Interpolation, Preservation, and Pebble Games

Jon Barwise and Johan van Benthem

August 12, 1996

Abstract

Preservation and interpolation results are obtained for $L_{\infty\omega}$ and sublogics $L \subseteq L_{\infty\omega}$ such that equivalence in L can be characterized by suitable back-and-forth conditions on sets of partial isomorphisms.

1 Introduction

In the heyday of infinitary logic in the 1960's and 70's, most attention was focused on $L_{\omega_1\omega}$ and its fragments (see e.g. Keisler [19]), since countable formulas seemed best behaved. The past decade has seen a renewed interest in $L_{\infty\omega}$ and its finite variable fragments $L^{(k)}_{\infty\omega}$ (for $2 \leq k < \omega$) and the modal fragment $L^{\diamond}_{\infty\omega}$ (see e.g. Ebbinghaus and Flum [17] on the former and Barwise and Moss [9] on the latter), due to various connections with topics in computer science. These logics form a hierarchy of increasingly powerful logics

$$L^{\diamond}_{\infty\omega} \subset L^{(2)}_{\infty\omega} \subset L^{(3)}_{\infty\omega} \subset \ldots \subset L^{(k)}_{\infty\omega} \subset \ldots \subset L_{\infty\omega},$$

with each of these inclusions being proper. Moreover, there is a useful and elegant algebraic characterization of equivalence in L in each of these logics L, from bisimilarity at the left end to potential isomorphism at the right. These algebraic relations on structures are often defined by what is called a "pebble game," a game between two players, one of whom wants to show that the structures are not equivalent, the other who wants to show they are.

Most of the recent work involving the finite variable fragments $L^{(k)}_{\infty\omega}$ has focused on finite structures. As a result, there are many open questions regarding these logics when one drops the restriction to finite structures – questions that should really have been settled in the 70's, but were overlooked, probably due to the preoccupation with $L_{\omega_1\omega}$. We address a certain class of these questions in this paper, namely, questions involving interpolation and preservation. The following is a sample result: if a sentence φ of $L_{\infty\omega}$ is preserved under bisimulation then it is logically equivalent to a sentence of the modal fragment $L^{\diamond}_{\infty\omega}$. We obtain this as a corollary to an interpolation theorem relating $L_{\infty\omega}$ and $L^{\diamond}_{\infty\omega}$.

A glance through this paper will show that we have obtained a number of results of this same general character: an interpolation theorem with a preservation theorem as consequence. The proofs all use a single method. We introduce this method by proving an interpolation theorem for $\mathrm{L}_{\infty\omega}$ in this section. In section 2 we extract a general method by analyzing this proof of interpolation in terms of "co-inductive pebble games." This analysis results in a certain "abstract interpolation theorem" which is then applied to yield other results in section 3. Section 4 discusses some open questions.

Notation

We assume the reader is familiar with first-order model theory. Our notation is fairly standard. By a *vocabulary* we mean a set L of relation and constant symbols. Most of what we do extends to function symbols, though not the results having to do with $\mathrm{L}^{(k)}_{\infty\omega}$. Given an L-structure $\boldsymbol{A}$, a formula φ, and a function s assigning elements of $\boldsymbol{A}$ to the free variables in φ, we write $\boldsymbol{A}, s \models \varphi$ and $\boldsymbol{A} \models \varphi[s]$ interchangeably to indicate that s satisfies the formula φ in the structure $\boldsymbol{A}$. We write A for the universe of $\boldsymbol{A}$ and $R^{\boldsymbol{A}}$ for the interpretation of the symbol R in $\boldsymbol{A}$.

Given a vocabulary L, the language $\mathrm{L}_{\infty\omega}$ is the usual language over L that allows finite strings of quantifiers, negations, and arbitrary conjunctions and disjunctions. $\mathrm{L}_{\omega\omega}$ is the fragment of $\mathrm{L}_{\infty\omega}$ allowing only finite conjunctions and disjunctions. $\mathrm{L}_{\omega_1\omega}$ is the fragment that allows only countable conjunctions and disjunctions. Given a natural number k, $2 \leq k < \omega$, the *k-variable fragment* $\mathrm{L}^{(k)}_{\infty\omega}$ of $\mathrm{L}_{\infty\omega}$ consists of those formulas of $\mathrm{L}_{\infty\omega}$ which have been constructed using at most k constants and variables, free or bound.

We prove results for $\mathrm{L}_{\infty\omega}$ and its fragments. Some of these results require that we be able to perform all the usual operations from first-order logic. We call such fragments *regular*. These include $\mathrm{L}_{\omega\omega}, \mathrm{L}_{\omega_1\omega}$, admissible fragments, fragments of the form $L_{\kappa\omega}$, and so forth. The finite variable fragments are not regular. By an *infinitary* fragment we mean one that is closed under at least recursive conjunctions and disjunctions.

Given a fragment L of $\mathrm{L}_{\infty\omega}$, we write $\Sigma^1_1(\mathrm{L})$ to indicate a second-order sentence obtained from a sentence of L by means of second-order existential quantification, or, if we are talking about a class $\mathcal{C}$ of structures, to indicate that $\mathcal{C}$ is definable by such a sentence. We write Σ^1_1 for $\Sigma^1_1(\mathrm{L}_{\omega\omega})$. We define $\Pi^1_1(\mathrm{L})$ and Π^1_1 dually.

Given a vocabulary L and L-structures $\boldsymbol{A}$ and $\boldsymbol{B}$ we write $(\boldsymbol{A}, \boldsymbol{B})$ for the structure that incorporates both structures in some standard way. Thus, for example, one could consider it as an L^* structure, where L^* contains L, a disjoint copy L' of L, and two unary predicates U, V such that A is the interpretation of U in $(\boldsymbol{A}, \boldsymbol{B})$ and similarly for B and V.

We write $f : \boldsymbol{A} \cong \boldsymbol{B}$ to mean that f is an isomorphism of the L-structures, and $\boldsymbol{A} \cong \boldsymbol{B}$ if $f : \boldsymbol{A} \cong \boldsymbol{B}$ for some f. By a *partial isomorphism* from $\boldsymbol{A}$ to $\boldsymbol{B}$

we mean a partial function with domain contained in A and range contained in B that preserves structure: if $a_1, \ldots, a_n \in dom(f)$ then for each n-ary predicate P of L,

$$\langle a_1 \ldots a_n \rangle \in P^{\boldsymbol{A}} \text{ iff } \langle f(a_1) \ldots f(a_n) \rangle \in P^{\boldsymbol{B}}$$

and if $a \in dom(f)$ then for each constant symbol c of L, $a = c^{\boldsymbol{A}}$ iff $f(a) = c^{\boldsymbol{B}}$. Not every interpretation of a constant of L need be in the domain of f.

The following definition allows us to express our results in a succinct form.

Definition 1 If R is a binary relation on L-structures then φ *entails* ψ *along* R, written $\varphi \models_R \psi$, iff for all L-structures $\boldsymbol{A}, \boldsymbol{B}$, if $\boldsymbol{A}\ R\ \boldsymbol{B}$ and $\boldsymbol{A} \models \varphi$ then $B \models \psi$. In particular, φ is *preserved under* R if $\varphi \models_R \varphi$.

For example, all sentences are preserved under isomorphism, and entailment along isomorphism just amounts to the usual notion of logical entailment. Existential sentences are preserved under embeddings. Positive sentences are preserved under homomorphic images.

1.1 Interpolation for $\mathrm{L}_{\infty\omega}$

There are several equivalent ways to formulate the interpolation property for a logic L. The one we find convenient is this. We say that L *has the standard interpolation property* if for all $\varphi \in \Sigma^1_1(\mathrm{L})$ and $\psi \in \Pi^1_1(\mathrm{L})$ in a common vocabulary L, if φ entails ψ, then there is a sentence $\theta \in \mathrm{L}$ in the same vocabulary L such that $\varphi \models \theta$ and $\theta \models \psi$. Such a θ will be called an *interpolant* between φ and ψ.

First-order logic $\mathrm{L}_{\omega\omega}$ and the infinitary logic $\mathrm{L}_{\omega_1\omega}$ have the standard interpolation property (due to Craig and to Lopez-Escobar [20], respectively), as do admissible fragments of $\mathrm{L}_{\omega_1\omega}$ (Barwise [5]). Not many other logics have the standard interpolation property, however. In particular, $\mathrm{L}_{\infty\omega}$ fails to have the standard interpolation property, a result due to Malitz [22].

The standard counterexample

It is instructive to recall one of the familiar counterexamples to interpolation for $\mathrm{L}_{\infty\omega}$. Let $\phi(<)$ be the sentence using $<$ and $=$ that says that the universe A of the structure $\boldsymbol{A}$ is wellordered by $<^{\boldsymbol{A}}$ in order type ω_1. Similarly, let $\psi(<)$ say that A is wellordered by $<^{\boldsymbol{A}}$ in of order type ω. Clearly $(\exists<)\varphi(<)$ entails $(\forall<)\neg\psi(<)$ since no set that can be wellordered in order type ω_1 can be wellordered in order type ω.

Suppose that θ is an $\mathrm{L}_{\infty\omega}$ interpolant for these two sentences, that is, a sentence implied by $(\exists<)\varphi(<)$ and implying $(\forall<)\neg\psi(<)$, and involving only the symbol $=$. Since any set of size $\aleph_1$ is a model of φ, θ is true in all models of size $\aleph_1$. Similarly, θ is true in no set of size $\aleph_0$. However, all infinite structures are known to be indistinguishable in $\mathrm{L}_{\infty\omega}$, if we use only $=$. So this is impossible.

Let's go over this last step a bit more carefully, since the reason it works is crucial to an understanding of the results of this paper.

Karp's Theorem

Let I be a non-empty set of partial isomorphims from $\boldsymbol{A}$ to $\boldsymbol{B}$. We say that I is a *potential isomorphism* between L-structures $\boldsymbol{A}$ and $\boldsymbol{B}$, written $I : \boldsymbol{A} \cong_p \boldsymbol{B}$, if I has the following back and forth property:

Back For every $f \in I$ and every $b \in B$ there is an $f' \supseteq f$ that is in I and has b in its range, and

Forth For every $f \in I$ and every $a \in A$ there is an $f' \supseteq f$ that is in I and has a in its domain.

We say that $\boldsymbol{A}$ is potentially isomorphic to $\boldsymbol{B}$, and write $\boldsymbol{A} \cong_p \boldsymbol{B}$, if there is an I such that $I : \boldsymbol{A} \cong_p \boldsymbol{B}$.

