An Interpolation Theorem*

Martin Otto

November 98

Abstract

Lyndon's Interpolation Theorem asserts that for any valid implication between two purely relational sentences of first-order logic, there is an interpolant in which each relation symbol appears only in those polarities in which it appears in both the antecedent and the succedent of the given implication. This note proves a similar interpolation result under the additional requirement that, for some fixed tuple $\mathbb{U}$ of unary predicates U, all formulae under consideration have all quantifiers explicitly relativized to one of the U. Under this stipulation, existential / universal quantification over U contributes a positive / negative occurrence of U.

It is shown how this single new interpolation theorem unifies a number of related results: namely, the classical preservation theorems concerning extensions / substructures with those concerning monotonicity, as well as a many-sorted interpolation theorem focusing on existentially / universally quantified sorts.

Consider first-order logic *with or without equality* in a finite, purely relational vocabulary τ. Boolean constants $\top$ and $\bot$ are taken to be atomic first-order formulae. We use the set $\mathrm{Occ} = \tau \times \{+, -\}$ to code polarities of predicate occurrences in formulae, through a mapping

$$occ\colon \varphi \longmapsto occ(\varphi) \subseteq \mathrm{Occ},$$

where $(R, +) \in occ(\varphi)$ if R occurs positively in φ, $(R, -) \in occ(\varphi)$ if R occurs negatively. As usual $free(\varphi)$ stands for the set of free variables in φ.

Let $\mathbb{U}$ be a tuple of designated unary predicates U in τ. We say that a formula φ is $\mathbb{U}$-*relativized*, if each quantifier in φ is explicitly relativized to some U, i.e. is of the form $\exists x(Ux \wedge \dots)$ or $\forall x(Ux \rightarrow \dots)$ for some U in $\mathbb{U}$. Up to logical equivalence, the $\mathbb{U}$-relativized formulae correspond exactly to the relativizations of first-order formulae to $\bigcup_{U\in\mathbb{U}} U$. This is actually even true up to a restricted form of logical equivalences which preserve polarities of predicate occurrences. We want to prove the following variant of Lyndon's Interpolation Theorem.

*This note arose out of the context of a Stanford Logic Seminar discussion with Johan van Benthem and Solomon Feferman, revolving around different accounts of a preservation result for Chu transforms and its relation to many-sorted interpolation. I am deeply indebted to both, Professor Feferman and Professor van Benthem, for their academic hospitality and personal kindness during my stay at Stanford as a visiting scholar in 1997/98.

Theorem 1 *Let φ and ψ be $\mathbb{U}$-relativized formulae such that $\varphi \models \psi$. Then there is a $\mathbb{U}$-relativized Lyndon interpolant χ for $\varphi \models \chi$:*

(i) χ is a $\mathbb{U}$-relativized formula.

(ii) $free(\chi) \subseteq free(\varphi) \cap free(\psi)$.

(iii) $occ(\chi) \subseteq occ(\varphi) \cap occ(\psi)$.

(iv) $\varphi \models \chi$ and $\chi \models \psi$.

The statement of the theorem has two readings, one for first-order with equality, and one for first-order without equality.

This strengthens Lyndon's Interpolation Theorem [7] (with or without $=$), which may be recovered from the theorem by trivializing the relativization.

Theorem 2 (Lyndon's Interpolation Theorem) *Let φ and ψ be formulae such that $\varphi \models \psi$. Then there is a Lyndon interpolant χ for $\varphi \models \chi$:*

(i) $occ(\chi) \subseteq occ(\varphi) \cap occ(\psi)$.

(ii) $free(\chi) \subseteq free(\varphi) \cap free(\psi)$.

(iii) $\varphi \models \chi$ and $\chi \models \psi$.

This follows from Theorem 1 if we put $\hat{\tau} = \tau \dot{\cup} \{U\}$ for a new unary U, $\mathbb{U} = \{U\}$ and pass from φ and ψ to their relativizations to U, $\hat{\varphi}$ and $\hat{\psi}$. By the theorem, there is a U-relativized Lyndon interpolant $\hat{\chi}$ for $\bigwedge U\boldsymbol{x} \wedge \hat{\varphi} \models \hat{\psi}$ (here $\bigwedge U\boldsymbol{x}$ is shorthand for $\bigwedge_{i=1}^{n} Ux_i$ if $\boldsymbol{x} = (x_1, \ldots, x_n)$. One obtains the desired Lyndon Interpolant χ for $\varphi \models \psi$ by replacing every atom of the form Uy in $\hat{\chi}$ by $\top$.

We turn to the proof of the theorem, and first introduce some terminology and notation. The cases with and without equality can be treated together. Actually, the case with equality requires only one minor systematic modification, as will be indicated immediately. Let $\mathfrak{A}$ and $\mathfrak{B}$ be τ-structures with universes A and B respectively. We consider certain subsets $p \subseteq A \times B$ which are to be viewed as weak partial isomorphisms. *This notion calls for the one crucial modification if we want to deal with equality: in that case, and in that case only, all the p under consideration are required to be the graphs of partial 1–1 functions; in the case without equality, we consider a priori arbitrary subsets of $A \times B$.*

If $p \subseteq A \times B$, we regard $\mathrm{dm}(p) = \{a \in A \,|\, (\exists b \in B)((a, b) \in p)\}$ as the domain of p, and $\mathrm{im}(p) = \{b \in B \,|\, (\exists a \in A)((a, b) \in p)\}$ as the image of p. If $\boldsymbol{a} = (a_1, \ldots, a_n)$, $\boldsymbol{b} = (b_1, \ldots, b_n)$, and if $(a_i, b_i) \in p$ for $i = 1, \ldots, n$, we simply indicate this situation by writing $\boldsymbol{a}\,\boldsymbol{b} \in p$.

