A NEW MODAL LINDSTRÖM THEOREM

Johan van Benthem, University of Amsterdam & Stanford University

December 2006

Abstract We prove new Lindström theorems for the basic modal propositional language, and for the guarded fragment of first-order logic. We find difficulties with such results for modal languages without a finite-depth property, high-lighting the difference between abstract model theory for fragments and for extensions of first-order logic.

1 What **is** *modal logic?*

I would broadly endorse the 'minimal design view' in van Benthem 1996, Andréka, van Benthem & Németi 1998, Blackburn, de Rijke & Venema 2000, which says that modal languages are *well-balanced fragments* of classical ones combining good expressive power with reasonable computational complexity for model checking and satisfiability. But one can also try to understand what makes modal logic tick in other ways. One obvious alternative format is that of a *Lindström theorem*. Indeed, de Rijke 1993 contains one such result – and we sketch its modern proof as our starting point.

2 A first modal Lindström theorem

Blackburn, de Rijke & Venema 2000 define an 'abstract modal language' as a formalism satisfying the usual base constraints from abstract model theory (Barwise & Feferman, eds., 1985), plus the modal characteristic of *bisimulation invariance* for all formulas. They then high-light the following semantic property, saying that modal formulas only look at models up to some finite depth:

> *Finite Depth Property*
> For any formula ϕ, there is a natural number k such that, for all models,
>
> $(\boldsymbol{M}, w) \models \phi$ iff $(\boldsymbol{M}|k, w) \models \phi$,
>
> where $\boldsymbol{M}|k$ is the model $\boldsymbol{M}$ with its domain restricted to just those points that can be reached from w in k or fewer successive R-steps.

Then we have the following 'maximality version' of a modal Lindström result, valid for abstract modal languages L with a finite vocabulary:

Theorem 1 Any abstract modal language extending the basic modal one which has the Finite Depth Property is the basic modal language itself.

The proof of this result revolves around the following fact.

Lemma 1 If an *L*-formula ϕ has the Finite Depth Property for distance *k*, then ϕ is preserved under modal equivalence up to operator depth *k*.

Proof Suppose that two models *(**M**, w), (**N**, v)* agree on modal formulas up to depth *k*, while *(**M**, w)* |= ϕ. By a standard technique, then, both models have bisimilar *tree unravelings Tree(**M**, w), Tree(**N**, v).* Moreover, since these tree models have the same modal theory up to depth *k* in their roots, there exists an obvious 'cut-off' bisimulation between them up to the first *k* tree levels. But this cut-off bisimulation is a *full* bisimulation between the 'cut-off models' *Tree(**M**, w)|k* and *Tree(**N**, v)|k*. By the Finite Depth Property plus invariance for bisimulation, we have that ϕ holds, successively, in

*(**M**, w), Tree(**M**, w), (Tree(**M**, w)|k, w), (Tree(**N**, v)|k*, v), and *(**N**, v).* ♠

The proof of Theorem 1 is then is clinched by a further well-known observation:

Lemma 2 If an *L*-formula ϕ is preserved under modal equivalence up to some finite operator depth *k*, it is definable by a modal formula of operator depth *k*.

Proof Recall that this fragment of the modal base language is logically finite, as the vocabulary of our abstract language *L* is finite. Thus, in particular, any class ***K*** of pointed models closed under modal *k*-equivalence is defined by the disjunction of all complete modal depth-*k* theories satisfied in ***K***. And this is a basic modal formula. ♠

Theorem 1, though informative, is not entirely satisfactory. It does not look like a standard Lindström result in abstract model theory, and the Finite Depth Property seems somewhat ad-hoc, and 'engineered' to capture the basic modal language. In order to improve on this, let us first consider a more standard model-theoretic proof.

3 The classical Lindström theorem for first-order logic

The original Lindström Theorem captured first-order logic (*FOL*) as the strongest extension of *FOL* satisfying the abstract version of the Compactness and Löwenheim-Skolem properties. Perhaps the most illuminating modern version, however, high-lights a basic model-existence property (Compactness) of *FOL* with an equally basic one of semantic invariance (often called the 'Karp Property'):

Theorem 2 An abstract logic *L* extending first-order logic coincides with *FOL* iff

(a) all formulas of *L* are invariant for potential isomorphism,

(b) *L* has the Compactness property.

Here a 'potential isomorphism' is a non-empty family of finite partial isomorphisms closed under the usual Back and Forth extension properties. Note that the crucial invariance property is not made part of the definition of abstract logics: these are just required to satisfy invariance for isomorphism. The version with Compactness and Löwenheim-Skolem follows from isomorphism invariance for abstract logics, plus the fact that potential isomorphisms between countable models are true isomorphisms.

The same style of thinking has proven its mettle for extended abstract logics, such as the infinitary language $L^{\infty}{}_{\omega}$, where the Barwise Characterization Theorem uses the Karp property (a) plus projective undefinability of well-ordering instead of Compactness (b). This result also leaves something to be desired, but for that, cf. Section 8 below.

