tag{4.8}$$

for then, by compactness, it follows that ψ is implied by $LC, C\boldsymbol{v}$ together with some finite conjunction ϑ of formulas in Θ. This ϑ then is an interpolant for φ, ψ which satisfies conditions 1 and 2 of 4.4.2. Finally, let $\vartheta' = \forall \boldsymbol{u}(Cv_1 \cdots v_k \to \vartheta)$, where $\forall \boldsymbol{u}$ quantifies over all variables in $\{v_1, \ldots, v_k\} \setminus (free(\varphi) \cup free(\psi))$. Then ϑ' is an interpolant satisfying all the conditions in the theorem.

To prove (4.8), consider an arbitrary $\mathcal{L}_\psi$-model $\mathcal{N}$, and $\boldsymbol{b} = b_1, \ldots, b_k \in N^k$ such that $\mathcal{N} \models \vartheta[\boldsymbol{b}]$, for every $\vartheta \in \Theta \cup LC \cup \{C\boldsymbol{v}\}$. Our task is to show that $\mathcal{N} \models \psi[\boldsymbol{b}]$.

We first note that there exists some $\mathcal{L}_\varphi$-model $\mathcal{M}$ and $\boldsymbol{a} = a_1, \ldots, a_k \in M^k$ such that $\mathcal{M} \models LC \wedge \varphi \wedge C\boldsymbol{v}[\boldsymbol{a}]$, and $\mathcal{M}, \boldsymbol{a} \equiv_{C/\mathcal{L}_{\bar{G}(\varphi)} \cap \mathcal{L}_{\bar{G}(\psi)}}^{MLR_k^C} \mathcal{N}, \boldsymbol{b}$. For, consider $\Phi = \{\vartheta \in MLR_k^C : \mathcal{L}_{\bar{G}(\vartheta)} \subseteq \mathcal{L}_{\bar{G}(\varphi)} \cap \mathcal{L}_{\bar{G}(\psi)}$ & $\mathcal{N} \models \vartheta[\boldsymbol{b}]\}$. In case a model $\mathcal{M}$

and $\boldsymbol{a} \in M^k$ as above do *not* exist, $\Phi \cup LC \cup \{C\boldsymbol{v}\} \models \neg\varphi$. By compactness it follows that $LC, C\boldsymbol{v} \models \varphi \to \neg \bigwedge \Phi_0$, for some finite conjunction of formulas in Φ. Therefore, $\neg \bigwedge \Phi_0 \in \Theta$, hence, $\mathcal{N} \models \neg \bigwedge \Phi_0[\boldsymbol{b}]$. Quod non.

By passing to ω-saturated elementary extensions of $\mathcal{M}$ and $\mathcal{N}$, we may without loss of generality assume that $\mathcal{M}$, $\mathcal{N}$ are ω-saturated, cf. [Andréka et al., 1998, Theorem 2.2.1].

Note that $\mathcal{M}, \boldsymbol{a} \equiv^{GF_k}_{C/\mathcal{L}_{\bar{G}(\varphi)} \cap \mathcal{L}_{\bar{G}(\psi)}} \mathcal{N}, \boldsymbol{b}$ and $\boldsymbol{a} \in I^{\mathcal{M}}(C)$ and $\boldsymbol{b} \in I^{\mathcal{N}}(C)$. By Lemma 4.4.13 there exists a guarded $C/(\mathcal{L}_{\bar{G}(\varphi)} \cap \mathcal{L}_{\bar{G}(\psi)}) \cup \{C\}$-bisimulation F between $\mathcal{M}$, $\mathcal{N}$ which links $\boldsymbol{a}$ and $\boldsymbol{b}$. As $(\mathcal{L}_{\bar{G}(\varphi)} \cap \mathcal{L}_{\bar{G}(\psi)}) \cup \{C\} = \mathcal{L}_\varphi \cap \mathcal{L}_\psi$, this F is a $C/\mathcal{L}_\varphi \cap \mathcal{L}_\psi$-bisimulation.

The aim of the rest of this proof is to amalgamate the models $\mathcal{M}$ and $\mathcal{N}$ in such a way that we can define guarded $C/\mathcal{L}_\varphi$ (resp. $C/\mathcal{L}_\psi$) -bisimulations between the amalgamated model and $\mathcal{M}$ (resp. $\mathcal{N}$) which, when composed, will map $\boldsymbol{a}$ to $\boldsymbol{b}$. Chasing the resulting diagram and using the fact that $LC, C\boldsymbol{v} \models \varphi \to \psi$ will yield the desired conclusion that $\mathcal{N} \models \psi[\boldsymbol{b}]$. This will be made precise in the sequel.

Define the model $\mathcal{MN}$ by setting

- $MN = \{\langle m, n\rangle \in M \times N \;:\; \mathcal{M}, m \sim_F \mathcal{N}, n\}$.
- For l-ary $R \in \mathcal{L}_\varphi$, set $\langle\langle m_1, n_1\rangle, \ldots, \langle m_l, n_l\rangle\rangle \in I^{\mathcal{MN}}(R)$ iff
  - $\mathcal{M}, m_1 \cdots m_l \sim_F \mathcal{N}, n_1 \cdots n_l$,
    (i.e., the m_i and n_i are not only pairwise bisimilar but jointly so), and
  - $\langle m_1, \ldots, m_l\rangle \in I^{\mathcal{M}}(R)$.
- The interpretation of relations in $\mathcal{L}_\psi$ is defined similarly.

Note that the universe MN is non-empty, as F is non-empty. Moreover, the interpretation of relations in the common language is well-defined thanks to the requirement on live subsets of $\mathcal{MN}$ to be jointly bisimilar. The upshot of amalgamating our models into a product is that we can take projection functions as building blocks for the desired bisimulations. This is the purport of the following lemma. Let π_i, $i = 1, 2$, denote the projection function to the i-th coordinate, and define

$$F_{\pi_i} \stackrel{\text{def}}{=} \{\pi_i : X \longrightarrow Y \;:\; X \subseteq^l_C \mathcal{MN}\}.$$

Amalgamation lemma *The set F_{π_1} is a guarded $C/\mathcal{L}_\varphi$-bisimulation between $\mathcal{MN}$ and $\mathcal{M}$. F_{π_2} is a guarded $C/\mathcal{L}_\psi$-bisimulation between $\mathcal{MN}$ and $\mathcal{N}$.*

Before proving the amalgamation lemma, let us first demonstrate its use and finish the proof of Lemma 4.4.2. Recall that $\mathcal{M}$ and $\boldsymbol{a} = a_1, \ldots, a_k \in M^k$ were chosen in such a way that $\mathcal{M} \models \varphi[\boldsymbol{a}]$. Note that $\{\langle a_1, b_1\rangle, \ldots, \langle a_k, b_k\rangle\} \in I^{\mathcal{MN}}(C)$. Hence, there exists some $f \in F_{\pi_1}$ with domain $\{\langle a_1, b_1\rangle, \ldots, \langle a_k, b_k\rangle\}$. By Lemma 4.4.11, φ is preserved under guarded $C/\mathcal{L}_\varphi$-bisimulations. Therefore, by the amalgamation lemma, $\mathcal{MN} \models \varphi[\langle a_1, b_1\rangle, \ldots, \langle a_k, b_k\rangle]$. Note that $\mathcal{MN} \models LC$. By assumption $LC, Cv_1, \ldots, v_k \models \varphi \to \psi$. This implies that $\mathcal{MN} \models \psi[\langle a_1, b_1\rangle, \ldots, \langle a_k, b_k\rangle]$. As $\{\langle a_1, b_1\rangle, \ldots, \langle a_k, b_k\rangle\}$ is also the domain of some projection in F_{π_2}, the second part of the amalgamation lemma allows us to conclude that $\mathcal{N} \models \psi[b_1, \ldots, b_k]$, as desired.

Now we turn to the proof of the amalgamation lemma. We will prove the first part of the lemma concerning F_{π_1}. The second statement about F_{π_2} can be shown similarly.

We already observed that F_{π_1} is non-empty. Let $\pi_1 \in F_{\pi_1}$. The domain of π_1 is a set of pairs $X = \{x_1, \ldots, x_l\}$, for some $l \leq k$. As X is a live set, it follows by definition of $\mathcal{MN}$ that $\mathcal{M}, \pi_1(x_1)\cdots\pi_1(x_l) \sim_F \mathcal{N}, \pi_2(x_1)\cdots\pi_2(x_l)$. By construction, this implies that for any m-ary $R \in \mathcal{L}_\varphi$, and any $\langle x_{i_1}, \ldots, x_{i_m}\rangle \in X^m$ it is the case that $\langle x_{i_1}, \ldots, x_{i_m}\rangle \in I^{\mathcal{MN}}(R)$ iff $\langle \pi_1(x_{i_1}), \ldots, \pi_1(x_{i_m})\rangle \in I^{\mathcal{M}}(R)$. In other words, π_1 preserves $\mathcal{L}_\varphi$-relations in both ways. Moreover, π_1 is 1-1, as F consists of 1-1 maps. We conclude that π_1 is a partial $\mathcal{L}_\varphi$-isomorphism.

The zig-condition is trivially fulfilled. For the zag-condition, consider some $\pi_1 \in F_{\pi_1}$, and $W \subseteq^l_C \mathcal{M}$. As we noted before, the domain of π_1 is a set of pairs $X = \{x_1, \ldots, x_l\}$, for some $l \leq k$, with $\mathcal{M}, \pi_1(x_1)\cdots\pi_1(x_l) \sim_F \mathcal{N}, \pi_2(x_1)\cdots\pi_2(x_l)$. By the zig-condition for F there exists some $g \in F$ with domain W such that $g(\pi_1(x)) = \pi_2(x)$, for $\pi_1(x) \in \pi_1[X] \cap W$. Let $W^* = \{\langle w, g(w)\rangle : w \in W\}$. Then $W^* \subseteq^l_C \mathcal{MN}$ and W^* is the desired pre-image for W. This finishes the proof of the amalgamation lemma, and hereby proves Theorem 4.4.5. ■

4.5 An excursion into definability

As we discussed in chapter 2, in general, an important reason to investigate the interpolation property is that it can be seen as an intermediate stage in proving the Beth definability theorem. In this section it will be shown that the limited form of interpolation expressed in Theorem 4.4.5 and Theorem 4.4.9 still serves this purpose for GF^k_k and PF^k_k. So, GF and PF present natural examples of logics without the interpolation property (as we saw in section 4.3), but with the Beth property.

The Beth property for predicate logics has been introduced in Definition 2.2.3. For the reader's convenience, we recall its formulation. Let $\mathcal{L}_0$ be a language and R and R' distinct relation symbols of the same arity that are not in $\mathcal{L}_0$. Let $\mathcal{L} = \mathcal{L}_0 \cup \{R\}$. Let Σ be a set of guarded sentences in the language $\mathcal{L}$, and let Σ' denote the result of renaming R to R' in Σ.

4.5.1. Theorem (Beth Theorem for $\mathbf{GF}^k_k$) *Let $\mathcal{L}_0$, $\mathcal{L}$, R, R', Σ and Σ' be as* *above. Let $k \in \omega$ be such that $\Sigma \cup \{R\boldsymbol{v}\} \subseteq \mathrm{GF}^k_k$. If Σ* implicitly defines *R, i.e., if*

$$\Sigma, \Sigma' \models \forall \boldsymbol{v}(R\boldsymbol{v} \leftrightarrow R'\boldsymbol{v}),$$

then there exists some $\varphi(\boldsymbol{v}) \in \mathrm{GF}^k_k$ in the language $\mathcal{L}_0$ such that

$$\Sigma \models \forall \boldsymbol{v}(R\boldsymbol{v} \leftrightarrow \varphi(\boldsymbol{v})).$$

This formula φ is called an explicit definition *for R relative to Σ. Moreover, φ can be chosen such that $\mathcal{L}_{G(\varphi)} \subseteq \mathcal{L}_{G(\Sigma)} \setminus R$.*

4.5.2. Remark By the above theorem every guarded finite variable fragment has the Beth property. Note the contrast with the situation for the full finite variable fragments of first order logic. For, as we have seen in Example 2.2.5, the Beth property fails in L_n, for every $n \geq 2$. ⊣

The proof of Theorem 4.5.1 uses the following lemma, due to Martin Otto.

4.5.3. Lemma *Let Σ be a set of guarded sentences that implicitly define R. Let $\mathcal{M}$ be a model of Σ, and let $Y \subseteq^l_{\mathcal{L}_{G(\Sigma)}} \mathcal{M}$. Then $Y \subseteq^l_{(\mathcal{L}_{G(\Sigma)} \setminus R)} \mathcal{M}$.*

Proof of Lemma 4.5.3: Let $\mathcal{M}$ be a model of Σ, and let $Y_0 \subseteq^l_{\mathcal{L}_{G(\Sigma)}} \mathcal{M}$. Suppose $Y_0 \not\subseteq^l_{(\mathcal{L}_{G(\Sigma)} \setminus R)} \mathcal{M}$. We will derive a contradiction from this.

Let $\mathcal{M}\times\underline{2}$ denote the model with domain $M \times \{0,1\}$ and the interpretation of the relations defined as follows. For every l-ary $P \in \mathcal{L}$ and every $\boldsymbol{s} = \langle s_1, \ldots, s_l \rangle \in \{0,1\}^l$, $\boldsymbol{s} \in I^{\mathcal{M}\times\underline{2}}(P)$ iff

1. $\langle \pi_1(s_1), \ldots, \pi_1(s_l) \rangle \in I^{\mathcal{M}}(P)$, and
2. for all $m, m' \in M$, if $\langle m, 0 \rangle = s_i$ and $\langle m', 1 \rangle = s_j$, for some $i, j \leq l$, then $m \neq m'$.

As the reader can easily verify, this implies that $F_1 \stackrel{\text{def}}{=} \{\pi_1 : X \longrightarrow Y \ : \ X \subseteq^l_{\mathcal{L}_{G(\Sigma)}} \mathcal{M} \times \underline{2}\}$ is a guarded $\mathcal{L}_{G(\Sigma)}/\mathcal{L}_\Sigma$-bisimulation between $\mathcal{M}\times\underline{2}$ and $\mathcal{M}$. Condition 2 guarantees that every projection in F_1 is 1-1. Since $\mathcal{M} \models \Sigma$, we conclude that $\mathcal{M}\times\underline{2} \models \Sigma$.

Our aim is to modify the interpretation of R on $\mathcal{M}\times\underline{2}$ in such a way that the resulting structure is again a model for Σ, contradicting the fact that Σ implicitly defines R. For this, we consider $X_0 \stackrel{\text{def}}{=} Y_0 \times \{0\}$. Let $(\mathcal{M}\times\underline{2})'$ be the model which differs from $\mathcal{M}\times\underline{2}$ only in that $X_0 \not\subseteq^l_R (\mathcal{M}\times\underline{2})'$. Note that $X_0 \subseteq^l_R \mathcal{M}\times\underline{2}$. Hence, the two models do really differ. We claim that $F_1' \stackrel{\text{def}}{=} \{\pi_1 : X \longrightarrow Y \ : \ X \subseteq^l_{\mathcal{L}_{G(\Sigma)}} (\mathcal{M}\times\underline{2})'\}$ is a guarded $\mathcal{L}_{G(\Sigma)}/\mathcal{L}_\Sigma$-bisimulation between $(\mathcal{M}\times\underline{2})'$ and $\mathcal{M}$.

F_1' is certainly not empty. Consider some $\pi_1 : X \longrightarrow Y$ in F_1'. If $X_0 \not\subseteq X$, then $\mathcal{L}_\Sigma$-relations are obviously preserved by π_1 in both ways. Let us see that it is the case that $X_0 \not\subseteq X$. For, we changed the interpretation of R such that X_0 is not R-live in $(\mathcal{M}\times\underline{2})'$. As Y_0 is not $(\mathcal{L}_{G(\Sigma)} \setminus R)$-live, it follows that X_0 is not $(\mathcal{L}_{G(\Sigma)} \setminus R)$-live in $(\mathcal{M}\times\underline{2})'$ either. Hence, X_0 is not $\mathcal{L}_{G(\Sigma)}$-live in $(\mathcal{M}\times\underline{2})'$. We conclude that no superset of X_0 is the domain of some $\pi_1 \in F_1'$.

The zig-condition needs no comment. For the zag-condition, consider some projection $\pi_1 : X \longrightarrow Y$ in F_1', and $W \subseteq^l_{\mathcal{L}_{G(\Sigma)}} \mathcal{M}$. If $W \subseteq Y$, the condition is trivially fulfilled. If not, consider $Z \stackrel{\text{def}}{=} \pi_1^{-1}[Y \cap W] \cup \{\langle m, 1\rangle : m \in W \setminus Y\}$. Note that $Z \subseteq^l_{\mathcal{L}_{G(\Sigma)}} \mathcal{M}\times\underline{2}$. As $W \setminus Y \neq \emptyset$ it follows that $\pi_2[Z] \neq \{0\}$. In particular, $Z \neq X_0$. We conclude that Z is also $\mathcal{L}_{G(\Sigma)}$-packed in $(\mathcal{M}\times\underline{2})'$. Therefore, Z fulfills the zag-condition for π_1, W.

This shows that $\mathcal{M}$ and $(\mathcal{M}\times\underline{2})'$ are $\mathcal{L}_{G(\Sigma)}/\mathcal{L}_\Sigma$-bisimilar. As $\mathcal{M} \models \Sigma$, also $(\mathcal{M}\times\underline{2})' \models \Sigma$. Summarizing, we see that $\mathcal{M}\times\underline{2} \models \Sigma$, $(\mathcal{M}\times\underline{2})' \models \Sigma$, $I^{\mathcal{M}\times\underline{2}}(P) = I^{(\mathcal{M}\times\underline{2})'}(P)$, for every $P \in \mathcal{L}_0$ but $I^{\mathcal{M}\times\underline{2}}(R) \neq I^{(\mathcal{M}\times\underline{2})'}(R)$. This contradicts the fact that Σ implicitly defines R. We conclude that Y_0 is indeed $(\mathcal{L}_{G(\Sigma)} \setminus R)$-live, as was to be shown. ■

In Theorem 2.3.10 we copied out the usual derivation of the Beth property from the interpolation property. The following proof is a variation on this theme, using the previous lemma and the formula $\mathsf{Live}_{\mathcal{L}}$ introduced at page 132.

Proof of Theorem 4.5.1: Let all data be as in the theorem, and assume that Σ implicitly defines R. By compactness we may assume Σ to be a single sentence, and we obtain that

$$\models \underbrace{(\Sigma \wedge R\boldsymbol{v})}_{\varphi} \rightarrow \underbrace{((\Sigma' \wedge \mathsf{Live}_{\mathcal{L}_{G(\Sigma')}}(\boldsymbol{v})) \rightarrow R'\boldsymbol{v})}_{\psi}. \tag{4.9}$$

We may assume $R \in \mathcal{L}_{G(\Sigma)}$ (else, add $\forall \boldsymbol{v}(R\boldsymbol{v} \rightarrow R\boldsymbol{v})$ to Σ). Hence, $\varphi, \neg\psi$ are self-guarded. Applying Theorem 4.4.5 to (4.9) we obtain an interpolant $\vartheta \in \mathrm{GF}^k_k$ uniformly guarded in $\mathsf{Live}_{\mathcal{L}_{G(\Sigma)} \cup \mathcal{L}_{G(\Sigma')}}$ with $\mathcal{L}_{\bar{G}(\vartheta)} \subseteq \mathcal{L}_0$ such that

$$\models (\Sigma \wedge R\boldsymbol{v}) \rightarrow \vartheta \text{ and } \models \vartheta \rightarrow ((\Sigma' \wedge \mathsf{Live}_{\mathcal{L}_{G(\Sigma')}}(\boldsymbol{v})) \rightarrow R'\boldsymbol{v}). \tag{4.10}$$

Substituting R for R' in (4.10) gives

$$\Sigma \models R\boldsymbol{v} \leftrightarrow (\vartheta[R/R'] \wedge \mathsf{Live}_{\mathcal{L}_{G(\Sigma)}}(\boldsymbol{v})). \tag{4.11}$$

Note that $\vartheta[R/R']$ is uniformly guarded in $\mathsf{Live}_{\mathcal{L}_{G(\Sigma)}}$. Let ϑ_0 be obtained from $\vartheta[R/R'] \wedge \mathsf{Live}_{\mathcal{L}_{G(\Sigma)}}(\boldsymbol{v})$ by replacing $\mathsf{Live}_{\mathcal{L}_{G(\Sigma)}}$ by $\mathsf{Live}_{(\mathcal{L}_{G(\Sigma)} \setminus R)}$. By Lemma 4.5.3, $\Sigma \models \mathsf{Live}_{\mathcal{L}_{G(\Sigma)}} \leftrightarrow \mathsf{Live}_{(\mathcal{L}_{G(\Sigma)} \setminus R)}$. Hence, by (4.11), $\Sigma \models \forall \boldsymbol{v}(R\boldsymbol{v} \leftrightarrow \vartheta_0)$. By Proposition 4.4.3, ϑ_0 is equivalent to some $\vartheta_0' \in \mathrm{GF}^k_k$ uniformly guarded in $\mathcal{L}_{G(\Sigma)} \setminus R$

with the same language for non-guard occurrences, i.e., $\mathcal{L}_{\bar{G}(\vartheta_0')} = \mathcal{L}_{\bar{G}(\vartheta_0)} \subseteq \mathcal{L}_0$. This ϑ_0' provides the desired explicit definition of R relative to Σ. ∎

Similar to the proof of Theorem 4.5.1 we obtain the Beth property for the packed finite variable fragments.

4.5.4. Theorem (Beth Theorem for $\mathbf{PF}_k^k$) *Let $\mathcal{L}_0$, $\mathcal{L}$, R, R', Σ, Σ' and k be as in Theorem 4.5.1 with this distinction that Σ and Σ' are sets of* packed *sentences. If Σ* implicitly defines *R, then there exists some $\varphi \in \mathrm{PF}_k^k$ in the language $\mathcal{L}_0$ which is an* explicit definition *for R, relative to Σ. Moreover, φ can be chosen such that $\mathcal{L}_{G(\varphi)} \subseteq \mathcal{L}_{G(\Sigma)} \setminus R$.*

This theorem is proven via the following version of Lemma 4.5.3 for the packed fragment.

4.5.5. Lemma *Let Σ be a set of packed sentences that implicitly define R. Let $\mathcal{M}$ be a model of Σ, and let $Y \subseteq^p_{\mathcal{L}_{G(\Sigma)}} \mathcal{M}$. Then $Y \subseteq^p_{(\mathcal{L}_{G(\Sigma)}\setminus R)} \mathcal{M}$.*

Proof of Lemma 4.5.5: We follow the proof of Lemma 4.5.3, making adaptations where necessary.

Let $\mathcal{M}$ be a model of Σ, and let $Y_0 \subseteq^p_{\mathcal{L}_{G(\Sigma)}} \mathcal{M}$. Suppose $Y_0 \not\subseteq^p_{\mathcal{L}_{G(\Sigma)}\setminus R} \mathcal{M}$. This implies the existence of $y, y' \in Y_0$ such that $\{y, y'\} \subseteq^l_R \mathcal{M}$ but $\{y, y'\} \not\subseteq^l_{\mathcal{L}_{G(\Sigma)}\setminus R} \mathcal{M}$. We will derive a contradiction.

Let $\mathcal{M}\times\underline{2}$ be the model defined in the proof of Lemma 4.5.3. It is not hard to see that $\{\pi_1 : X \longrightarrow Y \ : \ X \subseteq^p_{\mathcal{L}_{G(\Sigma)}} \mathcal{M}\times\underline{2}\}$ is a packed $\mathcal{L}_{G(\Sigma)}/\mathcal{L}_\Sigma$-bisimulation between $\mathcal{M}\times\underline{2}$ and $\mathcal{M}$. As $\mathcal{M} \models \Sigma$, also $\mathcal{M}\times\underline{2} \models \Sigma$.