As an example, any two dense linear orderings satisfying the same end point conditions are potentially isomorphic. (By contrast, it is not difficult to see that if $\boldsymbol{A} \cong_p \boldsymbol{B}$ and $<^{\boldsymbol{A}}$ is a wellordering of A, then $\boldsymbol{A} \cong \boldsymbol{B}$.) More generally, if $\boldsymbol{A}$ and $\boldsymbol{B}$ are any two infinite models of a countable $\aleph_0$-categorical theory (of $\mathrm{L}_{\omega\omega}$ or $\mathrm{L}_{\omega_1\omega}$) then $\boldsymbol{A} \cong_p \boldsymbol{B}$ (see Barwise [3]). In particular, in connection with the above counterexample to interpolation, if $\boldsymbol{A}$ and $\boldsymbol{B}$ be any two infinite sets (structures for the vocabulary that has only $=$ as a symbol), and I is the set of all finite one-one functions with domain a subset of $\boldsymbol{A}$ and range a subset of $\boldsymbol{B}$, then $I : \boldsymbol{A} \cong_p \boldsymbol{B}$. Thus, the last step in the proof of the counterexample to interpolation is an immediate consequence of the left to right half of the following famous result:

Theorem 1 (Karp's Theorem) *$\boldsymbol{A} \cong_p \boldsymbol{B}$ iff $\boldsymbol{A}$ and $\boldsymbol{B}$ are $\mathrm{L}_{\infty\omega}$-equivalent.*

The proof, originally given in Karp [18], can also be found in Chang [15] and Barwise [3, 5].

A simple back and forth argument shows that for countable $\boldsymbol{A}$ and $\boldsymbol{B}$, $\boldsymbol{A} \cong_p \boldsymbol{B}$ iff $\boldsymbol{A} \cong \boldsymbol{B}$. (Indeed, the reason for calling $\boldsymbol{A} \cong_p \boldsymbol{B}$ potential isomorphism comes from the fact that $\boldsymbol{A} \cong_p \boldsymbol{B}$ holds iff $\boldsymbol{A} \cong \boldsymbol{B}$ holds in some Boolean extension of the universe where $\boldsymbol{A}$ and $\boldsymbol{B}$ are countable.)

The relation $\boldsymbol{A} \cong_p \boldsymbol{B}$ is often formulated in game-theoretic terminology, since this provides a convenient way to think about it when trying to prove that structures are or are not potentially isomorphic. Imagine that two players, often called Spoiler and Duplicator, are playing a game. Spoiler wants to show that $\boldsymbol{A}$ and $\boldsymbol{B}$ are not similar, Duplicator wants to show they are. At each stage of the game the previous play has established a partial isomorphism f between substructures of $\boldsymbol{A}$ and $\boldsymbol{B}$ (the empty function at the start of play). Spoiler picks an element from one of the structures. The Duplicator must pick an element from the other structure so that $f \cup \{\langle a, b\rangle\}$ is still a partial isomorphism. If

Duplicator can do this, then play reverts to the Spoiler, if not, Spoiler wins. If the play goes on forever, then Duplicator wins. It is easy to see that $\boldsymbol{A} \cong_p \boldsymbol{B}$ is equivalent to Duplicator having a winning strategy in this game. (We'll give a similar proof later.)

This kind of game is often called a "pebble" game because one thinks of the players as placing variously colored pebbles on the elements of $\boldsymbol{A}$ and $\boldsymbol{B}$, with pebbles of the same color indicating the partial isomorphism. In the pebble game for $\mathrm{L}^{(k)}_{\infty\omega}$, each player has a stock of only k pebbles, but the players are allowed to pick them up from one place and put them elsewhere.

1.2 An interpolation theorem for $\mathrm{L}_{\infty\omega}$

Suppose $\varphi \in \Sigma^1_1(\mathrm{L}_{\infty\omega})$ and $\psi \in \Pi^1_1(\mathrm{L}_{\infty\omega})$, and there is an interpolant $\theta \in \mathrm{L}_{\infty\omega}$ such that $\varphi \models \theta$ and $\theta \models \psi$. A simple use of Karp's Theorem shows that φ entails ψ along *potential* isomorphism, not just along isomorphism. The counterexample to the standard interpolation reviewed above really amounts to an example that is preserved along potential isomorphism but not along isomorphism. Thus, in view of Karp's Theorem, for $\mathrm{L}_{\infty\omega}$ a stronger form of entailment is required:

Theorem 2 (Interpolation Theorem for $\mathrm{L}_{\infty\omega}$) *Given $\varphi \in \Sigma^1_1(\mathrm{L}_{\infty\omega})$ and $\psi \in \Pi^1_1(\mathrm{L}_{\infty\omega})$, the following are equivalent:*

1. *φ entails ψ along potential isomorphism.*
2. *There is a sentence θ of $\mathrm{L}_{\infty\omega}$ such that $\varphi \models \theta$ and $\theta \models \psi$.*

The proof of this result is not difficult, given some well-known facts about $\mathrm{L}_{\infty\omega}$ which we now review, starting with the following.

Theorem 3 (Boundedness Theorem) *If $\mathcal{C}$ is a $\Sigma^1_1(\mathrm{L}_{\infty\omega})$ definable class of wellorderings then some ordinal α is greater than the length of every wellordering in $\mathcal{C}$.*

The history of this result is a bit complicated. Using Karp's analysis of $\mathrm{L}_{\infty\omega}$-equivalence, Lopez-Escobar [21] gave an analysis of the $\mathrm{L}_{\infty\omega}$-definable well-orderings, from which the above result follows in the special case where $\mathcal{C}$ is explicitly definable using only $<$ and identity. The result for $\Sigma^1_1(\mathrm{L}_{\infty\omega})$ classes is much more complicated, and requires a model-construction argument. In the form stated here, the result is due to Barwise and Kunen [7], generalizing a similar result of Morley [23] for $\mathrm{L}_{\omega_1\omega}$; a proof can also be found in Barwise [5]. It is interesting to note that our Interpolation Theorem allows one to turn the tables and reduce the Boundedness Theorem to Lopez-Escobar's result. (Consider the Π^1_1 sentence $\psi(<)$ that asserts $<$ is a wellordering. Then any $\Sigma^1_1(\mathrm{L}_{\infty\omega})$ sentence defining a class of well-orderings entails ψ along $\cong_p$.)

It is also worth recalling that Karp's Theorem plus the Boundedness Theorem are characteristic for $\mathrm{L}_{\infty\omega}$, that is, $\mathrm{L}_{\infty\omega}$ is the strongest logic for which both these theorems hold. This was shown in Barwise [4]. As a consequence, we could obtain a new characterization of $\mathrm{L}_{\infty\omega}$ using this interpolation theorem plus Lopez-Escobar's result on the boundedness of definable classes of wellorderings. On the other hand, the proof of our results borrows from the proof of the characterization theorem for $\mathrm{L}_{\infty\omega}$, which in turn borrows from Lindström's proof of his theorem characterizing first-order logic in terms of the Compactness Theorem and the downward Löwenheim-Skolem Theorem.

In the following definition we use ∞ to represent the class of all ordinals and write $\alpha < \infty$ to mean that α is an ordinal.

Definition 2 Let L be some fragment of $\mathrm{L}_{\infty\omega}$. Let wo(L) be the least ordinal (or ∞), such that if $\mathcal{C}$ is a $\Sigma^1_1(\mathrm{L})$ definable class of wellorderings then some ordinal $\alpha < \mathrm{wo(L)}$ is greater than the length of every wellordering in $\mathcal{C}$.

The definition makes sense, by the Boundedness Theorem, which tells us that $\mathrm{wo}(\mathrm{L}_{\infty\omega}) = \infty$. By compactness, $\mathrm{wo}(\mathrm{L}_{\omega\omega}) = \omega$. Morley's result mentioned earlier says that $\mathrm{wo}(\mathrm{L}_{\omega_1\omega}) = \omega_1$. One can look at Morley's Theorem and the Boundedness Theorem as the remnant of compactness left over in $\mathrm{L}_{\omega_1\omega}$ and $\mathrm{L}_{\infty\omega}$, respectively.

Given an ordinal α and L-structures $\boldsymbol{A}$ and $\boldsymbol{B}$, we write $\boldsymbol{A} \equiv^{\alpha}_{\infty\omega} \boldsymbol{B}$ if for all sentences φ of quantifier rank at most α, $\boldsymbol{A} \models \varphi$ iff $B \models \varphi$. In her work on potential isomorphism, Karp also characterized $\boldsymbol{A} \equiv^{\alpha}_{\infty\omega} \boldsymbol{B}$ in terms of the existence of an "α-tower," that is, a sequence $\langle I_\beta \rangle_{\beta \leq \alpha}$ of non-empty sets of partial isomorphisms such that $I_\beta \subseteq I_\gamma$ when $\gamma \leq \beta$, and such that if $f \in I_{\beta+1}$ and $a \in A$ ($b \in \boldsymbol{B}$) then there is a $g \in I_\beta$ such that $f \subseteq g$ and $a \in dom(g)$ ($b \in rng(g)$, respectively). The proof of Theorem 2 also uses the following well-known result.

Theorem 4 (Scott's Theorem) *Let L be some vocabulary and let $\boldsymbol{A}$ be an L-structure. For every ordinal α there is a sentence $\sigma^{\alpha}_{\boldsymbol{A}}$ of $\mathrm{L}_{\infty\omega}$ of quantifier rank α such that for all L-structures $\boldsymbol{B}$, $\boldsymbol{B} \models \sigma^{\alpha}_{\boldsymbol{A}}$ if and only if $\boldsymbol{A} \equiv^{\alpha}_{\infty\omega} \boldsymbol{B}$. Moreover, the collection of all such sentences is a proper set.*

This result was implicit in Scott's proof that every countable structure can be characterized up to isomorphism by a sentence of $\mathrm{L}_{\omega_1\omega}$ in Scott [26]. As formulated here it can be found in Barwise [3]. In applying the general framework developed below, the main work will be in establishing analogues of Scott's result. This will usually be accomplished by fairly minor variations on Scott's proof, so we review it briefly here.

Scott's proof shows how to axiomatize the class of those $\boldsymbol{B}$ for which there is an α-tower relating $\boldsymbol{A}$ to $\boldsymbol{B}$. For every $\alpha < \infty$ and $a_1 \ldots a_\in A$ we define the formula $\sigma^{\alpha}_{(a_1 \ldots a_n)}(v_1 \ldots v_n)$ inductively. For $\alpha = 0$, define $\sigma^{0}_{(a_1 \ldots a_n)}(v_1 \ldots v_n)$ to be

$$\bigwedge\{\psi(v_1\dots v_n):\psi \text{ is a literal and } \boldsymbol{A}\models\psi[a_1\dots a_n]\}$$

For successors, define $\sigma^{\beta+1}_{(a_1\dots a_n)}(v_1\dots v_n)$ to be the conjunction of the following three formulas:

$$\sigma^{\beta}_{(a_1\dots a_n)}$$

$$\forall v_{n+1}\bigvee_{a\in A}\sigma^{\beta}_{(a_1\dots a_n,a)}(v_1\dots v_n,v_{n+1})$$

$$\bigwedge_{a\in A}\exists v_{n+1}\sigma^{\beta}_{(a_1\dots a_n,a)}(v_1\dots v_n,v_{n+1})$$

At limit ordinals $\lambda>0$ define $\sigma^{\lambda}_{(a_1\dots a_n)}(v_1\dots v_n)$ to be

$$\bigwedge_{\alpha<\lambda}\sigma^{\alpha}_{(a_1\dots a_n)}(v_1\dots v_n).$$

The sentence $\sigma^{\alpha}_{\boldsymbol{A}}$ we are after is $\sigma^{\alpha}_{()}$. For suppose $B\models\sigma^{\alpha}_{()}$. Let I_β consist of those finite functions

$$f=\{\langle a_1,b_1\rangle\dots\langle a_n,b_n\rangle\}$$

such that

$$\boldsymbol{B}\models\sigma^{\beta}_{(a_1\dots a_n)}[b_1\dots b_n]$$

The formulas $\sigma^{\beta}_{(a_1\dots a_n)}$ are defined so as to insure that this gives us an α tower. Notice that the empty function is in I_α by the assumption that $\boldsymbol{B}\models\sigma^{\alpha}_{()}$.