4 Problems with lifting standard proofs to first-order fragments

But what about, not extensions of first-order logic, but modal *fragments*? It is well-known that proofs of standard results may fail when we extend the first-order language to higher-order versions. But such proofs may also fail when we *restrict FOL*, because we have lost essential expressive resources. Let us inspect the proof of Theorem 2:

Proof Take any formula ϕ of the abstract logic L. To see that ϕ is first-order, it suffices to show that ϕ is preserved by elementary equivalence up to some quantifier depth k in the appropriate finite similarity type. So, suppose it is not. Use new predicate letters A, B (unary) and a countable family of $2k+1$-ary I to encode partial isomorphisms. Then, for each natural number k, we can find models of the following form:

(a) an A-part where ϕ holds, (b) a B-part where its negation $\neg\phi$ holds, and
(c) a chain of partial isomorphisms up to depth k between the A- and B-parts.

All this can be described in first-order terms, where we use a linear ordering $<$ (one more new predicate letter) over the distinguished argument of the I-predicates to encode the existence of the finite chain. Then, by the existence of all finite k-situations as described, Compactness gives us a model containing 'infinite objects' in the ordering $<$. Using the latter as indices gives an *infinite descending chain* of partial isomorphisms between a model A for ϕ and B for $\neg\phi$. The union of all stages in that chain is a potential isomorphism between the A- and B-models, with unrestricted Back and Forth clauses – and this contradicts the given invariance for potential isomorphisms. ♠

Lifting this type of argument to fragments of *FOL* runs into their *expressive weakness*. Standard proofs in abstract model theory rely throughout on coding up relevant properties in first-order formulas. This feature is often taken for granted, as it works for the usual richer languages than *FOL*. But this 'facility' may fail for poorer languages. The latter 'downward' direction has been studied much less: Garcia-Matos 2005 on negation-free abstract model theory is one noteworthy exception – and so is ten Cate 2005 on extended modal languages in between the basic modal language and *FOL*.

In particular, in the above proof, the encoding of potential isomorphism would now have to refer to bisimulations. And the latter still have a characteristic zigzag property which is first-order, but typically non-modal. Indeed, it defines a 'grid structure' related to tiling arguments for undecidability (van Benthem & Blackburn 2005).

5 A true Lindström theorem for basic modal logic

Still, there is an easy way around this particular problem for the basic modal language. The main point of this note is to give a standard modal Lindström theorem after all!

For this purpose, we first define an *abstract modal logic L* in the obvious way, with truth referring to pointed models *(**M**, w)*. But we add a base condition of

> *Relativization*
> For any *L*-formula ϕ and new unary proposition letter *p*, there exists an *L*-formula $Rel(\phi, p)$ which is true at a model $(\boldsymbol{M}, w)$ iff ϕ is true at $(\boldsymbol{M}|p, w)$: the submodel of ***M*** with just the points in ***M*** satisfying *p* for its domain.

Most usable logics satisfy Relativization, and indeed, it was presupposed in the above proof for *FOL*. Next, we do not build in invariance for bisimulation from the start, making it an explicit additional requirement instead, as in the earlier first-order analysis. For convenience, we take a vocabulary with one accessibility relation *R* and countably many proposition letters; but our argument also works for larger poly-modal vocabularies. Each *L*-formula only involves a finite part of the vocabulary, in the usual way.

Now we state our first main result:

Theorem 3 An abstract modal logic *L* extending the basic modal language equals the latter iff *L* satisfies (a) Invariance for Bisimulation, and (b) Compactness.

Proof The direction from left to right is obvious. Next, assume that *L* has the stated properties, and consider any formula ϕ in it. As before (Lemma 2), it suffices to show that ϕ is preserved under modal equivalence up to some finite operator depth *k*. And this will follow from the Finite Depth Property, as we have seen in Lemma 1. Now, all we do is note that Relativization, Compactness, and Bisimulation Invariance do the trick.

Lemma 3 In a compact abstract modal logic *L* which is invariant for bisimulation, any formula has the Finite Depth Property.

Proof Let ϕ be any formula in *L*. Suppose, for the sake of reductio ad absurdum, that it lacks the Finite Depth Property. Then for any natural number *k*, there exists a model

$(\boldsymbol{M}_k, w)$ and a cut-off version $(\boldsymbol{M}_k|k, w)$ which disagree on the truth value of ϕ. Without loss of generality, assume that the following happens for arbitrarily large k:

$$(\boldsymbol{M}_k|k, w) \models \phi, \text{ while } (\boldsymbol{M}_k, w) \models \neg\phi.$$

Here we use the fact that abstract modal logics are closed under negations. Now, take a new proposition letter p, and consider the following set Σ of L-formulas:

$$\neg\phi,\ Rel(\phi, p),\ \{[]^n p \mid \text{all natural numbers } n\}.$$

Given our assumptions, this set is clearly finitely satisfiable: we choose k sufficiently large, and make p true in the k-reachable part of one of the above sequence of models. But then, by Compactness for our abstract modal logic L, there must be a model $(\boldsymbol{N}, v)$ for the whole set Σ at once. But this leads to a contradiction:

We focus on the *generated submodel* $(\boldsymbol{N}_v, v)$ consisting of v and all points finitely reachable from it. Clearly, the identity is a bisimulation between any pointed model and its unique generated submodel. Hence by the assumed invariance for bisimulation, formulas of L have the same truth value in any pointed model and its generated submodel. Now, given the first formula in Σ, $\neg\phi$ holds in $(\boldsymbol{N}, v)$ and hence also in $(\boldsymbol{N}_v, v)$ *(i)*. On the other hand, since $(\boldsymbol{N}, v) \models Rel(\phi, p)$, we have $(\boldsymbol{N}|p, v) \models \phi$. But by the truth of the infinite third set of formulas, p holds in the whole generated submodel $(\boldsymbol{N}_v, v)$. Therefore, it is easy to see that the generated submodel of $(\boldsymbol{N}|p, v)$ is also just $(\boldsymbol{N}_v, v)$, and so we have that ϕ holds in $(\boldsymbol{N}_v, v)$ *(ii)*. This is a contradiction. ♠