Let $O \subseteq \mathrm{Occ}$. We say that p *preserves* O if for all $R \in \tau$, if R is n-ary and if $\boldsymbol{a} \in A^n$,$\boldsymbol{b} \in B^n$ are such that $\boldsymbol{a}\,\boldsymbol{b} \in p$, then

– if $(R,+) \in O$: $\mathfrak{A} \models R\boldsymbol{a} \Rightarrow \mathfrak{B} \models R\boldsymbol{b}$, and

– if $(R,-) \in O$: $\mathfrak{A} \models \neg R\boldsymbol{a} \Rightarrow \mathfrak{B} \models \neg R\boldsymbol{b}$.

If $O \subseteq \mathrm{Occ}$ and $\mathfrak{A} = (A, (U^{\mathfrak{A}})_{U\in\mathbb{U}}, \dots)$ we let $A^{+/O}$ and $A^{-/O}$ denote those parts of the universe A that lie within some $U^{\mathfrak{A}}$ which is positive or negative, respectively, according to O:

$$A^{+/O} = \bigcup_{(U,+)\in O} U^{\mathfrak{A}} \quad \text{and} \quad A^{-/O} = \bigcup_{(U,-)\in O} U^{\mathfrak{A}}.$$

Definition 3 A set $P \subseteq \{p \subseteq A \times B \mid p \text{ preserves } O\}$ [1] is a *back-and-forth-system with respect to O and $\mathbb{U}$*, if $P \neq \emptyset$ is such that for all $p \in P$, and all $\boldsymbol{a}\,\boldsymbol{b} \in p$:

– if $a \in A^{+/O}$, then there are $p' \in P$ and $b \in B$ such that $\boldsymbol{a}a\boldsymbol{b}b \in p'$.

– if $b \in B^{-/O}$, then there are $p' \in P$ and $a \in A$ such that $\boldsymbol{a}a\boldsymbol{b}b \in p'$.

The following are two related, asymmetric notions of similarity between structures, one algebraic in spirit, the other semantic. They have the same relationship between them as do partial isomorphism and elementary equivalence.

Definition 4 We write $(\mathfrak{A}, \boldsymbol{a}) \longrightarrow^{\mathbb{U}}_{O} (\mathfrak{B}, \boldsymbol{b})$ if there is a back-and-forth-system P w.r.t. O and $\mathbb{U}$ with $\boldsymbol{a}\,\boldsymbol{b} \in p$ for some $p \in P$. We also write $P\colon (\mathfrak{A}, \boldsymbol{a}) \longrightarrow^{\mathbb{U}}_{O} (\mathfrak{B}, \boldsymbol{b})$ in this situation, and $p\colon (\mathfrak{A}, \boldsymbol{a}) \longrightarrow^{\mathbb{U}}_{O} (\mathfrak{B}, \boldsymbol{b})$ if P consists of one single element $p \subseteq A \times B$.

Definition 5 We write $(\mathfrak{A}, \boldsymbol{a}) \Longrightarrow^{\mathbb{U}}_{O} (\mathfrak{B}, \boldsymbol{b})$, if for all $\mathbb{U}$-relativized formulae $\varphi(\boldsymbol{x})$ with $occ(\varphi) \subseteq O$: $\mathfrak{A} \models \varphi[\boldsymbol{a}] \quad \Rightarrow \quad \mathfrak{B} \models \varphi[\boldsymbol{b}]$. For $p \subseteq A \times B$, we write $p\colon \mathfrak{A} \Longrightarrow^{\mathbb{U}}_{O} \mathfrak{B}$, if $(\mathfrak{A}, \boldsymbol{a}) \Longrightarrow^{\mathbb{U}}_{O} (\mathfrak{B}, \boldsymbol{b})$ for all $\boldsymbol{a}\,\boldsymbol{b} \in p$.

Lemma 6 *If $P\colon \mathfrak{A} \longrightarrow^{\mathbb{U}}_{O} \mathfrak{B}$ then $p\colon \mathfrak{A} \Longrightarrow^{\mathbb{U}}_{O} \mathfrak{B}$ for all $p \in P$.*
In particular, $(\mathfrak{A}, \boldsymbol{a}) \longrightarrow^{\mathbb{U}}_{O} (\mathfrak{B}, \boldsymbol{b})$ implies $(\mathfrak{A}, \boldsymbol{a}) \Longrightarrow^{\mathbb{U}}_{O} (\mathfrak{B}, \boldsymbol{b})$.

This is proved by a straightforward induction on $\mathbb{U}$-relativized φ with $occ(\varphi) \subseteq O$ in negation normal form.

Lemma 7 *Let $\mathfrak{A}$ and $\mathfrak{B}$ be ω-saturated. Then $(\mathfrak{A}, \boldsymbol{a}_0) \Longrightarrow^{\mathbb{U}}_{O} (\mathfrak{B}, \boldsymbol{b}_0)$ implies $P\colon (\mathfrak{A}, \boldsymbol{a}_0) \longrightarrow^{\mathbb{U}}_{O} (\mathfrak{B}, \boldsymbol{b}_0)$, where $P = \{p \subseteq A \times B \mid p \text{ finite}, p\colon \mathfrak{A} \Longrightarrow^{\mathbb{U}}_{O} \mathfrak{B}\}$.*

Proof. P is nonempty: $(p\colon \boldsymbol{a}_0 \mapsto \boldsymbol{b}_0) \in P$ as $(\mathfrak{A}, \boldsymbol{a}_0) \Longrightarrow^{\mathbb{U}}_{O} (\mathfrak{B}, \boldsymbol{b}_0)$. All $p \in P$ preserve O by definition. It remains to check the O-$\mathbb{U}$-back-and-forth-property. We do the 'forth'-part. Let $\boldsymbol{a}\,\boldsymbol{b} \in p \in P$, and assume that $a \in U^{\mathfrak{A}}$, $(U,+) \in O$. We are seeking $b \in U^{\mathfrak{B}}$ for which $(\mathfrak{A}, \boldsymbol{a}a) \Longrightarrow^{\mathbb{U}}_{O} (\mathfrak{B}, \boldsymbol{b}b)$. Note that $(\mathfrak{A}, \boldsymbol{a}) \Longrightarrow^{\mathbb{U}}_{O} (\mathfrak{B}, \boldsymbol{b})$ holds, since $\boldsymbol{a}\,\boldsymbol{b} \in p$. Put

$$\Phi(\boldsymbol{a}, x) = \{\varphi(\boldsymbol{a}, x) \mid \varphi \ \mathbb{U}\text{-relativized}, occ(\varphi) \subseteq O, \mathfrak{A} \models \varphi[\boldsymbol{a}, a]\}.$$

[1] Recall that, for the reading with equality, these p are restricted to be 1–1 partial functions.

