Harmonious Logic

Harmonious Logic: Craig's Interpolation Theorem and its Descendants

Solomon Feferman

For Bill Craig, with great appreciation for his fundamental contributions to our subject, and for his perennially open, welcoming attitude and fine personality that enhances every encounter.

Abstract: Though deceptively simple and plausible on the face of it, Craig's interpolation theorem (published 50 years ago) has proved to be a central logical property that has been used to reveal a deep harmony between the syntax and semantics of first order logic. Craig's theorem was generalized soon after by Lyndon, with application to the characterization of first order properties preserved under homomorphism. After retracing the early history, this article is mainly devoted to a survey of subsequent generalizations and applications, especially of many-sorted interpolation theorems. Attention is also paid to methodological considerations, since the Craig theorem and its generalizations were initially obtained by proof-theoretic arguments while most of the applications are model-theoretic in nature. The article concludes with the role of the interpolation property in the quest for "reasonable" logics extending first-order logic within the framework of abstract model theory.

Key words: Craig's theorem, interpolation theorems, preservation theorems, many-sorted languages, extensions of first-order logic, abstract model theory.

1. Craig's Interpolation Theorem. A common statement of Craig's theorem (initially referred to by him as a lemma) goes as follows:

Suppose $\vdash \varphi(\underline{R}, \underline{S}) \rightarrow \psi(\underline{S}, \underline{T})$. *Then there is a* $\theta(\underline{S})$ *such that* $\vdash \varphi(\underline{R}, \underline{S}) \rightarrow \theta(\underline{S})$ *and* $\vdash \theta(\underline{S}) \rightarrow \psi(\underline{S}, \underline{T})$.

Here $\models$ is validity in classical first order logic with equality (FOL), φ, ψ, θ are sentences, and $\underline{R}$, $\underline{S}$, and $\underline{T}$ are sequences of relation symbols for which the sequence $\underline{S}$ is non-empty. Equality is treated as a logical symbol and not as one of the relation symbols. Here and in the following, I consider only languages with no function symbols, since not all generalizations of the interpolation theorem hold when those are present; the reader will be able easily to see when they do hold by use of the standard translation into a purely relational language. Constant symbols, on the other hand, *are* allowed, and for simplicity these are treated as 0-ary relation symbols. A fuller and more formal statement of Craig's theorem goes as follows, where we use Rel(φ) for the set of relation symbols in φ.

THEOREM 1 (Craig 1957). *Suppose φ, ψ are sentences with* $\models \varphi \rightarrow \psi$.
(i) *If* Rel(φ) ∩ Rel(ψ) *is non-empty then there exists a sentence θ such that* $\models \varphi \rightarrow \theta$ *and* $\models \theta \rightarrow \psi$ *and* Rel(θ) ⊆ Rel(φ) ∩ Rel(ψ).
(ii) *If* Rel(φ) *and* Rel(ψ) *are disjoint then* $\models \neg\varphi$ or $\models \psi$.

In the following we shall ignore exceptional cases like (ii) in generalizations of the interpolation theorem, i.e. we tacitly assume a hypothesis like that in (i) as needed to verify the given conclusion; that kind of hypothesis is always met in the applications of the interpolation theorem and its generalizations.

I first heard Bill Craig explain this result and his proof of it in a talk he gave for the Summer Institute for Symbolic Logic held at Cornell University in the month of July, 1957. But it would be some ten years before I began to appreciate its wider significance through its extension to many-sorted logics as described in sec. 4 below, and beyond that to abstract model theory (sec. 5).

The intuitive idea for Craig's proof of the Interpolation Theorem rests on the completeness theorem for FOL, in the form of equivalence of validity with provability in a suitable system of axioms and rules of inference. By "suitable" here is meant one in which there is a notion of a direct proof for which if φ → ψ is provable then there is a direct proof of ψ from φ. One would expect that in such a proof, the relation symbols of φ that are not in ψ would disappear in its middle. Such systems were devised by

Herbrand (1930) and Gentzen (1934); Hilbert-style systems enlarged by the axioms and rules of the epsilon-calculus can also serve this purpose.