This completes the proof of our new Lindström theorem for basic modal logic. ♠

6 A challenge: analyzing the Guarded Fragment

Here is an obvious test case for the generalizability of the preceding style of analysis. The *Guarded Fragment GF* is an expressive generalization of the basic modal language allowing for arbitrary first-order quantifications of the form

$$\exists \boldsymbol{y}\ (G(\boldsymbol{x}, \boldsymbol{y}) \ \&\ \phi(\boldsymbol{x}, \boldsymbol{y})),$$

where $\boldsymbol{x}$, $\boldsymbol{y}$ are tuples of variables, while the 'guard' G is an atomicpredicate with its variables occurring in any order and multiplicity. Andréka, van Benthem & Németi 1998 show that *GF* is decidable, while it has many 'modal' meta-properties thanks to its invariance for 'guarded bisimulation'. Van Benthem 2005 has further issues and results.

Because of its modal character, *GF* seems an obvious case for a Lindström-style analysis like the one we have given for the basic modal language. But there are some technical difficulties, to be explained below. Therefore, we focus on one special case.

In this section, we look only at the special case of a vocabulary with at most binary predicates. The *Binary Guarded Fragment* (GF_{bin}) has the following syntax rules

$$\text{atoms } P\mathbf{x} \mid \neg \mid \vee \mid \exists y\, (G(x, y)\ \&\ \phi(x, y))$$

Here, bold-face $\mathbf{x}$ indicate finite tuples of variables, and G is a binary predicate letter. In writing atoms, we do not care about the order of the variables. This language extends basic modal and temporal logic in an obvious sense. Also, GF_{bin} is easily seen to lie inside the *two-variable fragment FO2* of first-order logic. We will prove this result:

Theorem 4 An abstract logic L extending the Binary Guarded Fragment equals GF_{bin} iff L has (a) Invariance for Guarded Bisimulation, and (b) Compactness.

Before proving this, we start with some modal features that hold for *GF* in general.

First, there is a natural syntactic notion of *formula depth* – whose inductive definition counts the above polyadic quantifiers as single steps:

$$depth(P\mathbf{x}) = 0,\ \ depth\,(\neg\phi) = depth(\phi),\ \ depth(\phi\vee\psi) = maximum$$
$$(depth\,(\phi), depth\,(\psi)),\ \ depth(\exists\mathbf{y}(G(\mathbf{x}, \mathbf{y})\ \&\ \phi(\mathbf{x}, \mathbf{y}))) = depth(\phi)+1$$

Next, we define *distance* for points in models $(\mathbf{M}, \mathbf{s})$, where $\mathbf{s}$ is a tuple of worlds:

$dist(\mathbf{s}, s_i, 0)$ for all $s_i \in \mathbf{s}$, $dist(\mathbf{s}, t, n+1)$ if there is a u with $dist(\mathbf{s}, u, n)$ and $G(t, u, \mathbf{v})$ holds for some atomic predicate G and tuple of objects $\mathbf{v}$.

We write $Cut((\mathbf{M}, \mathbf{s}), n)$ for the submodel $\{t \in (\mathbf{M}, \mathbf{s}) \mid dist(\mathbf{s}, t, n)\}$ consisting of all points t in $\mathbf{M}$ lying at distance at most n from $\mathbf{s}$. The following result shows that *GF*, like the basic modal language, satisfies a *Finite Depth Property* – suitably defined:

Distance-Depth Lemma Let ϕ be any guarded formula of depth n, and let $(\mathbf{N}, \mathbf{s})$ be any submodel of $(\mathbf{M}, \mathbf{s})$ containing all of $Cut((\mathbf{M}, \mathbf{s}), n)$. Then $(\mathbf{M}, \mathbf{s}) \models \phi$ iff $(\mathbf{N}, \mathbf{s}) \models \phi$.

Next comes a generalization of modal bisimulation. A *guarded bisimulation* is a non-empty set $\mathbf{F}$ of *finite partial isomorphisms* between two models $\mathbf{M}$ and $\mathbf{N}$ which has the following back-and-forth conditions. Call a set of objects 'guarded' if some tuple with these objects stands in some atomic relation. Now, given any function $f{:}X\rightarrow Y$ in $\mathbf{F}$,

(i) for any guarded $Z\subseteq\mathbf{M}$, there is a $g\in\mathbf{F}$ with domain Z such that g and f agree on the intersection $X\cap Z$,

(ii) for any guarded $W\subseteq\mathbf{N}$, there is a $g\in\mathbf{F}$ with range W such that the inverses g^{-1} and f^{-1} agree on $Y\cap W$.

Also, 'rooted' guarded bisimulations $\boldsymbol{F}$ run between models $(\boldsymbol{M}, \boldsymbol{s})$ and $(\boldsymbol{N}, \boldsymbol{t})$ with given initial objects, where one requires that some match between $\boldsymbol{s}$, $\boldsymbol{t}$ is already a partial isomorphism in the set $\boldsymbol{F}$. By a simple inductive argument,

> *GF*-formulas ϕ are invariant for rooted guarded bisimulations.