Clearly any realization of $\Phi(\boldsymbol{b}, x)$ in $\mathfrak{B}$ will be such that $(\mathfrak{A}, \boldsymbol{a}a) \Longrightarrow_O^{\mathbb{U}} (\mathfrak{B}, \boldsymbol{b}b)$. It remains to show that $\Phi(\boldsymbol{b}, x)$ is consistent with $\mathrm{Th}(\mathfrak{B}, \boldsymbol{b})$. Assume to the contrary, that it were inconsistent. As Φ is closed under conjunctions, there would have to be a single $\varphi(\boldsymbol{a}, x) \in \Phi(\boldsymbol{a}, x)$ such that $(\mathfrak{B}, \boldsymbol{b}) \models \neg(\exists x \in U)\varphi(\boldsymbol{b}, x)$. But this contradicts $(\mathfrak{A}, \boldsymbol{a}) \Longrightarrow_O^{\mathbb{U}} (\mathfrak{B}, \boldsymbol{b})$, if we consider the formula $\psi(\boldsymbol{x}) = (\exists x \in U)\varphi(\boldsymbol{x}, x)$.

The 'back'-part is dealt with analogously, only that the partial type under consideration is $\Phi(\boldsymbol{b}, x) = \big\{\neg\varphi(\boldsymbol{b}, x) \;\big|\; \varphi\ \mathbb{U}\text{-relativized}, \mathit{occ}(\varphi) \subseteq O, \mathfrak{B} \models \neg\varphi[\boldsymbol{b}, b]\big\}$. $\square$

Corollary 8 *For $(\mathfrak{A}, \boldsymbol{a}_0) \Longrightarrow_O^{\mathbb{U}} (\mathfrak{B}, \boldsymbol{b}_0)$ there are countable $(\mathfrak{A}^*, \boldsymbol{a}_0)$ and $(\mathfrak{B}^*, \boldsymbol{b}_0)$ such that $(\mathfrak{A}^*, \boldsymbol{a}_0) \equiv (\mathfrak{A}, \boldsymbol{a}_0)$, $(\mathfrak{B}^*, \boldsymbol{b}_0) \equiv (\mathfrak{B}, \boldsymbol{b}_0)$, and $(\mathfrak{A}^*, \boldsymbol{a}_0) \longrightarrow_O^{\mathbb{U}} (\mathfrak{B}^*, \boldsymbol{b}_0)$.*

This follows from the previous lemma, if we first take arbitrary ω-saturated elementary extensions $\mathfrak{A} \preccurlyeq \mathfrak{A}'$ and $\mathfrak{B} \preccurlyeq \mathfrak{B}'$. By the lemma, $(\mathfrak{A}', \boldsymbol{a}_0) \longrightarrow_O^{\mathbb{U}} (\mathfrak{B}', \boldsymbol{b}_0)$. To obtain a countable version of this situation, it suffices to wrap up $(\mathfrak{A}', \boldsymbol{a}_0)$, $(\mathfrak{B}', \boldsymbol{b}_0)$, and the back-and-forth-system P in one first-order structure, and to apply the Löwenheim-Skolem-Tarski Theorem. [2]

Lemma 9 *If $\mathfrak{A}$ and $\mathfrak{B}$ are countable, and $(\mathfrak{A}, \boldsymbol{a}_0) \longrightarrow_O^{\mathbb{U}} (\mathfrak{B}, \boldsymbol{b}_0)$, then there is a $p \subseteq A \times B$ for which $p\colon \mathfrak{A} \longrightarrow_O^{\mathbb{U}} \mathfrak{B}$ and $\boldsymbol{a}_0\boldsymbol{b}_0 \in p$.*

Sketch of proof. Starting from a back-and-forth-system P (w.l.o.g. consisting of finite p, and closed under subsets) and enumerations of $A^{+/O} \subseteq A$ and of $B^{-/O} \subseteq B$ one finds (in the usual back-and-forth fashion) finite approximations to the desired p within P. Their union p satisfies $\mathrm{dm}(p) \supseteq A^{+/O}$ and $\mathrm{im}(p) \supseteq B^{-/O}$. $\square$

The following is the main proposition towards the proof of the theorem; indeed, it may be thought of as the *structural interpolation property* behind the theorem. The situation is depicted in the following sketch.

$$\begin{tikzcd}
\mathfrak{A} \arrow[rr, "p", "O"'] \arrow[dr, "p_1"', "O_1"] & & \mathfrak{B} \\
& \mathfrak{C} \arrow[ur, "p_2"', "O_2"] &
\end{tikzcd}$$

Proposition 10 (main proposition) *Let $O_1, O_2 \subseteq \mathrm{Occ}$, $O = O_1 \cap O_2$.*
If $p\colon \mathfrak{A} \longrightarrow_O^{\mathbb{U}} \mathfrak{B}$, then there are $\mathfrak{C}$, $p_1 \subseteq A \times C$, and $p_2 \subseteq C \times B$ such that $p = p_1 \circ p_2$ and $p_1\colon \mathfrak{A} \longrightarrow_{O_1}^{\mathbb{U}} \mathfrak{C}$ and $p_2\colon \mathfrak{C} \longrightarrow_{O_2}^{\mathbb{U}} \mathfrak{B}$. One may further require that $\mathrm{dm}(p_1) = A$ and $\mathrm{im}(p_2) = B$.

[2] Alternatively, one might want to work with encodings of pairs of structures and back-and-forth-systems from the start, and prove Corollary 8 directly from a model theoretic games perspective, via direct applications of compactness and Löwenheim-Skolem. This alternative approach, which avoids the detour through ω-saturated structures, has been elaborated on and applied to a uniform treatment of numerous preservation and characterization results in [6].