In Gentzen's calculus LK, one establishes formal combinations—called *sequents*—of the form $\varphi_1,\ldots,\varphi_n \mid\!- \psi_1,\ldots,\psi_m$ (abbreviated $\underline{\varphi} \mid\!- \underline{\psi}$) which are valid just in case $\varphi_1\wedge\ldots\wedge\varphi_n \rightarrow \psi_1\vee\ldots\vee\psi_m$ is valid. The axioms of LK are restricted to atomic formulas and its rules of inference govern each propositional operator and quantifier separately as rules for introduction of a formula with that connective or quantifier on the left or the right; in addition there are structural rules (permutation, weakening and contraction). Finally there is a *cut-rule*, which allows one to pass from $\underline{\varphi}' \mid\!- \underline{\psi}', \theta$ and $\theta, \underline{\varphi}'' \mid\!- \underline{\psi}''$ to $\underline{\varphi}', \underline{\varphi}'' \mid\!- \underline{\psi}', \underline{\psi}''$. Gentzen's cut-elimination theorem shows that every proof in LK can be transformed into another one which is *cut-free*. Cut-free proofs have a kind of directness property—in contrast to those which may involve cut—since in every application of one of its rules, each formula in one of the hypotheses of the rule is a subformula of some formula in its conclusion (with possible change of variables in the case of quantifier rules). In the case that all the formulas of $\underline{\varphi}$ and $\underline{\psi}$ are prenex in a derivable sequent $\underline{\varphi} \mid\!- \underline{\psi}$, a cut-free derivation can be reorganized to have all quantifier introduction rules below all propositional rules; this is Gentzen's generalization of Herbrand's fundamental theorem.

At the Cornell conference and then in Craig's paper (1957), "Linear reasoning. A new form of the Herbrand-Gentzen theorem", he introduced a calculus for passing directly from formulas of the form $(Q)(\varphi_1\wedge\ldots\wedge\varphi_n \rightarrow \psi_1\vee\ldots\vee\psi_m)$, where Q is a sequence of universally and existentially quantified variables and each of the φ_i's and ψ_j's is prenex, to new such formulas. Unlike Gentzen's system, each rule in Craig's system has exactly one hypothesis, so a derivation in it is indeed linear and is, moreover, direct as it stands. Any such derivation can be reorganized to have all quantificational inferences below all the propositional ones, just as in the Herbrand-Gentzen theorem. Then since interpolation is easily established for the propositional calculus, one can use that to ascend to an interpolant for a derivable sentence $\varphi \rightarrow \psi$, provided that both φ, ψ are prenex. To prove Theorem 1 in general, one simply uses the transformation of arbitrary formulas into equivalent prenex form.

Craig's system of linear reasoning is elegant and natural for the class of formulas considered, but it was not subsequently adopted by others as a tool for proof-theoretical work. The reason may be that Gentzen's approach has greater flexibility. To prove the interpolation theorem using LK, one can do it first, as Craig does, for prenex formulas via Gentzen's generalization of Herbrand's theorem. But it is not necessary to restrict to prenex formulas in LK. Given *any* sentences φ, ψ and a cut-free derivation D of $\varphi \vdash \psi$, one can inductively build up an interpolant step-by-step within D. Moreover, the same argument can be used to obtain an interpolation theorem for certain infinitary languages via a straightforward extension of LK to those languages (see sec. 5 below), while there is no obvious adaptation of Craig's system to those languages. (Aside from interpolation, the proof theory of infinitary languages has been of major importance in extensions of Hilbert's consistency program (see, e.g. Schütte (1977) and Pohlers (1989)). Further methodological considerations will be taken up below.