Andréka, van Benthem & Németi 1998 show that *GF* consists, up to logical equivalence, of just those first-order formulas which are invariant for guarded bisimulations.

Another 'modal' use of guarded bisimulation in the same paper is model unraveling. This is like standard modal unraveling, but the construction is a bit more delicate:

Definition The *tree unraveling* $Unr(\boldsymbol{M})$ of $\boldsymbol{M}$ has for its objects all pairs (π, d) – where the 'path' π is a finite sequence of guarded sets in $\boldsymbol{M}$, and the $\boldsymbol{M}$–object d is 'new' in π: i.e., it occurs in the final set of π but not in the one before that. The interpretation of predicate symbols Q on these objects (π, d) is as follows. $I(Q)$ holds for a finite sequence of objects $<(\pi_i, d_i)>_{1\leq i\leq k}$ iff $Q^{\boldsymbol{M}}<d_i>_{1\leq i\leq k}$ *and* there is some maximal path π^* among those listed of which all other π_i are initial segments in such a way that their new objects d_i remain present in each set until the end of π^*. For a model $(\boldsymbol{M}, \boldsymbol{s})$ this generalizes as follows. $Unr(\boldsymbol{M}, \boldsymbol{s})$ has paths π all starting from the initial set $\boldsymbol{s}$, but then continuing with guarded sets only. The objects (π, d) are defined as before.

The point here is that the set $\boldsymbol{F}$ of all restrictions of the finite maps sending (π_i, d_i) to d_i for all guarded finite domains in $Unr(\boldsymbol{M}, \boldsymbol{s})$ is a rooted guarded bisimulation between $(\boldsymbol{M}, \boldsymbol{s})$ and $(Unr(\boldsymbol{M}), \boldsymbol{s})$. Checking the zigzag conditions for the bisimulation will reveal the reason for the above technical definition of the predicate interpretations $I(Q)$.

Now we have the generalities in place for our Lindström Theorem – but it remains to make some adjustments. First, all these general *GF* notions and observations specialize to GF_{bin} in an obvious way, which we do not bother to spell out. Less generally, in the definition of tree unravelings, we make one simple change for the binary case:

> The finite paths of guarded sets always introduce one new object at each stage.
> At each continuation, one chooses a new object related to that new object, etc.

This allows paths starting with object *a* and then continuing with *Rab, Qcb, ...,* while ruling out paths like *Rab, Qac*. But the final atom is not omitted from the unraveled model, since one can have paths starting with *a* and then placing *Qac* immediately. Thus, even with these restricted paths, we still have a guarded isimulation between tree unravelings and their original models. The real point of this adjustment is the following.

The above definition of atomic formulas for path objects now makes binary relations hold only between objects $(\pi_1, d_1), (\pi_2, d_2)$ where π_2 is a one-step continuation of the path π_1, or vice versa. But then, counting distance as we did before,

> The new object at the end of a path of length *k*
> lies at distance *k* from the initial object of the path.

Put in more vivid terms, 'tree distance is true distance' in the original model. This is a non-obvious fact. E.g., with *ternary* guards *Rayz*, objects at the end of a path may keep links to the initial object *a* which might recur in the guarded sets building the path.

Proof of Theorem 4 As before, it suffices to show that an arbitrary formula ϕ in *L* is invariant for models that are equivalent for all GF_{bin}-formulas up to some finite depth *n*.

First, largely as in the earlier modal proof of Section 5, we first use the Compactness of *L*, together with its Relativization closure, to show that ϕ must have a Finite Distance Property at some level *n*. Before, universal prefix formulas $[]^k p$ (for all finite *k*) made sure that *p* holds in the generated submodel at the current world. This time, one uses all nested sequences of universal guarded quantifiers up to depth *k*, requiring that some new predicate *P* holds for all objects reached at the end. The *n* thus found for the local depth of the formula ϕ is the same *n* as needed for the following semantic invariance:

Next, given the above unraveling construction and Invariance for Guarded Bisimulation for *L*, we may assume, without loss of generality, that we have the following situation:

(a) $(Unr(\mathbf{M}), \mathbf{s}) \models \phi$,

(b) $(Unr(\mathbf{M}), \mathbf{s})$ is GF_{bin}-*n*-equivalent to $(unr(\mathbf{N}), \mathbf{t})$

Our aim is to show that $(Unr(\mathbf{N}), \mathbf{t}) \models \phi$.

Now, we cut the tree models to tree depth *n*, as we did before in our modal argument, obtaining $Cut((Unr(\mathbf{M}), \mathbf{s}), n)$, $Cut((Unr(\mathbf{N}), \mathbf{t}), n)$. Since tree depth is true depth, this does not change truth values of ϕ. in either model.

Next, we have as usual that there is an *n-tower* of partial isomorphisms $PI_n, ..., PI_0$, starting from the link **s**, **t**, which satisfies the guarded back and forth properties. Stage *j* contains tuples (in fact, guarded pairs) of objects satisfying the same GF_{bin} formulas up to syntactic depth *n-j*. The back and forth properties are proved by using, at stage *j+1*, guarded existential quantifiers describing the next object to be linked up to syntactic level *j*. In particular, working in our tree models, we can make sure that

(#) The finite partial isomorphisms at level *PIj* of the tower ($j<n$) are between guarded sets of two objects lying at distances *n-k*, *n–k-1* from the root objects, which satisfy the same formulas of GF_{bin}-*k*.