Proof. Let $* \notin A \cup B$. We let the universe C of the desired structure $\mathfrak{C}$ be a subset of $(A \dot\cup \{*\}) \times (B \dot\cup \{*\})$. Put

$$\begin{aligned} C &= p \;\dot\cup\; \Big(\big(A \setminus \mathrm{dm}(p)\big) \times \{*\}\Big) \;\dot\cup\; \Big(\{*\} \times \big(B \setminus \mathrm{im}(p)\big)\Big), \\ p_1 &= \Big\{(a,(a,b)) \;\Big|\; (a,b) \in p\Big\} \;\dot\cup\; \Big\{(a,(a,*)) \;\Big|\; a \in A \setminus \mathrm{dm}(p)\Big\}, \\ p_2 &= \Big\{((a,b),b) \;\Big|\; (a,b) \in p\Big\} \;\dot\cup\; \Big\{((*,b),b) \;\Big|\; b \in B \setminus \mathrm{im}(p)\Big\}. \end{aligned}$$

Note that $p = p_1 \circ p_2$, $\mathrm{dm}(p_1) = A$ and $\mathrm{im}(p_2) = B$ follow directly from the definitions. For the case with equality also note that the p_i are 1–1 partial functions if p is. We now have to fix the interpretation of predicates $R \in \tau$ over C with a view to make p_1 respect O_1, p_2 respect O_2, and to guarantee that $\mathrm{im}(p_1) \supseteq C^{-/O_1}$ and $\mathrm{dm}(p_2) \supseteq C^{+/O_2}$.

Let R be n-ary, $\boldsymbol{c} \in C^n$. We write $\boldsymbol{c} = \boldsymbol{a}\,\boldsymbol{b}$, where in general $\boldsymbol{a} \in (A \dot\cup \{*\})^n$ and $\boldsymbol{b} \in (B \dot\cup \{*\})^n$. We distinguish several cases.

(a) If $\boldsymbol{a} \in A^n$ and $\boldsymbol{b} \in B^n$ and $\mathfrak{A} \models R\boldsymbol{a} \Leftrightarrow \mathfrak{B} \models R\boldsymbol{b}$,
put: $\mathfrak{C} \models R\boldsymbol{c} \Leftrightarrow \mathfrak{A} \models R\boldsymbol{a}$.

(b) If $\boldsymbol{a} \in A^n$ and $\boldsymbol{b} \in B^n$ and $\mathfrak{A} \models R\boldsymbol{a}$ but $\mathfrak{B} \models \neg R\boldsymbol{b}$ (whence $(R,+) \notin O$):

- (i) if $(R,+) \in O_1 \setminus O_2$, put: $\mathfrak{C} \models R\boldsymbol{c}$.
- (ii) if $(R,+) \in O_2 \setminus O_1$, put: $\mathfrak{C} \models \neg R\boldsymbol{c}$.
- (iii) if $(R,+) \notin O_1 \cup O_2$, put whatever.

(c) If $\boldsymbol{a} \in A^n$ and $\boldsymbol{b} \in B^n$ and $\mathfrak{A} \models \neg R\boldsymbol{a}$ but $\mathfrak{B} \models R\boldsymbol{b}$ (which implies that $(R,-) \notin O$), proceed as in (b), based on a corresponding case distinction for $(R,-)$.

(d) If $\boldsymbol{a} \in A^n$ and $\boldsymbol{b} \notin B^n$ (i.e. $\boldsymbol{a} \notin \mathrm{dm}(p)$, and $\boldsymbol{c} \notin \mathrm{dm}(p_2)$),
put: $\mathfrak{C} \models R\boldsymbol{c} \Leftrightarrow (\mathfrak{A} \models R\boldsymbol{a}$ and $(R,+) \in O_1)$.

(e) If $\boldsymbol{a} \notin A^n$ and $\boldsymbol{b} \in B^n$ (i.e. $\boldsymbol{b} \notin \mathrm{im}(p)$, and $\boldsymbol{c} \notin \mathrm{im}(p_1)$),
put: $\mathfrak{C} \models R\boldsymbol{c} \Leftrightarrow (\mathfrak{B} \models R\boldsymbol{a}$ and $(R,-) \in O_2)$.

(f) If $\boldsymbol{a} \notin A^n$ and $\boldsymbol{b} \notin B^n$, put whatever.

Note that if e.g. $\boldsymbol{a}\,\boldsymbol{c} \in p_1$ then $\boldsymbol{a} \in A^n$ and $\boldsymbol{c} = \boldsymbol{a}\,\boldsymbol{b}$ for some $\boldsymbol{b} \in (B \dot\cup \{*\})^n$, where $\boldsymbol{a}'\boldsymbol{b}' \in p$ if the primed tuples are the projections of $\boldsymbol{a}$ and $\boldsymbol{b}$ to those positions in which $\boldsymbol{b}$ has entries from B. One checks that

- p_1 respects O_1: consider $\boldsymbol{a}\,\boldsymbol{c} \in p_1$ and $(R,+) \in O_1$; the relevant cases are those in which $\mathfrak{A} \models R\boldsymbol{a}$, and are dealt with in (a),(b), or (d). If $(R,-) \in O_1$, then we look at the situation where $\mathfrak{A} \models \neg R\boldsymbol{a}$, and find that the relevant cases are (a),(c), or (d).
- p_2 respects O_2: similarly.

- $\mathrm{im}(p_1) \supseteq C^{-/O_1}$: the only critical case is that of $c = (*, b)$, because all other elements of C are trivially in $\mathrm{im}(p_1)$. Note that in this case $b \notin \mathrm{im}(p)$; but if $(*, b) \in U^{\mathfrak{C}}$, then by (e), $(U, -) \in O_2$ and $b \in U^{\mathfrak{B}}$, and $(U, -) \notin O_1$, as otherwise $(U, -) \in O$ and $b \in \mathrm{im}(p)$.