2. Craig's applications of the Interpolation Theorem. First among the applications that Craig made of the Interpolation Theorem in his paper (1957a), "Three uses of the Herbrand-Gentzen theorem in relating model theory and proof theory", was to Beth's Definability Theorem. That result has an interesting history, beginning with a claim made by Alessandro Padoa in his article (1901) translated in part as "Logical introduction to any deductive theory" in van Heijenoort (1967), that we would rephrase in modern terms as follows:

> *To prove that a basic symbol* S *is independent of the other basic symbols in a system of axioms* ∑*, it is necessary and sufficient that there are two interpretations of* ∑ *which agree on all the basic symbols other than* S *and which differ at* S.[1]

As remarked in the introductory note to Padoa's article in van Heijenoort (1967), p. 118, the method to prove independence of the basic symbols of an axiomatic system "is merely stated, and Padoa seems to consider it intuitively evident. He does not offer a proof of its correctness, and such a proof could hardly have been undertaken then, because, first, Padoa's system (which is Peano's) is ill-defined, and, second, many results requisite for such a proof were still unkown at the time."

Sufficiency in Padoa's statement is obvious no matter which logic is considered, so the main issue is that of necessity. If the axiom system $\sum$ is finite, and its conjunction is a sentence $\varphi(\underline{R}, S)$ in a given language L, then necessity is equivalently expressed by saying that if the interpretation of S is the same for any two interpretations of φ which agree on $\underline{R}$, then S is explicitly definable from $\underline{R}$ in L on the basis of φ. Alfred Tarski and Adolf Lindenbaum (1926) pointed out that necessity holds when L is taken to be the (impredicative) theory of types, on the basis of a theorem by Tarski later published in his paper (1935) (cf. Hodges (2008) for discussion). Very simply, that is essentially done by taking $S(\underline{x}) \leftrightarrow \forall X(\varphi(\underline{R}, X) \rightarrow X(\underline{x}))$. Tarski had long been interested in Padoa's method through its applications in Euclidean geometry, for which Tarski had given an axiomatization in FOL (cf. Tarski and Givant 1999). But it was not until Evert Beth proved his Definability Theorem that it was done for FOL.

THEOREM 2 (Beth 1953). *Suppose* $\varphi(\underline{R}, S) \wedge \varphi(\underline{R}, S') \vdash S(\underline{x}) \leftrightarrow S'(\underline{x})$.
Then there is a $\theta(\underline{R}, \underline{x})$ such that $\varphi(\underline{R}, S) \vdash S(\underline{x}) \leftrightarrow \theta(\underline{R}, \underline{x})$.

Beth's first proof of this was by use of the completeness of Gentzen's LK and the cut-elimination theorem. He analyzed a cut-free proof of the sequent $\varphi(\underline{R}, S), S(\underline{x}) \vdash \varphi(R, S') \rightarrow S'(\underline{x})$. A draft of Beth's proof was sent to Tarski in May 1953, and Tarski discussed it with me at that time (I was at Berkeley then, working with him as a student). From Tarski's point of view, since the statement of Beth's definability theorem is model-theoretic, there ought to be a model-theoretic proof, and there was correspondence (via me) with Beth about how that might be accomplished (cf. van Ulsen 2000, pp. 136 ff). Beth made some efforts in that direction, but the published argument remains essentially proof-theoretic, making only cosmetic changes in the direction of a model-theoretic argument.

By compactness for FOL, Beth's result immediately extends to any set $\sum$ of sentences in place of φ.

Craig's Interpolation Theorem implies Beth's Definability Theorem by the following simple argument. By hypothesis, we have

$$\vdash \varphi(\underline{R}, S) \wedge S(\underline{x}) \rightarrow (\varphi(\underline{R}, S') \rightarrow S'(\underline{x})).$$

To treat this as an implication between sentences, replace $\underline{x}$ by corresponding new constant symbols $\underline{c}$, giving

$$\vdash \varphi(\underline{R}, S) \wedge S(\underline{c}) \rightarrow (\varphi(\underline{R}, S') \rightarrow S'(\underline{c})).$$