We now turn to the proof of our first Interpolation Theorem.

Proof of the Interpolation Theorem: The proof that (2) implies (1) is immediate, given (the easy half of) Karp's Theorem, as we have already indicated. We show that (1) implies (2) by showing that (1) implies (3) and (3) implies (2), where:

3. There is an ordinal α such that φ entails ψ along $\equiv^{\alpha}_{\infty\omega}$.

If $\varphi,\psi\in \mathrm{L}_{\infty\omega}$, then (3) is trivial; just choose α to be the larger of their two quantifier ranks. But φ and ψ are typically second-order sentences over $\mathrm{L}_{\infty\omega}$, so this does not work. So suppose that (3) fails. Then for every ordinal α there are structures $\boldsymbol{A}\models\varphi$, $\boldsymbol{B}\models\neg\psi$, and an α-tower $\langle I_\beta\rangle_{\beta\le\alpha}$ relating $\boldsymbol{A}$ and $\boldsymbol{B}$. Code all this up in a sentence $\pi(<,\dots)$, but one that is independent of α. The relation symbols here will include those quantified in φ and ψ. Since $\pi(<,\dots)$ has models where $<$ is arbitrarily long, it has to have a model that is not well-ordered. Any descending sequence in this ordering gives rise to a set of partial isomorphisms with the back and forth property, so we get a contradiction to (1). (This part of our proof is admittedly a bit sketchy but we will go into it in depth in our detailed analysis.)

The proof of (3) implies (2) is more straightforward. Let α be the bound in (3). Let θ be the disjunction of the set of all sentences of the form $\sigma^{\alpha}_{\boldsymbol{A}}$, the sentence that characterizes $\boldsymbol{A}$ up to $\equiv^{\alpha}_{\infty\omega}$, for all models $\boldsymbol{A}$ such that $\boldsymbol{A} \models \varphi$, from Scott's Theorem. There is at most a set of such sentences, so θ is a sentence. Clearly $\varphi \models \theta$. By (3), $\theta \models \psi$. QED

Entailment along potential isomorphism (or other natural relations) is an interesting logical notion as such – and hence there is a genuine issue as to what is the most appropriate formulation of the Interpolation Theorem, even for first-order logic itself. We continue with a related example. By imitating the usual proof of Beth's Theorem by way of interpolation as closely as possible, we obtain the following. This result was observed independently by Alexandru Baltag.

Corollary 5 *(A Beth Theorem for $\mathrm{L}_{\infty\omega}$) Let P be a predicate symbol not in L and let $L' = L \cup \{P\}$. The following are equivalent for all $\varphi(P)$ in $L'_{\infty\omega}$:*

1. *There is a formula $\delta(x) \in \mathrm{L}_{\infty\omega}$ such that $\varphi \models \forall x\,(P(x) \leftrightarrow \delta(x))$.*
2. *If $(\boldsymbol{A}, P^{\boldsymbol{A}})$ and $(\boldsymbol{B}, P^{\boldsymbol{B}})$ are L' structures that are models of $\varphi(P)$ and if $I : \boldsymbol{A} \cong_p \boldsymbol{B}$ then $I : (\boldsymbol{A}, P^{\boldsymbol{A}}) \cong_p (\boldsymbol{B}, P^{\boldsymbol{B}})$.*

Proof: The proof of (1) implies (2) is straightforward, given Karp's Theorem. For the converse, assume (2). Consider the sentences $\varphi'(c)$ given by

$$\exists P[\varphi(P) \wedge P(c)]$$

and ψ given by

$$\forall P[\varphi(P) \to P(c)].$$

We claim that by (2), φ' entails ψ along $\cong_p$. If so, there is a sentence $\delta(c)$ that is an interpolant and this easily yields (1). To establish our claim, let $I : (\boldsymbol{A}, c^{\boldsymbol{A}}) \cong_p (\boldsymbol{B}, c^{\boldsymbol{B}})$ where

$$(\boldsymbol{A}, c^{\boldsymbol{A}}) \models \exists P[\varphi(P) \wedge P(c)].$$

We need to see that

$$(\boldsymbol{B}, c^{\boldsymbol{B}}) \models \forall P[\varphi(P) \to P(c)].$$

Pick P so that $(\boldsymbol{A}, c^{\boldsymbol{A}}, P) \models \varphi(P) \wedge P(c)$ and assume that $(\boldsymbol{B}, c^{\boldsymbol{B}}, P') \models \varphi(P)$. We need to prove that $c^{\boldsymbol{B}} \in P'$. Since $I : (\boldsymbol{A}, c^{\boldsymbol{A}}) \cong_p (\boldsymbol{B}, c^{\boldsymbol{B}})$, $I : \boldsymbol{A} \cong_p \boldsymbol{B}$. By (2), $I : (\boldsymbol{A}, P^{\boldsymbol{A}}) \cong_p (\boldsymbol{B}, P^{\boldsymbol{B}})$. But then let $f \in I$ be defined on $c^{\boldsymbol{A}}$. Since $c^{\boldsymbol{A}} \in P$, $f(c^{\boldsymbol{A}})$ is in $P^{\boldsymbol{B}} = P'$. But since $I : (\boldsymbol{A}, c^{\boldsymbol{A}}) \cong_p (\boldsymbol{B}, c^{\boldsymbol{B}})$, $c^{\boldsymbol{B}} = f(c^{\boldsymbol{A}})$. QED

We have given this proof in some detail because we will be interested in analogues of it in other settings. Note how our theorem relates explicit definability of P in L to what may be called "safety" of potential L-isomorphisms for the extended language $L \cup \{P\}$. (Cf. van Benthem [13] for the analogous notion of safety for modal languages with respect to bisimulation.) If one looks at the reason we get this and not the standard statement of Beth's Theorem, a noteworthy fact emerges. Namely, entailment along potential isomorphism does not have the deduction property. In particular, it is possible to have $\varphi \wedge \eta$ entail ψ along potential isomorphisms without having φ entail $\eta \rightarrow \psi$ along potential isomorphisms.

2 An analysis of the proof

In this section we give an analysis of the above interpolation theorem and its proof. This will allow us to give a number of additional results with little further ado.

2.1 Some clues

We begin by making some disjointed remarks which, taken together, provide a set of clues for understanding what is at the heart of Theorem 2. By building on these ideas we will develop a general framework. This discussion will use some notions reviewed later in the paper.

Clue 1 If the vocabulary L is finite, then the condition $\boldsymbol{A} \cong \boldsymbol{B}$ is Σ^1_1 definable in the vocabulary L^* appropriate for pairs $(\boldsymbol{A}, \boldsymbol{B})$ of structures; the Σ^1_1 sentence just expresses the existence of an isomorphism.

Clue 2 Given any pair $\boldsymbol{A}, \boldsymbol{B}$ of L-structures there is always a largest set I of partial isomorphisms having the back-and-forth property, simply because the union of any family of such sets is another such set. Of course it could be that this largest family I is empty. Indeed, we see that $\boldsymbol{A} \cong_p \boldsymbol{B}$ iff this largest I is non-empty. This largest set is the greatest fixed point of the monotone operator Γ defined by

$$\Gamma(I) \quad = \quad \{f \mid \text{ the back and forth condition holds for } f \text{ and } I\}.$$

The α-th approximation I^α to this greatest fixed point is the largest α-tower of sets of partial isomorphisms that came up in the proof. Moreover, Scott's Theorem tells us, roughly, that the equivalence class associated with I^α is uniformly definable in $\mathrm{L}_{\infty\omega}$.

Clue 3 From recursion theory we recall that there is, over the natural numbers, an equivalence between Σ^1_1 definable predicates and greatest fixed points of first-order positive monotone operators. With a little bit of fiddling, the correspondence between Σ^1_1-definability and greatest fixed points of first-order elementary monotone operators extends from the natural numbers to arbitrary *countable* structures. On uncountable structures, the greatest fixed points are always Σ^1_1 but in general the converse fails.

This division into cases based on countable and uncountable structures is reminiscent of the following fact observed above:

Clue 4 The relation between $\cong$ and $\cong_p$ satisfies:

1. On all L-structures $\boldsymbol{A}, \boldsymbol{B}$, if $\boldsymbol{A} \cong \boldsymbol{B}$ then $\boldsymbol{A} \cong_p \boldsymbol{B}$.
2. On countable structures, the converse holds.

Looking at this suggests looking at Clue 3 in a bit more detail. Doing so, we recall the following:

Clue 5 When we look at the positive monotone operator Γ that is the counterpart to a given Σ_1^1-predicate $\varphi(x)$, it has the following properties:

1. on every structure $\boldsymbol{A}$, if $\boldsymbol{A}, a \models \varphi$ then $a \in \Gamma^{\infty}[A]$, the greatest fixed point of Γ on $\boldsymbol{A}$.
2. on countable structures $\boldsymbol{A}$, the converse holds.

Taken together, these clues suggests that the heart of Theorem 2 resides in the fact that $\cong_p$ is the coinductive counterpart of the Σ_1^1 relation $\cong$. This suggestion is strengthened by the observation that the proof of (1) $\Rightarrow$ (3) of Theorem 2 makes essential use of the stages of the coinductive definition of $\cong_p$. If this hypothesis is correct, we ought to be able to generalize this result from $\cong$ to other Σ_1^1 relations R on L-structures, and we will do so.

Where do pebble games come into all this? They arise from a well known correspondence between games and inductive definitions. Every monotone operator has associated with it open and closed games defining its least and greatest fixed points, respectively. The usual pebble game for $\cong_p$ is simply the closed game associated with the coinductive definition of $\cong_p$; see Proposition 10.

To make the interest of our results convincing, we need to be sure there really is something fundamental about the relationship between a Σ_1^1 relation R and its coinductive counterpart, which we will write R^{abs}. It is to this end that we devote the following digression.

2.2 Absolute versions of predicates

Since this section is more for motivation than for any results we will be using, we do not need a precise definition of absoluteness in this paper; an intuitive one will do the job.[1] The intuitive idea is that an n-ary predicate P is *absolute* if whether or not $P(a_1 \ldots a_n)$ holds is independent of the universe of set theory in which $a_1 \ldots a_n$ reside. In particular, if one were to "enlarge" the universe, so as to make some or all of $a_1 \ldots a_n$ countable, then $P(a_1 \ldots a_n)$ would hold

[1] A more careful exposition of these ideas is contained in Barwise [5].

in the new universe iff it held in the original universe. (One way to make this precise is to require that P be provably Δ_1 in the vocabulary of set theory.)