Let $X_0 \stackrel{\text{def}}{=} \{\langle y, 0\rangle, \langle y', 0\rangle\}$. Then $X_0 \subseteq^l_R \mathcal{M}\times\underline{2}$. Let $(\mathcal{M}\times\underline{2})'$ be the model which differs from $\mathcal{M}\times\underline{2}$ only in that $X_0 \not\subseteq^l_R (\mathcal{M}\times\underline{2})'$. Similar to the proof of Lemma 4.5.3 it can be verified that $F_1' \stackrel{\text{def}}{=} \{\pi_1 : X \longrightarrow Y \ : \ X \subseteq^p_{\mathcal{L}_{G(\Sigma)}} (\mathcal{M}\times\underline{2})'\}$ is a packed $\mathcal{L}_{G(\Sigma)}/\mathcal{L}_\Sigma$-bisimulation between $(\mathcal{M}\times\underline{2})'$ and $\mathcal{M}$. We only check the zag-condition. Let $\pi_1 : X \longrightarrow Y$ in F_1', and $W \subseteq^p_{\mathcal{L}_{G(\Sigma)}} \mathcal{M}$. Define $Z \stackrel{\text{def}}{=} \pi_1^{-1}[Y \cap W] \cup \{\langle m, 1\rangle : m \in W \setminus Y\}$. Note that $Z \subseteq^p_{\mathcal{L}_{G(\Sigma)}} \mathcal{M}\times\underline{2}$. We claim that $Z \subseteq^p_{\mathcal{L}_{G(\Sigma)}} (\mathcal{M}\times\underline{2})'$. Suppose $X_0 \subseteq X$. As X is $\mathcal{L}_{G(\Sigma)}$-packed in $(\mathcal{M}\times\underline{2})'$, also X_0 is $\mathcal{L}_{G(\Sigma)}$-packed in $(\mathcal{M}\times\underline{2})'$. But X_0 is not R-packed in $(\mathcal{M}\times\underline{2})'$ and, as $\{y, y'\}$ is not $\mathcal{L}_{G(\Sigma)}\setminus R$-live in $\mathcal{M}$, X_0 is not $\mathcal{L}_{G(\Sigma)}\setminus R$-packed in $(\mathcal{M}\times\underline{2})'$ either. Therefore, $X_0 \not\subseteq X$. By definition of Z, then $X_0 \not\subseteq Z$. This implies that $Z \subseteq^p_{\mathcal{L}_{G(\Sigma)}} (\mathcal{M}\times\underline{2})'$, and Z fulfills the zag-condition for π_1, W. This shows that $\mathcal{M}$ and $(\mathcal{M}\times\underline{2})'$ are packed $\mathcal{L}_{G(\Sigma)}/\mathcal{L}_\Sigma$-bisimilar. Reasoning as in the proof of Theorem 4.5.1, this leads to the desired contradiction. ∎

Proof of Theorem 4.5.4: Similar to the proof of Theorem 4.5.1, using Theorem 4.4.9, Lemma 4.5.5 and Proposition 4.4.7. ∎

4.6 Guarded finite variable fragments with interpolation

Let us return to the examples in section 4.3 which indicated the failure of interpolation in GF^k, GF_n and PF_n if $n, k \geq 3$ and in PF^k if $k \geq 2$. Recall that k denotes the restriction on the arity of the relation symbols and n denotes the restricted number of variables. In this section we establish the reverse of the earlier implication. That is, it will be shown that GF^k_n has the interpolation property if $k \leq 2$ or $n \leq 2$, and PF^k_n has the interpolation property if $k \leq 1$ or $n \leq 2$. As we will see, all these results follow from the fact that GF_2 has the interpolation property. In particular, the result for PF_2 follows since, according to Proposition 4.6.8, the expressive power of GF_2 and PF_2 coincide. The difference in the behavior of GF^2 and PF^2 lies in the fact that for any binary relation symbol R the size of R-live sets is at most 2 whereas R-packed sets may have arbitrary size.

The following theorem is the main result of the present section. A series of corollaries is given at the end of this section.

4.6.1. Theorem GF_2 *has the interpolation property.*

4.6.2. Remark Theorem 4.6.1 reveals an interesting asymmetry between GF_2 and the full 2-variable fragment of first order logic. For, not only does the latter fail to have interpolation, it even fails in a very strong sense: as we have seen in Example 2.3.7, for any $n \in \omega$ there exist first order formulas φ_n, ψ_n in just two variables and unary relation symbols such that $\varphi_n \rightarrow \psi_n$ is valid but a first order interpolant for φ_n, ψ_n in n variables does not exist. ⊣

In order to prove Theorem 4.6.1 we need some preliminary results, similar to the ones used in the proof of Lemma 4.4.2. Borrowing the terminology from [Andréka et al., 1998], by an $\mathcal{L}$-*guarded* set we understand a set whose elements are R-related, in any order or multiplicity, by some $R \in \mathcal{L}$. So, for any $n \in \omega$, a guarded set of size n is not only live, but moreover this fact can be expressed using n variables. We define an n-*guarded* $\mathcal{L}$-*bisimulation* as a guarded $\mathcal{L}$-bisimulation (cf. Definition 4.2.9) where the zigzag conditions are restricted to $\mathcal{L}$-*guarded* sets *of size at most* n. The relation $\equiv^{\mathrm{GF}_n}_{\mathcal{L}}$ has been introduced in Notation 4.4.12. Below, n denotes an arbitrary natural number.

4.6.3. Lemma (Preservation lemma) *Every* $\varphi \in \mathrm{GF}_n$ *is preserved under* n-*guarded* $\mathcal{L}_\varphi$-*bisimulations.*

Proof: By induction on the complexity of φ. ■

4.6.4. Lemma *Between* ω-*saturated models, the relation* $\equiv^{\mathrm{GF}_n}_{\mathcal{L}}$ *between* n-*tuples induces an* n-*guarded* $\mathcal{L}$-*bisimulation.*

Proof: Similar to the proof of Theorem 4.2.2 in [Andréka et al., 1998]. ∎

This finishes the preparations for the proof of Theorem 4.6.1.

Proof of Theorem 4.6.1: Let $k \in \omega$. Let $\varphi, \psi \in \mathrm{GF}_2^k$ be such that $\models \varphi \rightarrow \psi$ is valid. Write $\mathcal{L}_{\varphi\psi}$ for $\mathcal{L}_\varphi \cap \mathcal{L}_\psi$, and set $\Theta \stackrel{\text{def}}{=} \{\vartheta \in \mathrm{GF}_2^k : \mathcal{L}_\vartheta \subseteq \mathcal{L}_{\varphi\psi} \ \& \ \models \varphi \rightarrow \vartheta\}$. As in the proof of Theorem 4.4.5, our aim is to show that $\Theta \models \psi$. Consider an arbitrary $\mathcal{L}_\psi$-model $\mathcal{N}$, and $b_1, b_2 \in N$ such that $\mathcal{N} \models \vartheta[b_1, b_2]$, for every $\vartheta \in \Theta$. Our task is to show that $\mathcal{N} \models \psi[b_1, b_2]$.

Similar to the proof of Theorem 4.6.1, we first note the existence of some $\mathcal{L}_\varphi$-model $\mathcal{M}$ and $a_1, a_2 \in M$ such that $\mathcal{M} \models \varphi[a_1, a_2]$ and $\mathcal{M}, a_1a_2 \equiv^{\mathrm{GF}_2}_{\mathcal{L}_{\varphi\psi}} \mathcal{N}, b_1b_2$. As before, without loss of generality we may assume that $\mathcal{M}, \mathcal{N}$ are ω-saturated.

By Lemma 4.6.4 there exists a 2-guarded $\mathcal{L}_{\varphi\psi}$-bisimulation F between $\mathcal{M}$ and $\mathcal{N}$ such that $\mathcal{M}, a_1a_2 \sim_F \mathcal{N}, b_1b_2$. Define the $\mathcal{L}_\varphi \cup \mathcal{L}_\psi$-model $\mathcal{MN}$ as follows:

- $MN = \{\langle m, n\rangle \in M \times N \ : \ \mathcal{M}, m \sim_F \mathcal{N}, n\}$.
- For binary $R \in \mathcal{L}_{\varphi\psi}$, $\langle\langle m_1, n_1\rangle, \langle m_2, n_2\rangle\rangle \in I^{\mathcal{MN}}(R)$ iff
  - $\mathcal{M}, m_1m_2 \sim_F \mathcal{N}, n_1n_2$, and
  - $\langle m_1, m_2\rangle \in I^{\mathcal{M}}(R)$.
- For binary $R \in \mathcal{L}_\varphi \backslash \mathcal{L}_\psi$, $\langle\langle m_1, n_1\rangle, \langle m_2, n_2\rangle\rangle \in I^{\mathcal{MN}}(R)$ iff
  - Either $\mathcal{M}, m_1m_2 \sim_F \mathcal{N}, n_1n_2$, or $\{m_1, m_2\}$ is not $\mathcal{L}_{\varphi\psi}$-live, and
  - $\langle m_1, m_2\rangle \in I^{\mathcal{M}}(R)$.
- Similarly, we define the interpretation of binary relations in $\mathcal{L}_\psi \backslash \mathcal{L}_\varphi$. Also the interpretation of unary relations is defined in the same way. For relations of arbitrary arity we stipulate as an important side condition that we will not allow for live sets in $\mathcal{MN}$ with more than 2 elements. For example, for a ternary relation R, we set $\langle\langle m_1, n_1\rangle, \ldots, \langle m_3, n_3\rangle\rangle \in I^{\mathcal{MN}}(R)$ iff $|\{\langle m_1, n_1\rangle, \ldots, \langle m_3, n_3\rangle\}| \leq 2$ and the above conditions, modified in the obvious way, are fulfilled.

Note that the universe MN is non-empty, as F is non-empty. Also, the interpretation of relations in the common language is well-defined since any $f \in F$ preserves relations in the common language.

Let π_i denote the projection function to the i-th coordinate, and define $F_{\pi_1} = \{\pi_1 : X \longrightarrow Y \ : \ X \subseteq^l_{\mathcal{L}_\varphi} \mathcal{MN} \text{ or } X = \{\langle a_1, b_1\rangle, \langle a_2, b_2\rangle\}\}$. F_{π_2} is defined similarly. As before, the key-lemma is the following.

Amalgamation lemma *F_{π_1} is a 2-guarded $\mathcal{L}_\varphi$-bisimulation from $\mathcal{MN}$ to $\mathcal{M}$. F_{π_2} is a 2-guarded $\mathcal{L}_\psi$-bisimulation from $\mathcal{MN}$ to $\mathcal{N}$.*

Reasoning as in the proof of Lemma 4.4.2, the theorem follows from the amalgamation lemma via Lemma 4.6.3. We will prove the first part of the amalgamation lemma. The second statement follows analogously.

F_{π_1} is obviously non-empty. By definition of the interpretation of relations in $\mathcal{MN}$, every $f_{\pi_1} \in F_{\pi_1}$ preserves $\mathcal{L}_\varphi$-relations in both ways. To see that every $f_{\pi_1} \in F_{\pi_1}$ is 1-1, we observe that, by construction, the domain of f_{π_1} is a set of pairs $\{\langle m_1, n_1\rangle, \langle m_2, n_2\rangle\}$ such that either $\mathcal{M}, m_1m_2 \sim_F \mathcal{N}, n_1n_2$, or $\{m_1, m_2\}$ is not $\mathcal{L}_{\varphi\psi}$-live. In the first case, f_{π_1} is 1-1 since F consists of 1-1 maps. In the latter case, f_{π_1} is 1-1 since the size of the domain of f is at most 2 whereas its range (i.e., the set $\{m_1, m_2\}$) is not live, hence in particular not a singleton. We conclude that f_{π_1} is a partial $\mathcal{L}_\varphi$-isomorphism. The zig-condition is trivially fulfilled. To show the zag-condition, consider some $f_{\pi_1} : X \longrightarrow Y$ in F_{π_1} and some R-live subset $W = \{m_1, m_2\} \subseteq^l_R \mathcal{M}$, for some $R \in \mathcal{L}_\varphi$. We distinguish three cases.
(a). $|W \cap Y| = 2$. Then $W = Y$, and there is nothing to be shown.
(b). $|W \cap Y| = 1$. Say, $W \cap Y = \{m_1\}$. We distinguish as to whether or not W is $\mathcal{L}_{\varphi\psi}$-live. If not, consider some $\langle m_2, n\rangle \in MN$. Note that such pair exists, as singletons are live. Then by construction $\{f^{-1}_{\pi_1}(m_1), \langle m_2, n\rangle\}$ is an R-live subset of $\mathcal{MN}$, and the projection from this set fulfills the zag-condition for f_{π_1}, W.

If W is $\mathcal{L}_{\varphi\psi}$-live, then $\mathcal{M}, m_1 \sim_F \mathcal{N}, \pi_2(f^{-1}_{\pi_1}(m_1))$. By the zig-condition for F there exists some $n_2 \in N$ such that $\mathcal{M}, m_1m_2 \sim_F \mathcal{N}, \pi_2(f^{-1}_{\pi_1}(m_1))n_2$. By construction, $\{f^{-1}_{\pi_1}(m_1), \langle m_2, n_2\rangle\}$ is R-live in $\mathcal{MN}$. Therefore, the projection from this set is in F_{π_1} and fulfills the zag-condition for f_{π_1}, W.
(c). $W \cap Y = \emptyset$. Reason as in case (b).
This proves the amalgamation lemma, and hereby finishes the proof of Theorem 4.6.1. ■

4.6.5. Corollary *For any $n \in \omega$, GF^2_n has the interpolation property for self-guarded formulas.*

This corollary follows directly from Theorem 4.6.1 and the following proposition in which we observe that the maximum number of variables that is needed in a self-guarded formula is determined by the arity of the language.

4.6.6. Proposition *Let $k, n \in \omega$, and let φ be a self-guarded formula in GF^k_n. Then φ is equivalent to some $\varphi' \in \mathrm{GF}^k_k$ in the same language.*

Proof: Note that $|free(\varphi)| \leq k$, as all free variables in φ are guarded by some relation in $\mathcal{L}_{G(\varphi)}$ which is at most k-ary. Without loss of generality, assume $free(\varphi) \subseteq \{v_1, \ldots, v_k\}$. Suppose $v_{k+1} \in var(\varphi)$. Then φ has a subformula ϑ of the form $Qv_{k+1}(R\boldsymbol{v}, \psi(\boldsymbol{v}))$, where $Q \in \{\forall, \exists\}$. If $v_{k+1} \notin \boldsymbol{v}$, the quantification is vacuous and can be removed. If $v_{k+1} \in \boldsymbol{v}$, then v_i does not appear in $\boldsymbol{v}$, for some $i \in 1, \ldots, k$, as R is at most k-ary. In this case replace everywhere in ϑ the variable v_{k+1} by v_i. In both cases we obtain an equivalent of ϑ in at most the variables $v_1, \dots, v_k$. Repeating this procedure eventually yields the desired equivalent of φ. ■

For completeness' sake we mention that the restriction in Corollary 4.6.5 to self-guarded formulas can be removed. That is, we claim that GF^2_n has the interpolation property, for every $n \in \omega$. The interested reader is invited to supply a proof along the lines of the proof of Theorem 4.6.1.

The next corollary completes the announced chart on interpolation in guarded and packed finite variable fragments.

4.6.7. Corollary *The following fragments have the interpolation property:*

1. PF_2.
2. GF^1_n *and* PF^1_n*, for any* $n \in \omega$.
3. GF_1 *and* PF_1.

Corollary 4.6.7 will be proven via the following proposition.

4.6.8. Proposition *Let* $n \in \omega$*,* $i \in \{1, 2\}$*. Then*

1. $(\forall \varphi \in \mathrm{PF}_i)(\exists \varphi' \in \mathrm{GF}_i) \models \varphi \leftrightarrow \varphi'$*, and* $\mathcal{L}_\varphi = \mathcal{L}_{\varphi'}$.
2. $(\forall \varphi \in \mathrm{PF}^1_n)(\exists \varphi' \in \mathrm{GF}^1_n) \models \varphi \leftrightarrow \varphi'$*, and* $\mathcal{L}_\varphi = \mathcal{L}_{\varphi'}$.

Proof of Proposition 4.6.8: Both claims are proven by induction on the complexity of φ. **(1)**. The only non-trivial case is for φ of the form $\exists \boldsymbol{v}(\phi, \psi)$. We observe that all (i.e., at most 2) variables in $free(\psi)$ occur free in some conjunct ϕ_0 of ϕ. If ϕ_0 is an atom, then set $\varphi' = \exists \boldsymbol{v}(\phi_0, \phi \wedge \psi)$. Else, ϕ_0 is of the form $\exists \boldsymbol{u}\chi$, with χ an atom and we let $\varphi' = \exists \boldsymbol{vu}(\chi, \phi \wedge \psi)$. As $\psi \in \mathrm{GF}_i$ by induction hypothesis, in both cases $\varphi' \in \mathrm{GF}_i$. **(2)**. Again we only consider the case for φ of the form $\exists \boldsymbol{v}(\phi, \psi)$. By induction hypothesis, $\psi \in \mathrm{GF}^1_n$. Note that every two variables $u, v \in free(\psi)$ are guarded in ϕ by $u = v$. If $free(\psi) = \emptyset$, φ is equivalent to $\phi \wedge \psi \in \mathrm{GF}^1_n$. Else, fix some variable $v \in free(\psi)$ and replace in $\phi \wedge \psi$ all occurrences (free or bound) of variables from $free(\psi)$ by v. Let $(\phi \wedge \psi)'$ denote the result. Then $\exists v(v = v, (\phi \wedge \psi)')$ is an equivalent of φ in GF^1_n. ■

Proof of Corollary 4.6.7: (1). By Theorem 4.6.1 and Proposition 4.6.8.1. **(2)**. The interpolation property for GF^1_n for self-guarded formulas follows directly from Corollary 4.6.5. In general, it follows from our claim that the restriction in Corollary 4.6.5 to self-guarded formulas can be removed. Together with Proposition 4.6.8.2, this implies the result for PF^1_n. **(3)**. Recall that in GF_1 *all* quantifications are admissible as they may be regarded as being guarded by an identity. In other words, every first order formula in one variable is equivalent to a formula in GF_1 in the same language. Therefore, interpolation for GF_1 is equivalent to interpolation for the one variable fragment of first order logic. It is well-known that this latter fragment has indeed interpolation, cf. [Pigozzi, 1972]. Interpolation for PF_1 then follows via Proposition 4.6.8.1. ■

4.7 A Lyndon theorem for GF

In section 2.3.2 we discussed Lyndon's interpolation theorem for first order logic. Recall that this theorem is a refinement of Craig's theorem in which the positive and negative occurrences of relation symbols are taken into account. In this section we will refine Theorem 4.4.5 in a similar way which results in Theorem 4.7.6. The proof of this theorem is an adaptation of the proof of Theorem 4.4.5 which uses a construction similar to the one in [Otto, 2000, Theorem 1]. Independently, the manuscript [Otto, 1999] also presents a Lyndon interpolation property for the guarded fragment.

For technical reasons, let, for the moment, formulas be built up from atomic formulas using conjunction, disjunction, negation and existential and universal quantification. A given occurrence of a relation symbol is said to be *positive* iff it occurs within the scope of an even number of negation signs. It is called *negative* otherwise. In this section we distinguish predicate occurrences in two respects: whether a predicate appears positively or negatively, and whether a predicate occurs as a guard or not. For any guarded formula φ, we write $\mathcal{G}^+(\varphi)$ for the set of predicates which appear positively in φ at a guard position, $\bar{\mathcal{G}}^-(\varphi)$ for those predicates that appear negatively in φ at a non-guard position. $\mathcal{G}^-(\varphi)$ and $\bar{\mathcal{G}}^+(\varphi)$ are defined similarly. Moreover, we write $\bar{\mathcal{G}}(\varphi) = \bar{\mathcal{G}}^-(\varphi) \cup \bar{\mathcal{G}}^+(\varphi)$, and $^+(\varphi) = \bar{\mathcal{G}}^+(\varphi) \cup \mathcal{G}^+(\varphi)$. Similarly for $\mathcal{G}(\varphi)$, and $^-(\varphi)$.

In the present setting, the *type* of a formula is a four-tuple specifying for each formula φ the sets $\bar{\mathcal{G}}^+(\varphi), \bar{\mathcal{G}}^-(\varphi), \mathcal{G}^+(\varphi)$ and $\mathcal{G}^-(\varphi)$.

4.7.1. Definition [**Type**] For given sets of predicates $\bar{\mathcal{G}}^+, \bar{\mathcal{G}}^-, \mathcal{G}^+, \mathcal{G}^-$ we say that a guarded formula φ is of type $\langle \bar{\mathcal{G}}^+, \bar{\mathcal{G}}^-, \mathcal{G}^+, \mathcal{G}^- \rangle$ if $\bar{\mathcal{G}}^+(\varphi) \subseteq \bar{\mathcal{G}}^+, \bar{\mathcal{G}}^-(\varphi) \subseteq \bar{\mathcal{G}}^-, \mathcal{G}^+(\varphi) \subseteq \mathcal{G}^+$ and $\mathcal{G}^-(\varphi) \subseteq \mathcal{G}^-$. ⊣

The four different positions in which relation symbols may occur can be assimilated in the notion of bisimulation. Below, when we say that a model $\mathcal{M}$ is of type τ, we view τ as a similarity type in the usual sense.

4.7.2. Definition [**τ-Bisimulation**] Let $\tau = \langle \bar{\mathcal{G}}^+, \bar{\mathcal{G}}^-, \mathcal{G}^+, \mathcal{G}^- \rangle$ be a type. A non-empty set F of finite one-to-one partial maps between two τ-models $\mathcal{M}, \mathcal{N}$ is a τ-bisimulation from $\mathcal{M}$ to $\mathcal{N}$ if for any $f : X \longrightarrow Y \in F$ the following conditions hold:

1. If $R \in \bar{\mathcal{G}}^+$, $\boldsymbol{m} \in dom(f)$ and $\mathcal{M} \models R\boldsymbol{x}[\boldsymbol{m}]$, then $\mathcal{N} \models R\boldsymbol{x}[\boldsymbol{f(m)}]$.
2. If $R \in \bar{\mathcal{G}}^-$, $\boldsymbol{m} \in dom(f)$ and $\mathcal{N} \models R\boldsymbol{x}[\boldsymbol{f(m)}]$, then $\mathcal{M} \models R\boldsymbol{x}[\boldsymbol{m}]$.
3. For any $Z \subseteq^l_{\mathcal{G}^+} \mathcal{M}$ there is a $g \in F$ with domain Z such that g and f agree on the intersection of their domains. (The *zig*-condition)

4. For any $W \subseteq^l_{\mathcal{G}^-} \mathcal{N}$ there is a $g \in F$ with image W such that g^{-1} and f^{-1} agree on the intersection of their domains. (The *zag*-condition)

For any $k \in \omega$, $\boldsymbol{m} \in M^k$ and $\boldsymbol{n} \in N^k$, we write $F : \mathcal{M}, \boldsymbol{m} \sim_\tau \mathcal{N}, \boldsymbol{n}$ to denote that F is a τ-bisimulation from $\mathcal{M}$ to $\mathcal{N}$ which contains a map f such that $\boldsymbol{m} \in dom(f)$ and $f(m_i) = n_i$, for $i = 1, \ldots, k$. ⊣

Note that, contrary to the notion of a guarded bisimulation, the notion of a τ-bisimulation is *directed*. The following is a directed notion of equivalence.

4.7.3. Definition [**τ-Equivalence**] Let τ be a type. Let $\mathcal{M}, \mathcal{N}$ be two τ-models, $\boldsymbol{m} \in M^k$ and $\boldsymbol{n} \in N^k$, for some $k \in \omega$. Let $\boldsymbol{v} = v_1, \ldots, v_k$. We write $\mathcal{M}, \boldsymbol{m} \Rightarrow_\tau \mathcal{N}, \boldsymbol{n}$ if for every formula $\varphi(\boldsymbol{v})$ of type τ it is the case that if $\mathcal{M} \models \varphi(\boldsymbol{v})[\boldsymbol{m}]$, then $\mathcal{N} \models \varphi(\boldsymbol{v})[\boldsymbol{n}]$. ⊣

The aim of this section is to establish Theorem 4.7.6. This will be done by a proof of "amalgamation via bisimulation" similar to the proof of Theorem 4.4.5. This type of results hinges on finding the right notion of bisimulation which is such that

- it preserves the appropriate formulas
- between ω-saturated models, the relation of equivalence induces such a bisimulation.

These two conditions are indeed fulfilled by our notion of τ-bisimulation, as is the content of the following two lemmas.

4.7.4. Lemma (Preservation Lemma) *Let $\tau = \langle \bar{\mathcal{G}}^+, \bar{\mathcal{G}}^-, \mathcal{G}^+, \mathcal{G}^- \rangle$ be a type such that $\mathcal{G}^+ \subseteq \bar{\mathcal{G}}^+$, and $\mathcal{G}^- \subseteq \bar{\mathcal{G}}^-$. Let $\mathcal{M}$, $\mathcal{N}$ be models of type τ, and $\boldsymbol{m} \in M^k$ and $\boldsymbol{n} \in N^k$, for some $k \in \omega$. If $F : \mathcal{M}, \boldsymbol{m} \sim_\tau \mathcal{N}, \boldsymbol{n}$, then $\mathcal{M}, \boldsymbol{m} \Rightarrow_\tau \mathcal{N}, \boldsymbol{n}$.*

We emphasize that in the above preservation lemma it is required that $\mathcal{G}^+ \subseteq \bar{\mathcal{G}}^+$, and $\mathcal{G}^- \subseteq \bar{\mathcal{G}}^-$. Note that a similar requirement was needed in Lemma 4.4.11.

Proof: By a straightforward induction on the complexity of τ-formulas. As an example, let us treat the case for a formula φ of the form $\forall \boldsymbol{x}(R\boldsymbol{xy} \to \psi(\boldsymbol{x}, \boldsymbol{y}))$. Note that $R \in \mathcal{G}^- \subseteq \bar{\mathcal{G}}^-$, since φ is a formula of type τ. Let $\mathcal{M}, \mathcal{N}, F, \boldsymbol{m}, \boldsymbol{n}$ be as in the lemma, and suppose $\mathcal{M} \models \varphi[\boldsymbol{m}]$. Assume that $\mathcal{N} \models R\boldsymbol{xy}[\boldsymbol{n}, \boldsymbol{n'}]$, for some $\boldsymbol{n'} \in N$. We have to show that $\mathcal{N} \models \psi(\boldsymbol{x}, \boldsymbol{y})[\boldsymbol{n}, \boldsymbol{n'}]$. Consider some $f \in F$ with $\boldsymbol{m} \in dom(f)$ and $f(m_i) = n_i$, for $i = 1, \ldots, k$. Let $g \in F$ fulfill the zag-condition for $f, \boldsymbol{n}$. Such g exists, since $R \in \mathcal{G}^-$. Moreover, $R \in \bar{\mathcal{G}}^-$. Hence by condition 2, $\mathcal{M} \models R\boldsymbol{xy}[\boldsymbol{g^{-1}(n)}, \boldsymbol{g^{-1}(n')}]$. Recall that $g^{-1}(n_i) = f^{-1}(n_i) = m_i$, for $1 \leq i \leq k$. As $\mathcal{M} \models \varphi[\boldsymbol{m}]$, then $\mathcal{M} \models \psi(\boldsymbol{x}, \boldsymbol{y})[\boldsymbol{m}, \boldsymbol{g^{-1}(n')}]$. By induction hypothesis, then $\mathcal{N} \models \psi(\boldsymbol{x}, \boldsymbol{y})[\boldsymbol{g(m)}, \boldsymbol{n'}]$. That is, $\mathcal{N} \models \psi(\boldsymbol{x}, \boldsymbol{y})[\boldsymbol{n}, \boldsymbol{n'}]$. ■

With any type $\tau = \langle \bar{\mathcal{G}}^+, \bar{\mathcal{G}}^-, \mathcal{G}^+, \mathcal{G}^- \rangle$, we associate an extended type $\tau_{ext} \stackrel{\text{def}}{=} \langle \bar{\mathcal{G}}^+ \cup \mathcal{G}^+, \bar{\mathcal{G}}^- \cup \mathcal{G}^-, \mathcal{G}^+, \mathcal{G}^- \rangle$.