An interpolant for this implication is a sentence $\theta(\underline{R}, \underline{c})$ such that

$$\vdash \varphi(\underline{R}, S) \wedge S(\underline{c}) \rightarrow \theta(\underline{R}, \underline{c}) \text{ and } \vdash \theta(\underline{R}, \underline{c}) \rightarrow (\varphi(\underline{R}, S') \rightarrow S'(\underline{c})).$$

Replacing S' by S and reorganizing these implications it follows that

$$\vdash \varphi(\underline{R}, S) \rightarrow (S(\underline{c}) \leftrightarrow \theta(\underline{R}, \underline{c}))$$

and hence

$$\varphi(\underline{R}, S) \vdash S(\underline{x}) \leftrightarrow \theta(\underline{R}, \underline{x}).$$

We can avoid the detour here via the auxiliary constants $\underline{c}$ by means of a slight generalization of the Interpolation Theorem to formulas, in the form that an interpolant can be constructed for a given valid implication as one for which each of its free variables is free in both its antecedent and its consequent; this will be stated explicitly in sec. 4 below.

The first model-theoretic proof of Beth's definability theorem was given by Abraham Robinson in his paper (1956), "A result on consistency and its application to the theory of definition." Beth's theorem is shown there to follow from what is now called Robinson's Consistency Theorem:

THEOREM 3 (Robinson 1956). Suppose L_1 and L_2 are two languages and $L = L_1 \cap L_2$. If $\sum$ is a complete theory in L and $\sum_1$, $\sum_2$ are consistent extensions of T in L_1 and L_2 respectively, then $\sum_1 \cup \sum_2$ is consistent in $L_1 \cup L_2$.

Robinson's proof of this was by combination of the method of diagrams with the first use of a back-and-forth chain construction that was to become a standard tool later on in model-theory.

Craig referred to Robinson's work in (1957a) fn. 1, as follows: "For another interesting proof [of Beth's theorem], more along modeltheoretic lines, see A. Robinson

[1956]. I am grateful to him for oral and written suggestions regarding several points…" It seems from this that neither Craig nor Robinson was then aware that their theorems are equivalent by relatively easy model-theoretic proofs. It is not clear who first realized that; the equivalence was apparently folklore in Berkeley in the years directly following Craig's work.[2] Curiously, Craig himself speaks of Robinson's theorem as "a forerunner of more recent interpolation theorems" in his review of Robinson (1956) for *The Journal of Symbolic Logic* 25 (1960), p.174, but not in terms of their logical relationship. As far as I can determine, the first published proof of the equivalence was given in Robinson (1963), pp. 114-117.[3]

The second application that Craig gave of his Interpolation Theorem in the paper "Three uses of the Herbrand-Gentzen theorem in relating model theory and proof theory" made use of the following notions: a class K of models is called a *projective class* (PC) if it is the set of M = (A, $\underline{S}$) satisfying $\exists\underline{R}\ \varphi(\underline{R}, \underline{S})$ for some φ of FOL. K is an *elementary class* (EC) if it it consists of the models of a sentence $\theta(\underline{S})$ of FOL.

THEOREM 4. (Craig 1957a) Any two disjoint projective classes can be separated by an elementary class.

Proof. Under the assumption that $\vdash \neg[\exists\underline{R}\ \varphi(\underline{R}, \underline{S}) \wedge \exists\underline{T}\ \psi(\underline{T}, \underline{S})]$, where φ, ψ are both FOL formulas, we have $\vdash \varphi(\underline{R}, \underline{S}) \rightarrow \neg\psi(\underline{T}, \underline{S})$. Any interpolant $\theta(\underline{S})$ defines a separating EC.

As a corollary we have:

THEOREM 5. (Δ-Interpolation Theorem). *If* a class K of models *and its complement are both in* PC *then* K *is in* EC.