Now our crucial claim, as in the basic modal case, is that

The union of all sets PI_j is a GF_{bin}-bisimulation.

The only thing to be checked here is that the partial isomoprhisms in PI_0 still satisfy the guarded back and forth properties. In the purely modal case, this was because end points of the cut-off tree have no successors, and hence there is nothing to be proved. In the present case, points at distance *n* may have more significant relationships with objects in the cut-off tree, but given our definition of atomic predicates for path objects, these can only be of special forms. If the objects *u, v* in the guarded pair lie at distance $<n$ from a root object, we are done by (#). And if, say, *u* lies at distance *n* from the root, the only significant binary relationship it can have in our cut-off tree model is with its companion object *v* – which was unique, by our path construction. But then, the current partial isomorphism *itself* proveides the required back-and-forth match.

Finally, by invariance under guarded bisimulations, the truth of the given *L*-sentence ϕ is transferred from *Cut((Unr(**M**), **s**), n)* to *Cut((Unr(**N**), **t**), n)*, and hence to *(**N**, **t**)*.

This completes the proof of a Lindström Theorem for the Binary Guarded Fragment. ♠

There are also other generalizations of this tree unraveling and finite-depth argument. E.g., let the 'rightward Guarded Fragment' $GF^{\rightarrow}$ have atoms with arguments in left-to-right order, while quantification does not let initial worlds return in the matrix formula:

$$\exists \mathbf{y}\ (G\mathbf{x}, \mathbf{y}\ \&\ \phi(\mathbf{y})).$$

This language is close to so-called 'polyadic modal languages'. The semantic notions of guarded bisimulation and tree unraveling specialize to $GF^{\rightarrow}$ in an obvious manner.

Theorem 5 An abstract logic *L* extending the rightward Guarded Fragment equals $GF^{\rightarrow}$ iff *L* has (a) Invariance for Guarded Bisimulation, and (b) Compactness.

Proof (sketch) One shows that an arbitrary formula ϕ in *L* is invariant across models which are equivalent for all $GF^{\rightarrow}$-formulas up to some finite depth *n* (*).To find this number *n*, one first uses the Compactness of *L* and its closure under Relativization to show that ϕ has a Finite Distance Property at some level *n*. (Instead of modal formulas $[]^k p$ for all finite *k*, one now uses nested sequences of universal guarded quantifiers up to length *k*, requiring that some new predicate *P* holds for all objects thus reached.) To prove the semantic fact (*), we use invariance for rightward guarded bisimulation for *L*, and the matching tree unraveling. We start from: (a) *(Unr(**M**), **s**)* |= ϕ, (b) *(Unr(**M**), **s**)* is $GF^{\rightarrow}$-*n*-equivalent to *(unr(**N**), **t**)*, and show that *(Unr(**N**), **t**)* |= ϕ by essentially the 'cut-off bisimulation' construction at level *n* used in the basic modal case. ♠

Lifting this style of argument to the full *GF* raises complications. In particular, for arbitrary guarded formulas, the above finite distance property does not coincide with the natural 'tree distance' in models *(Unr(**M**), **s**)* measured from the root ***s*** – something which would be needed to make our base argument go through. The reason is that, even at greater 'tree distance' in unraveled models, objects may still enter into atomic relationships with the root ***s***, and hence stay at distance *0* from ***s*** according to our definition. This fact becomes relevant as soon as we allow general guarded quantifications $\exists \boldsymbol{y}\ (G(\boldsymbol{x}, \boldsymbol{y})\ \&\ \phi(\boldsymbol{x}, \boldsymbol{y}))$ where the initial ***x*** can occur in the matrix formula.

Martin Otto (p.c.) has high-lighted this difficulty with a formula which has finite distance *1* from the root *s*, and which is also invariant for guarded bisimulations, without being definable in *GF*. Otto's formula says that

> there exists a finite sequence of worlds $y_1, .., y_k$ such that Py_1, Qy_k, and Rxy_iy_{I+1} for all i, $1 \leq i < k$. Clearly, this is not in *GF*, by a simple compactness argument.

We have some further ideas how to overcome this, but nothing conclusive as yet. Indeed, could it be that the failure of Craig Interpolation for *GF*, discovered by Hoogland, Marx and Otto is relevant here, with a more sinister interpretation?

7 Connections with the modal invariance theorem

The modal Lindström theorem of Section 5 (let us call it *MLT* for short) looks a bit like the much earlier modal *Invariance Theorem* (*MIT*) which says that,

> Up to logical equivalence, the basic modal formulas are precisely
> those first-order formulas which are invariant for bisimulation.

Indeed, one implication between the two results can be proved in general terms:

Theorem 6 *MLT* implies *MIT*.

Proof Let $\phi(x)$ be any bisimulation-invariant first-order formula $\phi = \phi(x)$. Define an abstract logic *L* by adding ϕ to the basic modal language, and then closing off the result (in some suitable syntax) under (a) Boolean operations, (b) existential modalities <>, and (c) an operation of relativization in the inductive format $Rel(\alpha, \beta)$ where α, β are already formulas of the language. The semantic interpretation of this language on models *(**M**, s)* is obvious. This language contains the basic modal language, and also, it can be translated into a fragment of first-logic. The latter feature implies Compactness. Next, we prove the following observation by induction on *L*-formulas:

Claim All *L*-formulas are invariant for bisimulation.