4.7.5. Lemma *Let $\mathcal{M}, \mathcal{N}$ be ω-saturated models of type τ. For $k \in \omega$, and $\boldsymbol{m} \in M^k, \boldsymbol{n} \in N^k$ such that $\mathcal{M}, \boldsymbol{m} \Rightarrow_\tau \mathcal{N}, \boldsymbol{n}$ define the map $f : \{m_1, \ldots, m_k\} \longrightarrow \{n_1, \ldots, n_k\}$ by setting $f(m_i) = n_i$, for $i = 1, \ldots, k$. The collection F of all these maps is a τ_{ext}-bisimulation from $\mathcal{M}$ to $\mathcal{N}$.*

Proof: To show the first condition, let $R \in \bar{\mathcal{G}}^+ \cup \mathcal{G}^+$, $\mathcal{M} \models R[\boldsymbol{m}]$ and $\mathcal{M}, \boldsymbol{m} \Rightarrow_\tau \mathcal{N}, \boldsymbol{n}$. It needs to be shown that $\mathcal{N} \models R[\boldsymbol{n}]$. Suppose $R \in \bar{\mathcal{G}}^+$, then the formula $R\boldsymbol{x}$ is of type τ. Since $\mathcal{M} \models R[\boldsymbol{m}]$ and $\mathcal{M}, \boldsymbol{m} \Rightarrow_\tau \mathcal{N}, \boldsymbol{n}$, then $\mathcal{N} \models R[\boldsymbol{n}]$. If $R \notin \bar{\mathcal{G}}^+$, then $R \in \mathcal{G}^+$ and the formula $\exists y(R\boldsymbol{x} \wedge x = x)$ is a τ-formula. Since this formula is satisfied in $\mathcal{M}$ by $\boldsymbol{m}$, it is also satisfied in $\mathcal{N}$ by $\boldsymbol{n}$. Therefore, $\mathcal{N} \models R\boldsymbol{x}[\boldsymbol{n}]$. The second condition follows similarly, using the fact that for any $R \in \mathcal{G}^-$ the formula $\forall y(R\boldsymbol{x} \rightarrow \neg(x = x))$ is of type τ.

The zig-condition follows by a standard argument. Let us concentrate on the zag-condition. To this end, consider some $f : \boldsymbol{m} \longrightarrow \boldsymbol{n}$ in F, and $\{\boldsymbol{n}, \boldsymbol{n'}\} \in I^{\mathcal{N}}(R)$, for some $R \in \mathcal{G}^-$. Let Θ be the set of τ-formulas ϑ such that $\mathcal{N} \models \neg\vartheta[\boldsymbol{n}, \boldsymbol{n'}]$. For any finite $\Theta_0 \subseteq \Theta$, set $\chi_0 = \forall \boldsymbol{y}(R\boldsymbol{xy} \rightarrow \bigvee_{\vartheta \in \Theta_0} \vartheta)$. Note that χ_0 is a formula of type τ which is not satisfied in $\mathcal{N}$ by $\boldsymbol{n}$. Hence, $\mathcal{M} \not\models \chi_0[\boldsymbol{m}]$. That is, $\mathcal{M} \models \exists \boldsymbol{y}(R\boldsymbol{xy} \wedge \bigwedge_{\vartheta \in \Theta_0} \neg\vartheta)[\boldsymbol{m}]$, for every finite $\Theta_0 \subseteq \Theta$. By ω-saturatedness of $\mathcal{M}$, there exists some $\boldsymbol{m'} \in M$ such that $\{R\boldsymbol{xy}\} \cup \{\neg\vartheta : \vartheta \in \Theta\}$ is satisfied in $\mathcal{M}$ by $\boldsymbol{mm'}$. Then $\mathcal{M}, \boldsymbol{mm'} \Rightarrow_\tau \mathcal{N}, \boldsymbol{nn'}$. Define the map $g : \{\boldsymbol{m}, \boldsymbol{m'}\} \longrightarrow \{\boldsymbol{n}, \boldsymbol{n'}\}$, by $g(m_i) = n_i$ and $g(m'_j) = n'_j$. This g fulfills the zag-condition for $f, \{\boldsymbol{n}, \boldsymbol{n'}\}$. ■

We turn to the main result of the present section.

4.7.6. Theorem *Let $\varphi, \psi \in$ GF. If $\models \varphi \rightarrow \psi$, then there exists an interpolant $\vartheta \in$ GF satisfying the following conditions:*

1. *$\models \varphi \rightarrow \vartheta$ and $\models \vartheta \rightarrow \psi$.*
2. *$free(\vartheta) \subseteq free(\varphi) \cap free(\psi)$.*
3. *$\bar{\mathcal{G}}^+(\vartheta) \subseteq \bar{\mathcal{G}}^+(\varphi) \cap^+ (\psi)$, and*
   $\bar{\mathcal{G}}^-(\vartheta) \subseteq^- (\varphi) \cap \bar{\mathcal{G}}^-(\psi)$.
4. *$\mathcal{G}^+(\vartheta) \subseteq \mathcal{G}^+(\varphi)$, and*
   $\mathcal{G}^-(\vartheta) \subseteq \mathcal{G}^-(\psi)$.

Note that in case the sets $\bar{\mathcal{G}}(\varphi) \cup \bar{\mathcal{G}}(\psi)$ and $\mathcal{G}(\varphi) \cup \mathcal{G}(\psi)$ are disjoint, the above clause 3 reduces to: 3'. $\bar{\mathcal{G}}^+(\vartheta) \subseteq \bar{\mathcal{G}}^+(\varphi) \cap \bar{\mathcal{G}}^+(\psi)$, and $\bar{\mathcal{G}}^-(\vartheta) \subseteq \bar{\mathcal{G}}^-(\varphi) \cap \bar{\mathcal{G}}^-(\psi)$. This is in particular the case if φ and ψ are translations of modal formulas.

Proof: Consider $\varphi, \psi \in$ GF such that $\models \varphi \rightarrow \psi$. Let $\tau = \langle \bar{\mathcal{G}}^+(\varphi) \cap^+ (\psi), ^-(\varphi) \cap \bar{\mathcal{G}}^-(\psi), \mathcal{G}^+(\varphi), \mathcal{G}^-(\psi) \rangle$. Suppose there is no interpolant for φ, ψ of type τ. We will derive a contradiction.

Let $k \in \omega$ be big enough such that $free(\varphi) \cup free(\psi) \subseteq \{v_1, \ldots, v_k\}$, and define

$$\Theta = \{\vartheta(v_1, \ldots, v_k) \in \text{GF} \ : \ \vartheta \text{ is of type } \tau \ \& \ \models \varphi \rightarrow \vartheta\}.$$

By compactness, it follows from our assumption that $\Theta \not\models \psi$. Hence, there exists some model $\mathcal{N}$ and some $\boldsymbol{b} = b_1, \ldots, b_k \in N$ such that $\mathcal{N} \models \Theta \cup \neg\psi[\boldsymbol{b}]$.

We first show that there exists some model $\mathcal{M}$ and some $\boldsymbol{a} = a_1, \ldots, a_k \in M$ such that $\mathcal{M} \models \varphi[\boldsymbol{a}]$ and $\mathcal{M}, \boldsymbol{a} \Rightarrow_\tau \mathcal{N}, \boldsymbol{b}$. To this end, consider

$$\Xi = \{\neg\xi(v_1, \ldots, v_k) \in \mathrm{GF} \;:\; \xi \text{ is of type } \tau \;\&\; \mathcal{N} \models \neg\xi[\boldsymbol{b}]\}.$$

If $\mathcal{M}$, $\boldsymbol{a}$ as above do not exist, then $\Xi \cup \{\varphi\}$ is inconsistent. By compactness, $\models \varphi \to \bigvee \Xi_0$, for some finite disjunction of formulas of type τ such that $\mathcal{N} \models \neg \bigvee \Xi_0[\boldsymbol{b}]$. Since $\bigvee \Xi_0$ itself is a formula of type τ, $\bigvee \Xi_0 \in \Theta$. Therefore, $\mathcal{N} \models \bigvee \Xi_0[\boldsymbol{b}]$. Quod non.

By passing to ω-saturated elementary extensions of $\mathcal{M}$ and $\mathcal{N}$, we may without loss of generality assume that $\mathcal{M}, \mathcal{N}$ are ω-saturated. Let the type τ_{ext} be defined as in Lemma 4.7.5. That is, $\tau_{ext} = \langle(\bar{\mathcal{G}}^+(\varphi) \cap {}^+(\psi)) \cup \mathcal{G}^+(\varphi), ({}^-(\varphi) \cap \bar{\mathcal{G}}^-(\psi)) \cup \mathcal{G}^-(\psi), \mathcal{G}^+(\varphi), \mathcal{G}^-(\psi)\rangle$. By Lemma 4.7.5, there exists some τ_{ext}-bisimulation F from $\mathcal{M}$ to $\mathcal{N}$ which links $\boldsymbol{a}$ and $\boldsymbol{b}$.

Let $\tau_\varphi = \langle {}^+(\varphi), {}^-(\varphi), \mathcal{G}^+(\varphi), \mathcal{G}^-(\varphi)\rangle$ and $\tau_\psi = \langle {}^+(\psi), {}^-(\psi), \mathcal{G}^+(\psi), \mathcal{G}^-(\psi)\rangle$. Similar to the proof of Theorem 4.4.5, the aim of the rest of this proof is to amalgamate the models $\mathcal{M}$ and $\mathcal{N}$ in such a way that we can define a τ_φ-bisimulation from $\mathcal{M}$ to the amalgamated model $\mathcal{MN}$, and a τ_ψ-bisimulation from $\mathcal{MN}$ to $\mathcal{N}$ which, when composed, will map $\boldsymbol{a}$ to $\boldsymbol{b}$. This model $\mathcal{MN}$ is defined over the set MN consisting of pairs $\langle m, n\rangle \in M \times N$ whose components are F-bisimilar. More precisely,

- $MN = \{\langle m, f(m)\rangle \in M \times N \;:\; f \in F\}$.
- For l-ary R, $f \in F$ and $\boldsymbol{mn} = \langle m_1, f(m_1)\rangle, \ldots, \langle m_l, f(m_l)\rangle \in MN^l$, distinguish the following cases.
  - If $\mathcal{M} \models R\boldsymbol{x}[\boldsymbol{m}] \Leftrightarrow \mathcal{N} \models R\boldsymbol{x}[\boldsymbol{n}]$, put $\mathcal{MN} \models R\boldsymbol{x}[\boldsymbol{mn}] \Leftrightarrow \mathcal{M} \models R\boldsymbol{x}[\boldsymbol{m}]$.
  - If $\mathcal{M} \models R\boldsymbol{x}[\boldsymbol{m}]$ and $\mathcal{N} \models \neg R\boldsymbol{x}[\boldsymbol{n}]$, put
    - If $R \in \bar{\mathcal{G}}^+(\varphi) \setminus \bar{\mathcal{G}}^+(\psi)$, put $\mathcal{MN} \models R\boldsymbol{x}[\boldsymbol{mn}]$.
    - If $R \in {}^+(\psi) \setminus {}^+(\varphi)$, put $\mathcal{MN} \models \neg R\boldsymbol{x}[\boldsymbol{mn}]$.
    - Else, put whatever.
  - If $\mathcal{M} \models \neg R\boldsymbol{x}[\boldsymbol{m}]$ and $\mathcal{N} \models R\boldsymbol{x}[\boldsymbol{n}]$, put
    - If $R \in {}^-(\varphi) \setminus {}^-(\psi)$, put $\mathcal{MN} \models \neg R\boldsymbol{x}[\boldsymbol{mn}]$.
    - If $R \in \bar{\mathcal{G}}^-(\psi) \setminus \bar{\mathcal{G}}^-(\varphi)$, put $\mathcal{MN} \models R\boldsymbol{x}[\boldsymbol{mn}]$.
    - Else, put whatever.

Note that MN is non-empty, as F is non-empty. Moreover, the interpretation of the relations is well-defined. Next, let π_i denote the projection on the i-th coordinate, and set $F_1 = \{g : X \longrightarrow Y \;:\; X \subseteq dom(f),$ for some $f \in F,$ and $(\forall x \in$

$X)\ g(x) = \langle x, f(x)\rangle\}$, and $F_2 = \{\pi_2 : X \longrightarrow Y \ : \ X \subseteq \{\langle x, f(x)\rangle : x \in dom(f)\}$, for some $f \in F\}$.

Amalgamation lemma F_1 is a τ_φ-bisimulation from $\mathcal{M}$ to $\mathcal{MN}$. F_2 is a τ_ψ-bisimulation from $\mathcal{MN}$ to $\mathcal{N}$.

Before proving the lemma, let us first derive Theorem 4.7.6 from it. Recall that the model $\mathcal{M}$ and the sequence $\boldsymbol{a} \in M$ were chosen in such a way that $\mathcal{M} \models \varphi[\boldsymbol{a}]$. Moreover, some $f \in F$ links $\boldsymbol{a} = \langle a_1, \ldots, a_k\rangle$ to $\boldsymbol{b} = \langle b_1, \ldots, b_k\rangle$. Hence, there exists some $g \in F_1$ which links $\boldsymbol{a}$ to $\langle\langle a_1, b_1\rangle, \ldots, \langle a_k, b_k\rangle\rangle$. By the amalgamation lemma, F_1 is a τ_φ-bisimulation. From Lemma 4.7.4, we conclude that $\mathcal{MN} \models \varphi[\langle a_1, b_k\rangle, \ldots, \langle a_k, b_k\rangle]$. By assumption, then $\mathcal{MN} \models \psi[\langle a_1, b_1\rangle, \ldots, \langle a_k, b_k\rangle]$. Since we included $\pi_2 : \{\langle a_1, b_1\rangle, \ldots, \langle a_k, b_k\rangle\} \longrightarrow \{b_1, \ldots, b_k\}$ in F_2, the second part of the amalgamation lemma allows us to conclude that $\mathcal{N} \models \psi[\boldsymbol{b}]$. But $\mathcal{N}, \boldsymbol{b}$ were chosen such that $\mathcal{N} \models \neg\psi[\boldsymbol{b}]$, thus yielding the desired contradiction.

In the remaining part of this proof, we establish the amalgamation lemma. We first prove that F_1 is a τ_φ-bisimulation. Below, we verify that F_1 satisfies the four conditions in Definition 4.7.2.

Ad 1. Let $R \in {}^+(\varphi)$, $\boldsymbol{m} \subseteq dom(f)$, for some $f \in F$ and $\mathcal{M} \models R\boldsymbol{x}[\boldsymbol{m}]$. We have to show that $\mathcal{MN} \models R\boldsymbol{x}[\langle \boldsymbol{m}, \boldsymbol{f(m)}\rangle]$. If $\mathcal{N} \models R\boldsymbol{x}[\boldsymbol{f(m)}]$, then by construction we are done. Therefore, assume $\mathcal{N} \not\models R\boldsymbol{x}[\boldsymbol{f(m)}]$. As $\mathcal{M}, \boldsymbol{m} \Rightarrow_{\tau_{ext}} \mathcal{N}, \boldsymbol{f(n)}$, it follows that $R\boldsymbol{x}$ is not a τ_{ext}-formula. That is, $R \notin (\bar{\mathcal{G}}^+(\varphi) \cap {}^+(\psi)) \cup \mathcal{G}^+(\varphi)$. So we obtain that

$$R \notin \mathcal{G}^+(\varphi), \tag{4.12}$$

and

$$R \notin (\bar{\mathcal{G}}^+(\varphi) \cap {}^+(\psi)). \tag{4.13}$$

From (4.12) and the fact that $R \in {}^+(\varphi)$ we infer that $R \in \bar{\mathcal{G}}^+(\varphi)$. By (4.13), then $R \notin {}^+(\psi)$. Hence certainly, $R \notin \bar{\mathcal{G}}^+(\psi)$. Thus, $R \in \bar{\mathcal{G}}^+(\varphi) \setminus \bar{\mathcal{G}}^+(\psi)$. By construction then $\mathcal{MN} \models R\boldsymbol{x}[\langle \boldsymbol{m}, \boldsymbol{f(m)}\rangle]$.

Ad 2. Let $R \in {}^-(\varphi)$, $\boldsymbol{m} \subseteq dom(f)$, for some $f \in F$ and $\mathcal{M} \models \neg R\boldsymbol{x}[\boldsymbol{m}]$. We show that $\mathcal{MN} \models \neg R\boldsymbol{x}[\langle \boldsymbol{m}, \boldsymbol{f(m)}\rangle]$. This proves 2. First note that if $\mathcal{N} \models \neg R\boldsymbol{x}[\boldsymbol{f(m)}]$, then this follows immediately from the construction. Therefore, assume $\mathcal{N} \models R\boldsymbol{x}[\boldsymbol{f(m)}]$. As $\mathcal{M}, \boldsymbol{m} \Rightarrow_{\tau_{ext}} \mathcal{N}, \boldsymbol{f(n)}$, then $\neg R\boldsymbol{x}$ is not a τ_{ext}-formula. In other words, $R \notin ({}^-(\varphi) \cap \bar{\mathcal{G}}^-(\psi)) \cup \mathcal{G}^-(\psi)$. Summarizing,

$$R \notin \mathcal{G}^-(\psi), \tag{4.14}$$

and

$$R \notin (\bar{\mathcal{G}}^-(\psi) \cap {}^-(\varphi)). \tag{4.15}$$

Recall that $R \in {}^{-}(\varphi)$. By (4.15), $R \notin \bar{\mathcal{G}}^{-}(\psi)$. Together with (4.14), this implies that $R \notin {}^{-}(\psi)$. We conclude that $R \in {}^{-}(\varphi) \setminus {}^{-}(\psi)$. Again it follows from the construction that $\mathcal{MN} \models R\boldsymbol{x}[\langle \boldsymbol{m}, \boldsymbol{f(m)} \rangle]$.

Ad zig-condition. Consider some $g \in F_1$, $R \in \mathcal{G}^{+}(\varphi)$ and $Z \subseteq^{l}_{R} \mathcal{M}$. By definition of F_1, there exists some $f \in F$ such that $g(x) = \langle x, f(x) \rangle$, for all $x \in dom(g)$. As $R \in \mathcal{G}^{+}(\varphi)$, it follows from the zig-condition for f, Z that there exists some $f' \in F$ with domain Z such that f and f' agree on the intersection on their domains. Let $g' : Z \longrightarrow MN$ map any $z \in Z$ to $\langle z, f'(z) \rangle$. Then $g' \in F_1$ and this map fulfills the zig-condition for g, Z.

Ad zag-condition. Let $g : X \longrightarrow Y \in F_1$, $R \in \mathcal{G}^{-}(\varphi)$ and $W \subseteq^{l}_{R} \mathcal{MN}$. Let f be a map in F such that $g(x) = \langle x, f(x) \rangle$, for all $x \in X$. As before, such map exists by definition of F_1. By construction of the model $\mathcal{MN}$, there also exists some $h \in F$ such that any $w \in W$ is of the form $\langle m, h(m) \rangle$, for some $m \in M$. Recall that π_i denotes the projection on the i-th coordinate. Define the map $g' : \pi_1[W] \longrightarrow W$ by $g'(m) = \langle m, h(m) \rangle$. Consider some $u \in Y \cap W$. Then $u = \langle x, f(x) \rangle$, for some $x \in X$, and $u = \langle m, h(m) \rangle$, for some $m \in M$. In particular, $x = m$. Hence, $g^{-1}(u) = x = m = (g')^{-1}(u)$. We conclude that g' fulfills the zag-condition for g, W.

This shows that F_1 is a τ_φ-bisimulation. Similarly, it can be shown that F_2 is a τ_ψ-bisimulation. This finishes the proof of the amalgamation lemma and proves Theorem 4.7.6. ■

4.8 Appendix

In section 4.4 we claimed, for any relation R and $l \in \omega$, the existence of a guarded formula $\mathsf{Live}_R(v_1, \ldots, v_l)$ that is satisfied in a model $\mathcal{M}$ by $m_1, \ldots, m_l \in M$ if and only if the set $\{m_1, \ldots, m_l\}$ is R-live in $\mathcal{M}$. Let us see what this formula looks like.

Let s be the arity of R. Let e range over all complete equality types in variables $v_1, \ldots, v_l$. We regard e both as a quantifier-free formula $e(v_1, \ldots, v_l)$ in the empty vocabulary and as an equivalence relation on the set $\{1, \ldots, l\}$ according to $(j, i) \in e$ iff $e \models v_j = v_i$. Let ρ: $\{1, \ldots, s\} \rightarrow \{1, \ldots, l + s\}$ be a mapping that is onto $\{1, \ldots, l\}/e$, i.e., for every $j \in \{1, \ldots, l\}$ there is some $i \in \{1, \ldots, s\}$ such that $\rho(i)$ is in the same e equivalence class with j. Put, for any such pair of e and ρ,

$$\gamma_{e,\rho} = e(v_1, \ldots, v_l) \wedge \exists \boldsymbol{v} \big(R v_{\rho(1)} \ldots v_{\rho(s)}, true \big), \tag{4.16}$$

where $\boldsymbol{v}$ consists of those $v_{\rho(i)}$ for which $\rho(i) > l$ (if there are such; else no quantification is necessary and $\gamma_{e,\rho}$ is actually atomic). The desired formula $\mathsf{Live}_R(v_1, \ldots, v_l)$ is obtained as the disjunction over all $\gamma_{e,\rho}$ for matvarthetang pairs (e, ρ).

For any *finite* language $\mathcal{L}$ we further obtain a formula $\mathsf{L}ive_{\mathcal{L}}(v_1, \ldots, v_l)$ defining the set of $\mathcal{L}$-live l-tuples by putting

$$\mathsf{L}ive_{\mathcal{L}}(v_1, \ldots, v_l) = (\bigwedge_{1 \leq i,j \leq l} v_i = v_j) \vee \bigvee_{R \in \mathcal{L}} \mathsf{L}ive_R(v_1, \ldots, v_l), \tag{4.17}$$

where the first disjunct reflects the fact that all singleton sets are regarded as live (namely, as guarded by equality).

Now we are in a position to prove Proposition 4.4.3. Let us recall its formulation.

Proposition 4.4.3 *Let $\mathcal{L}$ be a language which contains at most k-ary relation symbols. Any first order formula φ in k variables which is uniformly guarded in $\mathsf{L}ive_{\mathcal{L}}(v_1, \ldots, v_k)$ is equivalent to some $\varphi' \in \mathrm{GF}_k^k$ such that $\mathcal{L}_{G(\varphi')} = \mathcal{L}$ and $\mathcal{L}_{\bar{G}(\varphi')} = \mathcal{L}_{\bar{G}(\varphi)}$.*

Proof: Proof is by induction on the complexity of φ. The only interesting case is the step for the quantifier. Thus, let us consider the guarded formula φ, uniformly guarded in $\mathsf{L}ive_{\mathcal{L}}$, of the form $\exists \boldsymbol{v}[\mathsf{L}ive_{\mathcal{L}}(\boldsymbol{u}, \boldsymbol{v}) \wedge \psi(\boldsymbol{u}, \boldsymbol{v})]$. Unraveling the definition of $\mathsf{L}ive_{\mathcal{L}}$, and using distributivity of $\exists$ over $\vee$, we see that φ is equivalent to $\exists \boldsymbol{v}[(\bigwedge_{x,y \in \boldsymbol{uv}} x = y) \wedge \psi] \vee \bigvee_{R \in \mathcal{L}} \bigvee_{e,\rho} (\exists \boldsymbol{v}(\gamma_{e,\rho} \wedge \psi))$, where e, ρ range over all matching pairs. The former disjunct can easily be rewritten in the desired format. Our remaining task is to conveniently rewrite the latter disjuncts.

To this end, fix $R \in \mathcal{L}$ and a matching pair e, ρ. By (4.16), $\exists \boldsymbol{v}(\gamma_{e,\rho}, \psi)$ is of the form

$$\exists \boldsymbol{v}(e(\boldsymbol{u}, \boldsymbol{v}) \wedge \exists \boldsymbol{w}(R\boldsymbol{u}'\boldsymbol{v}'\boldsymbol{w}) \wedge \psi(\boldsymbol{u}, \boldsymbol{v})), \tag{4.18}$$

for some $\boldsymbol{u}' \subseteq \boldsymbol{u}, \boldsymbol{v}' \subseteq \boldsymbol{v}$ such that any variable in $\boldsymbol{u}, \boldsymbol{v}$ has an e-equivalent among $\boldsymbol{u}', \boldsymbol{v}'$. Fix, for any e-class V of variables in $\boldsymbol{u}, \boldsymbol{v}$, a representative v_V such that $v_V \in \boldsymbol{u}$, if $V \cap \boldsymbol{u} \neq \emptyset$. In (4.18) we subsequently

- bring $e(\boldsymbol{u})$ outside the scope of the quantifier.
- replace each variable x (free or bound) in $\boldsymbol{u}, \boldsymbol{v}$ by the representative of its e-class, denoted by v_{x_e}. Don't do this in $e(\boldsymbol{u})$.
- make sure no variables from $\boldsymbol{u}$ appear in $\exists \boldsymbol{v}$ after this substitution. If they do, remove them from this quantification.