- $\mathrm{dm}(p_2) \supseteq C^{+/O_2}$: the only critical case is that of $c = (a, *)$, because all other elements of C are trivially in $\mathrm{dm}(p_2)$; but if $(a, *) \in U^{\mathfrak{C}}$, then by (d), $(U, +) \in O_1$ and $a \in U^{\mathfrak{A}}$, and $(U, +) \notin O_2$, as otherwise $(U, +) \in O$ and $a \in \mathrm{dm}(p)$.

This finishes the proof of the proposition. □

Proof of theorem 1. Let φ and ψ be $\mathbb{U}$-relativized formulae, $\varphi \models \psi$. Let $\boldsymbol{x}_0$ stand for the tuple of variables that are free in both φ and ψ, so that for appropriate tuples $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, we have pairwise disjoint tuples $\boldsymbol{x}_i$ such that $\varphi = \varphi(\boldsymbol{x}_0, \boldsymbol{x}_1)$ and $\psi = \psi(\boldsymbol{x}_0, \boldsymbol{x}_2)$ with all free variables displayed. Put $O_1 = occ(\varphi)$, $O_2 = occ(\psi)$, $O = O_1 \cap O_2$. Suppose there were no $\mathbb{U}$-relativized formula $\chi(\boldsymbol{x}_0)$ with $occ(\chi) \subseteq O$ such that $\varphi \models \chi$ and $\chi \models \psi$. Put

$$\Phi(\boldsymbol{x}_0) = \big\{\chi(\boldsymbol{x}_0) \;\big|\; \chi \ \mathbb{U}\text{-relativized}, occ(\chi) \subseteq O, \varphi \models \chi\big\}.$$

By assumption (and compactness), $\Phi \not\models \psi$. Let $(\mathfrak{B}, \boldsymbol{b}_0, \boldsymbol{b}_2) \models \Phi \cup \{\neg\psi\}$. Put

$$\Theta(\boldsymbol{x}_0) = \big\{\neg\chi(\boldsymbol{x}_0) \;\big|\; \chi \ \mathbb{U}\text{-relativized}, occ(\chi) \subseteq O, (\mathfrak{B}, \boldsymbol{b}_0) \models \neg\chi\big\}.$$

It follows that $\Theta \cup \{\varphi\}$ is consistent, too. Otherwise, and as Θ is closed under conjunctions, there would have to be a single $\neg\chi \in \Theta$ such that $\varphi \models \chi$. But then $\chi \in \Phi$, whence $\mathfrak{B} \models \chi$, a contradiction.

So we find some $(\mathfrak{A}, \boldsymbol{a}_0) \models \Theta \cup \{\varphi\}$. It follows that $(\mathfrak{A}, \boldsymbol{a}_0) \Longrightarrow^{\mathbb{U}}_{O} (\mathfrak{B}, \boldsymbol{b}_0)$. By Corollary 8 the situation can be upgraded to obtain $(\mathfrak{A}, \boldsymbol{a}_0) \longrightarrow^{\mathbb{U}}_{O} (\mathfrak{B}, \boldsymbol{b}_0)$, where $\mathfrak{A} \models \varphi[\boldsymbol{a}_0, \boldsymbol{a}_1]$ and $\mathfrak{B} \models \neg\psi[\boldsymbol{b}_0, \boldsymbol{b}_2]$.

Using countable $\mathfrak{A}$ and $\mathfrak{B}$ and applying Lemma 9, we may assume that $p \colon \mathfrak{A} \longrightarrow^{\mathbb{U}}_{O} \mathfrak{B}$ for a single $p \subseteq A \times B$ with $\boldsymbol{a}_0\boldsymbol{b}_0 \in p$. But now the proposition yields a structure $\mathfrak{C}$ and relations p_1 and p_2 such that $p = p_1 \circ p_2$, $\boldsymbol{a}_0\boldsymbol{a}_1 \in \mathrm{dm}(p_1)$, $\boldsymbol{b}_0\boldsymbol{b}_2 \in \mathrm{im}(p_2)$, and $p_1 \colon \mathfrak{A} \longrightarrow^{\mathbb{U}}_{O_1} \mathfrak{C}$, $p_2 \colon \mathfrak{C} \longrightarrow^{\mathbb{U}}_{O_2} \mathfrak{B}$. Let $\boldsymbol{c}_0$ be such that $\boldsymbol{a}_0\boldsymbol{c}_0 \in p_1$ and $\boldsymbol{c}_0\boldsymbol{b}_0 \in p_2$ (recall that $\boldsymbol{a}_0\boldsymbol{b}_0 \in p$ and $p = p_1 \circ p_2$). Let further $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ be such that $\boldsymbol{a}_0\boldsymbol{a}_1\, \boldsymbol{c}_0\boldsymbol{c}_1 \in p_1$ and $\boldsymbol{c}_0\boldsymbol{c}_2\, \boldsymbol{b}_0\boldsymbol{b}_2 \in p_2$ (these exist, since $\boldsymbol{a}_1 \in dom(p_1)$ and $\boldsymbol{b}_2 \in \mathrm{im}(p_2)$, respectively). It follows that $(\mathfrak{A}, \boldsymbol{a}_0, \boldsymbol{a}_1) \Longrightarrow^{\mathbb{U}}_{O_1} (\mathfrak{C}, \boldsymbol{c}_0, \boldsymbol{c}_1)$ and $(\mathfrak{C}, \boldsymbol{c}_0, \boldsymbol{c}_2) \Longrightarrow^{\mathbb{U}}_{O_2} (\mathfrak{B}, \boldsymbol{b}_0, \boldsymbol{b}_2)$. We find that by preservation $\mathfrak{C} \models \varphi[\boldsymbol{c}_0, \boldsymbol{c}_1]$, therefore, and as $\varphi \models \psi$, we have $\mathfrak{C} \models \psi[\boldsymbol{c}_0, \boldsymbol{c}_2]$, whence by preservation also $\mathfrak{B} \models \psi[\boldsymbol{b}_0, \boldsymbol{b}_2]$, a contradiction.