Gerald Sacks [24] observed that " ... the central notions of model theory are absolute, and absoluteness, unlike cardinality, is a logical concept." Because cardinality is not absolute, the relation $\boldsymbol{A} \cong \boldsymbol{B}$ of being isomorphic structures is not absolute. For example, two infinite sets of different cardinality are not isomorphic, but if they both became countable, by enlarging the universe of sets to include countings of them, then they would be isomorphic. By contrast, the relation $\boldsymbol{A} \cong_p \boldsymbol{B}$ of being potentially isomorphic is absolute. Not only that, the two relations agree as long as $\boldsymbol{A}$ and $\boldsymbol{B}$ are countable. This is a special case of a very general phenomenon.

Suppose we have two relations P and P^{abs} on sets. We say that P^{abs} is *the absolute version* of P provided that P is a subrelation of P^{abs}, P and P^{abs} agree on hereditary countable sets and that P^{abs} is absolute.[2] Thus, as we have just noted, the relation $\boldsymbol{A} \cong_p \boldsymbol{B}$ is the absolute version of $\boldsymbol{A} \cong \boldsymbol{B}$.

The relation "$\boldsymbol{A}$ is embeddable in $\boldsymbol{B}$," is not absolute but it is Σ^1_1 so does have a natural absolute version. Similarly for "$\boldsymbol{B}$ is a homomorphic image of $\boldsymbol{A}$." Once one knows to look for them, it is not hard to find the absolute versions of these relations. We will obtain preservation and interpolation theorems for these absolute relations and in so doing obtain analogues of the Łos-Tarski Theorem and the Lyndon Theorem for $\mathrm{L}_{\infty\omega}$.

One can obtain the usual preservations and interpolation results for $\mathrm{L}_{\omega_1\omega}$ and $\mathrm{L}_{\omega\omega}$ from the results we obtain below by observing that because these logics satisfy the downward Löwenheim-Skolem Theorem, entailment along a Σ^1_1 relation R and along its absolute version amount to the same thing, in the following sense. We say that L has the *Löwenheim property* if every sentence in $\Sigma^1_1(\mathrm{L})$ that has a model has a finite or countable model.

Proposition 6 *Let R be some Σ^1_1 relation on L-structures, R^{abs} its absolute version. Let* L *be a regular logic having the Löwenheim property, and let $\varphi \in \Sigma^1_1(\mathrm{L})$ and $\psi \in \Pi^1_1(\mathrm{L})$. Then the following are equivalent:*

1. *φ entails ψ along R.*
2. *φ entails ψ along R^{abs}.*

PROOF: Since $R \subseteq R^{abs}$, the proof of (2) implies (1) is trivial. The proof of the converse is easy. For suppose that (1) holds but (2) fails. Then there are L-structures $\boldsymbol{A}, \boldsymbol{B}$ such that $\boldsymbol{A} \models \varphi$, $\boldsymbol{A}\ R^{abs}\ \boldsymbol{B}$, and $\boldsymbol{B} \models \neg\psi$. But then this whole thing can be expressed as a sentence in $\Sigma^1_1(\mathrm{L})$, which has a countable model. But that would violate (1) since R and R^{abs} agree on countable models. QED

[2] If we make the notion of absoluteness precise by requiring P to be Δ_1 definable in the vocabulary of set theory, then a predicate can have at most one absolute version, up to provable equivalence in set theory, thus justifying calling P^{abs} "the" absolute version of P; see Theorem VI.6.2 in Barwise [5].

As a corollary, we note that if $\varphi \in \Sigma_1^1(\mathrm{L}_{\omega_1\omega})$ and $\psi \in \Pi_1^1(\mathrm{L}_{\omega_1\omega})$, then φ entails ψ (along isomorphism) iff φ entails ψ along $\cong_p$.

In a similar way, one can prove the following:

Proposition 7 *Let $F \subseteq \mathrm{L}_{\infty\omega}$ be such that the predicate "$\varphi \in F$" is absolute. If the sentences in F are (provably) preserved along the Σ_1^1 relation R, they must also be preserved along its absolute version R^{abs}.*

These results demonstrate that the natural place to look for preservation theorems is along absolute relations among structures. Moreover, any Σ_1^1 relation on R on structures has an associated absolute version, given as the largest fixed point of an inductive definition. Our strategy for generalizing Theorem 2, then, is as follows.

Strategy: Given some interesting Σ_1^1 binary relation R on structures:

1. Find its absolute version R^{abs}.
2. Find a set of $\mathrm{L}_{\infty\omega}$ sentences F that are preserved under R for obvious syntactic reasons. They will also be preserved under R^{abs}.
3. Show that if $\exists S_1\varphi(S_1)$ entails $\forall S_2\psi(S_2)$ along R^{abs} then there is a sentence $\theta \in F$ such that $\exists S_1\varphi(S_1) \models \theta$ and $\theta \models \forall S_2\psi(S_2)$. (The converse follows from (2).)
4. Note that as a consequence we get the following preservation result: a sentence of $\mathrm{L}_{\infty\omega}$ is preserved under R^{abs} iff it is equivalent to a sentence in F.

In Theorem 2, R was the relation of isomorphism and F is the class of all sentences of $\mathrm{L}_{\infty\omega}$. In this case, step (4) degenerates into triviality, but in other examples it won't. For example, when we start with R being the embeddability relation we obtain a semantic characterization of the set of universal sentences of $\mathrm{L}_{\infty\omega}$. (Some of our applications start with a relation R that is given outright as a coinductive relation on structures. In these cases the first step is trivial, since R is its own absolute version.)

Step (3) is the main one, and typically involves two parts, one corresponding to each of the two main steps in our proof of Theorem 2. The main tool for both steps is to realize that the relation R^{abs} can be expressed as the greatest fixed point of a first-order positive operator Γ, and so can be approximated from above by means of relations R^α:

$$R^0 \supseteq R^1 \supseteq \ldots \supseteq R^\alpha \supseteq \ldots \supseteq R^{abs}.$$

Let us review this basic construction.

2.3 Co-inductive definitions

Suppose Γ is a monotone operator on the power set of some set X, that is if $Y_1 \subseteq Y_2 \subseteq X$ then $\Gamma(Y_1) \subseteq \Gamma(Y_2) \subseteq X$. A fixed point of Γ is any set $Y \subseteq X$ such that $\Gamma(Y) = Y$. Every monotone operator Γ has a least fixed point Γ_∞ and a greatest fixed point Γ^∞. These are naturally defined by means of iterates as follows:

$$\begin{array}{rcl} \Gamma_\alpha & = & \Gamma(\bigcup_{\beta<\alpha} \Gamma_\beta) \\ \Gamma_\infty & = & \bigcup_\alpha \Gamma_\alpha \\ & & \\ \Gamma^\alpha & = & \Gamma(\bigcap_{\beta<\alpha} \Gamma_\beta) \\ \Gamma^\infty & = & \bigcap_\alpha \Gamma_\alpha \end{array}$$

where $\bigcup \emptyset = \emptyset$ and $\bigcap \emptyset = X$ are used to get the recursions going at $\alpha = 0$.

Example 1 Let X be the set of finite partial isomorphisms between two structures $\boldsymbol{A}$ and $\boldsymbol{B}$ and let $\Gamma(I) =$

$$\{f \in X \mid \forall a \in A[\text{or } b \in B]\exists f' \in I, f' \supseteq f, a \in dom(f'))[\text{or } b \in rng(f')]\}$$

This is clearly a monotone operator. Its smallest fixed point is the empty set. Any non-empty fixed point is a set I with the back and forth property. Thus the largest fixed point is non-empty iff $\boldsymbol{A} \cong_p \boldsymbol{B}$. Similarly, $\langle \Gamma^\beta \rangle_{\beta \leq \alpha}$ is the largest α-tower so Γ^α is non-empty iff $\boldsymbol{A} \equiv^\alpha_{\infty\omega} \boldsymbol{B}$.

The smallest and largest fixed points of a monotone operator Γ also have useful non-constructive characterizations: Γ_∞ is the smallest $Y \subseteq X$ such that $\Gamma(Y) \subseteq Y$ and Γ^∞ is the largest $Y \subseteq X$ such that $Y \subseteq \Gamma(Y)$. When speaking of the smallest fixed point, Γ is called an inductive definition. When speaking of the largest fixed point, it is called a co-inductive definition. We are primarily interested in the coinductive definitions, since that is the form our pebble games take, but since inductive definitions are more familiar, we state and prove some of our results in both versions.

Suppose that L is some vocabulary. Let L' expand L by a single unary relation symbol U. Let $\varphi(x)$ be a formula of $L'_{\infty\omega}$ in one free variable, and in which the predicate U occurs positively. Since U is positive in φ, φ is monotone in the sense that $(\boldsymbol{A}, U) \models \varphi[a]$ and $U \subseteq U' \subseteq A$ entails that $(\boldsymbol{A}, U') \models \varphi[a]$. Consequently, if we define an operator ${}^{\boldsymbol{A}}\Gamma$ on the universe A of an L structure $\boldsymbol{A}$ by

$${}^{\boldsymbol{A}}\Gamma(Y) \quad = \quad \{a \in A \ \mid (A, Y) \models \varphi[a]\}$$

then ${}^{\boldsymbol{A}}\Gamma$ is monotone. The sets ${}^{\boldsymbol{A}}\Gamma_\infty$ and ${}^{\boldsymbol{A}}\Gamma^\infty$ are the sets *inductively* and *coinductively* defined by φ, respectively. Such an operator is called an $\mathrm{L}_{\infty\omega}$-inductive or coinductive operator.

The following is an abstract version of the proof of (1) implies (3) in the proof of Theorem 2. (Actually, it's the dual result, for coinductive operators that applies in that proof.) It is an analogue of Theorem A on uniform inductive definability in Barwise and Moschovakis [8].[3] Its proof replaces the compactness argument of that paper by a use of the Boundedness Theorem.

Theorem 8 *Let* L *be a regular fragment of* $\mathrm{L}_{\infty\omega}$*, let* $\mathcal{C}$ *be a* $\Sigma^1_1(\mathrm{L})$ *class of structures, and let* Γ *be an* L *positive operator. The following are equivalent:*

1. ${}^{\boldsymbol{A}}\Gamma_\infty = A$ *for all* $\boldsymbol{A} \in \mathcal{C}$.
2. *There is an ordinal* $\alpha < \mathrm{wo}(\mathrm{L})$ *such* ${}^{\boldsymbol{A}}\Gamma_\alpha = A$ *for every* $\boldsymbol{A} \in \mathcal{C}$.

Before giving the proof, we state the following lemma.