This yields the following equivalent of (4.18),

$$e(\boldsymbol{u}) \wedge \exists \boldsymbol{v}_{\boldsymbol{v_e}} \exists \boldsymbol{w}[R\boldsymbol{v}_{\boldsymbol{u}'_e}\boldsymbol{v}_{\boldsymbol{v}'_e}\boldsymbol{w}, ((\bigwedge_{v \in \boldsymbol{v}, u \in \boldsymbol{uv}, <u,v> \notin e} v_{v_e} \neq v_{u_e}) \wedge \psi(\boldsymbol{v}_{\boldsymbol{u_e}}, \boldsymbol{v}_{\boldsymbol{v_e}}))]. \tag{4.19}$$

Note that the formula φ' displayed in (4.19) is guarded. For this it is crucial that any variable in $\boldsymbol{u}, \boldsymbol{v}$ has an e-equivalent among $\boldsymbol{u}', \boldsymbol{v}'$. Observe that if the arity of the relation symbols in $\mathcal{L}$ is at most k, and our original φ is a formula in at most k variables, then φ' contains at most k variables. Moreover, $\mathcal{L}_{G(\varphi')} = \{R\} \cup \mathcal{L}_{G(\psi)}$ and $\mathcal{L}_{\bar{G}(\varphi')} = \mathcal{L}_{\bar{G}(\psi)}$. By induction hypothesis we conclude that φ' is an equivalent of the original formula φ which fulfills the conditions of the proposition. ■

Recall that for a finite language $\mathcal{L}$, and $n \in \omega$, we defined

$$\mathsf{P}ack_{\mathcal{L}}(v_1, \ldots, v_n) = \bigwedge_{i,j \leq n} \mathsf{L}ive_{\mathcal{L}}(v_i, v_j).$$

According to Proposition 4.4.7, $\mathsf{P}ack_{\mathcal{L}}$ is allowed at guard positions in packed formulas.

Proposition 4.4.7 *Let $\mathcal{L}$ be a language which contains at most k-ary relation symbols. Any first order formula φ in k variables which is uniformly guarded in $\mathsf{P}ack_{\mathcal{L}}(v_1, \ldots, v_k)$ is equivalent to some $\varphi' \in \mathrm{PF}^k_k$ such that $\mathcal{L}_{G(\varphi')} = \mathcal{L}$ and $\mathcal{L}_{\bar{G}(\varphi')} = \mathcal{L}_{\bar{G}(\varphi)}$.*

Proof: Proof is by induction on the complexity of φ. The only interesting case is the step for the quantifiers. Therefore, consider the packed formula φ, uniformly guarded in $\mathsf{P}ack_{\mathcal{L}}$, of the form $\exists v_1 \cdots v_l[\mathsf{P}ack_{\mathcal{L}}(v_1, \ldots, v_n) \wedge \psi(v_1, \ldots, v_n)]$. By definition of $\mathsf{P}ack_{\mathcal{L}}$, φ is equivalent to

$$\exists v_1 \cdots v_l[(\underbrace{\bigwedge_{i,j \leq n} \bigvee_{R \in \mathcal{L}} \bigvee_{\gamma_{\langle e,\rho \rangle}} \gamma_{e,\rho}}_{\chi}) \wedge \psi]. \tag{4.20}$$

Write χ in disjunctive normal form. This yields a huge disjunction χ' ranging over all possibilities for v_i, v_j to be $\mathcal{L}$-live, for $i, j \leq n$. To make this more precise, define the set F of maps $f : \{v_1, \ldots, v_n\} \times \{v_1, \ldots, v_n\} \longrightarrow \{\gamma_{R,e,\rho} : R \in \mathcal{L}, \langle e, \rho \rangle$ a matching pair for $R\}$. Then χ' is of the form

$$\bigvee_{f \in F} (\bigwedge_{i,j \leq n} f(v_i, v_j)). \tag{4.21}$$

Hence, (4.20) is equivalent to

$$\bigvee_{f \in F} \exists v_1 \cdots v_l (\bigwedge_{i,j \leq n} f(v_i, v_j) \wedge \psi). \tag{4.22}$$

By definition of $\gamma_{e,\rho}$, (cf. (4.16)), each $f(v_i, v_j)$ is either the identity $v_i = v_j$ or the conjunction of $v_i \neq v_j$ and a formula of the form $\exists \boldsymbol{u}(Rv_iv_j\boldsymbol{u})$. Moving these inequalities inside ψ, we obtain a formula φ' in PF_n which is equivalent to our original φ, such that $\mathcal{L}_{G(\varphi')} = \mathcal{L} \cup \mathcal{L}_{G(\psi)}$ and $\mathcal{L}_{\bar{G}(\varphi')} = \mathcal{L}_{\bar{G}(\psi)}$. By induction hypothesis we are done. ■

Chapter 5

Interpretability logics

Outline of the chapter

In this chapter we study interpolation, definability and fixed points in interpretability logics.

In section 5.2 we recall the definition of the basic system of interpretability logic IL together with the complete Kripke semantics for this logic first presented in [de Jongh and Veltman, 1990]. This enables us to tackle the question of interpolation for IL via a model-theoretic (Henkin-style) construction. This proof, in section 5.3, forms the heart of the first part of this chapter. Since the Beth property can be derived in IL from the interpolation property as usual, we also obtain the Beth theorem for IL. Another corollary is the interpolation theorem for the system ILP. This proof, in section 5.4, is due to Hájek. Section 5.5 contains an example which shows that interpolation *fails* in the system ILW. In the second part, i.e., section 5.6, we explore an interesting interplay between Beth definability and fixed points. For a general class of logics these two properties will be shown to be interderivable. This class includes all provability and interpretability logics. Combined with our earlier result that IL has the Beth theorem this yields an alternative proof for the fixed point theorem for IL (cf. [de Jongh and Visser, 1991]). Moreover, it implies that all extensions of the basic system of provability logic GL and all extensions of IL have the Beth definability property. This extends the result in [Maksimova, 1989] concerning the Beth property for provability logics.

5.1 Introduction to interpretability logics

Provability logics and interpretability logics *Provability logics* are normal modal logics that have been introduced out of an interest in formal systems of arithmetic: it turns out that the study of these modal logics sheds light on concepts like 'provability' and 'consistency'. In the words of G. Boolos, the founding father of provability logics, "One of the principal aims of this study is to investigate the effects of interpreting the box of modal logic to mean *it is provable (in a certain formal theory) that...*". Important references in this connection include [Smoryński, 1985], [Boolos, 1993] and [Japaridze and de Jongh, 1998].

In this chapter we study expansions of provability logics that were introduced under the name of *interpretability logics* in [Visser, 1990]. In that paper the modal logics IL, ILM and ILP are defined by expanding the object language of the basic system of provability logic GL with a binary operator $\rhd$. This modality is to be read, relative to an (arithmetical) theory T, as: $A \rhd B$ iff $T + B$ is relatively interpretable in $T + A$. To put it simply, there is a function f (the interpretation) on the formulas of the language of T such that $T + B \vdash C \to T + A \vdash f(C)$. (Obviously this translation function should satisfy certain further requirements.) The main importance of interpretability logics is that they permit a finer analysis of arithmetical theories than provability logics. For example, whereas the provability operators $\Box_{\mathsf{PA}}$ and $\Box_{\mathsf{GB}}$[1] have the same properties, the interpretability operator for PA and the one for GB differ: $\rhd_{\mathsf{PA}}$ satisfies the axiom $M : A \rhd B \to (A \wedge \Box C) \rhd (B \wedge \Box C)$, whereas $\rhd_{\mathsf{GB}}$ satisfies the axiom $P : A \rhd B \to \Box(A \rhd B)$.

Interpretability logics are useful and powerful tools for the study of the strength of different theories. However, we are only interested in interpretability logics as systems of (non-standard) modal logic. In the present chapter we establish purely theoretic results about systems of interpretability logic, like the interpolation property for IL and ILP. To this end, a simple modal reading of $\rhd$ over Kripke models suffices.

Interpolation and definability in interpretability logics For systems of interpretability logic, some (positive and negative) results about interpolation are known. For the axiomatization of the systems directly discussed here, namely IL, ILP and ILW, see respectively sections 5.2, 5.4 and 5.5 For the other systems and more information, consult [Visser, 1998]. In that same paper a proof by Ignatiev of failure of interpolation for ILM is adapted, showing that systems between ILM_0 and ILM do not have interpolation. It follows for example that ILW^* does not have interpolation. In [de Rijke, 1992] unary interpretability logic, i.e., the logic of $(\top \rhd \psi)$ is studied. De Rijke shows that the restricted systems il, ilp and ilm, all satisfy interpolation.

[1] PA is Peano's formalization of Arithmetic and GB is the Gödel-Bernays formalization of set theory.

The question of interpolation for the basic system of interpretability logic IL was raised by Baaz. [Hájek, 1992] gave a positive answer to this question, but unfortunately overlooked some cases as was pointed out by Ignatiev. The latter fixed some of the cases in [Ignatiev, 1992], but the proof remained incomplete for years. In this article we provide a full proof. The techniques developed for this proof also serve to establish interpolation for the system ILP. An alternative way of settling this question was given by Hájek who showed interpolation for ILP assuming that this property holds for IL (cf. [Hájek, 1992]). By using the model-theoretic notion of bisimulation we will furthermore prove failure of interpolation for ILW.

In interpretability logics, the Beth property can be derived from the interpolation property as usual. However, as we will show in this chapter, interpolation is much stronger than the Beth property in the context of interpretability logics. It turns out that every normal extension of the basic system IL (that is, every interpretability logic), has the Beth definability property.

The following table summarizes the results in the field after our contribution.

Binary Systems	IL	ILP	ILM	ILF	ILW	ILW*
Interpolation	yes	yes	no	open	no	no
	Thm 5.3.1	Hájek 1992 Thm 5.4.3	Ignatiev		Thm 5.5.4	Visser 1997
Beth	yes	yes	yes	yes	yes	yes
	Thm 5.6.7	Thm 5.6.7	Thm 5.6.7	Thm 5.6.7	Thm 5.6.7	Thm 5.6.7

Unary Systems	il	ilp	ilm			
Interpolation	yes	yes	yes			
	de Rijke 1992	de Rijke 1992	de Rijke 1992			

In this chapter we assume the reader is familiar with basic notions of modal logic in general (see the appendix, if desired), but we develop in detail the necessary concepts specifically devised in the context of provability and interpretability logics (section 5.2). For a thorough introduction to this topic covering the arithmetical interest of the project we refer to [Japaridze and de Jongh, 1998] and [Visser, 1998].

5.2 Preliminaries

In this section we gather some definitions and preliminary results needed for our main theorem. We start by recalling from [Visser, 1990] the definition of the basic system of interpretability logic IL.

5.2.1. Definition [**The system IL**] The *basic system for interpretability logic* IL is defined by the following axiom schemes:

$L1$ All classical tautologies,
$L2$ $\Box(A \to B) \to (\Box A \to \Box B)$,
$L3$ $\Box A \to \Box\Box A$,

$L4$ $\Box(\Box A \to A) \to \Box A$,
$J1$ $\Box(A \to B) \to A \rhd B$,
$J2$ $(A \rhd B \wedge B \rhd C) \to A \rhd C$,
$J3$ $(A \rhd C \wedge B \rhd C) \to (A \vee B) \rhd C$,
$J4$ $A \rhd B \to (\Diamond A \to \Diamond B)$,
$J5$ $\Diamond A \rhd A$,

together with the rules of Modus Ponens and Necessitation (i.e., $\vdash A \Rightarrow \vdash \Box A$). The notions of proof in IL and of theorems and rules are defined as usual. ⊣

For some intuitions about the role of the above axioms let us turn for a moment to their arithmetical interpretation. Axioms $L1$ to $L4$ are the principles of the basic system of provability logic that we call GL, for Gödel and Löb; $J1$ says that the identity is an interpretation; $J2$ expresses transitivity of the $\rhd$-modality, reflecting that interpretations can be composed. By $J3$ two different interpretations can be joined in a definition by cases; $J4$ states that relative interpretability implies relative consistency; $J5$ is the 'Interpretation Existence Lemma' (cf. [Visser, 1998]), a formalization in arithmetic of Henkin's completeness theorem.

In the proof of our main theorem, Theorem 5.3.1, the following facts will be useful. Proofs can be found in [Japaridze and de Jongh, 1998] and [Visser, 1998].

5.2.2. Proposition *In* IL *the following theorems are derivable:*

1. $\vdash \Box D \leftrightarrow \neg D \rhd \bot$.
2. $\vdash (D \vee \Diamond D) \rhd D$.
3. $\vdash D \rhd (D \wedge \Box\neg D)$.
4. $\vdash ((D \wedge E) \rhd F) \to (\neg D \rhd F \to E \rhd F)$.

Proof: Part (1), (2) and (4) are easy; (3) follows from the fact that in the system of provability logic GL we can derive $\vdash_{\mathsf{GL}} \Diamond D \to \Diamond(D \wedge \Box\neg D)$, and hence, $\vdash_{\mathsf{GL}} D \to (D \wedge \Box\neg D) \vee \Diamond(D \wedge \Box\neg D)$. Now apply (2). ■

We turn to semantics. A Kripke semantics (in this case also called *Veltman semantics*) for IL was first presented in [de Jongh and Veltman, 1990].

5.2.3. Definition [**IL-frame, IL-model, forcing relation**] A tuple $\langle W, R, S\rangle$ is an IL-*frame* if:

- W is a non-empty set.
- R is a transitive, upwards well-founded binary relation on W.
- For each $w \in W$,
  - S_w is a binary relation defined on $w\!\uparrow \stackrel{\text{def}}{=} \{u \in W : wRu\}$.
  - S_w is transitive and reflexive.
  - $wRuRv \Rightarrow uS_wv$.

An IL-*model* is a structure $\langle\langle W, R, S\rangle, V\rangle$, where $\langle W, R, S\rangle$ is an IL-frame and V is a modal valuation assigning subsets of W to proposition letters. A *forcing relation* $\models$

on an IL-model satisfies the usual clauses for atomic formulas, Boolean connectives and $\Box$-modality (with R as the accessibility relation), plus the following extra clause:

- $w \models A \rhd B \Leftrightarrow \forall u((wRu \wedge u \models A) \Rightarrow \exists v(uS_w v \wedge v \models B))$. ⊣

A *modal completeness theorem* for IL with respect to finite IL-models is provided in [de Jongh and Veltman, 1990].

Note that the clause for the $\rhd$-modality in the definition of the forcing relation above, is unlike the clause for the usual $\Box$-modality. This is why we consider interpretability logics to be *non-standard* systems of modal logic.

5.2.4. Convention In the remaining part of this section we will tacitly assume that we are working in IL. Hence, all the notions defined below are to be read relative to this system. For example, when we speak about a set of formulas it will be understood that these are IL-formulas, etc. ⊣

The method we will use for showing interpolation is a standard model-theoretic Henkin style proof as can be found, e.g., in the proof of interpolation for provability logic in [Smoryński, 1978]. The aim of this kind of proofs is to construct a model of the logic under consideration whose worlds are based on maximal consistent sets of formulas. However, since IL is not compact, maximal consistent sets should be confined to finite adequate subsets of the language. Our first task is to specify this notion of adequateness (see [de Jongh and Veltman, 1990]).

5.2.5. Definition [$\sim A$**, Adequate set**] If the formula A is not a negation, then $\sim A$ is $\neg A$. Otherwise, if A is $\neg B$, then $\sim A$ is B. A set X of formulas is called *adequate* if X is closed under subformulas and the $\sim$-operation, $\bot \rhd \bot \in X$ and X contains $A \rhd B$ whenever A, B are antecedent or succedent of a $\rhd$-formula in X. ⊣

From this point onwards it is best to consider $\Box A$ as an abbreviation of $\sim A \rhd \bot$. This is allowed by the first part of Proposition 5.2.2. In particular, this convention implies that whenever formulas of the form $\Box\neg A, \Box\neg B$ are contained in an adequate set X, then also $A \rhd B \in X$.

5.2.6. Notation For any set of formulas X there exists a smallest adequate set containing X, denoted by $\mathcal{A}_X$. As usual, we omit brackets when appropriate. By $\mathcal{L}_X$ (read: *the language of* X) we denote the set of IL-formulas built up from proposition letters occurring in formulas in X. For X a finite set of formulas, we interchangeably write X for its conjunction: e.g. $\vdash \bigwedge X \rightarrow A$ will be written simply as $\vdash X \rightarrow A$. ⊣

5.2.7. Remark Note that if X is finite, then so is $\mathcal{A}_X$, as desired. In order to ensure this, the set X in Definition 5.2.5 was required to be closed under negation of non-negated formulas only. ⊣

In modal logic, proofs of interpolation are in general close in spirit to completeness proofs. The central role played by *maximal consistent sets* in the latter is in the former taken over by *complete inseparable pairs*.

5.2.8. Definition [**Inseparable pair**] A pair $\langle X, Y\rangle$ of finite sets of formulas is called *separable* if there exists a formula $A \in \mathcal{L}_X \cap \mathcal{L}_Y$ such that $\vdash X \to A$ and $\vdash Y \to \neg A$. A pair is called *inseparable* if it is not separable. ⊣

Note that for any inseparable pair $\langle X, Y\rangle$, the sets X and Y are each consistent.

5.2.9. Definition [**Complete pair**] Let $\langle X, Y\rangle$ be an inseparable pair. We say that $\langle X, Y\rangle$ is *complete* if

1. For each $A \in \mathcal{A}_X$, either $A \in X$ or $\sim A \in X$.
2. For each $A \in \mathcal{A}_Y$, either $A \in Y$ or $\sim A \in Y$. ⊣

In e.g. [Smoryński, 1985] the following analogue of Lindenbaum's Lemma can be found.

5.2.10. Proposition *Let $\langle X, Y\rangle$ be an inseparable pair. Then there exist sets X', Y' such that $X \subseteq X' \subseteq \mathcal{A}_X$, $Y \subseteq Y' \subseteq \mathcal{A}_Y$ and $\langle X', Y'\rangle$ is a complete pair.*

The preparations up to now suffice to define the worlds of the construction we are after. To define the relations in this model the following notion is needed.

5.2.11. Definition [$\prec$ **Relation**] Let $\langle X, Y\rangle$, $\langle X', Y'\rangle$ be two complete pairs such that $\mathcal{A}_X = \mathcal{A}_{X'}$, $\mathcal{A}_Y = \mathcal{A}_{Y'}$. We put $\langle X, Y\rangle \prec \langle X', Y'\rangle$ if

1. For each A, if $\Box A \in X \cup Y$ then $\Box A, A \in X' \cup Y'$.
2. There exists some A such that $\Box A \notin X \cup Y$ but $\Box A \in X' \cup Y'$. ⊣

The above is the canonical definition of the accessibility relation for the $\Box$-modality which takes care of the conditions of transitivity and upward well-foundedness.

In order to motivate the next definition, let us jump a little bit ahead of ourselves, and ask what this entire enterprise should amount to. As usual in Henkin-style proofs for interpolation, the idea is the following. On the assumption that some two formulas B and C (such that $\vdash B \to C$) *do not* have an interpolant, the pair $\langle\{B\}, \{\neg C\}\rangle$ can be extended to a complete pair which will be a world in the model that is now to be constructed. The key point is then to prove a truth lemma for the eventual model saying that a formula is valid in a world if and only if that formula is contained in one component of the complete pair which constitutes that world. This lemma implies that we have constructed a world in which B and $\neg C$ holds, contrary to the fact that $B \to C$ is a theorem and we are done. Now, for proving the truth lemma we will in particular have to show that, if a formula of the form $\neg(G \rhd A)$ is contained in some world w, then $w \not\models (G \rhd A)$. According to the truth definition, we should in that case produce an R-successor

u of w which contains G and which 'avoids' A in the sense that any S_w-successor of u does not contain A.

What makes this concept of 'A-avoiding' hard to grasp, is the fact that avoiding a formula A involves other formulas D as well. Let us see why. Consider a world w which contains a formula of the form $D \rhd A$. Hence, by the truth lemma, $w \models D \rhd A$. In this case any truly A-avoiding successor u of w is not allowed to contain D, nor to have an R-successor v containing D. In the first case it follows directly from the truth definition that u has an S_w-successor satisfying A, contrary to u being A-avoiding. In the second case we reason as follows. Since wRv (by transitivity of R) it follows again from the truth-definition that v has an S_w-successor z which contains A. Moreover, $wRuRv$ and hence, by the definition of IL-frame, uS_wv. Since S_w is transitive, this shows that z is a S_w-successor of u, and again we end up with an S_w-successor of u containing A. Bearing this in mind, a first attempt to formalize the intuitive notion of 'A-avoiding successor' would be via the following concept of A-criticality (see [Hájek, 1992]).

5.2.12. Definition [A**-Critical, preliminary**] Let $\langle X, Y\rangle$, $\langle X', Y'\rangle$ be two complete pairs such that $\mathcal{A}_X = \mathcal{A}_{X'}$, $\mathcal{A}_Y = \mathcal{A}_{Y'}$. Let $\Box\neg A \in \mathcal{A}_X \cup \mathcal{A}_Y$. We say that $\langle X', Y'\rangle$ is an *A-critical successor* of $\langle X, Y\rangle$ if the following conditions are met.

1. $\langle X, Y\rangle \prec \langle X', Y'\rangle$.
2. $X_1 \overset{\text{def}}{=} \{\neg D, \Box\neg D : D \rhd A \in X\} \subseteq X'$.
   $Y_1 \overset{\text{def}}{=} \{\neg E, \Box\neg E : E \rhd A \in Y\} \subseteq Y'$. ⊣

However complicated as the above definition may seem, it does not yet suffice since it does not reckon with a possible interplay between formulas from $\mathcal{A}_X$ and $\mathcal{A}_Y$. To make this point more precise, let us imagine the situation where $A \in \mathcal{A}_X \setminus \mathcal{A}_Y$ and $B \in \mathcal{A}_Y \setminus \mathcal{A}_X$. Although the formulas A and B come from entirely different adequate sets, still B can turn out to be an undesirable member of any A-critical successor of a pair $\langle X, Y\rangle$. For it can be the case that $\vdash X \to C \rhd A$ and $\vdash Y \to B \rhd C$, for some $C \in \mathcal{L}_X \cap \mathcal{L}_Y$ but *not necessarily in $\mathcal{A}_X$ or $\mathcal{A}_Y$*. By soundness then $\langle X, Y\rangle \models B \rhd A$, and B should henceforth be avoided as not to run in the same trouble as before. However, since $B \rhd A$ is not contained in any of the adequate sets $\mathcal{A}_X, \mathcal{A}_Y$, and hence $B \rhd A \notin X \cup Y$, Definition 5.2.12 does not give any restrictions in this case. On these grounds we exchange our preliminary definition for the one below.

5.2.13. Definition [A**-Critical**] Let $\langle X, Y\rangle$, $\langle X', Y'\rangle$ be two complete pairs such that $\mathcal{A}_X = \mathcal{A}_{X'}$, $\mathcal{A}_Y = \mathcal{A}_{Y'}$. Let $\Box\neg A \in \mathcal{A}_X \cup \mathcal{A}_Y$. We say that $\langle X', Y'\rangle$ is an *A-critical successor of* $\langle X, Y\rangle$ (notation: $\langle X, Y\rangle \prec_A \langle X', Y'\rangle$), if the following conditions are met.

1. $\langle X, Y\rangle \prec \langle X', Y'\rangle$.

2. If $\Box\neg A \in \mathcal{A}_X$, then
$X_1 \stackrel{\text{def}}{=} \{\neg D, \Box\neg D : D \rhd A \in X\} \subseteq X'$.
$Y_1 \stackrel{\text{def}}{=} \{\neg E, \Box\neg E : \Box\neg E \in \mathcal{A}_Y \&$
$\exists C \in \mathcal{L}_X \cap \mathcal{L}_Y [\vdash Y \rightarrow (E \rhd C) \ \& \ \vdash X \rightarrow (C \rhd A)]\} \subseteq Y'$.
3. If $\Box\neg A \in \mathcal{A}_Y$, then
$X_2 \stackrel{\text{def}}{=} \{\neg D, \Box\neg D : \Box\neg D \in \mathcal{A}_X \&$
$\exists C \in \mathcal{L}_X \cap \mathcal{L}_Y [\vdash X \rightarrow (D \rhd C) \ \& \ \vdash Y \rightarrow (C \rhd A)]\} \subseteq X'$.
$Y_2 \stackrel{\text{def}}{=} \{\neg E, \Box\neg E : E \rhd A \in Y\} \subseteq Y'$. ⊣

Note that the complications described above only occur in case A and B are contained in different adequate sets. That is why the sets X_1 and Y_2 in Definition 5.2.13 remain unaltered as compared to the sets X_1, Y_1 in Definition 5.2.12.

Summarizing, the difficulties in finding the above notion of criticality which will turn out to be the one needed for the interpolation proof were twofold. First, the non-standard character of the $\rhd$-modality brought on the problem that avoiding one formula involves other formulas. Second, the fact that we are interested in interpolation made us pay attention to the languages. The next claim implies that the above notion is well-defined.