Applications We show how Theorem 1 actually combines several interpolation and preservation results in one. Namely, it directly implies not only Lyndon's interpolation theorem [7] (see Theorem 2 above), but also a number of related interpolation and preservation results. Among them the classical theorems concerning preservation under extensions and substructures [3], a variant of Feferman's many-sorted interpolation theorem [4, 5], as well as a very recent preservation theorem of van Benthem's [2] concerning Chu transforms.

Classical preservation theorems Consider a relational first-order formula $\varphi(\boldsymbol{x}) = \varphi(x_1, \dots, x_k)$ that is preserved under extensions: $\mathfrak{A} \subseteq \mathfrak{B}$ and $\mathfrak{A} \models \varphi[\boldsymbol{a}]$ together imply that $\mathfrak{B} \models \varphi[\boldsymbol{a}]$. Let U_1, U_2 be two new unary predicates, not in the vocabulary of φ, and put $\mathbb{U} = \{U_1, U_2\}$. Let φ^{U_i} be the result of relativizing φ to U_i. Then preservation under extensions for φ is expressed by the following validity:

$$\forall x(U_1 x \to U_2 x) \wedge \bigwedge U_1 \boldsymbol{x} \wedge \varphi^{U_1}(\boldsymbol{x}) \models \varphi^{U_2}(\boldsymbol{x}).$$

Note that all formulae occurring here are $\mathbb{U}$-relativized. Theorem 1 provides a $\mathbb{U}$-relativized interpolant $\chi(\boldsymbol{x})$, with no occurrence of U_1 and only positive occurrences of U_2. It follows that χ is purely existential. Considering the special case that $U_1 = U_2$ in the antecedent, we find that $\bigwedge U_2 \boldsymbol{x} \wedge \varphi^{U_2} \models \chi \models \varphi^{U_2}$. Further restricting attention to the case that $\forall x U_2 x$ in these implications, we replace in χ all atoms $U_2 y$ by $\top$ to obtain a formula χ', which is in the vocabulary of φ, purely existential, and equivalent with φ, since in this special case both $\bigwedge U_2 \boldsymbol{x} \wedge \varphi^{U_2}$ and φ^{U_2} are logically equivalent with φ.

We note that a similar connection between existential preservation and a *many-sorted* interpolation property (to be dealt with in the next section) is prominently discussed as an application for many-sorted interpolation by Feferman in [5].

As usual, a similar argument gives the classical preservation theorem concerning monotonicity and positivity (which of course directly follows from Lyndon's Interpolation Theorem). The point in the above is rather that the extension/existential and the monotone/positive preservation phenomena are attributed to the same source, as it were, in the present relativized picture.

A many-sorted interpolation theorem Consider many-sorted first-order logic with n sorts in a relational vocabulary. We here assume that all predicates and variables are typed w.r.t. the sorts and that the sorts are disjoint. [3] The standard translation into a one-sorted framework transforms a many-sorted relational structure $(A_1, \dots, A_n, R, \dots)$ with disjoint sorts A_i into a structure whose universe A is (an arbitrary superset of) $\bigcup A_i$ having new unary predicates $U_1, \dots, U_n$ to indicate the different sub-domains corresponding to the different sorts. Putting $\mathbb{U} = (U_1, \dots, U_n)$, the many-sorted first-order formulae naturally translate into $\mathbb{U}$-relativized formulae. Care has to be taken if free variables are around: while the many-sorted framework restricts each free variable to its particular sort implicitly, the corresponding semantic restriction has to be made explicitly in the translation. Care has also to be taken w.r.t. the translation of implications. A priori the validity of a many-sorted implication implies the validity of the $\mathbb{U}$-relativized one-sorted translated implication only in restriction to those one-sorted structures that arise as encodings of the original many-sorted structures. Those are the structures for which the U_s are disjoint, each relation R is restricted to

[3] However, we do *not* a priori require all sorts to be non-empty; corresponding stipulations would have to be made explicitly by means of existential statements. If typed-ness of predicates or disjoint-ness of sorts seem too restrictive, there are straightforward modifications which can deal with those other settings; in fact, some precautions that have to be taken here can be avoided in those more liberal settings.

a product of sub-domains U_s according to its specified type, and each free variable x_i is interpreted as an element of that $U_{s(i)}$ that corresponds to its sort. For well-formed many-sorted formulae, however, the restriction concerning the interpretation of relations R is irrelevant because R-atoms in what would be inadequate sorts do not affect the semantics of the translated formula (as long as the free variables are restricted to the appropriate sorts). The restriction of the free variables to their respective sorts will be made explicit below. So what remains is the apparent problem about disjoint versus overlapping sub-domains U_s, which is resolved in the light of the following reverse transformation. Let $\mathfrak{A} = (A, U_1, \ldots, U_n, R, \ldots)$ be an arbitrary structure of the indicated type, $\boldsymbol{a} = (a_1, \ldots, a_k) \in A^k$ such that $a_i \in U^{\mathfrak{A}}_{s(i)}$ for $i = 1, \ldots, k$. Let $\check{U}_s = U^{\mathfrak{A}}_s \times \{s\}$, and $\check{R} = \{\boldsymbol{a} = ((a_1, s_1), \ldots, (a_r, s_r)) \mid (a_1, \ldots, a_r) \in R^{\mathfrak{A}}\}$ for an r-ary relation R of type $(s_1, \ldots, s_r)$. If $\varphi(x_1, \ldots, x_k)$ is a well-formed many-sorted formula with x_i of sort $S_{s(i)}$, and $\varphi^{\mathbb{U}}$ the result of relativizing every quantifier in φ to the appropriate U_s, then

$$\mathfrak{A} \models \varphi^{\mathbb{U}}[\boldsymbol{a}] \quad \text{iff} \quad (\check{U}_1, \ldots, \check{U}_n, \check{R}, \ldots) \models \varphi[(a_1, s_1), \ldots, (a_k, s_k)].$$

Note that this is even true if φ has equality, because the typed-ness of φ implies that equality atoms can only link variables from the same sort. With these considerations in mind, we find that a valid many-sorted implication $\varphi \models \psi$, with combined free variables $x_1, \ldots, x_k$, x_i of sort $s(i)$, gives rise to the following valid one-sorted implication between $\mathbb{U}$-relativized formulae:

$$\bigwedge_{x_i \in \mathit{free}(\varphi)} U_{s(i)} x_i \wedge \varphi^{\mathbb{U}} \models \bigwedge_{x_i \in \mathit{free}(\psi)} U_{s(i)} x_i \longrightarrow \psi^{\mathbb{U}}, \tag{1}$$

where $\varphi^{\mathbb{U}}$ and $\psi^{\mathbb{U}}$ are the results of relativizing every quantifier in these formulae to the appropriate sort. We specialize to the case in which φ and ψ share the same free variables $x_1, \ldots, x_k$. This assumption makes the analysis more straightforward.