Lemma 1 *Let* Γ *be a monotone operator on a set* X*. Suppose there are sets* $I_n \subseteq X$ *such that* $\Gamma(I_{n+1}) \subseteq I_n$*, for each* $n < \omega$*. If* $I = \bigcap_{n<\omega} I_n$ *then* $\Gamma_\infty \subseteq I$*.*

Proof: We need only prove that $\Gamma(I) \subseteq I$. Let $a \in \Gamma(I)$. Since $I \subseteq I_{n+1}$ and Γ is monotone, $a \in \Gamma(I_{n+1})$, for each n. But $\Gamma(I_{n+1}) \subseteq I_n$, by assumption, so $a \in I_n$, for each n, hence $a \in I$, as desired. QED

Proof of Theorem 8: The proof that (2) implies (1) is immediate. The proof of (1) implies (2) is a generalization of the proof that (1) implies (3) in the proof of Theorem 2. Our proof there was rather sketchy since we give a detailed proof here. Again, we proceed by contradiction. Thus we suppose that (1) holds but that (2) fails. Then for every $\alpha < \mathrm{wo}(\mathrm{L})$ there is a structure $\boldsymbol{A} \in \mathcal{C}$ such that ${}^{\boldsymbol{A}}\Gamma_\alpha \neq A$. By cardinality considerations, the cardinality of $\boldsymbol{A}$ must be at least as large as the cardinality of α. Hence there is a subset $\boldsymbol{A}_0 \subseteq A$ and a wellordering $<$ of $\boldsymbol{A}_0$ of order type $\alpha + 1$. For each $a \in A_0$ let $\rho(a)$ be the order type of the predecessors of $\boldsymbol{A}$ which is an ordinal $\leq \alpha$. Now define a relation R on $\boldsymbol{A}$ by

$$R = \{\langle a, b\rangle \mid a \in A_0, b \in^{\boldsymbol{A}} \Gamma_{\rho(a)}\}.$$

Now let $\boldsymbol{A}^*$ be the structure $(A, A_0, <, R)$, a structure for a vocabulary L^* extending L. The structure $\boldsymbol{A}^*$ satisfies the following conditions:

1. $<$ is a linear ordering of $\boldsymbol{A}_0$ with a last element.
2. For every $x \in A_0$ and every y, $R(x, y)$ iff

   $$\varphi(y, \{z \mid \exists x_0 < xR(x_0, z)\})$$

   (where the set expression is substituted for the unary U in the definition $\varphi(x, U)$ of Γ).

[3] We could strengthen this result to make the result even more analogous but only the cost of space and perspicuity. We leave the generalization to the reader familiar with Barwise and Moschovakis [8].

3. $\exists y \forall x \in A_0 \neg R(x,y)$.

Recalling that $\mathcal{C}$ is $\Sigma^1_1(\mathrm{L})$ definable, let $\mathcal{C}$ be defined by, say, $\exists S\varphi(S)$. Note that each of the above is expressible by a sentence of L and that none of these sentences depends on $\boldsymbol{A}$ or on α. Let $\pi_0(<)$ be the conjunction of these sentences, as well as the sentence $\varphi(S)$, and let $\pi(<)$ be the result of quantifying out all the symbols except $<$. Then for each $\alpha < \mathrm{wo}(\mathrm{L})$, $\pi(<)$ has models $\boldsymbol{A}$ where $<^{\boldsymbol{A}}$ is a wellordering of order type $\alpha+1$. Hence $\pi(<)$ has a model that is not a wellordering. But then there are A_0, R such that $(A, A_0, <, R)$ makes all the above true. Let $a_0, \ldots, a_{n+1} \in A_0$ such that

$$\ldots a_{n+1} < a_n < \ldots a_1 < a_0.$$

Define $I_n = \{b \mid R(a_n, b)\}$. We claim that $\Gamma(I_{n+1}) \subseteq I_n$, for each n. To see this, note that

$$\begin{array}{rcl}
\Gamma(I_{n+1}) & = & \{b \mid R(a_{n+1}, b)\} \\
 & = & \{b \mid \boldsymbol{A} \models \varphi(b, \{z \mid \exists x < a_{n+1} R(x,z)\})\} \\
 & \subseteq & \{b \mid \boldsymbol{A} \models \varphi(b, \{z \mid \exists x < a_n \ \ R(x,z)\})\} \\
 & = & \{b \mid R(a_n, b)\} \\
 & = & I_n
\end{array}$$

By the lemma, we have ${}^{\boldsymbol{A}}\Gamma_\infty \subseteq I_n$ for each n, but I_0 is a proper subset A, contradicting (1). QED

By duality we obtain the following, which is what we are really after.

Corollary 9 *Let* L *be a regular fragment of* $\mathrm{L}_{\infty\omega}$*, let* $\mathcal{C}$ *be a* $\Sigma^1_1(\mathrm{L})$ *class of structures, and let* Γ *be an* L *positive operator. The following are equivalent:*

1. ${}^{\boldsymbol{A}}\Gamma^\infty = \emptyset$ *for all* $\boldsymbol{A} \in \mathcal{C}$.
2. *There is an ordinal* $\alpha < \mathrm{wo}(\mathrm{L})$ *such* ${}^{\boldsymbol{A}}\Gamma^\alpha = \emptyset$ *for every* $\boldsymbol{A} \in \mathcal{C}$.

To apply this result in a wide class of cases, we want a framework where we can see the various pebble games as given by coinductive operators.

2.4 Pebble games

Given sets X and Y, let $X \otimes Y$ be the set of all finite relations r on $X \times Y$; that is, $X \otimes Y$ is the finitary power set of $X \times Y$. For the most part, we are interested in the elements of $X \otimes Y$ that are partial one-one functions, but not exclusively.

A *pebble game on* X *and* Y is a monotone operator Γ on $X \otimes Y$. We call any $r \in X \otimes Y$ a *position.* (To maintain the pebble metaphor, think of elements of X and Y as related by r when they have associated with them pebbles of the same color. An element may have one, none, or many pebbles associated with it, in general.) Call the players *I* and *II.* A round, or hand, of the game is played as follows.

1. Player I begins by playing a position r_0, called the *initial position* of the round.

2. Player II responds by playing a set R_0 of positions such that $r_0 \in \Gamma(R_0)$.

3. Player I responds by playing a position $r_1 \in R_0$.

4. Player II responds by playing a set R_1 of positions such that $r_1 \in \Gamma(R_1)$.

$\vdots$

If at a given stage of play, one of the players is unable to play, the other player wins. If the play goes on forever, then player II is the winner.

The following is a (special case of a) well-known observation. We include it only to explain why we are justified in thinking of any monotone operator on $X \otimes Y$ as a pebble game.

Proposition 10 *A position r is in Γ^∞ iff II has a winning strategy whenever I plays r.*

Proof: If $r \in \Gamma^\infty$ then let II always play Γ^∞ for a winning strategy. For the converse, let J be the set of all plays r from which II has some winning strategy. Clearly $J \subseteq \Gamma(J)$ and so $J \subseteq \Gamma^\infty$. QED

Given L-structures $\boldsymbol{A}$ and $\boldsymbol{B}$, let $\boldsymbol{A} \otimes \boldsymbol{B}$ be the structure

$$(\boldsymbol{A}, \boldsymbol{B}, A \otimes B, Holds)$$

where $Holds(r, a, b)$ iff $r \in A \otimes B$, $a \in A$, $b \in B$ and $\langle a, b\rangle \in r$. One can either think of this as a three-sorted structure, or as a single sorted structure for a richer vocabulary, depending on one's preferences. We will call these *pebble structures*. A L-*definable pebble game* is a monotone operator on structures of the form $\boldsymbol{A} \otimes \boldsymbol{B}$ that, as a coinductive definition, is definable by a positive formula in L. That is Γ is given by

$$\Gamma(R) = \{r \mid (\boldsymbol{A} \otimes \boldsymbol{B}, R) \models \eta[r]\}$$

where $\eta(r, R) \in$ L is R-positive. By the above, Γ^∞ is the set of positions from which II can win. We write Γ^∞ as $\Gamma^\infty[\boldsymbol{A}, \boldsymbol{B}]$ when we need to indicate its dependence on $\boldsymbol{A}$ and $\boldsymbol{B}$.

Generalizing the notation $f : \boldsymbol{A} \cong_p \boldsymbol{B}$, write $r : \boldsymbol{A} \prec_\Gamma \boldsymbol{B}$ if $r \in \Gamma^\infty[A, B]$ and $\boldsymbol{A} \prec_\Gamma \boldsymbol{B}$ if there is a position r such that $r : \boldsymbol{A} \prec_\Gamma \boldsymbol{B}$. Similarly, we write $r : \boldsymbol{A} \prec_\Gamma^\alpha \boldsymbol{B}$ if $r \in \Gamma^\alpha[\boldsymbol{A}, \boldsymbol{B}]$ and $\boldsymbol{A} \prec_\Gamma^\alpha \boldsymbol{B}$ if there is a position r such that $r : \boldsymbol{A} \prec_\Gamma^\alpha \boldsymbol{B}$. The following is an abstract version of the equivalence of (1) and (3) in Theorem 2. For formulas containing free variables we say that φ entails ψ along $\prec_\Gamma$ if $r : \boldsymbol{A} \prec_\Gamma \boldsymbol{B}$, $a_1 r b_1, \ldots, a_n r b_n$ and $\boldsymbol{A} \models \varphi[a_1 \ldots a_n]$ entails $\boldsymbol{B} \models \psi[b_1 \ldots b_n]$.

Theorem 11 *Let* L *be a regular, infinitary fragment of* $\mathrm{L}_{\infty\omega}$. *Let* Γ *be a* L-*definable pebble game on* L-*structures. Let* $\varphi \in \Sigma^1_1(\mathrm{L})$ *and* $\psi \in \Pi^1_1(\mathrm{L})$. *The following are equivalent.*

1. φ *entails* ψ *along* $\prec_\Gamma$.

2. *There is an* $\alpha < \mathrm{wo}(\mathrm{L})$ *such that* φ *entails* ψ *along* $\prec^\alpha_\Gamma$.

PROOF: (2) clearly implies (1) so we need only prove the converse. Let $\mathcal{C}$ be the class of all structures isomorphic to those of the form $\boldsymbol{A} \otimes \boldsymbol{B}$ where $\boldsymbol{A} \models \varphi$ and $\boldsymbol{B} \models \neg\psi$. It is easy to see (using the fact that we have an infinitary fragment) that this class is $\Sigma^1_1(\mathrm{L})$. The key point here is that we can axiomatize the class of finite subsets of $A \times B$ in terms of *Holds*. This is just a slight variant on the well-known observation that one can axiomatize the finite power set operation in $\mathrm{L}_{\omega_1\omega}$. Given this, the proof follows in a straightforward manner from Corollary 9. For suppose that (1) holds. Then for each $\boldsymbol{A} \otimes \boldsymbol{B}$ in $\mathcal{C}$, $\boldsymbol{A} \not\prec_\Gamma \boldsymbol{B}$ so $\Gamma^\infty[\boldsymbol{A}, \boldsymbol{B}] = 0$. Hence, there is an α such that $\Gamma^\alpha[\boldsymbol{A}, \boldsymbol{B}] = 0$ for each $\boldsymbol{A} \otimes \boldsymbol{B}$ in $\mathcal{C}$. Then for $\boldsymbol{A} \otimes \boldsymbol{B}$ in $\mathcal{C}$, $\boldsymbol{A} \not\prec^\alpha_\Gamma \boldsymbol{B}$. But this implies (2). QED

2.5 An Abstract Interpolation Theorem

We are narrowing in on our generalization of Theorem 2. What we need is a class of interpolating formulas.