5.2.14. Claim *If $\Box\neg A \in \mathcal{A}_X \cap \mathcal{A}_Y$ in Definition 5.2.13, then $X_1 = X_2$ and $Y_1 = Y_2$.*

Proof: Let $\Box\neg A \in \mathcal{A}_X \cap \mathcal{A}_Y$. Obviously $X_1 \subseteq X_2$. For the other inclusion, consider a formula D such that $\neg D, \Box\neg D \in X_2$. That is, $\Box\neg D \in \mathcal{A}_X$ and there exists some $C \in \mathcal{L}_X \cap \mathcal{L}_Y$ such that (*) $\vdash X \rightarrow (D \rhd C)$ and $\vdash Y \rightarrow (C \rhd A)$. We want to show that $\neg D, \Box\neg D \in X_1$, i.e., $D \rhd A \in X$. Let us assume for contradiction that $D \rhd A \notin X$. Since $D \rhd A \in \mathcal{A}_X$, by completeness of $\langle X, Y\rangle$ this assumption implies that (**) $\neg(D \rhd A) \in X$. By (*), $\vdash X \rightarrow [(C \rhd A) \rightarrow (D \rhd A)]$. From (**) it now follows that $\vdash X \rightarrow \neg(C \rhd A)$. We conclude that $C \rhd A$ separates X and Y. Contradiction. To show that $Y_1 = Y_2$, one proceeds analogously. ■

Note that for any $\langle X, Y\rangle$, $\langle X', Y'\rangle$, $\langle X'', Y''\rangle$ and any formula A we have that

$$\text{If } \langle X, Y\rangle \prec_A \langle X', Y'\rangle \prec \langle X'', Y''\rangle, \text{ then } \langle X, Y\rangle \prec_A \langle X'', Y''\rangle.$$

This finishes the necessary preliminaries for the next section.

5.3 The interpolation theorem for IL

This section is devoted to proving that the basic system of interpretability logic IL has the interpolation property CIP$^{\rightarrow}$. By the completeness of IL, this is equivalent to the following theorem.

5.3.1. Theorem (CIP$^{\rightarrow}$ **for** IL) *Let* D_0, E_0 *be* IL*-formulas. Assume* $D_0 \rightarrow E_0$ *is valid in* IL*. Then there exists an* IL*-formula* $I \in \mathcal{L}_{D_0} \cap \mathcal{L}_{E_0}$ *such that* $\vdash_{\mathsf{IL}} D_0 \rightarrow I$ *and* $\vdash_{\mathsf{IL}} I \rightarrow E_0$.

Proof of Theorem 5.3.1: Let $\vdash_{\mathsf{IL}} D_0 \rightarrow E_0$. Assume there is no interpolant. In the next few pages it will be shown that this assumption enables us to construct an IL-model which contains a world satisfying both D_0 and $\neg E_0$. From the soundness of IL a contradiction follows. Now let us get to work.

By assumption $\vdash_{\mathsf{IL}} D_0 \rightarrow E_0$ has no interpolant. In other words, $\langle \{D_0\}, \{\neg E_0\}\rangle$ is inseparable. By Proposition 5.2.10 there exist sets X_0, Y_0 such that $\{D_0\} \subseteq X_0 \subseteq \mathcal{A}_{D_0}$, $\{\neg E_0\} \subseteq Y_0 \subseteq \mathcal{A}_{E_0}$ and $\langle X_0, Y_0\rangle$ is a complete pair. We define the model $\mathcal{M} \stackrel{\text{def}}{=} \langle\langle W, R, S\rangle, V\rangle$ as follows.

Begin construction of model.

- Each world in W will be a sequence of 2-tuples consisting of a complete pair together with a sequence of formulas recording 'how we arrived at that pair'. Let [] represent the empty sequence and $*$ stand for concatenation. Formally, W is the smallest set satisfying the following two conditions:
  - $w_0 \stackrel{\text{def}}{=} [(\langle X_0, Y_0\rangle, [])] \in W$.
  - Let $[(\langle X_0, Y_0\rangle, []), \dots, (\langle X_n, Y_n\rangle, \tau_n)] \in W$. Let $\langle X, Y\rangle$ be a complete pair such that $X \subseteq \mathcal{A}_{D_0}$, $Y \subseteq \mathcal{A}_{E_0}$ and $\langle X_n, Y_n\rangle \prec_A \langle X, Y\rangle$, for some A. Then $[(\langle X_0, Y_0\rangle, []), \dots, (\langle X_n, Y_n\rangle, \tau_n), (\langle X, Y\rangle, \tau_n * [A])] \in W$.

  5.3.2. Notation For all $w \in W$, $w = [(\langle X_0, Y_0\rangle, []), \dots, (\langle X_n, Y_n\rangle, \tau_n)]$ we will write X_w (resp. Y_w, τ_w) for the set X_n (resp. Y_n, τ_n). For $w, u \in W$, the notation $w \subseteq u$ (resp. $w \subset u$) indicates that w is an initial (resp. *proper* initial) segment of u. ⊣

- For all $w, u \in W$, we define wRu iff $w \subset u$.

- For all $w, u, v \in W$, we define uS_wv iff there exists some formula A and complete pairs $\langle X', Y'\rangle$, $\langle X'', Y''\rangle$ such that $w * [(\langle X', Y'\rangle, \tau_w * [A])] \subseteq u$, $w * [(\langle X'', Y''\rangle, \tau_w * [A])] \subseteq v$.

We leave it to the reader to check that $\langle W, R, S\rangle$ is an IL-frame. That is, W is finite, R is transitive and irreflexive, and S_w is a transitive and reflexive relation defined over the set $\{u \in W \ : \ wRu\}$ such that for every $w', w'' \in W$ we have that $wRw'Rw''$ implies $w'S_ww''$.

Finally, for every $w \in W$ and every proposition letter $p \in \mathcal{L}_{D_0} \cup \mathcal{L}_{E_0}$, we set the valuation V to

$$w \in V(p) \overset{\text{def}}{\Longleftrightarrow} p \in X_w \cup Y_w.$$

End of construction.

The proof of Theorem 5.3.1 now reduces to the following truth lemma.

Truth lemma *Let $\mathcal{M} = \langle\langle W, R, S\rangle, V\rangle$ be the model defined above. Then for any $w \in W$,*

1. *$B \in \mathcal{A}_{D_0}$ implies $w \models B \leftrightarrow B \in X_w$, and*
2. *$B \in \mathcal{A}_{E_0}$ implies $w \models B \Leftrightarrow B \in Y_w$.*

Note that the truth lemma in particular implies that $w_0 \models D_0$ and $w_0 \models \neg E_0$, for $w_0 \in W$ defined above. Hence, this lemma is all that stands between us and a proof of Theorem 5.3.1.

The hard part of proving the truth lemma is summarized in the two lemmas below, the proof of which is postponed till their use has been demonstrated.

5.3.3. Notation For all $w, u \in W$, and any formula A,

$$wR_A u \overset{\text{def}}{\Longleftrightarrow} \text{ there exists } \langle X', Y'\rangle \text{ such that } w * [(\langle X', Y'\rangle, \tau_w * [A])] \subseteq u.$$

⊣

In particular, $wR_A u$ implies that $\langle X_w, Y_w\rangle \prec_A \langle X_u, Y_u\rangle$.

5.3.4. Lemma *Let $\neg(G \rhd F) \in X_w$ (resp. Y_w). Then there exists some $u \in W$ such that $wR_F u$ and $G \in X_u$ (resp. Y_u).*

5.3.5. Lemma *Let $G \rhd F \in X_w$ (resp. Y_w). Let $u \in W$ be such that $wR_A u$ and $G \in X_u$ (resp. Y_u). Then there exists $v \in W$ such that $wR_A v$ and $F \in X_v$ (resp. Y_v).*

Proof of truth lemma: This proof is by induction on the complexity of B. The atomic case is given by definition, the Boolean cases are an easy exercise and the $\Box$-case is an instance of the $\rhd$-case. Hence, let us concentrate on the latter.

Let B be of the form $G \rhd F \in \mathcal{A}_{D_0} \cup \mathcal{A}_{E_0}$. Let us assume that $G \rhd F \in \mathcal{A}_{D_0}$ (in case that $(G \rhd F) \in \mathcal{A}_{E_0}$ we reason similarly).

CASE " ⇒ ": Let $G \rhd F \notin X_w$. By completeness of $\langle X_w, Y_w\rangle$, then $\neg(G \rhd F) \in X_w$. By Lemma 5.3.4, no S_w-successor v of the element u produced there, satisfies

F: for, wR_Fu and uS_wv imply that wR_Fv. Since $F \rhd F \in X_w$, it follows that $v \not\models F$. We conclude that $w \not\models G \rhd F$, and we are done.

CASE "⇐": Let $G \rhd F \in X_w$. Let $u \in W$ be such that wRu and $u \models G$. Then wR_Au, for some formula A. By induction hypothesis, $G \in X_u$. By Lemma 5.3.5 there exists some $v \in W$ such that uS_wv and $F \in X_v$. Again by the induction hypothesis $v \models F$, and it follows that $w \models G \rhd F$. ■

Now let us prove the two auxiliary lemmas. Both lemmas will be shown to hold for X_w, X_u, X_v. For Y_w, Y_u, Y_v, the proofs are similar.

Proof of Lemma 5.3.4: Let $\neg(G \rhd F) \in X_w$. We define

$$X^- \stackrel{\text{def}}{=} \boxdot X_w \cup \{G, \Box\neg G\} \cup \{\neg D, \Box\neg D : D \rhd F \in X_w\},$$
$$Y^- \stackrel{\text{def}}{=} \boxdot Y_w \cup \{\neg E, \Box\neg E : \Box\neg E \in \mathcal{A}_{E_0} \ \& \ \exists C \in \mathcal{L}_{D_0} \cap \mathcal{L}_{E_0} [\vdash Y_w \to (E \rhd C) \ \& \ \vdash X_w \to (C \rhd F)]\},$$

where here, as elsewhere in the proof, for any set of formulas X,

$$\boxdot X \stackrel{\text{def}}{=} \{D, \Box D : \Box D \in X\}.$$

We will show that X^- and Y^- are inseparable. For then, by Proposition 5.2.10, we can extend $\langle X^-, Y^-\rangle$ to a complete pair $\langle X_u, Y_u\rangle$, and the element $u \stackrel{\text{def}}{=} w * [(\langle X_u, Y_u\rangle, \tau_w * [F])]$ will satisfy all our requirements.

Let us assume for contradiction that X^- and Y^- are separable. That is, there exists some $I \in \mathcal{L}_{D_0} \cap \mathcal{L}_{E_0}$ such that

$$\vdash X^- \to I \quad \text{and} \quad \vdash Y^- \to \neg I.$$

Now we can derive the following:

$$\vdash \boxdot X_w \to [(G \wedge \Box\neg G \wedge \neg I) \to \bigvee_{D \rhd F \in X_w} (D \vee \Diamond D)].$$

Henceforth we will simply omit the index set (in this case X_w) over which a disjunction is taken, in case this set is clear from the context. Reasoning as in provability logic, we obtain from the definition of $\boxdot X_w$ and axiom $J1$ that

$$\vdash X_w \to [(G \wedge \Box\neg G \wedge \neg I) \rhd \bigvee (D \vee \Diamond D)].$$

By the second part of Proposition 5.2.2 and the fact that $D \rhd F \in X_w$, then

$$\vdash X_w \rightarrow [(G \wedge \Box\neg G \wedge \neg I) \rhd F].$$

With the help of clause 4 in Proposition 5.2.2 we derive that

$$\vdash X_w \rightarrow [(I \rhd F) \rightarrow ((G \wedge \Box\neg G) \rhd F)]. \tag{5.1}$$

On the other hand,

$$\vdash \boxdot Y_w \rightarrow [I \rightarrow \bigvee(E_j \vee \Diamond E_j)],$$

for some finite index set J. The formulas E_j are such that there exists $C_j \in \mathcal{L}_{D_0} \cap \mathcal{L}_{E_0}$ for which

$$\vdash Y_w \rightarrow [I \rhd (\bigvee C_j)] \qquad \text{and} \qquad \vdash X_w \rightarrow [(\bigvee C_j) \rhd F] \qquad \text{holds.}$$

It follows that

$$\vdash X_w \rightarrow [(I \rhd (\bigvee C_j)) \rightarrow (I \rhd F)].$$

Together with (5.1) and the fact that $\neg(G \rhd F) \in X_w$ this implies via part 3 of Proposition 5.2.2 that

$$\vdash X_w \rightarrow [\neg(I \rhd (\bigvee C_j))].$$

Hence, $I \rhd (\bigvee C_j)$ separates X_w and Y_w. A contradiction. We conclude that X^- and Y^- are indeed inseparable, as was to be shown. This completes the proof of Lemma 5.3.4. ■

Proof of Lemma 5.3.5: Let $G \rhd F \in X_w$. Let $u \in W$ be such that wR_Au and $G \in X_u$. By definition of criticality, $\Box\neg A \in \mathcal{A}_{D_0} \cup \mathcal{A}_{E_0}$. In this proof we distinguish as to whether $\Box\neg A \in \mathcal{A}_{D_0}$ or $\Box\neg A \in \mathcal{A}_{E_0}$.

CASE 1: Let $\Box\neg A \in \mathcal{A}_{D_0}$. Analogously to the proof of Lemma 5.3.4 we define

$$X^- \stackrel{\text{def}}{=} \boxdot X_w \cup \{F, \Box\neg F\} \cup \{\neg D, \Box\neg D : D \rhd A \in X_w\},$$
$$Y^- \stackrel{\text{def}}{=} \boxdot Y_w \cup \{\neg E, \Box\neg E : \Box\neg E \in \mathcal{A}_{E_0} \;\&\; \exists C \in \mathcal{L}_{D_0} \cap \mathcal{L}_{E_0} [\vdash Y_w \rightarrow (E \rhd C) \;\&\; \vdash X_w \rightarrow (C \rhd A)]\}.$$

Again we will show that X^-, Y^- are inseparable. As before, this implies that the pair $\langle X^-, Y^-\rangle$ can be extended to a complete pair $\langle X_v, Y_v\rangle$. Then, the element $v \stackrel{\text{def}}{=} w * [(\langle X_v, Y_v\rangle, \tau_w * [A])]$ will have all the required properties.

Assume for contradiction that there exists some $I \in \mathcal{L}_{D_0} \cap \mathcal{L}_{E_0}$ such that

$$\vdash X^- \to I \quad \text{and} \quad \vdash Y^- \to \neg I.$$

Again we derive that

$$\vdash \boxdot X_w \to [(F \wedge \Box\neg F \wedge \neg I) \to \bigvee(D \vee \Diamond D)].$$

Reasoning as before we see that

$$\vdash X_w \to [(F \wedge \Box\neg F \wedge \neg I) \rhd A],$$

and

$$\vdash X_w \to [(I \rhd A) \to (F \wedge \Box\neg F \rhd A)]. \tag{5.2}$$

Since $G \rhd F \in X_w$ one immediately sees that

$$\vdash X_w \to [(F \rhd A) \to (G \rhd A)]. \tag{5.3}$$

Now assume that $(G \rhd A) \in X_w$. Since $wR_A u$, then $\neg G \in X_u$, which by assumption is not the case. We conclude that $(G \rhd A) \notin X_w$, hence by completeness of $\langle X_w, Y_w \rangle$

$$\neg(G \rhd A) \in X_w. \tag{5.4}$$

On the other hand,

$$\vdash \boxdot Y_w \to [I \to \bigvee(E_j \vee \Diamond E_j)],$$

for some finite index set J. The formulas E_j are chosen in such a way that there exist formulas $C_j \in \mathcal{L}_{D_0} \cap \mathcal{L}_{E_0}$ such that

$$\vdash Y_w \to [I \rhd (\bigvee C_j)] \quad \text{and} \quad \vdash X_w \to [(\bigvee C_j) \rhd A].$$

It follows that

$$\vdash X_w \to [(I \rhd (\bigvee C_j)) \to (I \rhd A)]. \tag{5.5}$$

(5.2), (5.3), (5.4), (5.5) and item 3 of Proposition 5.2.2 together imply that

$$\vdash X_w \to [\neg(I \rhd (\bigvee C_j))].$$

This shows that $(I \rhd (\bigvee C_j))$ separates X_w and Y_w, which is a contradiction. We conclude that the pair $\langle X^-, Y^- \rangle$ is inseparable.

CASE 2: Let $\Box\neg A \in \mathcal{A}_{E_0}$. This time we define

$$X^{-} \stackrel{\text{def}}{=} \boxdot X_w \cup \{F, \Box\neg F\} \cup \{\neg D, \Box\neg D : \Box\neg D \in \mathcal{A}_{D_0} \; \& \\ \exists C \in \mathcal{L}_{D_0} \cap \mathcal{L}_{E_0} [\vdash X_w \to (D \rhd C) \; \& \; \vdash Y_w \to (C \rhd A)]\},$$
$$Y^{-} \stackrel{\text{def}}{=} \boxdot Y_w \cup \{\neg E, \Box\neg E : E \rhd A \in Y_w\}.$$

Again we assume for contradiction that there exists some $I \in \mathcal{L}_{D_0} \cap \mathcal{L}_{E_0}$ such that

$$\vdash X^{-} \to I \qquad \text{and} \qquad \vdash Y^{-} \to \neg I.$$

Now we reason as follows. First note that

$$\vdash \boxdot Y_w \to [I \to \bigvee (E \vee \Diamond E)],$$

where for every E it is the case that $(E \rhd A) \in Y_w$. Hence,

$$\vdash Y_w \to [I \rhd A]. \tag{5.6}$$

Also,

$$\vdash \boxdot X_w \to [(F \wedge \Box\neg F) \to (I \vee \bigvee (D_j \vee \Diamond D_j))],$$

for some finite index set J. Since $G \rhd F \in X_w$ this implies by item 3 of Proposition 5.2.2 that

$$\vdash X_w \to [G \rhd (I \vee \bigvee (D_j \vee \Diamond D_i))]. \tag{5.7}$$

The formulas D_j are such that there exist formulas $C_j \in \mathcal{L}_{D_0} \cap \mathcal{L}_{E_0}$ for which

$$\vdash Y_w \to [(\bigvee C_j) \rhd A], \qquad \text{and} \tag{5.8}$$

$$\vdash X_w \to [(\bigvee D_j) \rhd (\bigvee C_j)].$$

Then also $\vdash X_w \to [(I \vee (\bigvee D_j)) \rhd (I \vee (\bigvee C_j))]$, hence by (5.7),

$$\vdash X_w \to [G \rhd (I \vee (\bigvee C_j))]. \tag{5.9}$$

From (5.8) and (5.6) it follows that

$$\vdash Y_w \to [(I \vee (\bigvee C_j)) \rhd A]. \tag{5.10}$$

By definition of A-criticality, (5.9) and (5.10) imply that $\neg G \in X_{w'}$, for every A-critical successor w' of w. But $wR_A u$, and $G \in X_u$. Contradiction. We conclude that the pair $\langle X^-, Y^- \rangle$ is inseparable. Hence, we can extend it to a complete pair $\langle X_v, Y_v \rangle$. The element $v \stackrel{\text{def}}{=} w * [(\langle X_v, Y_v \rangle, \tau_w * [A])]$ has all the required properties. This finishes the proof of Case 2, and hereby completes the proof of Lemma 5.3.5. ∎

We have just proven the two auxiliary lemmas 5.3.4 and 5.3.5. As we saw, these imply the truth lemma which in its turn implies Theorem 5.3.1. Therefore, we have finished the proof of Theorem 5.3.1. ∎

Theorem 5.3.1 has several interesting corollaries. First of all, via a standard argument (as e.g., in the proof of Theorem 2.3.10), it entails the Beth property for IL formulated in Definition 5.6.2. This theorem itself has some important consequences, as we will see in section 5.6. Moreover, in the next section we will derive the interpolation theorem for the system ILP from Theorem 5.3.1. We finish this section with another, simple, corollary which is an interpolation theorem for IL with regard to the $\rhd$-modality. For, if you think about it, the $\rhd$-modality can be viewed as a conditional. The following proposition specifies its relation with the material implication '$\to$'.

5.3.6. Proposition *$\vdash_{\mathsf{IL}} D \rhd E$ if and only if $\vdash_{\mathsf{IL}} D \to E \vee \Diamond E$.*

Proof: The direction "$\Leftarrow$" follows from item 2 in Proposition 5.2.2. For "$\to$", assume $\nvdash_{\mathsf{IL}} D \to E \vee \Diamond E$. By completeness there exists an IL-model $\langle\langle W, R, S\rangle, V\rangle$ and some world $w_1 \in W$ such that $w_1 \models D$ and $w_1 \not\models E \vee \Diamond E$. Let $W' \stackrel{\text{def}}{=} \{w \in W : w_1 R w\} \cup \{w_1, w_0\}$, where w_0 is some fresh element. By R' we denote the transitive closure of $(R_{\upharpoonright(W' \setminus \{w_0\})} \cup \langle w_0, w_1 \rangle)$. Here, by $R_{\upharpoonright(W' \setminus \{w_0\})}$ we understand the restriction of the relation R to the set $W' \setminus \{w_0\}$. Let S'_{w_0} be the reflexive closure of $R_{\upharpoonright(W' \setminus \{w_0\})}$, and $S'_w = S_w$, for $w \in W' \setminus \{w_0\}$. The so obtained $\langle\langle W', R', S'\rangle, V'\rangle$, where V' is any valuation extending V, is an IL-model. Moreover, w_1 is an R'-successor of w_0 satisfying D without a S'_{w_0}-successor satisfying E. In other words, $w_0 \not\models D \rhd E$. ∎

If we think of the $\rhd$-modality as a conditional, then the following interpolation property suggests itself. The proof of Corollary 5.3.7 follows immediately from Proposition 5.3.6 and Theorem 5.3.1.

5.3.7. Corollary ($\rhd$-Interpolation for IL) *Let D_0, E_0 be IL-formulas. Assume $\vdash_{\mathsf{IL}} D_0 \rhd E_0$. Then there exists an IL-formula $I \in \mathcal{L}_{D_0} \cap \mathcal{L}_{E_0}$ such that $\vdash_{\mathsf{IL}} D_0 \rhd I$ and $\vdash_{\mathsf{IL}} I \rhd E_0$.*

5.4 Corollary: interpolation for ILP

The system ILP is defined by adding to IL the *persistence principle*, $P : A \rhd B \to \Box(A \rhd B)$ (i.e., if $T+B$ is relatively interpretable in $T+A$, then this can be proved in T). A direct proof of interpolation for ILP can be obtained using the techniques introduced in the previous section. More elegantly, the question of interpolation for ILP can be reduced to a corollary of Theorem 5.3.1 by observing, as was done in [Hájek, 1992], that ILP is *strongly interpretable* in IL.

5.4.1. Definition [**Strong interpretation of ILP in IL**] We define the translation $^{\#}$ for a formula A in ILP as follows: for A atomic, $A^{\#}$ is A, $^{\#}$ commutes with Boolean connectives and with $\Box$ and $(B \rhd C)^{\#}$ is $(B^{\#} \rhd C^{\#}) \wedge \Box(B^{\#} \rhd C^{\#})$. $\dashv$

Given the P-axiom it is immediate that $\vdash_{\mathsf{ILP}} A \leftrightarrow A^{\#}$.

5.4.2. Proposition $\vdash_{\mathsf{IL}} A^{\#}$ *if and only if* $\vdash_{\mathsf{ILP}} A$*.*

Proof: The direction form left to right is trivial. The other direction is proved by induction on the length of the proof in ILP. The core of the proof consists of establishing that the translation of all the axioms of ILP are theorems of IL. ■

5.4.3. Theorem (CIP$^{\to}$ **for ILP**) *Let* D_0, E_0 *be* ILP*-formulas. Assume* $D_0 \to E_0$ *is valid in* ILP*. Then there exists an* ILP*- formula* $I \in \mathcal{L}_{D_0} \cap \mathcal{L}_{E_0}$ *such that* $\vdash_{\mathsf{ILP}} D_0 \to I$ *and* $\vdash_{\mathsf{ILP}} I \to E_0$*.*

The proof below is due to Hájek.

Proof: We reduce interpolation for ILP to interpolation for IL. Assume $\vdash_{\mathsf{ILP}} E_0 \to D_0$. Then, by Proposition 5.4.2, $\vdash_{\mathsf{IL}} E_0^{\#} \to D_0^{\#}$. Applying the interpolation result for IL, i.e., Theorem 5.3.1, we obtain an interpolant I such that $\vdash_{\mathsf{IL}} E_0^{\#} \to I$ and $\vdash_{\mathsf{IL}} I \to D_0^{\#}$. Obviously, $\vdash_{\mathsf{ILP}} E_0^{\#} \to I$ and $\vdash_{\mathsf{ILP}} I \to D_0^{\#}$. As $\vdash_{\mathsf{ILP}} A^{\#} \leftrightarrow A$, it follows that $\vdash_{\mathsf{ILP}} E_0 \to I$ and $\vdash_{\mathsf{ILP}} I \to D_0$. Note that I is in the common language of E_0, D_0, since the translation $^{\#}$ does not alter languages. Therefore, I is an interpolant for $\vdash_{\mathsf{ILP}} E_0 \to D_0$ ■

Reasoning as we did for IL it is straightforward to prove the following corollary.