Craig's third application in his (1957a) is a little more complicated to explain; it concerns axiomatizability with prescribed groupings of axioms according to which basic symbols they contain. Let $\prod_1$ and $\prod_2$ be two finite sets of basic symbols ("parameters" in Craig's terminology). A ($\prod_1$, $\prod_2$) axiom system $\sum$ is one for which $\sum = \sum_1 \cup \sum_2$ and all the symbols of $\sum_i$ are contained in $\prod_i$. Craig gives a necessary and sufficient condition that one can add a given sentence φ to a given axiom system so that the result is

($\prod_1$, $\prod_2$) axiomatizable; the same is done more generally for any finite sequence of finite sets of parameters. The reader is referred to Craig (1957a) pp. 282ff for the statement and proof.

3. Lyndon's Interpolation Theorem and properties preserved under homomorphism. In Roger Lyndon's article (1959), "An interpolation theorem in the predicate calculus" he gave a generalization of Craig's Interpolation Theorem in terms of the polarities of the relation symbols involved, and then applied that in the article (1959a) to establish a characterization of the properties expressed in FOL that are preserved under homomorphism. As we shall see, this was paradigmatic for further uses of suitably generalized interpolation theorems. Like Craig's results, Lyndon's work was first announced at the 1957 Cornell conference.

In preparation for both Lyndon's and more general interpolation theorems, given any map F from formulas to sets, *a formula* θ *is called an interpolant for a valid implication* $\varphi \rightarrow \psi$ *with respect to* F if

$$\vdash \varphi \rightarrow \theta \text{ and } \vdash \theta \rightarrow \psi \text{ and } F(\theta) \subseteq F(\varphi) \cap F(\psi) .$$

For Lyndon's theorem we make use of the two functions Rel^+ and Rel^- such that for each formula φ, $Rel^+(\varphi)$ ($Rel^-(\varphi)$) is the set of relation symbols with *at least one positive* (*at least one negative*) *occurrence* in φ. These functions can be defined inductively, or via the negation normal form of φ, obtained by driving the negation symbol down to atomic formulas using the de Morgan laws.

THEOREM 6. (Lyndon 1959) *If* $\vdash \varphi \rightarrow \psi$ *then it has an interpolant w.r.t.* Rel^+ *and* Rel^-.[4]

Of the argument for this, Lyndon writes (1959, p.130), "our first proof of [this] Interpolation Theorem used the Gentzen calculus; it did not differ essentially from Craig's proof, at that time unpublished, of his lemma." He then describes a new proof which "serves as a substitute" for the methods due to Herbrand and Gentzen, motivated by conversations with Henkin and Tarski in which Tarski "emphasized the desirability of establishing the Interpolation Theorem by methods independent of the theory of proof."

But the substitute appears to be a model-theoretic version of a sharpened form of the Herbrand-Gentzen theorem. The first proof of a more widely applicable model-theoretic character of Lyndon's Interpolation Theorem (and still stronger results) was due to H. Jerome Keisler (1960). This extended the back-and-forth chain constructions combined with the method of diagrams inaugurated by Robinson (1956). In the same period, ultrapower and ultraproduct constructions began to be used as an alternative to the method of diagrams. The first such use was for the proof of the theorem, announced independently by Keisler (1959) and Simon Kochen (1959), that any two elementarily equivalent structures have isomorphic ultralimits (or "strong limit ultrapowers"); that implies Robinson's theorem. For basic model-theoretic methods used in establishing the interpolation theorems, cf. Chang and Keisler (1973) and Hodges (1993); other such methods will be noted below.