Let *free-sorts* be the set of sorts of the free variables of φ and ψ. Identifying a sort with its encoding predicate U, we may think of *free-sorts* as a subset of $\mathbb{U}$. A direct application of Theorem 1 to the valid implication (1) yields an interpolant $\chi(x_1, \ldots, x_k)$, which is $\mathbb{U}$-relativized and has

$$\mathit{occ}(\chi) \subseteq \big(\mathit{occ}(\varphi^{\mathbb{U}}) \cup \mathit{free\text{-}sorts} \times \{+\}\big) \cap \big(\mathit{occ}(\psi^{\mathbb{U}}) \cup \mathit{free\text{-}sorts} \times \{-\}\big). \tag{2}$$

Now χ is not at first of the form that would immediately translate back into a many-sorted formula. But note that, since $\bigwedge_i U_{s(i)} x_i \wedge \varphi^{\mathbb{U}} \models \chi(x_1, \ldots, x_k)$, all free variables of χ may be assumed to be appropriately typed. It remains to remove in χ all wrongly typed occurrences of predicates R, and all occurrences of predicates U_s apart from occurrences in explicit relativizations. W.l.o.g. each bound variable in χ is quantified exactly once, and therefore also has a unique sort attributed to it. Any R-atom in χ that may in this fashion acquire an inappropriate type is now replaced by $\perp$, the same goes for atoms $U_s y$, whenever y is a variable attributed to any sort different from U_s. Clearly, the resulting formula χ' is still good as an interpolant for the implication (1) in restriction to all structures that arise as one-sorted encodings of many-sorted structures.

Also observe that none of these manipulations introduces new occurrences. Thus χ' may be regarded as a well-formed many-sorted formula, for which $\varphi \models \chi'$ and $\chi' \models \psi$ are valid many-sorted implications.

Now for occurrences of predicates in χ (or χ'). W.r.t. predicates in the original vocabulary of φ and ψ, we clearly have $occ(\chi') \subseteq occ(\chi) \subseteq occ(\varphi) \cap occ(\psi)$, by (2) — as is to be expected for a many-sorted variant of Lyndon's Interpolation. Moreover, however, $\mathbb{U}$-occurrences can be used to analyze which sorts are existentially or universally quantified.

Let $\exists\text{-}sorts(\varphi)$ stand for the set of those sorts that are existentially quantified over in φ, $\forall\text{-}sorts(\varphi)$ the set of those that are universally quantified, and similarly for ψ and χ (assuming, of course, negation normal form). Then (2) implies that

$$\begin{array}{rcl} \exists\text{-}sorts(\chi') & \subseteq & (\exists\text{-}sorts(\varphi) \cup free\text{-}sorts) \cap \exists\text{-}sorts(\psi) \\ \forall\text{-}sorts(\chi') & \subseteq & \forall\text{-}sorts(\varphi) \cap (\forall\text{-}sorts(\psi) \cup free\text{-}sorts). \end{array}$$

We thus have the following many-sorted interpolation theorem, which is closely related to Feferman's many-sorted interpolation theorem [4].

Proposition 11 *Suppose φ and ψ are many-sorted relational formulae with the same free variables. If $\varphi \models \psi$ is a valid implication, then there is a many-sorted Lyndon interpolant χ with*

(i) $occ(\chi) \subseteq occ(\varphi) \cap occ(\psi)$.

(ii) $\exists\text{-}sorts(\chi) \subseteq (\exists\text{-}sorts(\varphi) \cup free\text{-}sorts) \cap \exists\text{-}sorts(\psi)$,
$\forall\text{-}sorts(\chi) \subseteq \forall\text{-}sorts(\varphi) \cap (\forall\text{-}sorts(\psi) \cup free\text{-}sorts)$.

(iii) $\varphi \models \chi$ *and* $\chi \models \psi$.

Note that in the case of sentences (i.e. without free variables) condition (ii) simply says that a sort is existentially/universally quantified in χ only if it is existentially/universally quantified in both φ and ψ.

In comparison with [4], our variant concerns just first-order, rather than fragments of infinitary logic which are a main issue in Feferman's theorem; on the other hand it goes slightly beyond Feferman's theorem in two other respects (apart from being obtained along very different lines):

(a) the Lyndon condition on the polarities of predicate occurrences ((i) above) is not considered by Feferman in [4].

(b) the condition on existentially and universally quantified sorts ((ii) above) is more symmetric and in general stronger than the corresponding condition in [4]. [4]

[4] It is worth noting that this is not only due to the fact that our sorts are not required to be non-empty; the use of explicit conditions to this effect in the constituent formulae still yields a more restrictive condition (ii) than the one discussed in [4, 5].

Van Benthem's preservation theorem This new preservation result [2] arises in the context of Chu spaces (over $\{0,1\}$), regarded as two-sorted relational structures in one binary E of type $(1,2)$. A Chu transform between two such two-sorted E-structures $\mathfrak{A} = (A_1, A_2, E^{\mathfrak{A}})$ and $\mathfrak{B} = (B_1, B_2, E^{\mathfrak{B}})$ is a pair of mappings $f\colon A_1 \to B_1$ and $g\colon B_2 \to A_2$ such that for all $a \in A_1$ and $\beta \in B_2$:

$$(a, g(\beta)) \in E^{\mathfrak{A}} \quad \Leftrightarrow \quad (f(a), \beta) \in E^{\mathfrak{B}}.$$

Note that the first sort transforms as under a homomorphism, while the second sort transforms as under an inverse homomorphism. Special Chu transforms are the following:

(a) extensions in the first sort: $A_1 \subseteq B_1$, $A_2 = B_2$, $\mathfrak{A} \subseteq \mathfrak{B}$, $f, g = \mathrm{id}$.