5.4.4. Corollary ILP *has* $\rhd$*-interpolation.*

5.5 Failure of interpolation in ILW

We finish the part of this chapter on interpolation with a negative result: ILW, the system obtained by extending IL with the axiom $W : A \rhd B \to A \rhd (B \wedge \Box\neg A)$, does not have interpolation. To establish this failure we should exhibit a pair of formulas D, E such that $\vdash_{\mathsf{ILW}} D \to E$ whereas no interpolant exists for D and E. We propose the following implication

$$D \to E \stackrel{\text{def}}{=} (\Box(s \leftrightarrow \Box\neg p) \wedge (p \rhd q)) \to (q \rhd r \to r \rhd (r \wedge s)).$$

5.5.1. Claim $\vdash_{\mathsf{ILW}} D \to E$.

Proof: That $D \to E$ is a theorem of ILW follows from $J2 : p \rhd q \to (q \rhd r \to p \rhd r)$ and (*) $\vdash_{\mathsf{ILW}} p \rhd r \to r \rhd (r \wedge \Box\neg p)$. To prove this last theorem, reason as follows. By propositional logic, $\vdash_{\mathsf{ILW}} r \to ((r \wedge \Box\neg p) \vee (r \wedge \Diamond p))$ and with the aid of $J1$ we derive

$$\vdash_{\mathsf{ILW}} r \rhd ((r \wedge \Box\neg p) \vee (r \wedge \Diamond p)). \tag{5.11}$$

On the other hand, from $W : p \rhd r \to p \rhd (r \wedge \Box\neg p)$, by $J2$ and $J5$ we obtain $\vdash_{\mathsf{ILW}} p \rhd r \to \Diamond p \rhd (r \wedge \Box\neg p)$ and hence,

$$\vdash_{\mathsf{ILW}} p \rhd r \to (r \wedge \Diamond p) \rhd (r \wedge \Box\neg p). \tag{5.12}$$

(5.11) and (5.12) immediately imply (*). ■

What remains to be proven is that $D \to E$ does not have an interpolant in ILW. In our proof, the following notion of bisimulation, introduced in [Visser, 1990], is crucial.

5.5.2. Definition [P**-bisimulation**] Let P be a set of proposition letters. A *P-bisimulation* between the models $\mathcal{M} = \langle\langle W, R, S\rangle, V\rangle$ and $\mathcal{M}' = \langle\langle W', R', S'\rangle, V'\rangle$ is a nonempty relation $Z \subseteq W \times W'$ such that

atom $wZw' \to (w \in V(p)$ iff $w' \in V'(p))$, for all $p \in P$.

zig If wZw' and wRv, then there is a v' with vZv' and $w'R'v'$ and, for all u' with $v'S_{w'}u'$, there is an u with uZu' and vS_wu.

zag If wZw' and $w'R'v'$, then there is an v with vZv' and wRv and, for all u with vS_wu, there is an u' with uZu' and $v'S_{w'}u'$. ⊣

Recall that by $\mathcal{L}_P$ we denote the set of IL-formulas built up from proposition letters in the set P. The important result about P-bisimulations (cf. [Visser, 1990]) is that they preserve truth of formulas in $\mathcal{L}_P$.

5.5.3. Proposition *Let $\mathcal{M}$ and $\mathcal{M}'$ be two* IL*-models and Z a P-bisimulation between them. Then for any formula $A \in \mathcal{L}_P$, $wZw' \Rightarrow (\mathcal{M}, w \models A$ iff $\mathcal{M}', w' \models A)$.*

ILW-frames are IL-frames such that for each w, the composition $R \circ S_w$ is upwards well-founded. [de Jongh and Veltman, 1999] proves completeness for ILW with respect to finite ILW-models.

Consider the two ILW-models in Figure 5.1 and Figure 5.2. We use the following conventions. Worlds are labeled with the proposition letters which hold in them.

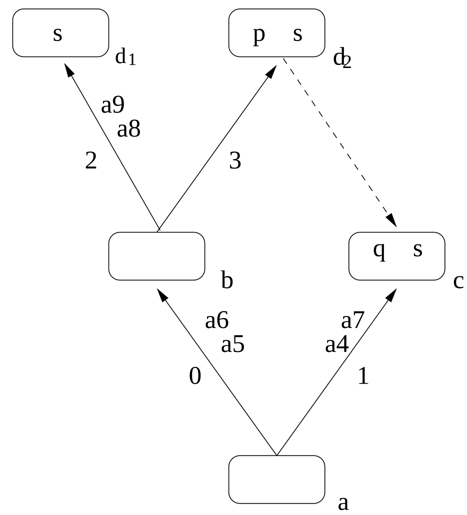


Figure 5.1: The model $\mathcal{M}$.

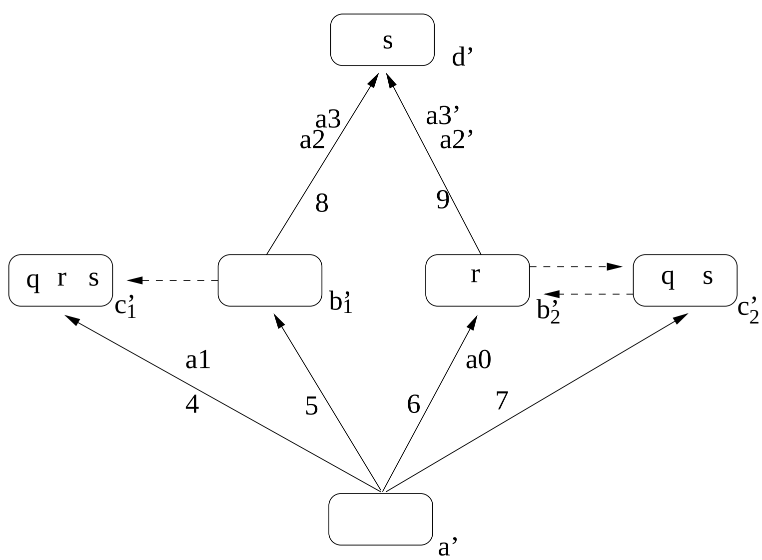


Figure 5.2: The model $\mathcal{M}'$.

Filled arrows stand for both R and S relations, while dashed arrows are S relations only. Whenever we have wRv, wRu and vSu then actually vS_wu. Finally, we should consider the transitive closure of the filled arrows and the reflexive-transitive closure of the dashed ones.

We claim that the relation linking pairs of worlds labeled by the same letter (disregarding subindices and $'$) is a $\{q,s\}$-bisimulation. Condition **atom** is easily checked. Verifying **zig** and **zag** requires more work.

To this aim, we point out that we can interpret the **zig** (and similarly the **zag**) condition on bisimulations as a rule in a game which requires that whenever wZw' and a 'move' wRv has been played, we should be able to answer with a 'counter-move' $w'R'v'$ which fulfills the necessary conditions on S and Z. We have labeled the arrows in the models according to this idea. An arrow marked n (for $n \in \{0,\dots,9\}$) is 'answered' by the arrow marked an in the other model. Note that some arrows are played twice because the bisimulation is not injective. For example, arrow 2 is answered by $a2$ when played from the position bZb'_1 and by $a2'$ when played from bZb'_2.

Once the fact that $\mathcal{M}$ and $\mathcal{M}'$ are $\{q,s\}$-bisimilar has been established, what rests is simple. Suppose $D \to E$ above has an interpolant I. Note that $\mathcal{M}, a \models D$. As $\vdash_{\mathsf{ILW}} D \to I$, we have $\mathcal{M}, a \models I$. But, as shown, there is a $\{q,s\}$-bisimulation linking a and a'. Hence by Proposition 5.5.3, $\mathcal{M}', a' \models I$. As $\vdash_{\mathsf{ILW}} I \to E$, we have $\mathcal{M}', a' \models E$, which is not the case. We conclude that no interpolant for $D \to E$ exists, as was to be proven.

In this section, we have established the following theorem.

5.5.4. Theorem *The interpretability logic* ILW *does not have the interpolation property* CIP$^{\to}$.

Finally, let us remark that our failure result is of a less general kind than the result in [Visser, 1998] mentioned in section 5.1. As the model in Figure 5.2 is not an ILM-model we cannot extend the failure result to all logics between ILW and ILM.

5.6 Beth definability and fixed points

The Beth property has been formulated in Definition 2.2.6. For technical purposes, which will become apparent in due course, we use a slightly different formulation in the present section.

5.6.1. Notation In this section we will, if useful, denote formulas in such a way that the proposition letters from which they are built up are displayed. For example, writing $\boldsymbol{p} = p_1,\dots,p_k$, the notation $A(\boldsymbol{p}, r)$ implies that the proposition letters that occur in the formula A are among $p_1,\dots,p_k, r$. Also, for any formula

A we write $\boxdot A$ to abbreviate $A \wedge \Box A$. Moreover, a formula A is said to be *modalized* in r if every occurrence of the proposition letter r in A is in the scope of a modality. ⊣

5.6.2. Definition [**Beth property for interpretability logics**] A system of interpretability logic S has the *Beth definability property* iff for all formulas $A(\boldsymbol{p}, r)$ the following holds:

$$\text{If } \vdash_S \boxdot A(\boldsymbol{p}, r) \wedge \boxdot A(\boldsymbol{p}, r') \rightarrow (r \leftrightarrow r'),$$

(in words, if $A(\boldsymbol{p}, r)$ *implicitly defines* r in terms of $\boldsymbol{p}$) then there exists a formula $C(\boldsymbol{p})$ (called an *explicit definition*) such that

$$\vdash_S \boxdot A(\boldsymbol{p}, r) \rightarrow (C(\boldsymbol{p}) \leftrightarrow r).$$

⊣

The main difference between the above formulation of the Beth property and the one in Definition 2.2.6 is that the above implicit definition $A(\boldsymbol{p}, r)$ is a single formula whereas the implicit definition $\Gamma(\boldsymbol{p}, r)$ in 2.2.6 may be a (possibly infinite) set of formulas. As IL is not compact, these two formulations really differ. Definition 5.6.2 formulates the definability property that is usually studied in provability logics, e.g., in [Smoryński, 1978] and [Maksimova, 1989].

Using the argument in the proof of Theorem 2.3.10, one easily derives the Beth definability property for IL from Theorem 5.3.1. But, as we will shortly see (cf. Corollary 5.6.10), we can infer much more. To this end, we will make a detour via fixed points.

5.6.3. Definition [**Fixed Point Property**] A logic S has the *fixed point property* iff for any formula $A(\boldsymbol{p}, r)$ which is modalized in r, there exists a formula $F(\boldsymbol{p})$ (called a *fixed point*) such that

$$\vdash_S F(\boldsymbol{p}) \leftrightarrow A(\boldsymbol{p}, F(\boldsymbol{p})) \text{ (existence), } \textit{and}$$
$$\vdash_S \boxdot(r \leftrightarrow A(\boldsymbol{p}, r)) \wedge \boxdot(r' \leftrightarrow A(\boldsymbol{p}, r')) \rightarrow r \leftrightarrow r' \text{ (uniqueness).}$$

⊣

Outline of this section First we will show in Theorem 5.6.4 that for a general class of logics the fixed point property can be derived from the Beth property. Second it will be proven in Theorem 5.6.7 that for these logics the Beth property is in its turn derivable from the fixed point property. Since any extension of IL is such a logic, we can reason as follows. As noted above, IL has the Beth property. Hence, by Theorem 5.6.4, IL has the fixed point property. The nature of the fixed point property is such that it is inherited by any extension. Via Theorem 5.6.7 we then reach the general conclusion that any extension of IL has the Beth property.

Let us finally note that Theorem 5.6.4 and Theorem 5.6.7 also apply to all extensions of the provability logic GL, hereby subsuming known results in that area (see [Smoryński, 1978] and [Maksimova, 1989]).

5.6.1 From Beth definability to fixed points

One of the well-known applications of the Beth definability property can be found in the literature on provability logic. In [Smoryński, 1978], C. Smoryński derives for the provability logic GL the existence of fixed points —the more interesting half of the fixed point theorem— from the uniqueness of fixed points via an application of the Beth property. Theorem 5.6.4 generalizes this result.

5.6.4. Theorem *Let S be a normal modal logic in which*

1. $\vdash_S \Box A \to \Box\Box A$,
2. $\vdash_S \boxdot B \to (\Box A \to A)$ *implies* $\vdash_S \boxdot B \to A$,
3. *the Beth theorem holds.*

Then S has the fixed point property.

Proof of Theorem 5.6.4: Let the logic S satisfy the conditions in the theorem, and let $A(\boldsymbol{p}, r)$ be an S-formula which is modalized in r. For brevity, let us write $A(r)$. As every occurrence of r in A is in the scope of a modality, we have

$$\vdash_S \Box(r \leftrightarrow r') \to (A(r) \leftrightarrow A(r')).$$

Hence,

$$\vdash_S \boxdot((r \leftrightarrow A(r)) \wedge (r' \leftrightarrow A(r'))) \to (\Box(r \leftrightarrow r') \to (r \leftrightarrow r')).$$

An application of the second condition on the logic S shows that fixed points of $A(r)$ are unique.

In order to construct a fixed point for this formula, we note that uniqueness of fixed points of $A(r)$ is equivalent to $A(r) \leftrightarrow r$ being an implicit definition of r in terms of $\boldsymbol{p}$. As S has the Beth property, this implies the existence of some formula C built up from propositional variables in $\boldsymbol{p}$ such that

$$\vdash_S \boxdot(A(r) \leftrightarrow r) \ \to \ (r \leftrightarrow C). \tag{5.13}$$

We will show that C is a fixed point for $A(r)$. We first substitute $A(C)$ for r in (5.13), yielding

$$\vdash_S \boxdot(A(A(C)) \leftrightarrow A(C)) \;\rightarrow\; (A(C) \leftrightarrow C). \tag{5.14}$$

Reasoning in the modal system $\mathsf{K4}$, we then infer that

$$\vdash_S \Box(A(A(C)) \leftrightarrow A(C)) \;\rightarrow\; \Box(A(C) \leftrightarrow C).$$

That is, $A(C)$ and C are equivalent under the $\Box$-operator, given $\Box(A(A(C)) \leftrightarrow A(C))$. As r is modalized in $A(r)$ this implies that

$$\vdash_S \Box(A(A(C)) \leftrightarrow A(C)) \;\rightarrow\; (A(A(C)) \leftrightarrow A(C)).$$

By the second condition on the logic S this suffices to conclude that

$$\vdash_S A(A(C)) \leftrightarrow A(C).$$

Hence, $\vdash_S \boxdot(A(A(C)) \leftrightarrow A(C))$. Recalling (5.14) we conclude that

$$\vdash_S A(C) \leftrightarrow C.$$

■

5.6.5. Remark Consider the following weakening of condition 2 in 5.6.4,

$$2'. \quad \vdash_S \Box A \rightarrow A \text{ implies } \vdash_S A.$$

We note that in the above proof the existence of fixed points is actually derived from conditions 1 and 3 together with this weakened version of condition 2. Theorem 5.6.4 could therefore be rephrased as saying that any normal modal logic $\mathcal{L}$ has the fixed point property if the following requirements are met: $\vdash_S \Box A \rightarrow \Box\Box A$, condition $2'$ holds, S has the Beth property, and fixed points in S are unique. ⊣

Let us verify that Theorem 5.6.4 is indeed a generalization of Smoryński's aforementioned result. Moreover, some more efforts will yield the fixed point theorem for IL, a direct proof of which was already given in [de Jongh and Visser, 1991].

5.6.6. Corollary *Let S be an extension of GL, or an extension of IL. Then S has the fixed point property.*

Proof: We will check that GL and IL satisfy the conditions in Theorem 5.6.4. The first condition needs no comment. The Beth theorem for GL is proven in [Smoryński, 1978]. As we noted before, the Beth theorem for IL can be derived from Theorem 5.3.1 as usual. With regard to the second condition, we note that in any logic S which satisfies the provability axioms (cf. $L1$–$L4$ in Definition 5.2.1), $\vdash_S \boxdot B \to \Box A$ can be inferred from $\vdash_S \boxdot B \to (\Box A \to A)$. An application of modus ponens yields condition 2. We conclude from Theorem 5.6.4 that GL and IL have the fixed point property. This obviously implies that all extensions of L and IL have the fixed point property. ∎

5.6.2 From fixed points to Beth definability

Another angle on the Beth property and fixed points has first been taken in [Maksimova, 1989] where it was shown that for provability logics the fixed point property in its turn implies the Beth property. In what follows, we will generalize this result.

5.6.7. Theorem *Let S be a normal modal logic in which*

1. $\vdash_S \Box A \to \Box\Box A$,
2. $\vdash_S \boxdot B \to (\Box A \to A)$ *implies* $\vdash_S \boxdot B \to A$,
3. *the fixed point theorem holds.*

Then S has the Beth property.

A first difficulty that arises in proving the Beth theorem from the fixed point theorem, is the more general character of the former. For, the fixed point theorem that is at our disposal is a statement about *modalized* formulas, whereas the Beth theorem is about *arbitrary* formulas. The next lemma, from [Maksimova, 1989], reduces arbitrary formulas to ones which are 'largely modalized', and thereby provides a starting point for proving the Beth theorem from the fixed point theorem.

5.6.8. Lemma *Let S be a normal modal logic, and let $A(\boldsymbol{p}, r)$ be an arbitrary S-formula. Then there exist S-formulas $A_1(\boldsymbol{p}, r)$, $A_2(\boldsymbol{p}, r)$ which are modalized in r such that*

$$\vdash_S A(\boldsymbol{p}, r) \leftrightarrow [(r \wedge A_1(\boldsymbol{p}, r)) \vee (\neg r \wedge A_2(\boldsymbol{p}, r))].$$

This observation rests on some syntactic considerations: writing an arbitrary formula in disjunctive normal form and collecting the disjuncts containing r and the ones containing $\neg r$ will give the form required by Lemma 5.6.8.

Proof of Theorem 5.6.7: Let the logic S satisfy the conditions in the theorem. Consider an implicit S-definition $A(\boldsymbol{p}, r)$ of r in terms of $\boldsymbol{p}$. Abbreviating $A(\boldsymbol{p}, r)$ to $A(r)$, this can be expressed by

$$\vdash_S \boxdot A(r) \wedge \boxdot A(r') \rightarrow (r \leftrightarrow r'). \tag{5.15}$$

Let us gather some facts. By the previous lemma, there exist formulas $A_1(r)$, $A_2(r)$ which are modalized in r such that

$$\vdash_S A(r) \leftrightarrow [(r \wedge A_1(r)) \vee (\neg r \wedge A_2(r))]. \tag{5.16}$$

As S has the fixed point property, there exists a formula F_1 built up from propositional variables in $\boldsymbol{p}$ which is a fixed point of $A_1(r)$, i.e.,

$$\vdash_S F_1 \leftrightarrow A_1(F_1). \tag{5.17}$$

Moreover, fixed points are unique. Hence,

$$\vdash_S \boxdot(r \leftrightarrow A_1(r)) \rightarrow (r \leftrightarrow F_1). \tag{5.18}$$

Our aim is to show the following claim.

5.6.9. Claim $\vdash_S \boxdot A(r) \rightarrow [\Box(A_1(r) \rightarrow r) \rightarrow (A_1(r) \rightarrow r)]$.

From this claim it follows by the second condition on the logic S that

$$\vdash_S \boxdot A(r) \rightarrow (A_1(r) \rightarrow r). \tag{5.19}$$

On the other hand, from (5.16) it is obvious that $\vdash_S A(r) \rightarrow (r \rightarrow A_1(r))$. Hence from (5.19) we conclude that $\vdash_S \boxdot A(r) \rightarrow (r \leftrightarrow A_1(r))$, and therefore,

$$\vdash_S \boxdot A(r) \rightarrow \boxdot (r \leftrightarrow A_1(r)).$$

From the uniqueness of fixed points (see (5.18) above), it then follows that

$$\vdash_S \boxdot A(r) \rightarrow (r \leftrightarrow F_1).$$

Ergo, F_1 is an explicit definition of r. What remains is to prove Claim 5.6.9.

Proof of Claim 5.6.9: From (5.16) we obtain that $\vdash_S A(r) \rightarrow (r \rightarrow A_1(r))$, and hence, $\vdash \Box A(r) \rightarrow \Box(r \rightarrow A_1(r))$. Therefore,

$$\vdash_S \boxdot A(r) \wedge \Box(A_1(r) \to r) \to \Box(r \leftrightarrow A_1(r)). \tag{5.20}$$

For notational convenience, let us denote the formula $\boxdot A(r) \wedge \Box(A_1(r) \to r)$ by C. So (5.20) amounts to

$$\vdash_S C \to \Box(r \leftrightarrow A_1(r)). \tag{5.21}$$

From (5.18) it follows that $\vdash_S \Box(r \leftrightarrow A_1(r)) \to \Box(r \leftrightarrow F_1)$ Hence, by (5.21)

$$\vdash_S C \to \Box(r \leftrightarrow F_1).$$

In other words, r and F_1 are equivalent under the $\Box$-operator (relative to C). In particular,

$$\vdash_S C \to \Box(A(F_1)) \quad \text{and} \tag{5.22}$$
$$\vdash_S C \to (A_1(r) \to A_1(F_1)), \tag{5.23}$$

where (5.23) holds by virtue of A_1 being modalized in r, and (5.22) by definition of C. Let us note for future reference that from (5.23) and the fact that F_1 is a fixed point for A_1 (cf. (5.17)) it follows that

$$\vdash_S C \to (A_1(r) \to F_1), \tag{5.24}$$

and $\vdash_S C \to [A_1(r) \to (F_1 \wedge A_1(F_1))]$. By (5.16), this latter implication shows that $\vdash_S C \to (A_1(r) \to A(F_1))$ which together with (5.22) implies

$$\vdash_S C \to (A_1(r) \to \boxdot A(F_1)). \tag{5.25}$$

Recall from (5.15) that $A(r)$ is an implicit definition of r. In particular, $\vdash_S \boxdot A(r) \wedge \boxdot A(F_1) \to (r \leftrightarrow F_1)$. From (5.25) we then derive that

$$\vdash_S C \to (A_1(r) \to (r \leftrightarrow F_1)).$$

By (5.24), we obtain the claim. ■

This completes the proof of Theorem 5.6.7 ■

In the proof of Corollary 5.6.6 it has already been shown that all extensions of the basic system of provability logic GL and all extensions of the basic system of interpretability logic IL satisfy conditions 1–2 in Theorem 5.6.7. Hence, from Theorem 5.6.7 and Corollary 5.6.6 we obtain the following result.

5.6.10. Corollary *Let S be an extension of* GL, *or an extension of* IL. *Then S has the Beth property.*

This corollary reveals a striking contrast between interpolation and definability properties in the context of interpretability logics. For example, as was mentioned in the introduction, all systems between ILM_0 and ILM lack interpolation. Or, as was shown in section 5.5, ILW does not have this property either. On the other hand, by Corollary 5.6.10 they all have the Beth property.

Appendix A

Appendix

This appendix lists the notions and terminology that we used, but did not explain, in the thesis. For more information, the reader is advised to consult any of the following sources.

- [Chang and Keisler, 1990] for classical logic.
- [Blackburn et al., 2001] for modal logic.
- [Burris and Sankappanavar, 1981] for universal algebra.
- [Balbes and Dwinger, 1974] for lattice theory.
- [Blok and Pigozzi, 1989] and [Andréka et al., 1994] for algebraic logic.

We assume the reader is familiar with the basic concepts of first order logic. Below, F_{fol} denotes the set of first order formulas over some fixed, but arbitrary, first order language $\mathcal{L}$. M_{fol} is the class of first order models of the corresponding similarity type. The relation $\models_{fol}$ denotes the usual notion of truth in a first order model.

Logic In this appendix, we view a *logic* as a triple $S = \langle F_S, M_S, \models_S \rangle$, where F_S is a set of *formulas*, M_S is a class of *models* and $\models_S \subseteq M_S \times \mathcal{P}(F_S)$. Usually, the formulas in F_S are recursively defined from some set of *atomic formulas* by means of some fixed set $Cn(S)$ of *logical connectives*. In case $Cn(S)$ contains conjunction (resp. implication), the logic S is said to be *conjunctive* (resp. *implicative*).

For example, first order logic on finite structures is the triple $\langle F_{fol}, M_{fin}, \models_{fol} \rangle$, where M_{fin} is the class of finite first order models. Other examples are discussed below.

With any logic S we associate a *semantic consequence relation* $\models_S \subseteq \mathcal{P}(F_S) \times F_S$ in the usual way.[1] That is, for $\Gamma \cup \{\varphi\} \subseteq F_S$,

[1] Note that, as is the standard practice, the symbol $\models_S$ denotes both the semantic consequence relation and the relation of a formula being true in a model.

$$\Gamma \models_S \varphi \text{ iff for every } \mathcal{M} \in M_S \text{ [if } \mathcal{M} \models_S \Gamma, \text{ then } \mathcal{M} \models_S \varphi].$$

For modal logics, at least two semantic consequence relations are considered in the literature. These are discussed below, when we have formally introduced the notion of a normal modal logic.

A.0.3 Normal modal logics

Similarity type A *modal similarity type* is a set t of modal operators. Each operator in t has a given finite *rank*. The similarity types considered in this chapter only contain unary modalities. Therefore, we will restrict ourselves in the following discussion to such types.

Modal formulas Fix a set of propositional variables. Given a similarity type t, the set of *t-formulas* F_t is recursively defined as follows:

1. Propositional variables are t-formulas.
2. If $\varphi, \psi \in F_t$, then also $\neg\varphi, \varphi \wedge \psi \in F_t$.
3. If $\varphi \in F_t$ and $\Diamond$ is an (unary) operator in t, then $\Diamond\varphi \in F_t$.

We write $\Box\varphi \stackrel{\text{def}}{=} \neg\Diamond\neg\varphi$. We sometimes say that type t contains the operator $\Box$, when properly speaking t contains the dual operator $\Diamond$. Also, we use the usual abbreviations $\varphi \vee \psi$, $\varphi \rightarrow \psi$ and $\varphi \leftrightarrow \psi$.