Let Γ be a pebble game defined on structures of the form $\boldsymbol{A} \otimes \boldsymbol{B}$, where $\boldsymbol{A}, \boldsymbol{B}$ are L-structures. A formula φ of $\mathrm{L}_{\infty\omega}$ is *preserved along* $\prec_\Gamma$ if for all $\boldsymbol{A}, \boldsymbol{B}$, if $r : A \prec_\Gamma \boldsymbol{B}$, if $a_1 r b_1, \ldots, a_n r b_n$, and if $\boldsymbol{A} \models \varphi[a_1 \ldots a_n]$ then $\boldsymbol{B} \models \varphi[b_1 \ldots b_n]$.

Definition 3 A pebble game Γ has the *Scott property* with respect to a class F of formulas if

1. every formula in F is preserved along $\prec_\Gamma$

2. for every L structure $\boldsymbol{A}$ and every α there is a sentence $\sigma^\alpha_{\boldsymbol{A}} \in F$ such that for all L-structures $\boldsymbol{B}$,

$$\boldsymbol{B} \models \sigma^\alpha_{\boldsymbol{A}} \text{ iff } \boldsymbol{A} \prec^\alpha_\Gamma \boldsymbol{B},$$

   and

3. if the class of all such $\sigma^\alpha_{\boldsymbol{A}}$, for fixed α, is a set.

Theorem 12 (Abstract Interpolation Theorem) *Let* L *be a regular infinitary fragment of* $\mathrm{L}_{\infty\omega}$ *and* Γ *be an* L-*definable pebble game with the Scott Property relative to* F. *For any* $\varphi \in \Sigma^1_1(\mathrm{L})$ *and* $\psi \in \Pi^1_1(\mathrm{L})$, *the following are equivalent:*

1. φ *entails* ψ *along* $\prec_\Gamma$.

2. There is a disjunction θ of a set of sentences in F such $\varphi \models \theta$ and $\theta \models \psi$.

In particular, a sentence φ of $\mathrm{L}_{\infty\omega}$ is preserved under $\prec_\Gamma$ iff it is equivalent to a disjunction of sentences from F.

PROOF: The proof that (2) entails (1) is straightforward. The converse is just like the proof of the special case, given our earlier results. QED

After this paper was finished more or less finished, Alexandru Baltag obtained a very interesting analysis of a wide class of $\mathrm{L}_{\infty\omega}$-definable pebble games Γ that lets one read off from the form of Γ a class F of formulas such that Γ has the Scott property with respect to F. Using Baltag's result, one could replace our case by case analysis give in the various results below by simply reading off the Scott property from the form of the pebble game.

3 Applications

The method developed in the preceding section yields many interpolation and preservation results. The main step involves finding a pebble game that have the Scott Property relative to some natural class F. We begin with an example where the application of our method is quite straightforward.

3.1 A Łos-Tarski Theorem

Our first example is where we start with the relation "$\boldsymbol{A}$ is embeddable in $\boldsymbol{B}$." This is a Σ^1_1 relation on structures. The *existential formulas*, those constructed from literals using set conjunctions and disjunctions, as well as existential quantifiers, are clearly preserved under this relation.

The pebble game $\Gamma_\exists$ defining the absolute version of this relation is just like that for $\boldsymbol{A} \cong_p \boldsymbol{B}$ except that it restricts Spoiler's choices to the universe of $\boldsymbol{A}$. The coinductive notion thus defined is that of a *potential embedding* of $\boldsymbol{A}$ into $\boldsymbol{B}$, being a potential isomorphism with the "back" clause of the back-and-forth property clipped off. Since existential formulas are preserved along embeddings, they are also preserved under potential embeddings.

Lemma 2 *The existential sentences have the Scott property with respect to the pebble game $\Gamma_\exists$.*

PROOF: This is a routine modification of the proof of Scott's Theorem. For successors, define $\sigma^{\beta+1}_{(a_1 \ldots a_n)}(v_1 \ldots v_n)$ to be the conjunction:

$$\sigma^{\beta}_{(a_1 \ldots a_n)} \wedge \bigwedge_{a \in A} \exists v_{n+1} \sigma^{\beta}_{(a_1 \ldots a_n, a)}(v_1 \ldots v_n, v_{n+1})$$

Otherwise the proof is the same. QED

This yields:

Theorem 13 *The following are equivalent for all $\varphi \in \Sigma_1^1(\mathrm{L}_{\infty\omega})$ and $\psi \in \Pi_1^1(\mathrm{L}_{\infty\omega})$:*

1. *φ implies ψ along potential embeddings.*

2. *There exists some existential sentence θ with $\varphi \models \theta$ and $\theta \models \psi$.*

In particular, a sentence of $\mathrm{L}_{\infty\omega}$ is preserved under potential embeddings iff it is logically equivalent to an existential sentence.

This all dualizes to preservation downward along potential embeddings and universal sentences.

3.2 Lyndon's Theorem

Very similar considerations apply to other standard preservation theorems, recast as interpolation under suitably generalized model relations. Many of these may be found by systematic variation on the inductive scheme for generating the formulas of $\mathrm{L}_{\infty\omega}$ from literals. We sample three.

The formulas $\varphi(U^+)$ that involve a predicate U only *positively* play an important role in the theory of inductive and coinductive definability, so one would like a semantic characterization of such formulas. As usual, this implies that $\varphi(U^+)$ is semantically monotone, i.e., preserved under increases in the extension of U. More generally, such a sentence will be preserved under U-homomorphism images. The notion f is a U-homomorphism from $\boldsymbol{A}$ onto $\boldsymbol{B}$ is defined just as with *isomorphism*, except that with respect to U it need only be a homomorphism, that is, we require

$$\langle a_1, \dots, a_n \rangle \in U^{\boldsymbol{A}} \text{ implies } \langle f(a_1), \dots, f(a_n) \rangle \in U^{\boldsymbol{B}},$$

but not the converse.

There is an obvious pebble game Γ_{U+}: the partial functions are isomorphisms with respect to the symbols of L other than U, while for U, we require only the one direction displayed above. Formulas $\varphi(U^+)$ with positive U are clearly preserved along the functions $f \in \Gamma_{U+}^{\infty}$, too. *Mutatis mutandis,* this leads to the following result.

Theorem 14 *The following are equivalent for all $\varphi \in \Sigma_1^1(\mathrm{L}_{\infty\omega})$ and $\psi \in \Pi_1^1(\mathrm{L}_{\infty\omega})$ (in the vocabulary $L(U)$):*

1. *φ implies ψ along U-preserving potential L-isomorphisms.*

2. *There exists a U-positive formula χ with $\varphi \models \chi$ and $\chi \models \psi$.*

Proof: Given our general results, we need only verify that the U-positive formulas have the Scott Property with respect to the pebble game just defined. The proof is just like the proof of Scott's Theorem except at stage 0 we do not allow any negated literals involving U. QED

We could similarly prove a version for homomorphisms and positive formulas, of course, but turn instead to a different topic.

3.3 A modal preservation theorem

Modal logic is usually formulated as a propositional logic with additional operators. However, one can also treat it as a limited fragment of quantificational logic, as stressed in van Benthem [10, 11]. Doing so facilitates the discovery of illuminating relationships between quantificational logic and modal logic.

Example 2 Consider the formula $\Box p_1 \rightarrow p_1$ of modal logic. Under the possible-worlds interpretation of the necessity operator, this has an implicit universal quantification over possible worlds y accessible from some given possible world x. This is written as $\forall y \in x\ P_1(y) \rightarrow P_1(x)$ in $\mathrm{L}^{\diamond}_{\infty\omega}$, where the quantification is made explicit.

Thus, in this section, the vocabularies always consist of a set of unary relation symbols P_j (where j ranges over some index set J) and a special designated binary symbol $\in$. Structures are of the form $\boldsymbol{A} = \langle A, E, P_j \rangle_{j \in J}$. Here A is thought of as a set of sets or worlds, P_j defines the set of those sets containing the urelement p_j or the worlds where the atomic sentence p_j is true, and aEb can mean either that a is an element of b or that a is accessible from b. (Note that we are reversing the accessibility relation from the usual way of doing it. This makes the formulas of $\mathrm{L}^{\diamond}_{\infty\omega}$ more readable, and makes a better fit in the case where we want to interpret the predicate $\in$ as membership. This interpretation was studied extensively in Barwise and Moss [9] in the context of non-wellfounded set theory. It is also explored in d'Agostino et. al. [1].

The modal fragment $\mathrm{L}^{\diamond}_{\infty\omega}$ of $\mathrm{L}_{\infty\omega}$ consists of all formulas of $\mathrm{L}_{\infty\omega}$ containing exactly one free variable built up from unary literals $P_j(x), \neg P_j(x)$, using arbitrary conjunctions and disjunctions of formulas with the same free variable, as well as bounded quantification: if $\varphi(y) \in \mathrm{L}^{\diamond}_{\infty\omega}$ and x is a variable other than y then $\exists y \in x\ \varphi(y)$ and $\forall y \in x\ \varphi(y)$ are in $\mathrm{L}^{\diamond}_{\infty\omega}$. When we speak of a $\Sigma^1_1(\mathrm{L}^{\diamond}_{\infty\omega})$ or $\Pi^1_1(\mathrm{L}^{\diamond}_{\infty\omega})$ formula, we allow only quantification over unary predicate symbols, not over $\in$. (This corresponds to quantifying over propositions in the usual way for formulating modal logic.)

In this setting, a bisimulation is a one-pebble game between two models testing for the same atoms on related points, and allowing only back-and-forth choices for the two players that are $\in$-successors of the current positions. More formally, given $\boldsymbol{A}, \boldsymbol{B}$ and $I \subseteq A \times B$ define $\Gamma_{\diamond}(I)$ to be the set of those pairs $\langle a, b \rangle$ such that $P_j(a)$ iff $P_j(b)$ (for all $j \in J$), for every $a_0 \in^{\boldsymbol{A}} a$ there is a $b_0 \in^{\boldsymbol{B}} b$

such that $\langle a_0, b_0 \rangle \in I$, and vice versa. This is easily seen to be equivalent to the more usual definition. Stated in this coinductive fashion, again, our previous analysis applies, including approximation via modal fragments up to quantifier rank α (see Barwise and Moss [9] for details).

Theorem 15 *The following are equivalent for all $\varphi \in \Sigma_1^1(\mathrm{L}_{\infty\omega})$ and $\psi \in \Pi_1^1(\mathrm{L}_{\infty\omega})$ both with one free variable x:*

1. *φ entails ψ along bisimulation.*
2. *There exists some modal formula χ with $\varphi \models \chi$ and $\chi \models \psi$.*

The following consequence of our result was obtained by different means in van Benthem [14].

Corollary 16 (Modal Invariance Theorem) *A formula $\varphi(x)$ of of $\mathrm{L}_{\infty\omega}$ is preserved under bisimulation iff it is equivalent to a formula of $\mathrm{L}^{\diamond}_{\infty\omega}$.*

The finitary version of this result was already known; see van Benthem [10]. (See Andréka, et. al. [2] on this topic for more general "modal fragments" of first-order languages.)