Proof The inductive cases except that for relativization are obvious, following the usual proof of bisimulation invariance for modal formulas, plus the given invariance for ϕ. Next consider a formula $Rel(\alpha, \beta)$, where we already assume invariance for α, β. Let *E* now be a bisimulation between two models $(\boldsymbol{M}, s)$ and $(\boldsymbol{N}, t)$, while $(\boldsymbol{M}, s) \models Rel(\alpha, \beta)$. By definition, we have that $(\boldsymbol{M}|\alpha, s) \models \beta$. We observe that

> The relation $E|\alpha$ consisting of all pairs in *E* which connect α-worlds in ***M*** to α-worlds in ***N*** *is itself a bisimulation* between $(\boldsymbol{M}|\alpha, s)$ and $(\boldsymbol{N}|\alpha, t)$.

To check the zigzag clause here, suppose that $(\boldsymbol{M}, u) \models \alpha$ and $u\ E\ v$, with $(\boldsymbol{N}, v) \models \alpha$. Let *Ruu'* in ***M*** with $(\boldsymbol{M}, u') \models \alpha$. Since *E* is a bisimulation, there exists a world *v'* in ***N*** with *Rvv'* and *u'Ev'*. But by the inductive hypothesis, then, $(\boldsymbol{N}, v') \models \alpha$.. Thus, we have shown the required zigzag property for the relation $E|\alpha$. We may then conclude, again by the inductive hypothesis, that $(\boldsymbol{N}|\alpha, t) \models \beta$ – and hence that $(\boldsymbol{N}, t) \models Rel(\alpha, \beta)$.

Thus, *L* satisfies all conditions of Theorem 3, and hence, in particular, the *L*-formula ϕ must be equivalent to a modal formula. ♠

Incidentally, this also seems to be a new proof of the *MIT*!

It is unclear, however, if *MLT* is implied by *MIT*. To get this converse, one would first have to show that any formula ϕ in the abstract logic *L* is first-order, and then apply the *MIT*. But, even though bisimulation invariance implies invariance for potential isomorphism, first-orderness for ϕ does not follow obviously from the earlier Lindström theorem for *FOL*. The difficulty is, again, that one cannot use the full coding power of first-order logic, as we only know that *L* contains the basic modal language.

Remark Even so, the key property in the usual proofs of *MIT* looks 'Lindström-like'. It says that any two models $(\boldsymbol{M}, s)$ and $(\boldsymbol{N}, t)$ which are modally equivalent have elementary equivalent models $(\boldsymbol{M}^+, s)$ and $(\boldsymbol{N}^+, t)$ with some bisimulation between them connecting the distinguished world *s* to *t*. Replacing elementary extension by *L*-equivalence would do the trick here. We leave this line for further investigation.

8 Connections with interpolation

Our modal Lindström theorem is in the spirit of Barwise's work on abstract model theory of infinitary languages, in that it replaces the Löwenheim-Skolem property used in the original result by a semantic invariance condition. This reformulation also has a disadvantage. The original Lindström conditions yield an *interpolation theorem*:

> Any abstract logic *L* containing *FOL* which has Compactness and Löwenheim-Skolem also satisfies 'first-order separation of projective classes': if ***K***, ***K'*** are two disjoint *PC* classes for *L*, then there exists a first-order formula ϕ with $\boldsymbol{K} \subseteq MOD(\phi)$ and $\boldsymbol{K'} \cap MOD(\phi) = \varnothing$.

As it happens, we can prove a similar result in our modal setting: this time, relying on the Finite Depth Property. We consider only one of the possible versions of modal interpolation here, viz. with respect to shared *proposition letters*. The standard version of modal interpolation in this setting reads as follows:

> If $\phi(\boldsymbol{p}, \boldsymbol{q}) \models \psi(\boldsymbol{q}, \boldsymbol{r})$, then there exists a formula $\alpha(\boldsymbol{q})$ with $\phi \models \alpha$ and $\alpha \models \psi$

But the basic modal language has an even stronger version of this property (which does not hold, e.g., for first-order logic), called *uniform interpolation*:

> The modal interpolant $\alpha(\boldsymbol{q})$ depends only on the antecedent formula $\phi(\boldsymbol{p}, \boldsymbol{q})$ and the specified sublanguage $\boldsymbol{q}$.

Theorem 7 Any abstract modal logic with Relativization which extends the basic modal language, while also satisfying Compactness and Bisimulation Invariance, satisfies Uniform Interpolation.

Proof Consider any *L*-formula $\phi(\boldsymbol{p}, \boldsymbol{q})$ with similarity type $\boldsymbol{p}$, $\boldsymbol{q}$. By the proof of Theorem 3, ϕ is invariant for some finite successor depth *k*. Now, define $Cons(\phi, \boldsymbol{q}, k)$ as the disjunction of all complete modal theories in the language with $\boldsymbol{q}$ only up to depth *k* of all models for $\phi(\boldsymbol{p}, \boldsymbol{q})$. This can be done with one finite formula, given the logical finiteness of the basic modal language. This is the required uniform interpolant for the vocabulary $\boldsymbol{q}$. To see this, let $\phi(\boldsymbol{p}, \boldsymbol{q}) \models \psi(\boldsymbol{q}, \boldsymbol{r})$ – meaning that this implication holds in models whose interpreted proposition letters include those mentioned.