Normal modal logic, syntactic definition Usually, the notion of a *normal modal logic* is defined syntactically as follows. A normal modal logic S of type t is a set of modal formulas of type t that contains

1. all classical tautologies,
2. the axioms $\Box(p \rightarrow q) \rightarrow (\Box p \rightarrow \Box q)$, for all $\Box \in t$,

and that is closed under the rules of modus ponens, necessitation and substitution. That is,

1. If $\varphi \in S$ and $\varphi \rightarrow \psi \in S$, then $\psi \in S$.
2. If $\varphi \in S$, and $\Box \in t$, then $\Box\varphi \in S$.
3. If $\varphi \in S$, and σ is a substitution, then $\sigma\varphi \in S$.

Below, we give a semantic description of this notion.

Frames Let $t = \{\Diamond_i \ : \ i \in I\}$ be a modal similarity type. A *t-frame* is a tuple $\mathcal{F} = \langle W, R_i \rangle_{i \in I}$, where W is a non-empty set and R_i is a binary relation on W, for all $i \in I$. W is called the *universe* of $\mathcal{F}$ and R_i is the *accessibility relation* corresponding to the modal operator $\Diamond_i$.

Models A *t-model* is a pair $\mathcal{M} = \langle \mathcal{F}, V \rangle$, where $\mathcal{F}$ is a t-frame and V is a valuation on the universe of $\mathcal{F}$. That is, V is a function that maps propositional variables to subsets of the universe of $\mathcal{F}$. In this case, we say that $\mathcal{M}$ is *based* on the frame $\mathcal{F}$.

Truth and validity Consider the t-frame $\mathcal{F} = \langle W, R_i \rangle_{i \in I}$ and the model $\mathcal{M} = \langle \mathcal{F}, V \rangle$. Let $w \in W$. The notion of a t-formula φ being *true in* $\mathcal{M}$ *at state* w (notation: $\mathcal{M}, w \models \varphi$) is inductively defined as follows:

1. $\mathcal{M}, w \models p$ iff $w \in V(p)$, for all propositional variables p.

2. $\mathcal{M}, w \models \neg\varphi$ iff $\mathcal{M}, w \not\models \varphi$.

3. $\mathcal{M}, w \models \varphi \wedge \psi$ iff $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$.

4. $\mathcal{M}, w \models \Diamond_i \varphi$ iff there exists some $v \in W$ such that $R_i wv$ and $\mathcal{M}, v \models \varphi$.

The formula φ is *true in the model* $\mathcal{M}$ (notation: $\mathcal{M} \models \varphi$) if $\mathcal{M}, w \models \varphi$, for all $w \in W$. The formula φ is *valid on the frame* $\mathcal{F}$ (notation: $\mathcal{F} \models \varphi$) if $\langle \mathcal{F}, V \rangle \models \varphi$, for all valuations V.

Global versus local consequence For modal logics, two possible consequence relations suggest itself, a *global* and a *local* one. These are defined as follows. Given t-formulas φ, ψ and a class of t-models M_S,

$$\varphi \models_S^{glob} \psi \Leftrightarrow \forall \mathcal{M} \in M_S \text{ [if } \mathcal{M} \models \varphi \text{ then } \mathcal{M} \models \psi].$$

$$\varphi \models_S^{loc} \psi \Leftrightarrow \forall \mathcal{M} = \langle \langle W, R_i \rangle_{i \in I}, V \rangle \in M_S, \forall w \in W \text{ [if } \mathcal{M}, w \models \varphi \text{ then } \mathcal{M}, w \models \psi.]$$

Note that the global and the local version of a modal logic S have the same set of theorems. That is, for any formula φ,

$$\models_S^{loc} \varphi \Leftrightarrow \models_S^{glob} \varphi. \tag{A.1}$$

General frames Let $\mathcal{F} = \langle W, R_i \rangle_{i \in I}$ be a t-frame. The unary operation m_i on the powerset $\mathcal{P}(W)$ of W, is defined by

$$m_i(X) = \{w \ : \ \text{there exists some } v \in X \text{ such that } R_i wv\}.$$

A *general frame* is a pair $\mathbb{F} = \langle \mathcal{F}, A \rangle$, where $\mathcal{F}$ is a t-frame and $A \subseteq \mathcal{P}(W)$ is closed under the Boolean operations and the operations m_i, for all R_i. A valuation V is called *admissible* for $\mathbb{F}$ if for each propositional variable p, $V(p) \in A$. A *model based on a general frame* $\mathbb{F}$ is a pair $\langle \mathbb{F}, V \rangle$, where V is an admissible valuation for $\mathbb{F}$.

Normal modal logic, semantic definition By a *normal modal logic S* of type t we understand the triple $\langle F_t, M_S, \models_S^{glob} \rangle$, where F_t is the set of modal formulas of type t and M_S is the class of all models that are based on a general frame $\mathbb{F} \in \mathit{Gfr}_S$, for some given class Gfr_S of general frames of type t. Note that, if not explicitly stated otherwise, with any normal modal logic we always associate the *global* consequence relation.

A.0.11. Remark In the concrete cases we consider in chapter 2, the modal logics are *frame complete*. That is, there exists some class of frames Fr such that M_S is the class of all models that are based on a frame in Fr. For example, M_{K4} is the class of all models that are based on a transitive frame. Another example is the logic S defined in Example 2.2.11, where M_S is the set of all models that are based on the frame $\mathcal{F}$. In the latter case we say that S is *the logic of the frame* $\mathcal{F}$. However, in general a normal modal logic need not be frame complete. On the other hand, we do have a fundamental completeness result for normal modal logics (seen as syntactic objects) and general frames. This implies that the above semantic definition of a normal modal logic in terms of general frames coincides with the earlier, and more familiar, syntactic definition. ⊣

Deduction property Define $\boxdot^0 \varphi = \varphi$, and $\boxdot^{n+1}\varphi = \varphi \wedge \Box \boxdot^n \varphi$. That is, $\boxdot^n \varphi = \varphi \wedge \Box\varphi \wedge \Box\Box\varphi \wedge \cdots \wedge \underbrace{\Box \cdots \Box}_{n \text{ times}} \varphi$. We note that any normal modal logic S has the following local deduction property. For $\varphi, \psi \in F_S$,

$$\varphi \models_S^{glob} \psi \Leftrightarrow \text{ there exists some } n \in \omega \text{ such that } \models_S^{glob} \boxdot^n \varphi \to \psi. \tag{A.2}$$

For any extension of $\mathsf{K4}$ this reduces to

$$\varphi \models_S^{glob} \psi \Leftrightarrow \models_S^{glob} (\varphi \wedge \Box\varphi) \to \psi.$$

For extensions of $\mathsf{S5}$ we even have

$$\varphi \models_S^{glob} \psi \Leftrightarrow \models_S^{glob} \Box\varphi \to \psi.$$

Moreover, the local consequence relation has the following deduction theorem.

$$\varphi \models_S^{loc} \psi \Leftrightarrow \models_S^{loc} \varphi \to \psi. \tag{A.3}$$

Global versus local consequence, revisited In chapter 2 we associate with any modal logic the global consequence relation. However, this chapter also contains results on modal logics with a local consequence relation. To see this, we note the following equivalences. For any modal logic S, the relation $\models_S^{loc}$ has CIP$^{\models}$ iff $\models_S^{loc}$ has CIP$^{\rightarrow}$ (by (A.3)) iff $\models_S^{glob}$ has CIP$^{\rightarrow}$ (by (A.1)). Similarly, for any compact modal logic S, the relation $\models_S^{loc}$ has the Beth property iff $\models_S^{loc}$ has the $\rightarrow$-Beth property (this notion has been introduced on page 33), iff $\models_S^{glob}$ has the $\rightarrow$-Beth property. By [Maksimova, 1992b, Theorem 1], this is equivalent to the fact that $\models_S^{glob}$ has CIP$^{\rightarrow}$.

Some important modal systems Below we list some important modal axioms together with the names we used for them. The basic system of modal logic is denoted by K. Usually, the system K + T + 4 is abbreviated to S4, and the system K + T + 4 +5 to S5.

Logic	**axiom**
4	$\Box p \rightarrow \Box\Box p$
5	$\Diamond p \rightarrow \Box\Diamond p$
T	$\Box p \rightarrow p$
D	$\Diamond\top$
B	$p \rightarrow \Box\Diamond p$
GL	$\Box(\Box p \rightarrow p) \rightarrow \Box p$
Grz	$\Box(\Box(p \rightarrow \Box p) \rightarrow p) \rightarrow p$
3	$(\Diamond p \wedge \Diamond q) \rightarrow (\Diamond(p \wedge \Diamond q) \vee \Diamond(q \wedge \Diamond p) \vee \Diamond(p \wedge q)$
M	$\Box\Diamond p \rightarrow \Diamond\Box p$

A.0.4 Some other logics

Intermediate logics We assume that the reader is familiar with classical propositional logic (CPC) and intuitionistic propositional logic (IPC). An *intermediate logic* S is usually syntactically defined as a set of formulas that

1. contains all the theorems of IPC,
2. is contained in the set of theorems of CPC, and
3. is closed under the rule of substitution.

Along the same lines as for normal modal logics, a semantic rendering of this definition can be given. We will not do this.

We remark that in the literature also the name *superintuitionistic* logics appears.

Finite variable fragments Fix countably many variables $\{v_i : i \in \omega\}$, and let $k \in \omega$. By $\mathsf{L}^k = \langle F_{\mathsf{L}^k}, M_{fol}, \models_{fol}\rangle$ we denote the *k-variable fragment of first order logic*, where F_{L^k} is the set of formulas from F_{fol} which only contain the variables $\{v_1, \ldots, v_k\}$.

Infinitary logics The infinitary logic $\mathsf{L}_\infty = \langle F_{\mathsf{L}_\infty}, M_{fol}, \models_{fol}\rangle$ is an extension of first order logic where infinite conjunctions and disjunctions are allowed. That is, F_{L_∞} is the closure of F_{fol} under the following condition:

$$\text{If } \Phi \subseteq F_{\mathsf{L}_\infty}, \text{ then } \bigwedge \Phi \in F_{\mathsf{L}_\infty} \text{ and } \bigvee \Phi \in F_{\mathsf{L}_\infty}. \tag{A.4}$$

For any $\alpha \geq \omega$, the logic L_α is defined similar to L_∞ with condition (A.4) restricted to sets Φ of size less then α. In particular, L_ω is usual first order logic.

In the infinitary logics considered in this chapter, only formulas of *finite* quantifier depth are permitted (unlike some other logics one may find in the literature). Sometimes this is reflected in the name of the logic, by adding an extra 'ω' at the bottom. That is, our L_∞ (resp. L_α) is sometimes denoted by $\mathsf{L}_{\infty\omega}$ (resp. $\mathsf{L}_{\alpha\omega}$).

Bibliography

[Anderson and Belnap, 1975] Anderson, A. R. and Belnap, N. D. (1975). *Entailment. The logic of relevance and necessity.* Princeton University Press.

[Andréka et al., 1982] Andréka, H., Comer, S., and Németi, I. (1982). Epimorphisms of finite dimensional cylindric algebras need not be surjective. Unpublished.

[Andréka et al., 1995] Andréka, H., Düntsch, I., and Németi, I. (1995). Expressibility of properties of relations. *J. Symbolic Logic*, 60:970–991.

[Andréka et al., 1999] Andréka, H., Hodkinson, I., and Németi, I. (1999). Finite algebras of relations are representable on finite sets. *J. Symbolic Logic*, 64:243–267.

[Andréka et al., 1994] Andréka, H., Kurucz, A., Németi, I., and Sain, I. (1994). Applying algebraic logic to logic. In Nivat, M., Rus, T., and Scollo, G., editors, *Algebraic Methodology and Software Technology*, pages 3–26. Springer-Verlag. Extended version appeared as *Applying Algebraic Logic: A General Methodology,* in the Proc. of the Summer School of Algebraic Logic, 1994.

[Andréka et al., 1995] Andréka, H., Kurucz, A., Németi, I., and Sain, I. (1995). General algebraic logic including algebraic model theory: an overview. In *Logic Colloquium '92 (Veszprém, 1992)*, pages 1–60. CSLI Publications., Stanford, CA.

[Andréka and Németi, 1994] Andréka, H. and Németi, I. (1994). General algebraic logic: A perspective on 'what is logic'. In Gabbay, D., editor, *What is a logical system?*, pages 485–569. Clarendon Press.

[Andréka and Németi, 1996] Andréka, H. and Németi, I. (1996). Craig interpolation does not imply amalgamation after all. Manuscript, Math. Inst. Hungar. Acad. Sci., Budapest.

[Andréka et al., 1998] Andréka, H., Németi, I., and van Benthem, J. (1998). Modal languages and bounded fragments of predicate logic. *J. Philosophical Logic*, 27(3):217–274.

[Areces et al., 2000a] Areces, C., Blackburn, P., and Marx, M. (2000a). Repairing the interpolation lemma in quantified modal logic. Unpublished.

[Areces et al., 2000b] Areces, C., de Jongh, D., and Hoogland, E. (2000b). Interpolation, definability and fixed points in interpretability logics. In Zakharyaschev, M., Segerberg, K., de Rijke, M., and Wansing, H., editors, *Advances in Modal Logic, Volume 2*, Lecture Notes, pages 53–76, Stanford, CA. CSLI Publications.

[Bacsich, 1975] Bacsich, P. (1975). Amalgamation properties and interpolation theorems for equational theories. *Algebra Universalis*, 5:45–55.

[Balbes and Dwinger, 1974] Balbes, R. and Dwinger, P. (1974). *Distributive lattices*. University of Missouri Press, Columbia, Mo.

[Barwise and Feferman, 1985] Barwise, J. and Feferman, S., editors (1985). *Model-Theoretic Logics*. Springer-Verlag.

[Barwise and van Benthem, 1999] Barwise, J. and van Benthem, J. (1999). Interpolation, preservation, and pebble games. *J. Symbolic Logic*, 64(2):881–903.

[Beth, 1953] Beth, E. W. (1953). On Padoa's method in the theory of definition. *Nederl. Akad. Wetensch. Proc. Ser. A.* **56** = *Indagationes Math.*, 15:330–339.

[Blackburn et al., 2001] Blackburn, P., de Rijke, M., and Venema, Y. (2001). *Modal logic*. Cambridge University Press. Forthcoming.

[Blok and Hoogland, 2001] Blok, W. J. and Hoogland, E. (2001). The Beth property in equivalential logics. Technical report, Institute for Logic, Language and Computation, University of Amsterdam. To appear.

[Blok and Pigozzi, 1986] Blok, W. J. and Pigozzi, D. (1986). Protoalgebraic logics. *Studia Logica*, 45(4):337–369.

[Blok and Pigozzi, 1989] Blok, W. J. and Pigozzi, D. (1989). Algebraizable logics. *Memoirs of the American Mathematical Society*, 77, 396:vi+78 pp.

[Blok and Pigozzi, 1991] Blok, W. J. and Pigozzi, D. (1991). Local deduction theorems in algebraic logic. In Andréka, H., Monk, J. D., and Németi, I., editors, *Algebraic logic*, volume 54 of *Colloq. Math. Soc. János Bolyai*, pages 75–109. North-Holland, Amsterdam.

[Blok and Pigozzi, 1992] Blok, W. J. and Pigozzi, D. (1992). Algebraic semantics for universal Horn logic without equality. In Smith, J. and Romanowska, A., editors, *Universal algebra and quasigroup theory (Proc. Conf. Jadwisin, Poland, May 23–28, 1989)*, volume 19 of *Research and Exposition in Mathematics*, pages 1–56, Berlin. Heldermann Verlag.

[Blok and Pigozzi, 2001] Blok, W. J. and Pigozzi, D. (2001). Abstract algebraic logic and the deduction theorem. *Bull. Symbolic Logic*. To appear.

[Boolos, 1980] Boolos, G. (1980). On systems of modal logic with provability interpretations. *Theoria*, 46(1):7–18.

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

[Brignole and Monteiro, 1967] Brignole, D. and Monteiro, A. (1967). Caractérisation des algèbres de Nelson par des égalités. I, II. *Proc. Japan Acad.*, 43:279–283; 284–285.

[Bruns and Harding, 1999] Bruns, G. and Harding, J. (1999). Amalgamation of ortholattices. Unpublished.

[Burgess, 1965] Burgess, W. (1965). The meaning of mono and epi in some familiar categories. *Canad. Math. Bull.*, 8:759–769.

[Burris and Sankappanavar, 1981] Burris, S. and Sankappanavar, H. P. (1981). *A course in universal algebra*. Springer-Verlag, New York.

[Chang, 1958] Chang, C. C. (1958). Algebraic analysis of many valued logics. *Trans. Amer. Math. Soc.*, 88:467–490.

[Chang and Keisler, 1990] Chang, C. C. and Keisler, H. (1990). *Model theory*, volume 73 of *Studies in Logic and the Foundations of Mathematics*. North-Holland Publishing Co., Amsterdam, third edition.

[Comer, 1969] Comer, S. (1969). Classes without the amalgamation property. *Pacific J. Math.*, 28:309–318.

[Cornish, 1982] Cornish, W. (1982). On Iseki's BCK-algebras. *Lecture notes in pure and applied math.*, 74:101–122.

[Craig, 1957] Craig, W. (1957). Three uses of the Herbrand-Gentzen theorem in relating model theory and proof theory. *J. Symbolic Logic*, 22:269–285.

[Czelakowski, 1981] Czelakowski, J. (1981). Equivalential logics, I and II. *Studia Logica*, 40:227–236, 355–372.

[Czelakowski, 1982] Czelakowski, J. (1982). Logical matrices and the amalgamation property. *Studia Logica*, 41(4):329–341.

[Czelakowski and Pigozzi, 1999] Czelakowski, J. and Pigozzi, D. (1999). Amalgamation and interpolation in abstract algebraic logic. In Caicedo, X. and Montenegro, C., editors, *Models, Algebras and Proofs, selected papers of the X Latin American symp. on math. logic held in Bogotá*, volume 203 of *Lecture Notes in Pure and Applied Mathematics*, pages 187–265, New York. Marcel Dekker, Inc.

[Czermak, 1975] Czermak, J. (1975). Interpolation theorem for some modal logics. In Rose, H. and Shepherdson, J., editors, *Logic Colloquium '73 (Proc. of the Logic Colloquium, Bristol, 1973)*, volume 80 of *Stud. Logic Foundations of Math.*, pages 381–393. North-Holland, Amsterdam.

[Daigneault, 1964] Daigneault, A. (1964). Freedom in polyadic algebras and two theorems of Beth and Craig. *Michigan Math. J.*, 11:129–135.

[Day, 1972] Day, A. (1972). Varieties of Heyting algebras, II. Manuscript.

[Day and Jezek, 1984] Day, A. and Jezek, J. (1984). The amalgamation property for varieties of lattices. *Trans. Amer. Math. Soc.*, 286(1):251–256.

[de Jongh and Veltman, 1990] de Jongh, D. and Veltman, F. (1990). Provability logics for relative interpretability. In *Mathematical Logic*, pages 31–42. Plenum, New York.

[de Jongh and Veltman, 1999] de Jongh, D. and Veltman, F. (1999). The modal completeness of ILW. In Gerbrandy, J., Marx, M., de Rijke, M., and Venema, Y., editors, *JFAK. Essays Dedicated to Johan van Benthem on the Occasion of his 50th Birthday*, Vossiuspers. Amsterdam University Press, Amsterdam.

[de Jongh and Visser, 1991] de Jongh, D. and Visser, A. (1991). Explicit fixed points in interpretability logic. *Studia Logica*, 50:39–50.

[de Rijke, 1992] de Rijke, M. (1992). Unary interpretability logic. *Notre Dame J. of Formal Logic*, 33:249–272.

[Feferman, 1974] Feferman, S. (1974). Applications of many-sorted interpolation theorems. In *Proc. of the Tarski Symposium (Proc. Sympos. Pure Math., Vol. XXV, Univ. of California, Berkeley, Calif., 1971)*, pages 205–223, Providence, R. I. Amer. Math. Soc.

[Fine, 1979] Fine, K. (1979). Failures of the interpolation lemma in quantified modal logic. *J. Symbolic Logic*, 44(2):201–206.

[Fitting, 1983] Fitting, M. (1983). *Proof methods for modal and intuitionistic logics*. D. Reidel Publishing Co., Dordrecht.

[Fitting, 2000] Fitting, M. (2000). Interpolation for first-order S5. Submitted for publication.

[Font and Jansana, 1994] Font, J. M. and Jansana, R. (1994). On the sentential logics associated with strongly nice and semi-nice general logics. *Bulletin of the Interest Group in Pure and Applied Logics*, 2(1):55–67.

[Font and Jansana, 1996] Font, J. M. and Jansana, R. (1996). *A general algebraic semantics for sentential logics*, volume 7 of *Lecture Notes in Logic*. Springer-Verlag.

[Font et al., 2000] Font, J. M., Jansana, R., and Pigozzi, D., editors (2000). *Abstract algebraic logic. I.* Kluwer Academic Publishers Group, Dordrecht. Papers from the workshop held in Barcelona, July 1–5, 1997, Studia Logica **65** (2000), no. 1.

[Font and Rodríguez, 1990] Font, J. M. and Rodríguez, G. (1990). Note on algebraic models for relevance logic. *Zeitschr. f. math. Logik und Grundlagen d. Math.*, 36(6):535–540.

[Fraïssé, 1954] Fraïssé, R. (1954). Sur l'extension aux relations de quelques propriétés des ordres. *Ann. Sci. École Norm. Sup.*, 71(3):363–388.

[Freese, 1979] Freese, R. (1979). The variety of modular lattices is not generated by its finite members. *Trans. Amer. Math. Soc.*, 255:277–300.

[Friedman, 1973] Friedman, H. M. (1973). Beth's theorem in cardinality logics. *Israel J. Mathematics*, 14:205–212.

[Gabbay, 1972] Gabbay, D. (1972). Craig's interpolation theorem for modal logics. In Hodges, W., editor, *Conference in Mathematical Logic—London '70 (Proc. Conf., Bedford Coll., London, 1970)*, volume 255 of *Lecture Notes in Math.*, pages 111–127. Springer-Verlag.

[Gabbay, 1971] Gabbay, D. M. (1971). Semantic proof of Craig's interpolation theorem for intuitionistic logic and extensions, II. In *Logic Colloquium '69 (Proc. Summer School and Colloq., Manchester, 1969)*, pages 403–410. North-Holland, Amsterdam.

[Ghilardi and Zawadowski, 1995] Ghilardi, S. and Zawadowski, M. (1995). Undefinability of propositional quantifiers in the modal system S4. *Studia Logica*, 55:259–271.

[Glivenko, 1929] Glivenko, V. (1929). Sur quelque points de la logique de M. Brouwer. *Bull. Acad. des Sci. de Belgique*, 15:183–188.

[Grädel, 1999] Grädel, E. (1999). On the restraining power of guards. *J. Symbolic Logic*, 64(4):1719–1742.

[Grädel and Walukiewicz, 1999] Grädel, E. and Walukiewicz, I. (1999). Guarded fixed point logic. In *Proc. 14th Symp. on Logic in Computer Science, LICS'99*, pages 45–54.

[Grätzer et al., 1973] Grätzer, G., Jónsson, B., and Lakser, H. (1973). The amalgamation property in equational classes of modular lattices. *Pacific J. Math*, 45:507–524.

[Grätzer and Lakser, 1971] Grätzer, G. and Lakser, H. (1971). The structure of pseudocomplemented distributive lattices. ii: Congruence extension and amalgamation. *Transactions of the American Mathematical Society*, 156:343–358.

[Grätzer and Schmidt, 1957] Grätzer, G. and Schmidt, E. T. (1957). On a problem of M. H. Stone. *Acta Math. Acad. Sci. Hungar.*, 8:455–460.

[Gregory, 1974] Gregory, J. (1974). Beth definability in infinitary languages. *J. Symbolic Logic*, 39:22–26.

[Gurevich, 1984] Gurevich, Y. (1984). Toward logic tailored for computational complexity. In *Computation and proof theory (Aachen, 1983)*, pages 175–216. Springer, Berlin.

[Hájek, 1977] Hájek, P. (1977). Generalized quantifiers and finite sets. *Prace Nauk. Inst. Mat. Politech. Wrocław.*, 14 Ser. Konfer. No. 1:91–104.

[Hájek, 1992] Hájek, P. (1992). IL satisfies interpolation. Unpublished.

[Hendriks, 2000] Hendriks, L. (2000). Doing logic by computer: Interpolation in fragments of intuitionistic propositional logic. *Annals of pure and applied logic*, 104(1–3):97–112.

[Henkin, 1963] Henkin, L. (1963). An extension of the Craig-Lyndon interpolation theorem. *J. Symbolic Logic*, 28:201–216.

[Henkin et al., 1985] Henkin, L., Monk, J. D., and Tarski, A. (1971 & 1985). *Cylindric Algebras, Parts I & II.* North-Holland Publishing Co.

[Herrmann, 1997] Herrmann, B. (1997). Characterizing equivalential and algebraizable logics by the Leibniz operator. *Studia Logica*, 58(2):305–323.

[Hodkinson, 1993] Hodkinson, I. (1993). Finite variable logics. *Bull. Europ. Assoc. Theor. Comp. Sci.*, 51:111–140.

[Hodkinson, 2000] Hodkinson, I. (2000). Loosely guarded fragment has finite model property. Submitted for publication.

[Hoogland, 1996] Hoogland, E. (1996). Algebraic characterizations of two Beth definability properties. Master thesis.