(b) substructures in the second sort: $A_1 = B_1$, $A_2 \supseteq B_2$, $\mathfrak{A} \supseteq \mathfrak{B}$, $f, g = \mathrm{id}$.

Let the variables of the first sort be denoted x, those of the second sort ν. A two-sorted first-order formula $\varphi(\boldsymbol{x}, \boldsymbol{\nu})$ is preserved under Chu transforms if, whenever f, g constitute a Chu transform from $\mathfrak{A}$ to $\mathfrak{B}$, and if $\boldsymbol{a} \in (A_1)^n$ and $\boldsymbol{\beta} \in (B_2)^m$, then

$$\mathfrak{A} \models \varphi[\boldsymbol{a}, g(\boldsymbol{\beta})] \quad \Rightarrow \quad \mathfrak{B} \models \varphi[f(\boldsymbol{a}), \boldsymbol{\beta}].$$

Suppose $\varphi(\boldsymbol{x}, \boldsymbol{\nu})$ is an equality-free [5] two-sorted first-order formula that is preserved under the special Chu transforms of types (a) and (b). According to van Benthem's preservation theorem [2], φ is equivalent to a formula φ^* that is purely existential in the first sort and purely universal in the second. We can here obtain φ^* as an interpolant in an application of our main theorem, in the context of the natural one-sorted encodings of Chu spaces. Let us use unary predicates U and V for the sub-domains corresponding to the first and second sort, respectively. Let $\varphi^{\mathbb{U}}$ be the result of relativizing all quantifiers in φ accordingly, to either U or V. In the light of the treatment of the many-sorted interpolation theorem outlined above, it will suffice to show that $\varphi^{\mathbb{U}}$ is equivalent with a $\mathbb{U}$-relativized formula in which U occurs only positively and V only negatively. This is achieved in two separate steps, the first one using Chu transforms of type (a) to eliminate negative occurrences of U, the second one using Chu transforms of type (b) to eliminate positive occurrences of V. For the first step let $\varphi_1 = \varphi^{\mathbb{U}}$, φ_2 the result of renaming U to U' in $\varphi^{\mathbb{U}}$. Then preservation of φ under Chu transforms of type (a) corresponds to the validity of

$$\varphi_2(\boldsymbol{x}, \boldsymbol{\nu}) \wedge \bigwedge U'\boldsymbol{x} \wedge \bigwedge V\boldsymbol{\nu} \wedge \forall x(U'x \to Ux) \models \varphi_1(\boldsymbol{x}, \boldsymbol{\nu}).$$

This is an implication between $\{U, V, U'\}$-relativized formulae, and Theorem 1 yields a formula $\chi(\boldsymbol{x}, \boldsymbol{\nu})$ with no occurrence of U' and only positive occurrences of U. Identifying U and U', we find that $\varphi^{\mathbb{U}}$ and χ are indeed equivalent if $\boldsymbol{x}$ and $\boldsymbol{\nu}$ respect the

[5] To recover the full content of van Benthem's theorem, one should not require the absense equality, but obtain it in a preliminary step based on preservation under onto-homomorphisms in the first sort (which goes to eliminate inequalities over the first sort) and inverse onto-homorphisms in the second sort (which similarly allows us to eliminate inequalities over the second sort); we leave that part out, as it has no direct relevance for our concerns.

sorts. For the second step assume that $\varphi^{\mathbb{U}}$ is positive in U, and let $\varphi_1 = \varphi^{\mathbb{U}}$, φ_2 the result of renaming V to V' in $\varphi^{\mathbb{U}}$. Preservation of φ under Chu transforms of type (b) means that

$$\varphi_2(\boldsymbol{x}, \boldsymbol{\nu}) \wedge \bigwedge U\boldsymbol{x} \wedge \forall x(Vx \to V'x) \models \bigwedge V\boldsymbol{\nu} \longrightarrow \varphi_1(\boldsymbol{x}, \boldsymbol{\nu}).$$

Theorem 1 now yields a formula $\chi(\boldsymbol{x}, \boldsymbol{\nu})$, still positive in U, with no occurrence of V' and only negative occurrences of V. Identifying V and V', we find that $\varphi^{\mathbb{U}}$ and χ are indeed equivalent if $\boldsymbol{x}$ and $\boldsymbol{\nu}$ respect the sorts. This new χ then can essentially serve as φ^*. Up to some necessary syntactic clean-up as we saw in the treatment of many-sorted interpolation, φ^* may be translated back into a two-sorted formula in the original vocabulary involving just E with only existential quantification over the first and only universal quantification over the second sort.

References

1. J. Barwise, *A preservation theorem for Interpretations*, in Proc. Cambridge Summer School in Mathematical Logic 1971, A. Mathias et al. (ed.), LNM, vol. 337, Springer 1973, pp. 618–621.
2. J. van Benthem, *Information transfer across Chu spaces*, talk in the Stanford Logic Seminar, May 98.
3. C.C. Chang and H.J. Keisler, Model Theory, 3rd ed., North-Holland 1990.
4. S. Feferman, *Lectures on proof theory*, in Proc. Summer School in Logic, Leeds 67, M. Löb (ed.), LNM, vol. 70, Springer 1968, pp. 1–107.
5. S. Feferman, *Applications of many-sorted interpolation theorems*, in Proc. Taarski Symposium, L. Henkin et al. (ed.), AMS Proc. of Symposia in Pure Mathematics, vol. XXV, 1974, pp. 205–223.
6. J. Flum, *First-order logic and its extensions*, in Proc. Int. Summer Inst. and Logic Colloquium, Kiel 1974, G. Müller et al. (ed.), LNM, vol. 499, Springer 1975, pp. 248–310.
7. R.C. Lyndon, *An interpolation theorem in the predicate calculus*, Pacific Journal of Mathematics, 9, 1959, pp. 129–142.