Lyndon (1959a) applied his theorem as follows. Given $M = (A, \underline{R})$, and $M' = (A', \underline{R}')$ of the same similarity type, a map $h: A \to A'$ is said to be a *homomorphism of* M *onto* M′ if h is onto and for each i, $h(R_i) \subseteq R'_i$. (When the relations are functions, this is the usual notion of homomorphism.) A sentence φ is said to be *preserved under homomorphisms* if whenever $M \models \varphi$ and M′ is a homomorphic image of M then $M' \models \varphi$. A sentence is called *positive* if it has no negative occurrences of relation symbols, i.e. if $Rel^{-}(\varphi)$ is empty. It is easily seen that every positive sentence is preserved under homomorphisms. Lyndon used his interpolation theorem to characterize the sentences preserved under homomomorphism as just those equivalent to positive sentences.
To see how this is established, consider the special case that we are dealing with structures with just one binary relation symbol R. In place of homomorphisms, we may deal with congruence relations. For this, let E be a new binary relation symbol and then let Cong(R, E) express that E is an equivalence relation, together with
$\forall x_1 \forall x_2 \forall y_1 \forall y_2\, [\, E(x_1, y_1) \wedge E(x_2, y_2) \wedge R(x_1, x_2) \to R(y_1, y_2)\,]$.
For a sentence φ expressed in FOL with one binary relation symbol R, write φ(R, E) for the result of replacing atomic formulas $x = y$ in φ by E(x, y). Given new relation symbols R′ and E′, write $R \subseteq R'$ for $\forall x \forall y[R(x,y) \to R'(x,y)]$, and similarly $E \subseteq E'$.

<u>Lemma</u>. φ *is preserved under homomorphisms iff*

$\mathrm{Cong}(R, E) \wedge \mathrm{Cong}(R', E') \wedge R \subseteq R' \wedge E \subseteq E' \wedge \varphi(R, E) \rightarrow \varphi(R', E')$ *is valid*.

THEOREM 7. (Lyndon 1959a) φ *is preserved under homomorphisms iff it is equivalent to a positive sentence.*

Proof Apply Lyndon's Interpolation Theorem to

$\vdash \mathrm{Cong}(R, E) \wedge R \subseteq R' \wedge E \subseteq E' \wedge \varphi(R, E) \rightarrow [\, \mathrm{Cong}(R', E') \rightarrow \varphi(R', E')\,]$.

Note that R' and E' have no negative occurrences in the antecedent and R and E have no occurrences in the consequent. An interpolant θ meeting the conditions of Lyndon's theorem can thus be chosen to have the form $\theta(R', E')$, positive in both R' and E'. By then specializing to the case that $R' = R$ and $E' = E$, it follows that

$$\mathrm{Cong}(R, E) \vdash \varphi(R,E) \leftrightarrow \theta(R, E).$$

Finally, take E to be the equality relation.

Note on negative and positive results in finite model theory. The basic results on interpolation theorems and their consequences for preservation characterizations described in this and the preceding section have all been revisited in finite model theory, which studies validity and satisfiability conditions for various languages—including that of FOL—when restricted entirely to finite structures. As shown by Tait, Gurevich and others, and exposited in Ebbinghaus and Flum (1999), sec. 3.5, most of the statements have turned out to be negative when restricted to finite models, including the Beth Definability Theorem, the Lyndon Interpolation Theorem, the Los-Tarski theorem (see next section) and the Lyndon theorem for preservation under onto homomorphisms. A recent surprising exception is the characterization due to Benjamin Rossman (2005 and n.d.) of the first-order sentences preserved in the finite under *into* homomorphisms.

4. Many-sorted interpolation theorems and their applications. Many-sorted languages and associated structures are ubiquitous in mathematical practice; familiar examples are geometries (points, lines, planes,…) and vector spaces (vectors, scalars). Standard examples from logic are type theory and typed lambda calculi. What distinguishes many-sorted languages is that each variable is of a definite sort: for a non-empty set J of sorts and for each $j \in J$, we have infinitely many variables $x_j, y_j, z_j, \ldots$ of sort j, and for j, j′ distinct, the corresponding sets of variables are disjoint. A structure for such a language is thus of the form $M = (\langle A_j \rangle_{j \in J}, \ldots)$ where each of the basic domains A_j is non-empty and the variables of sort j range over A_j. As to the constraints on the relations, including the relation of equality, there are two basic options, *strict* and *liberal.* Under the strict option, each n-ary relation R is of a specified arity, given by an n-tuple of sorts $\langle j_1, \ldots, j_n \rangle$, and atomic formulas are of the form $R(\underline{x})$ where x is an n-tuple of sorted variables, where the sort of x_i is j_i for each $i = 1, \ldots, n$. In particular, there is an equality relation $=_j$ for each $j \in J$, whose arity is $\langle j, j \rangle$. A typical example of the strict option is provided by the simple theory of types, where J is the set of natural numbers and, in addition to the relations $=_j$ we have the membership relations $\in_j$ of arity $\langle j, j+1 \rangle$ for each $j \in J$. Under the liberal option, each relation, including equality, may relate objects of arbitrary sort; the only constraint is the specification of the number of its arguments. For example, in Galois theory one considers fields as vector spaces over subfields as the scalars. Other examples of the liberal option are given by certain versions of ramified type theory or the theory of constructible sets and forcing extensions.