Claim $Cons(\phi, \boldsymbol{q}, k) \models \psi(\boldsymbol{q}, \boldsymbol{r})$.

Let $(\boldsymbol{M}, s) \models Cons(\phi, \boldsymbol{p}, k)$ where $\boldsymbol{M}$ interprets $\boldsymbol{q}$, $\boldsymbol{r}$. By the definition of $Cons(\phi, \boldsymbol{p}, k)$, there is some model $(\boldsymbol{N}, t) \models \phi$ interpreting $\boldsymbol{p}$, $\boldsymbol{q}$, whose modal theory for $\boldsymbol{p}$-formulas up to depth *k* equals that of $(\boldsymbol{M}, s)$. Now we unravel both models to bisimilar trees $Tree(\boldsymbol{M}, s)$ and $Tree\ (\boldsymbol{N}, t)$, where the bisimulations refer to the appropriate similarity types, with $\boldsymbol{q}$, $\boldsymbol{r}$ and $\boldsymbol{p}$, $\boldsymbol{q}$, respectively. These trees still have the same modal $\boldsymbol{q}$-theory up to depth *k* in their roots, and hence there is an obvious *k*-step descending bisimulation *E* (w.r.t. the modal language of $\boldsymbol{q}$) between their roots *s* and *t*. Now cut off both trees at depth *k*: the result is a genuine bisimulation *E* between the cut-off trees, while ϕ still holds in *(Cut(Tree(**N**, t)), k), t)* by the finite depth property at *k* for the formula ϕ:

***M**, s*	***q**, k*-equivalent	***N**, t* |= ϕ
bisimulation ***q**, **r***		bisimulation ***p**, **q***
*Tree (**M**, s), s*	***q**, k*-equivalent	*Tree (**N**, t), t* |= ϕ
*Cut(Tree (**M**, s), k), s*	**q**-bisimulation *E*	*Cut(Tree (**N**, t), k), t* |= ϕ

The essential proof step is as follows. The ***q***-bisimulation *E* between the cut-off trees is not by itself a ***p**, **q***-bisimulation, since it only respect proposition letters in ***q***. E.g., on the left hand side, there may be one tree successor at a node, while the matching node under *E* on the right has two successors: one with *p* and one with ¬*p*. But, we can *improve E* to a bisimulation for the whole language with ***p**, **q*** between *expanded models* by a familiar modal technique of expanding models to bisimilar ones by copying successor nodes (there are various technical ways of doing this):

> We multiply nodes in *Cut(Tree (**M**, s), k)* until each node has enough successors to copy the ***p***-propositions from nodes of *Cut(Tree (**N**, t), k)*. In doing so, we adapt the bisimulation *E* in the obvious way, while also keeping track of the copy relation on the left. In this way, we get a tree model *(**K**, s)* which is (a) ***p**, **q***–bisimilar to *Cut(Tree (**N**, t), k), t,* and also (b) ***q**, **r***-bisimilar to *Cut(Tree (**M**, s), k), s.*

Here we can assume that the tree *(**K**, s)* has branch depth at most *k*. Next. we can make this tree fully ***q**, **r***-bisimilar to *Tree(**M**, s)* by suitably extending it at levels beyond *k*. In particular, at each end node *x* of *(**K**, s),* corresponding to some node *y* at level *k* in *Tree(**M**, s*) by the bisimulation of clause (b) above, place a copy of the subtree of *Tree(**M**, s*) starting at *y*. Call this extended tree model *($\boldsymbol{K}^+$, s).*

Now we analyze transfer of *L*-formulas in this setting. First, since *Cut(Tree (**N**, t), k), t* \|= ϕ, we also have *(**K**, s)* \|= ϕ, by the invariance of all *L*-formulas for bisimulation in their similarity type. Moreover, by the *k*-depth property of ϕ, we still have that *($\boldsymbol{K}^+$, s)* \|= ϕ. Next, by the original assumption that ϕ*(**p**, **q**)* \|= ψ*(**q**, **r**),* we have that *($\boldsymbol{K}^+$, s)* \|= ψ. Then, composing the ***q**, **r***-bisimulation between *($\boldsymbol{K}^+$, s)* and the model *Tree (**M**, s), s,* and the further unraveling ***q**, **r***-bisimulation from that to the original model *(**M**, s),* we have that *(**M**, s)* \|= ψ – again by the invariance of *L*-formulas under bisimulation. ♠

MLT obviously follows from this for languages which are closed under negations.

The above proof really distinguishes between different vocabulary types for formulas, and demands their invariance only for bisimulations with respect to that vocabulary. It should be obvious how the preceding text can be sharpened up along those lines.

There is more to interpolation in a modal setting. Ten Cate's 2005 dissertation shows that even standard Craig Interpolation is scarce among (hybrid) fragments of *FOL*. This suggests that we may be fishing in a very small pond here. Also, our analysis has not addressed the stronger form of modal interpolation in the case of shared relations for indexed modalities *<a>, <b>, ...* We leave this as an open question here.