If we restrict attention to formulas in the modal fragment, we can do even better.

Theorem 17 *The following are equivalent for all $\varphi \in \Sigma_1^1(\mathrm{L}^{\diamond}_{\infty\omega})$ and $\psi \in \Pi_1^1(\mathrm{L}^{\diamond}_{\infty\omega})$ both with one free variable x:*

1. *φ entails ψ.*
2. *φ entails ψ along bisimulation.*
3. *There exists some modal formula χ with $\varphi \models \chi$ and $\chi \models \psi$.*

PROOF: The equivalence of (2) and (3) is a special case of the previous theorem. Clearly (2) entails (1). However, it turns out that (1) entails (2). To see this, suppose that Q_1 and Q_2 are distinct unary symbols, and that $\varphi(w) = \exists Q_1 \varphi_1(Q_1, w)$ and $\psi(w) = \forall Q_2 \psi_2(Q_2, w)$. (The case with more second-order quantifiers over unary predicates is similar.) Suppose that $\varphi(w) \models \psi(w)$, that $\boldsymbol{A}, a \models \varphi(w)$, $\boldsymbol{B}, b \models \neg\psi$, and that R is a bisimulation between $\boldsymbol{A}$ and $\boldsymbol{B}$ with aRb. Using (1), we need to get a contraction. With a slight use-mention confusion, let $(\boldsymbol{A}, Q_1, a) \models \varphi_1$ and $(\boldsymbol{B}, Q_2, b) \models \neg\psi_2$. Define a structure $\boldsymbol{C}$ as follows:

$$C = \{\langle w, v\rangle \in A \times B \mid wRv\}$$

$$\langle w, v\rangle \in^{\boldsymbol{C}} \langle w', v'\rangle \text{ iff } w \in^{\boldsymbol{A}} w' \text{ and } v \in^{\boldsymbol{B}} v'$$

If P is one of the unquantified predicates, and so defined in both structures, we know that the structures agree about P along the bisimulation relation so that we can define:

$$\langle w, v\rangle \in P^{\boldsymbol{C}} \text{ iff } w \in P^{\boldsymbol{A}} \text{ iff } v \in P^{\boldsymbol{B}}$$

As for the new predicates, we define

$$\langle w, v\rangle \in Q_1^{\boldsymbol{C}} \text{ iff } w \in Q_1$$

and

$$\langle w, v\rangle \in Q_2^{\boldsymbol{C}} \text{ iff } v \in Q_2.$$

Now define relations $\sim_1$ and $\sim_2$ on $C\times A$ and $C\times B$ by $\langle w, v\rangle \sim_1 w$ and $\langle w, v\rangle \sim_2 v$. It is easy to see that these are bisimulations on the obvious structures. (They are, in fact, what is known as p-morphisms). Since $\langle a, b\rangle \sim_1 a$, we see that $\boldsymbol{C}, \langle a, b\rangle \models \varphi_1$. Hence $\boldsymbol{C}, \langle a, b\rangle \models \psi_2$ by (1). But since $\langle a, b\rangle \sim_2 b$ we see that $(\boldsymbol{B}, Q_2, b) \models \psi_2$, which is a contradiction. QED

To demonstrate another instance of this technique, we state a Łos-Tarski result.

Theorem 18 *For infinitary modal formulas φ, the following are equivalent:*

1. *φ is preserved under submodels.*
2. *φ is equivalent to some* universal *modal formula, i.e., one constructed without using $\exists$ (i.e, without using $\diamond$).*

PROOF: This time, our general strategy makes (2) equivalent to

3. φ is downward preserved along *simulations* (i.e., bisimulations without the backward zigzag clause).

A simple model-theoretic argument establishes what we need:

Lemma 3 *For any pair of models $\boldsymbol{A}, \boldsymbol{B}$, the following are equivalent:*

1. *There exists a simulation from $\boldsymbol{A}$ into $\boldsymbol{B}$.*
2. *There exist models $\boldsymbol{A}^+, \boldsymbol{B}^+$ such that $\boldsymbol{A}$ is bisimilar to $\boldsymbol{A}^+$, $\boldsymbol{B}$ is bisimilar to $\boldsymbol{B}^+$, and $\boldsymbol{A}^+$ is isomorphic to a submodel of $\boldsymbol{B}^+$.*

The models$\boldsymbol{A}^+$ and $\boldsymbol{B}^+$ arise from $\boldsymbol{A}$ and $\boldsymbol{B}$ by "unwinding" the latter into trees, with suitable *duplication* of nodes. In more detail, let κ be the cardinality of A. The nodes of the new models are finite sequences $\langle a_1, \ldots, a_k\rangle$ of $\in$-successors. However, in constructing $\boldsymbol{B}^+$ we duplicate nodes so that each node has κ copies of each former successor. This can be done so as to maintain a bisimulation. One can then define an embedding of $\boldsymbol{A}^+$ into $\boldsymbol{B}^+$ by induction on the length k of the nodes in the tree, using the simulation. It is easy to see that this gives an embedding that preserves all literals. QED

These positive results raise a general question. Can we extend all standard interpolation and preservation theorems to suitably *bounded-quantifier* fragments of $\mathrm{L}_{\infty\omega}$? (Cf. Andréka, et. al. [2] for very general fragments of this kind.)

3.4 Finite Variable Fragments

The finite variable fragments $\mathrm{L}^{(k)}_{\infty\omega}$ of $\mathrm{L}_{\infty\omega}$ have become a standard tool in the study of finite model theory in computer science. For these languages, one assumes that there are only relation symbols and constant symbols in the vocabularies, no function symbols. We also require that there be at most k constants and (free or bound) variables in any formula of $\mathrm{L}^{(k)}_{\infty\omega}$. The right pebble game for $\mathrm{L}^{(k)}_{\infty\omega}$ has only partial isomorphisms up to size k, with back and forth conditions up to that size (see Barwise [6] for details). The partial isomorphisms are not required to be defined on all constants of the vocabulary. Again, this fits all conditions of our analysis. The Scott property for this fragment was observed by Dawar [16]. Thus we obtain the following:

Theorem 19 *The following are equivalent for formulas $\varphi \in \Sigma^1_1(\mathrm{L}_{\infty\omega}), \psi \in \Pi^1_1(\mathrm{L}_{\infty\omega})$ with at most k constants or free variables between them:*

1. *φ implies ψ along potential isomorphisms of size k.*

2. *There exists some $\theta \in \mathrm{L}^{(k)}_{\infty\omega}$ with $\varphi \models \theta$ and $\theta \models \psi$.*

In particular, a sentence of $\mathrm{L}_{\infty\omega}$ is preserved under potential isomorphisms of size k iff it is logically equivalent to to a sentence of $\mathrm{L}^{(k)}_{\infty\omega}$.

The preservation part of this is an extension of a semantic characterization for first-order k-variable fragments given in van Benthem [12].

Notice that the counterexample to standard interpolation for $\mathrm{L}_{\infty\omega}$ given above actually gives us sentences in $\varphi \in \Sigma^1_1(\mathrm{L}^{(2)}_{\infty\omega})$ and $\psi \in \Pi^1_1(\mathrm{L}^{(2)}_{\infty\omega})$ such that $\varphi \models \psi$ but with no interpolant in $\mathrm{L}_{\infty\omega}$; see, e.g, Ebbinghaus and Flum [17]. The following example, adapted from Andréka, et. al [2], shows that there are *finitary* Σ^1_1 and Π^1_1 sentences in two variables with no interpolant in $\mathrm{L}^{(2)}_{\infty\omega}$.

Example 3 Let φ be $\exists P \varphi_0(P)$ and ψ be $\forall P, Q \neg\psi_0(P, Q)$ where φ_0 is the conjunction of

$$\forall x, y\ ((P(x) \wedge P(y)) \rightarrow x = y)$$
$$\forall x, y\ ((\neg P(x) \wedge \neg P(y)) \rightarrow x = y)$$

and ψ_0 is

$$\exists x((P(x) \wedge Q(x)) \wedge \exists x(Px \wedge \neg Q(x)) \wedge \exists x(\neg P(x) \wedge Q(x)$$

Both φ and ψ hold in $\boldsymbol{A}$ iff A has at most two elements. Hence $\varphi \models \psi$. The natural first-order interpolant, using only $=$, is

$$\forall xyz(x = y \vee x = z \vee y = z),$$

a sentence with three variables. But φ does not entail ψ along potential isomorphisms of size two, since any set of size two is potentially 2-isomorphic to any set of size at least two. Hence there can be no interpolant in the language $\mathrm{L}^{(2)}_{\infty\omega}$.

Given the interest in these languages, it is perhaps worth calling attention to the following consequence of our theorem. The proof is just like our earlier proof.

Corollary 20 *(A Beth Theorem for $\mathrm{L}^{(k)}_{\infty\omega}$) Let $n \leq k$, let P be an n-ary predicate symbol not in L and let $L' = L \cup \{P\}$. The following are equivalent for all $\varphi(P)$ in $L'_{\infty\omega}$:*

1. *There is a formula $\delta(x_1 \dots x_n) \in \mathrm{L}^{(k)}_{\infty\omega}$ such that*

$$\varphi \models \forall x_1 \dots x_n \, (P(x_1 \dots x_n) \leftrightarrow \delta(x_1 \dots x_n))$$

2. *If $(\boldsymbol{A}, P^{\boldsymbol{A}})$ and $(\boldsymbol{B}, P^{\boldsymbol{B}})$ are L' structures that are models of $\varphi(P)$ and if $I : \boldsymbol{A} \cong^{(k)}_p \boldsymbol{B}$ then $I : (\boldsymbol{A}, P^{\boldsymbol{A}}) \cong^{(k)}_p (\boldsymbol{B}, P^{\boldsymbol{B}})$.*

This may be contrasted with the failure of the standard Beth Theorem noted in Sain [25].

4 Further Directions

4.1 Generalized semantics

Andréka, et. al. [2] have suggested that by using a more general semantics, the consequence $\varphi \models \psi$ becomes stronger, and hence possibly a better match for interpolation or preservation results. Combining this with the notion of potential isomorphism suggests the following notion of a *generalized model*. Given an L-structure $\boldsymbol{A}$, let S be a set of partial assignments on $\boldsymbol{A}$. Define $(\boldsymbol{A}, S), s \models \varphi$ as usual, where $s \in S$, $dom(s) \supseteq FV(\varphi)$, with the following clause for the existential quantifier (and a dual clause for the universal quantifier):

$$(\boldsymbol{A}, S), s \models \exists x\varphi \;\text{ iff for some } d \in A, s^x_d \in S \;\text{ and } (\boldsymbol{A}, S), s^x_d \models \varphi$$

Ordinary model-theoretic notions extend to generalized models. For instance, $(\boldsymbol{A}, S)$ is a *generalized submodel* of $(\boldsymbol{B}, S')$ if $\boldsymbol{A} \subseteq \boldsymbol{B}$ and $S \subseteq S'$.