[Hoogland, 2000] Hoogland, E. (2000). Algebraic characterizations of various Beth definability properties. *Studia Logica*, 65:91–112.

[Hoogland and Marx, 2000] Hoogland, E. and Marx, M. (2000). Interpolation in guarded fragments. Technical Report PP-2000-11, Institute for Logic, Language and Computation, University of Amsterdam. Submitted for publication.

[Hoogland et al., 1999] Hoogland, E., Marx, M., and Otto, M. (1999). Beth definability for the guarded fragment. In Ganzinger, H., McAllester, D., and Voronkov, A., editors, *Logic for Programming and Automated Reasoning, 6th Int. Conf. LPAR99, Tbilisi, Georgia*, volume 1705 of *LNAI*, pages 273–285. Springer.

[Ignatiev, 1992] Ignatiev, K. (1992). Private comunication. Unpublished.

[Isbell, 1966] Isbell, J. (1966). Epimorphisms and dominions. In *Proc. Conf. Categorical Algebra (La Jolla, Calif.)*, pages 232–246. Springer.

[Isbell, 1973] Isbell, J. R. (1973). Epimorphisms and dominions, V. *Algebra Universalis*, 3:318–320.

[Jansana, 2000] Jansana, R. (2000). Lecture notes on abstract algebraic logic (given at the ILLC in the spring 2000). Unpublished.

[Japaridze and de Jongh, 1998] Japaridze, G. and de Jongh, D. (1998). The logic of provability. In Buss, S., editor, *Handbook of Proof Theory*, pages 475–546. Elsevier Science B.V.

[Johnson, 1970] Johnson, J. (1970). Amalgamation of polyadic algebras. *Trans. Amer. Math. Soc.*, 149:627–652.

[Jónsson, 1956] Jónsson, B. (1956). Universal relational systems. *Math. Scand.*, 4:193–208.

[Jónsson, 1962] Jónsson, B. (1962). Algebraic extensions of relational systems. *Math. Scand.*, 11:179–205.

[Jónsson, 1965] Jónsson, B. (1965). Extensions of relational structures. In *Theory of Models (Proc. 1963 Internat. Sympos. Berkeley)*, pages 146–157. North-Holland, Amsterdam.

[Kalman, 1958] Kalman, J. (1958). Lattices with involution. *Trans. Amer. Math. Soc.*, 87:485–491.

[Keisler, 1971] Keisler, H. (1971). *Model theory for infinitary logic. Logic with countable conjunctions and finite quantifiers.* North-Holland Publishing Co., Amsterdam. Studies in Logic and the Foundations of Mathematics, Vol. 62.

[Kiss et al., 1983] Kiss, E., Márki, L., Pröhle, P., and Tholen, W. (1983). Categorical algebraic properties. A compendium on amalgamation, congruence extension, epimorphisms, residual smallness and injectivity. *Studia Sci. Math. Hungar.*, 18:79–141.

[Kreisel, 1960] Kreisel, G. (1960). Explicit definability in intuitionistic logic. *J. Symbolic Logic*, 25:389–390.

[Kreisel and Krivine, 1967] Kreisel, G. and Krivine, J. (1967). *Elements of mathematical logic. Model theory.* North-Holland Publishing Co., Amsterdam. Studies in Logic and the Foundations of Mathematics.

[Krzystek and Zachorowski, 1977] Krzystek, P. and Zachorowski, S. (1977). Łukasiewicz logics have not the interpolation property. *Rep. Math. Logic*, 9:39–40.

[Lopez-Escobar, 1965] Lopez-Escobar, E. (1965). An interpolation theorem for denumerably long formulas. *Fund. Math.*, 57:253–272.

[Łoś and Suszko, 1958] Łoś, J. and Suszko, R. (1958). Remarks on sentential logics. *Nederl. Akad. Wetensch. Proc. Ser. A 61 = Indag. Math.*, 20:177–183.

[Łukasiewicz and Tarski, 1930] Łukasiewicz, J. and Tarski, A. (1930). Untersuchungen über den Aussagenkalkül. *Comptes rendus de la Société des Sciences et des Lettres de Varsovie*, 23(cl. iii):51–77.

[Lyndon, 1959a] Lyndon, R. (1959a). An interpolation theorem in the predicate calculus. *Pacific J. Math.*, 9:129–142.

[Lyndon, 1959b] Lyndon, R. (1959b). Properties preserved under homomorphism. *Pacific J. Math.*, 9:143–154.

[Madarász, 1998] Madarász, J. (1998). Interpolation in algebraizable logics; semantics for non-normal multi-modal logic. *J. Appl. Non-Classical Logics*, 8(1-2):67–105.

[Maehara and Takeuti, 1961] Maehara, S. and Takeuti, G. (1961). A formal system of first-order predicate calculus with infinitely long expressions. *J. Math. Soc. Japan*, 13:357–370.

[Makkai, 1969] Makkai, M. (1969). On the model theory of denumerably long formulas with finite strings of quantifiers. *J. Symbolic Logic*, 34:437–459.

[Makowsky and Shelah, 1979] Makowsky, J. and Shelah, S. (1979). The theorems of Beth and Craig in abstract model theory, I. The abstract setting. *Trans. Amer. Math. Soc.*, 256:215–239.

[Maksimova, 1979] Maksimova, L. (1979). Interpolation properties of superintuitionistic logics. *Studia Logica*, 38(4):419–428.

[Maksimova, 1980] Maksimova, L. (1980). Interpolation theorems in modal logics and amalgamable varieties of topological Boolean algebras. *Algebra and Logic*, 18(5):348–370.

[Maksimova, 1982] Maksimova, L. (1982). The Lyndon interpolation theorem in modal logics. In *Mathematical logic and the theory of algorithms*, pages 45–55. "Nauka" Sibirsk. Otdel., Novosibirsk.

[Maksimova, 1989] Maksimova, L. (1989). Definability theorems in normal extensions of the provability logic. *Studia Logica*, 48(4):495–507.

[Maksimova, 1991] Maksimova, L. (1991). Amalgamation and interpolation in normal modal logic. *Studia Logica*, 50(3-4):457–471.

[Maksimova, 1992a] Maksimova, L. (1992a). Beth's property, interpolation and amalgamation in varieties of modal algebras. *Soviet Math. Dokl.*, 44(1):327–331.

[Maksimova, 1992b] Maksimova, L. (1992b). Definability and interpolation in classical modal logics. In *Proc. of the International Conference on Algebra, Part 3 (Novosibirsk, 1989)*, pages 583–599, Providence, RI. Amer. Math. Soc.

[Maksimova, 1993] Maksimova, L. (1993). An analog of Beth's theorem in normal extensions of the modal logic K4. *Siberian Mathematical J.*, 33(6):1052–1065.

[Maksimova, 1999a] Maksimova, L. (1999a). Projective Beth properties in modal and superintuitionistic logics. *Algebra and Logic*, 38(6):171–180.

[Maksimova, 1999b] Maksimova, L. (1999b). Superintuitionistic logics and the projective Beth property. *Algebra and Logic*, 38(6):374–382.

[Malinowski and Michalczyk, 1981] Malinowski, G. and Michalczyk, M. (1981). Interpolation properties for a class of many-valued propositional calculi. *Polish Acad. Sci. Inst. Philos. Sociol. Bull. Sect. Logic*, 10(1):9–16.

[Malinowski, 1989] Malinowski, J. (1989). *Equivalence in intensional logics*. PhD thesis, The polish academy of sciences, institute of philosophy and sociology.

[Malitz, 1969] Malitz, J. (1969). Universal sentences and infinitary logic. *Duke Math. J.*, 36:621–630.

[Malitz, 1971] Malitz, J. (1971). Infinitary analogues of theorems from first-order model theory. *J. Symbolic Logic*, 36:216–228.

[Marx, 1995] Marx, M. (1995). *Algebraic Relativization and Arrow Logic*. PhD thesis, Institute for Logic, Language and Computation. University of Amsterdam.

[Marx, 1997] Marx, M. (1997). Complexity of modal logics of relations. Technical Report ML–97–02, Institute for Logic, Language and Computation, University of Amsterdam.

[Marx, 1999a] Marx, M. (1999a). Interpolation in modal logic. In *Algebraic methodology and software technology (Amazonia, 1999)*, pages 154–163. Springer, Berlin.

[Marx, 1999b] Marx, M. (1999b). Tolerance logic. Technical Report IR–469, Free University. To appear in *J. of Logic, Language and Information*.

[McLarty, 1992] McLarty, C. (1992). *Elementary categories, elementary toposes*. The Clarendon Press Oxford University Press, New York. Oxford Science Publications.

[Mekler and Shelah, 1986] Mekler, A. and Shelah, S. (1986). Stationary logic and its friends, II. *Notre Dame J. Formal Logic*, 27(1):39–50.

[Moisil, 1935] Moisil, G. C. (1935). Recherches sur l'algèbre de la logique. *Ann. Sci. Univ. Jassy. Sect. I.*, 22:1–117.

[Németi, 1984] Németi, I. (1984). Beth definability is equivalent with surjectiveness of epis in general algebraic logic. Technical report, Math. Inst. Hungar. Acad. Sci., Budapest.

[Németi, 1985] Németi, I. (1985). Cylindric relativised set algebras have strong amalgamation. *J. Symbolic Logic*, 50(3):689–700.

[Ono, 1986] Ono, H. (1986). Interpolation and the Robinson property for logics not closed under the Boolean operations. *Algebra Universalis*, 23(2):111–122.

[Orłowska, 1999] Orłowska, E., editor (1999). *Logic at Work. Essays Dedicated to the Memory of Elena Rasiowa*. Studies in Fuzziness and Soft Computing. Springer Verlag, Berlin/Heidelberg.

[Otto, 1999] Otto, M. (1999). More on interpolation for the guarded fragment. Unpublished.

[Otto, 2000] Otto, M. (2000). An interpolation theorem. *The Bulletin of Symbolic Logic*, 6(4):447–462.

[Pałasińska, 1988] Pałasińska, K. (1988). The failure of strong amalgamation property in certain classes of BCK-algebras. *Math. Japonica*, 33(6):913–917.

[Pierce, 1968] Pierce, R. S. (1968). *Introduction to the theory of abstract algebras*. Holt, Rinehart and Winston, New York.

[Pigozzi, 1972] Pigozzi, D. (1972). Amalgamation, congruence extension and interpolation properties in algebras. *Algebra Universalis*, 1(3):269–349.

[Pitts, 1992] Pitts, A. (1992). On an interpretation of second-order quantification in first-order intuitionistic propositional logic. *J. Symbolic Logic*, 57(1):33–52.

[Prucnal and Wroński, 1974] Prucnal, T. and Wroński, A. (1974). An algebraic characterization of the notion of structural completeness. *Polish Acad. Sci. Inst. Philos. Sociol. Bull. Sect. Logic*, 3(1):30–33.

[Rasiowa, 1974] Rasiowa, H. (1974). *An algebraic approach to non-classical logics*. North-Holland Publishing Co., Amsterdam. Studies in Logic and the Foundations of Mathematics, Vol. 78.

[Rautenberg, 1983] Rautenberg, W. (1983). Modal tableau calculi and interpolation. *J. Philosophical Logic*, 12(4):403–423.

[Renardel de Lavalette, 1981] Renardel de Lavalette, G. (1981). The interpolation theorem in fragments of logics. *Nederl. Akad. Wetensch. Indag. Math.*, 43(1):71–86.

[Renardel de Lavalette, 1989a] Renardel de Lavalette, G. (1989a). Interpolation in fragments of intuitionistic propositional logic. *J. Symbolic Logic*, 54(4):1419–1430.

[Renardel de Lavalette, 1989b] Renardel de Lavalette, G. (1989b). Modularisation, parametrisation and interpolation. *J. Inf. Process. Cybern. EIK*, 25(5/6):283–292.

[Robinson, 1956] Robinson, A. (1956). A result on consistency and its application to the theory of definition. *Nederl. Akad. Wetensch. Proc. Ser. A.* **59** = *Indag. Math.*, 18:47–58.

[Routley et al., 1982] Routley, R., Plumwood, V., Meyer, R. K., and Brady, R. T. (1982). *Relevant logics and their rivals. Part I. The basic philosophical and semantical theory*. Ridgeview Publishing Co., Atascadero, Calif.

[Sain, 1990] Sain, I. (1990). Beth's and Craig's properties via epimorphisms and amalgamation in algebraic logic. In *Algebraic logic and universal algebra in computer science*, volume 24 of *Lecture Notes in Computer Science*, pages 209–225. Springer.

[Sain, 1998] Sain, I. (1998). On characterizations of definability properties in abstract algebraic logic. In Font, J. M., Jansana, R., and Pigozzi, D., editors, *Workshop on Abstract Algebraic Logic*, volume 10 of *Quaderns*, pages 162–175.

[Sánchez Valencia,] Sánchez Valencia, V. Lyndon interpolation for some modal logics. Unpublished.

[Schreier, 1927] Schreier, O. (1927). Die Untergruppen der freien Gruppen. *Abh. Math. Sem. Univ. Hamburg*, 5:161–183.

[Schumm, 1976] Schumm, G. (1976). Interpolation in S5 and some related systems. *Rep. Math. Logic*, 6:107–109.

[Schütte, 1962] Schütte, K. (1962). Der Interpolationssatz der intuitionistischen Prädikatenlogik. (German). *Math. Ann.*, 148:192–200.

[Smoryński, 1978] Smoryński, C. (1978). Beth's theorem and self-referential sentences. In Macintyre, A., Pacholski, L., and Paris, J., editors, *Logic Colloquium '77 (Proc. Conf., Wrocław, 1977)*, volume 96 of *Stud. Logic Foundations Math.*, pages 253–261. North-Holland, Amsterdam.

[Smoryński, 1985] Smoryński, C. (1985). *Self-reference and modal logic*. Springer-Verlag.

[Thistlewaite et al., 1988] Thistlewaite, P., McRobbie, M., and Meyer, R. (1988). *Automated theorem-proving in non-classical logics*. Pitman.

[Urquhart, 1984] Urquhart, A. (1984). Many-valued logic. In Gabbay, D. and Guenthner, F., editors, *Handbook of Philosophical Logic. Volume III: Alternatives in Classical Logic*, Synthese library, pages 71–117. D. Reidel Publishing Company, Dordrecht.

[Urquhart, 1999] Urquhart, A. (1999). Beth's definability theorem in relevant logics. In Orłowska, E., editor, *Logic at Work. Essays Dedicated to the Memory of Elena Rasiowa*, Studies in Fuzziness and Soft Computing, pages 229–234. Springer Verlag, Berlin/Heidelberg.

[van Benthem, 1997] van Benthem, J. (1997). Dynamic bits and pieces. Technical Report LP-97-01, Institute for Logic, Language and Computation, University of Amsterdam.

[van Benthem, 1998] van Benthem, J. (1998). Modal logic in two gestalts. Technical Report ML-98-12, Institute for Logic, Language and Computation, University of Amsterdam.

[van Benthem, 1999] van Benthem, J. (1999). Modal foundations for predicate logic. In [Orłowska, 1999], pages 39–55.

[van Ulsen, 2000] van Ulsen, P. (2000). *E.W. Beth als logicus*. PhD thesis, Institute for Logic, Language and Computation. University of Amsterdam.

[Vardi, 1998] Vardi, M. (1998). Why is modal logic so robustly decidable? In Immerman, N. and Kolaitis, P., editors, *Descriptive Complexity and Finite Models: Proceedings of a DIMACS Workshop*, volume 31 of *Series in Discrete Mathematics and Theoretical Computer Science*, pages 149–184. American Math. Soc.

[Veloso, 1996] Veloso, P. (1996). On pushout consistency, modularity and interpolation for logical specifications. *Inform. Process. Lett.*, 60(2):59–66.

[Venema and Marx, 1999] Venema, Y. and Marx, M. (1999). A modal logic of relations. In [Orłowska, 1999], pages 124–167.

[Visser, 1990] Visser, A. (1990). Interpretability logic. In *Mathematical logic, Proc. of the Heyting 1988 summer school in Varna, Bulgaria*, pages 175–209. Plenum Press, Boston.

[Visser, 1996] Visser, A. (1996). Uniform interpolation and layered bisimulation. In Hájek, P., editor, *Gödel '96, Logical Foundations of Mathematics, Computer Science and Physics — Kurt Gödel's Legacy*, pages 139–164. Springer, Berlin.

[Visser, 1998] Visser, A. (1998). An overview of interpretability logic. In Kracht, M., de Rijke, M., Wansing, H., and Zakharyaschev, M., editors, *Advances in Modal logic I*, Lecture Notes. CSLI Publications, Stanford, CA.

[Werner, 1978] Werner, H. (1978). *Discriminator-algebras*, volume 6 of *Studien zur Algebra und ihre Anwendungen.* Akademie-Verlag.

Index

List of symbols

Samenvatting

In dit proefschrift bestuderen we *Beth definieerbaarheid* en *interpolatie*. Dit zijn eigenschappen van logicas die, net als bijvoorbeeld compactheid of beslisbaarheid, meten in hoeverre een logica zich netjes gedraagt. In Hoofdstuk 2 worden deze eigenschappen gedefinieerd en uitgebreid besproken. Kort gezegd is de Beth eigenschap op te vatten als een volledigheidsstelling in de definitieleer. Vergelijkbaar met Gödels volledigheidsstelling die zegt dat alles wat waar is ook daadwerkelijk een bewijs heeft, zegt Beths stelling dat alles wat vastgelegd kan worden ook een expliciete definitie heeft. Het bewijs dat een specifieke logica S deze eigenschap heeft, verloopt meestal via het aantonen van de interpolatie eigenschap voor S. Deze eigenschap eist dat elke implicatie $\varphi \to \psi$ die geldig is in S een *interpolant* heeft. D.w.z., er bestaat een formule ϑ in de gemeenschappelijke taal van φ, ψ zodanig dat $\vdash_S \varphi \to \vartheta$ en $\vdash_S \vartheta \to \psi$. Los van het verband met definieerbaarheid is interpolatie ook zelf een interessante notie die duidt op een net deductief systeem.

De bijdrage van dit proefschrift is vierledig. Opeenvolgend wordt

1. Een inleidend overzicht gepresenteerd waarin in grote lijnen het onderzoek op het gebied van definieerbaarheid en interpolatie in kaart wordt gebracht.

2. Definieerbaarheid in verband gebracht met de algebraïsche eigenschap van het surjectief zijn van epimorphismen.

3. Gereedschappen ontwikkeld om definieerbaarheidsstellingen en interpolatiestellingen te bewijzen en te weerleggen.

4. Voorbeelden gegeven die aantonen dat interpolatie een veel sterkere eigenschap is dan de Beth eigenschap. Dit wordt gedaan door middel van gedetailleerde studies naar *bewaakte fragmenten* en naar *interpreteerbaarheidslogicas.*

Hoofdstuk 2 In dit inleidende hoofdstuk wordt de lezer vertrouwd gemaakt met de twee themas van dit proefschrift: definieerbaarheid en interpolatie. Aan de hand van eenvoudige voorbeelden worden intuïties verschaft. Ook worden de belangrijkste resultaten op dit gebied samengevat en worden verscheidene bewijsmethoden besproken. Verder wordt helderheid gegeven in de menagerie van interpolatie- en definieerbaarheidseigenschappen die in de literatuur te vinden is: we motiveren de keuzes voor de verschillende eigenschappen en stellen de preciese relaties tussen hen vast.

Daarnaast benadert dit hoofdstuk de zaak vanuit een algebraïsch perspectief. Er wordt uitgelegd dat interpolatie en definieerbaarheid nauwe verwantschap hebben met bepaalde algebraïsche eigenschappen, te weten *amalgamatie* en *surjectiviteit van epimorphismen*. Deze eigenschappen worden besproken en er wordt nader ingegaan op hun relatie met logica.

Hoofdstuk 3 Binnen het gebied van de abstracte algebraïsche logica wordt onderzocht hoeveel informatie over een logica S te verkrijgen is uit een klasse van algebras Mod^*S. Hoofdstuk 3 behandelt deze vraag specifiek voor de Beth eigenschap. Het belangrijkste resultaat uit dit hoofdstuk zegt dat, onder zwakke aannamen op de logica S, S de Beth eigenschap heeft dan en slechts dan als Mod^*S de eigenschap heeft van surjectiviteit van epimorphismen. Met dezelfde algemeenheid geven we soortgelijke karakteriseringen van de *zwakke Beth eigenschap* en de *projectieve Beth eigenschap*. Deze karakteriseringen zijn in termen van het surjectief zijn van een bepaald soort epimorphismen.

Daarnaast bestuderen we het al dan niet surjectief zijn van epimorphismen voor een aantal concrete klassen van algebras. Toepassing van onze hoofdstelling leert vervolgens ondermeer dat de Beth eigenschap niet opgaat voor de meerwaardige logicas van Łukasiewicz en voor een groot aantal relevantie logicas.

Hoofdstuk 4 Dit hoofdstuk is een case-study waarin interpolatie en definieerbaarheid onderzocht wordt voor *bewaakte fragmenten*. Dit zijn fragmenten van eerste orde logica waarin kwantoren op een bepaalde manier gerelativiseerd worden. D.w.z., kwantoren zijn van de vorm $\exists x(Gx \wedge \varphi(x))$ en $\forall x(Gx \rightarrow \varphi(x))$, waar Gx aan bepaalde syntactische specificaties voldoet. Bijvoorbeeld, in het *guarded fragment* (GF) is Gx een atomaire formule, in het *packed fragment* (PF) is Gx een bepaalde conjunctie van atomen.

Er wordt aangetoond dat GF en PF *niet* de interpolatie eigenschap hebben. Het blijkt dat deze eigenschap niet de correcte notie van interpolatie is voor bewaakte fragmenten. Dit leidt tot een alternatieve notie van interpolatie die in deze context beter lijkt te passen. We bewijzen dat GF en PF wel deze alternatieve eigenschap hebben. Verder blijkt dit een vrij sterke eigenschap te zijn. In ieder geval sterk genoeg om er de Beth eigenschap uit af te kunnen leiden. Beter nog, alle bewaakte n-variabele fragment hebben de Beth eigenschap. Dit laat zien dat deze fragmenten zich veel netter gedragen m.b.t. definieerbaarheid dan de volledige n-variabele fragmenten van eerste orde logica. Ook wordt interpolatie voor deze fragmenten onderzocht. Hieruit blijkt dat het bewaakte 2-variabele fragment interpolatie heeft. Tenslotte wordt de alternatieve interpolatiestelling voor GF verfijnd tot een Lyndon-stelling die rekening houdt met de (positieve en negatieve) voorkomens van kwantoren.

Hoofdstuk 5 Het laatste hoofdstuk is eveneens een case-study waarin ditmaal interpolatie en definieerbaarheid bestudeerd worden voor *interpreteerbaarheidslogicas*. Dit zijn modale systemen met een extra, non-standaard, modale operator.

Het voornaamste resultaat van het eerste deel van dit hoofdstuk is de interpolatiestelling voor het basissysteem van interpreteerbaarheidslogica IL. Het bewijs van deze stelling maakt gebruik van een model-theoretische constructie die sterk lijkt op de constructie zoals die te vinden is in een Henkin-volledigheidsbewijs. Directe gevolgen van deze stelling zijn de Beth stelling voor IL en, gebruik makend van een redenering van Hájek, interpolatie voor het systeem ILP. Er wordt tenslotte een voorbeeld gegeven van een interpreteerbaarheidslogica *zonder* interpolatie, namelijk ILW.

In het tweede deel van het hoofdstuk komt een interessant samenspel aan het licht tussen de Beth eigenschap en het bestaan (en uniek zijn) van dekpunten. Het blijkt dat voor een grote klasse van logicas deze twee eigenschappen equivalent zijn. Deze klasse omvat alle interpreteerbaarheidslogicas. Samen met de Beth stelling voor IL leidt dit tot een nieuw bewijs van de dekpuntseigenschap voor IL. Maar aangezien de dekpuntseigenschap wordt bewaard onder het nemen van extensies, toont ons resultaat daarnaast aan dat *alle* extensies van IL (d.w.z., alle interpreteerbaarheidslogicas) de Beth eigenschap hebben.

Titles in the ILLC Dissertation Series:

ILLC DS-2000-01: **Renata Wasserman**
Resource Bounded Belief Revision

ILLC DS-2000-02: **Jaap Kamps**
A Logical Approach to Computational Theory Building (with applications to sociology)

ILLC DS-2000-03: **Marco Vervoort**
Games, Walks and Grammars: Problems I've Worked On

ILLC DS-2000-04: **Paul van Ulsen**
E.W. Beth als logicus

ILLC DS-2000-05: **Carlos Areces**
Logic Engineering. The Case of Description and Hybrid Logics

ILLC DS-2000-06: **Hans van Ditmarsch**
Knowledge Games

ILLC DS-2000-07: **Egbert L.J. Fortuin**
Polysemy or monosemy: Interpretation of the imperative and the dative-infinitive construction in Russian

ILLC DS-2001-01: **Maria Aloni**
Quantification under Conceptual Covers

ILLC DS-2001-02: **Alexander van den Bosch**
Rationality in Discovery - a study of Logic, Cognition, Computation and Neuropharmacology.

ILLC DS-2001-03: **Erik de Haas**
Logics For OO Information Systems: a Semantic Study of Object Orientation from a Categorial Substructural Perspective

ILLC DS-2001-04: **Rosalie Iemhoff**
Provability Logic and Admissible Rules

ILLC DS-2001-05: **Eva Hoogland**
Definability and Interpolation: Model-theoretic investigations

ILLC DS-2001-06: **Ronald de Wolf**
Quantum Computing and Communication Complexity