Interpolation is often considered a sort of optimal design feature for logical languages. Languages without it somehow have not reached expressive maturity yet, and may therefore need to be extended. Likewise, our Lindström theorem states a property expressing a sort of general *design adequacy* for any modal-like first-order language. The formalism has to be the strongest one within the first-order area (this is where Compactness puts us) satisfying invariance for the chosen notion of bisimulation.

9 Discussion: extensions and challenges

Our results still leave many questions unanswered. Here are a few.

Does our type of proof work for still further modal languages? One obvious direction are other formalisms allowing for Lindström-type characterizations, such as *infinitary modal logic* (cf. Barwise & van Benthem 1999). In this case, we lose Compactness, but we have a substitute, viz. the projective undefinability of well-ordering. On the other hand, *definability* of well-ordering is typical for the widespread modal languages with *fixed-point operators* (such as the μ–calculus) – though Lindström-type results seem unknown there in general, also for the extended first-order language *LFP(FO).*

Next, consider first-order languages, and in particular, *extended modal* ('hybrid') *languages* in between basic modal logic and full *FOL*. Our proof still works for such languages if they have a decent Finite Depth Property (*FDP*). E.g., Balder ten Cate has given a Lindström theorem for such a language with additional 'graded modalities'. But our method fails for extended modal languages lacking *FDP*. The above issues with the Guarded Fragment were a clear illustration. For a simpler illustration, consider the basic modal language extended with a *universal modality (MLU)* saying that

$$\boldsymbol{M}, s \models U\phi \quad \text{iff} \quad \boldsymbol{M}, t \models \phi \text{ for all worlds } t.$$

An extended modal formula like *U<>T* says that every world should have a successor, and this requirement is active at any depth of the model. So, we have to find a different way of encoding things, closer to the original first-order proof, using the additional strength of the unbounded modality *U*. The preceding draft of this paper had a proof for the obvious conjecture here, viz. that an abstract modal logic *L* extending *MLU* equals the latter language iff *L* satisfies (a) Invariance for Total Bisimulation, and (b) Compactness. Here a *total bisimulation* between two models is an ordinary modal bisimulation having one model for its domain and the other for its range. But our proof has run aground – so we must leave this result open.

Finally, Balder ten Cate has observed that our Lindström theorem even holds on *finite models*. To achieve this, we need to replace Compactness (which fails on finite models) by the following condition implied by it, which does hold on finite models:

> If ϕ globally implies ψ (that is, all models where ϕ holds throughout also have ψ true throughout) , then there exists some finite number *k* such that $\phi, []\phi, ..., []^k\phi$ locally imply ψ, in all models *(**M**, s)* at the special world *s*.

Going back to our proof in Section 5, this does the main trick in establishing the *FDP*. After that, the rest of our argument also works for finite models only.

Summarizing, we seem to have made a few steps beyond what was known in the area, making nice Lindström theorems for weaker languages than *FOL* more of a reality. But given the narrow scope of our analysis so far, clearly, we need to think more about the right 'spectacles' to chart the area of formalisms below first-order logic.

10 Acknowledgment

I thank Balder ten Cate, MariCarmen Martinez, Maarten Marx, Tomasz Sadzik, and Jouko Väänänen for their helpful comments on earlier drafts of this paper. In particular, Jouko asked me the initial question about finding 'better' modal Lindström theorems, Maarten asked the one about connections with the modal invariance theorem, while Balder supplied further proofs and new ideas, e.g., about extensions to finite models. Finally, special thanks are due to Martin Otto for pointing out the delicate situation with Lindstrom results for the full Guarded Fragment.

11 References

H. Andréka, J. van Benthem & I. Németi, 1998, 'Modal Languages and Bounded Fragments of Predicate Logic', *Journal of Philosophical Logic* 27, 217–274.

J. Barwise & J. van Benthem, 1999, 'Interpolation, Preservation, and Pebble Games', *Journal of Symbolic Logic* 64:2, 881–903.

J. Barwise & S. Feferman, eds., 1985, *Model-Theoretic Logics,* Springer, New York.

J. van Benthem, 1996, *Exploring Logical Dynamics*, CSLI Publications, Stanford.

J. van Benthem, 2005, 'Guards, Bounds, and Generalized Semantics', *Journal of Logic, Language and Information* 14, 263 - 279.

J. van Benthem & P. Blackburn, 2005, 'Modal Logic, a semantic perspective', draft chapter for P. Blackburn, J. van Benthem & F. Wolter, eds., *Handbook of Modal Logic*, Elsevier, Amsterdam, to appear.

P. Blackburn, M. de Rijke & Y. Venema, 2000, *Modal Logic*, Cambridge University Press, Cambridge.

B. ten Cate, 2005, *Model Theory for Extended Modal Languages*, dissertation, Institute for Logic, Language and Computation, University of Amsterdam.

M. Garcia-Matos, 2005, *Abstract Model Theory for Negation-Free Languages*, Mathematical Institute, University of Helsinki.

E. Grädel, 1999, 'The Decidability of Guarded Fixed Point Logic', in *JFAK. Essays Dedicated to Johan van Benthem on the Occasion of his 50th Birthday* (J. Gerbrandy, M. Marx, M. de Rijke, and Y. Venema, eds.), CD-ROM http://turing.wins.uva.nl/~j50/cdrom/, Amsterdam University Press.

M. de Rijke, 1993, *Extending Modal Logic*, dissertation, Institute for Logic, Language and Computation, University of Amsterdam.