Theorem 21 *For all $\mathrm{L}_{\infty\omega}$-formulas φ not containing $=$, the following are equivalent:*

1. *φ is preserved under generalized submodels.*

2. *φ is equivalent to a universal formula.*

We do not know if this result holds true for sentences with $=$.

The proof uses a notion of partial isomorphism appropriate to generalized models. A *potential isomorphism* between generalized models $(\boldsymbol{A}, S)$ and $(\boldsymbol{B}, S')$ is a binary relation I on $S \times S'$ such that sIt implies:

1. s and t has the same domain;

2. there is a unique partial isomoprhism f between $rng(s)$ and $rng(t)$ such that $t = s \circ f$;

3. for every $s' \in S$ with $s =_x s'$ there is a $t' \in S'$ such that $t =_x t'$ and $s'It'$; and

4. for every $t' \in S'$ with $t =_x t'$ there is a $s' \in S$ such that $s =_x s'$ and $s'It'$.

We write $I : (\boldsymbol{A}, S) \cong_p (\boldsymbol{B}, S')$ to indicate that I is a potential isomorphism of generalized models. Notice the similarity of this notion to that of bisimulation, when one thinks of variable assignments as states. It is easy to show that if $I : (\boldsymbol{A}, S) \cong_p (\boldsymbol{B}, S')$ and sIs' then s and s' satisfy the same formulas in the respective generalized models.

For any ordinary L-structure $\boldsymbol{A}$, we write $\mathrm{fa}(\boldsymbol{A})$ for the set of all finite assignments taking values in A, and write $\boldsymbol{A}^*$ for the generalized model $(\boldsymbol{A}, \mathrm{fa}(\boldsymbol{A}))$. It is clear that truth and satisfaction in $\boldsymbol{A}$ and $\boldsymbol{A}^*$ agree.

Proof of Theorem 21: The direction from (2) to (1) is obvious. For the converse, by our earlier result it suffices to show that (1) implies that φ entails φ downward along potential embeddings. So, let $\boldsymbol{A} \models \varphi$, where $\boldsymbol{B}$ has a potential embedding E into $\boldsymbol{A}$. We need to prove that $\boldsymbol{A} \models \varphi$. To do this we construct an L structure $\boldsymbol{B}_E$ and a set S of variable assignments in $\boldsymbol{B}_E$ satisfying:

1. $\boldsymbol{B}_E \models \varphi$

2. $\boldsymbol{A}^* \cong_p (\boldsymbol{B}_E, S)$

Since φ is preserved under generalized submodels and $(\boldsymbol{B}_E, S)$ is a generalized submodel of $\boldsymbol{B}_E^*$, we have $(\boldsymbol{B}_E, S) \models \varphi$, by (1), and so $\boldsymbol{A}^* \models \varphi$ by (2) and hence $\boldsymbol{A} \models \varphi$. So we need only construct $\boldsymbol{B}_E$ and S having these properties. The structure $\boldsymbol{B}_E$ has domain

$$B_E \;=\; \{\langle b, f\rangle \mid f \in E, b \in rng f\} \cup (B - \bigcup_{f \in E} (rng(f)).$$

Let $F : B_E \to B$ be defined in the obvious way: $F(\langle b, f\rangle) = b$ on the first kind of element and $F(b) = b$ on the second kind of element. The interpretation of a predicate Q in $\boldsymbol{B}_E$ copies its behavior in $\boldsymbol{B}$ so that F is a strong homomorphism from $\boldsymbol{B}_E$ to $\boldsymbol{B}$. Strong homomorphisms preserve truth of =-free sentences so $\boldsymbol{B}_E \models \varphi$, as desired.

The set S of partial assignments consists of all finite assignments t such that each $t(x)$ is a pair $\langle b_x, f_x\rangle$ and the set $\{f_x \mid x \in dom(t)\}$ is linearly ordered by inclusion; this insures that the f_x are all subfunctions of a common partial embedding. To prove that $\boldsymbol{A}^* \cong_p (\boldsymbol{B}_E, S)$, define a relation I on $\mathrm{fa}(\boldsymbol{A}) \times S$ as follows:

> sIt iff $dom(s) = dom(t)$ and for all x in this domain, if $t(x) = \langle b, f\rangle$ then $f(s(x)) = b$.

We need to verify that $I : \boldsymbol{A}^* \cong_p (\boldsymbol{B}_E, S)$. So assume sIt. Clauses (1) and (2) are true by definition of I. We need to verify the back and forth conditions. For (3), suppose $s =_x s'$. Let f be the largest function associated with t and extend it using the Forth property of E. This gives a $t' \in S$ such that $s'It'$ and $t =_x t'$. For (4), suppose that $t =_x t'$. Let $t'(x) = \langle b, g\rangle$. Since $t' \in S$, g has to be compatible with the remaining functions associated with t'. Define s' by $s'(x) = g^{-1}(b)$, $s' =_x s$. Clearly $s'It'$, as desired. QED

4.2 Generalized consequence relations

Our notion of *consequence along* R may be seen as a generalized notion of logical consequence, which shares many features with its classical counterpart. For instance, for arbitrary relations R, it satisfies Strengthening and Disjunction of antecedents, as well as Weakening and Conjunction of consequents. Nevertheless, these are also differences. E.g., we observed that consequence along potential isomorphisms did not satisfy a Deduction Theorem. This raises an obvious natural problem: find a complete system of deduction for this relation for the languages considered here. The second author has some results in this direction which he hopes to pursue in a later paper. These involve calculi of sequents $\Gamma \Rightarrow_L \Delta$, where the inference rules also manipulate vocabulary sets L. Complete calculi of this kind may exist even for finite variable fragments, where standard proof theory is quite complex (cf. Andréka, et. al. [2]).

4.3 Interpolation for $\mathrm{L}_{\omega_1\omega}$

All the above results may be generalized to regular fragments L of $\mathrm{L}_{\infty\omega}$, but the results do not always give one as much as one would like. For example, if we apply the same method to $\mathrm{L}_{\omega_1\omega}$ the interpolation result obtained is weaker than Lopez-Escobar's. Namely, we obtain that a $\Sigma^1_1(\mathrm{L}_{\omega_1\omega})$ sentence φ entails a $\Pi^1_1(\mathrm{L}_{\omega_1\omega})$ sentence ψ iff there is a θ that is a disjunction of set sentences of $\mathrm{L}_{\omega_1\omega}$. The disjunction is not insured to be countable, however, and so θ may not be a sentence of $\mathrm{L}_{\omega_1\omega}$. On the other hand, interpolation is known to hold for $\mathrm{L}_{\omega_1\omega}$, by other methods. The analogous result for $\mathrm{L}_{\omega\omega}$ works, on the other hand, by compactness. This raises the obvious question: can one obtain the usual interpolation theorem for $L_{\omega_1\omega}$ by some improvement on the methods of this paper?

4.4 Other infinitary languages

It seems one should be able to extend the results here to certain infinitary languages with generalized quantifiers, as long as the Boundedness Theorem continues to hold, since such languages typically have pebble games. It is more challenging to reconsider the infinitary language $L_{\infty\kappa}$ for $\kappa > \omega$ in light of the results obtained here. Do any of these languages admit of results similar to those for $L_{\infty\omega}$ presented here? Chang [15] shows there are analogues of the pebble game and Scott's Theorem for such languages, so that is promising. There is nothing analogous to the Boundedness Theorem for for such languages, however, since the class of well-orderings is definable in these languages. Consequently the proof method of this paper breaks down at a seemingly crucial step.

References

1. d. Agostino, G., J. van Benthem, A. Montanari and A. Policriti, "Modal deduction in second-order logic and set theory." Report ML-95-02, ILLC, University of Amsterdam. **J. Logic and Computation** (to appear)

2. Andréka, H., J van Benthem and I. Németi, "Modal Logics and Bounded First-Order Fragments." Report ML-96-03, ILLC, Amsterdam. **J. Phil. Logic** (to appear)

3. Barwise, J. "Back and forth through infinitary logic." *Studies in Model Theory.* (ed. by M. Morley) 5-35, MAA Studies in Mathematics (1973)

4. Barwise, J. "Axioms for abstract model theory." **Annals of Math Logic**. 7:221-265 (1974)

5. Barwise, J. *Admissible Sets and Structures.* Springer-Verlag (1975)

6. Barwise, J. "On Moschovakis closure ordinals." **J. Symbolic Logic**. 42:292-98 (1977)

7. Barwise, J. and K. Kunen, "Hanf numbers for fragments of $L_{\infty\omega}$." **Israel J. Math.** 10:306-320 (1971)

8. Barwise, J. and Y. Moschovakis, "Global inductive definability." **J. Symbolic Logic**. 43:521-534 (1978)

9. Barwise, J. and L. Moss, *Vicious Circles.* CSLI Publications, Stanford (1996)

10. van Benthem, J. *Modal Correspondence Theory.* Ph.D. Thesis, Mathematics Institute, University of Amsterdam. (1976)

11. van Benthem, J. *Modal Logic and Classical Logic.* Bibliopolis, Naples. (1985)

12. van Benthem, J. *Language in Action.* N. Holland, Amsterdam. (1991)

13. van Benthem, J. "Programming operations that are safe for bisimulation." Logic Colloquium 1994, Clermont-Ferrand. **Studia Logica** (to appear)

14. van Benthem, J. *Exploring Logical Dynamics.* Studies in Logic, Language and Information, CSLI Publications, Stanford. (to appear)

15. Chang, C.C. "Some remarks on the model theory of infinitary languages." *The Syntax and Semantics of Infinitary Languages.* (ed. by J. Barwise) LNM 72, Springer-Verlag. (1968)

16. Dawar, A. "Feasible computation through model theory." Ph.D. dissertation. University of Pennsylvania. (1993)

17. Ebbinghaus, H-D, and J. Flum, *Finite Model Theory.* Springer-Verlag (1995)

18. Karp, C. "Finite quantifier equivalence." *The Theory of Models.* (ed. by J.Addison, et. al.) N. Holland 407-412 (1965)

19. Keisler, H. J. *Model Theory for Infinitary Logic*, N. Holland (1971)

20. Lopez-Escobar, E. "An interpolation theorem for denumerably long sentences." **Fund. Math.** LVII 253-272 (1965)

21. Lopez-Escobar, E. "On definable well-orderings." **Fund. Math.** LVIX 13-21, 299-300 (1966)

22. Malitz, J. "Infinitary analogues of theorems from first-order model theory." **J. Symbolic Logic** 36:216-228 (1971)

23. Morley, M. "Omitting classes of elements." *The Theory of Models.* (ed. by J.Addison, et. al.) N. Holland 265-273 (1965)

24. Sacks, G. *Saturated Model Theory.* Benjamin (1972)

25. Sain, I. "Beth's and Craig's properties via epimorphism and amalgamation in algebraic logic." *Algebraic Logic and Universal Algebra in Computer Science* (ed. by C. Bergman et. al.) LNCS, Springer-Verlag 209-226 (1990)

26. Scott, D."Invariant Borel sets." **Fund. Math**. 56:117-128 (1964)