The model theory of many-sorted structures is standardly reduced to that of single-sorted structures by the *unification of domains*. Associated with the many-sorted language is a single-sorted language with new unary predicates U_j for each $j \in J$, and associated with each many-sorted structure $M = (\langle A_j \rangle_{j \in J}, \ldots)$ is a single sorted structure $M^{(U)} = (A, \langle A_j \rangle_{j \in J}, \ldots)$, where A is the union of the A_j for $j \in J$ and U_j is interpreted as A_j. Finally, with each sentence φ of the many-sorted language is associated a sentence $\varphi^{(U)}$ of the associated single-sorted language, obtained by successively relativizing quantifiers, i.e. replacing each quantified occurrence $\forall x_j(\ldots)$ in φ, resp. $\exists x_j(\ldots)$, by $\forall x(U_j(x) \rightarrow \ldots)$, resp. $\exists x_j(U_j(x) \wedge \ldots)$ (replacing distinct variables by distinct variables). Then $M \models \varphi$ iff $M^{(U)} \models \varphi^{(U)}$. Note that the distinction between the strict option and the liberal option is partially wiped out in the process, since now we have just a single equality relation.

In Feferman (1968) I established a generalization of Lyndon's interpolation theorems to many-sorted languages making use of the following additional functions:

$\mathrm{Sort}(\varphi) = \{j \in J \mid \text{a variable of sort j occurs in } \varphi\}$

$\mathrm{Free}(\varphi)$ = the set of free variables of φ

$\mathrm{Un}(\varphi) = \{j \in J \mid \text{a } \forall x_j \text{ occurs in nnf}(\varphi)\}$

$\mathrm{Ex}(\varphi) = \{j \in J \mid \text{an } \exists x_j \text{ occurs in nnf}(\varphi)\}$,

where nnf(φ) = the negation normal form of φ. Note that an occurrence $\exists x_j$ in nnf(φ) corresponds in $\varphi^{(U)}$, under the process of relativization of quantifiers, to a positive occurrence of U_j in a context $\exists x(U_j(x) \wedge \ldots)$ while an occurrence $\forall x_j$ in nnf(φ) corresponds to a negative occurrence of U_j in a context $\forall x(U_j(x) \rightarrow \ldots)$.

THEOREM 8. (Feferman 1968). *If* $\vdash \varphi \rightarrow \psi$ *then it has an interpolant* θ *w.r.t.* Rel^+, Rel^-, Sort, *and* Free, *for which*

(†) $\quad\quad \mathrm{Un}(\theta) \subseteq \mathrm{Un}(\varphi)$ and $\mathrm{Ex}(\theta) \subseteq \mathrm{Ex}(\psi)$.

By the *basic form of many-sorted interpolation* is meant the same statement without (†).

It might be thought that Theorem 8 can be inferred from Lyndon's Interpolation Theorem, by applying it to $\vdash \varphi^{(U)} \rightarrow \psi^{(U)}$, i.e. to the result of unifying domains and relativizing quantifiers. However, there is no assurance from the statement of Lyndon's theorem that an interpolant for this implication can be chosen to be in the form $\theta^{(U)}$, since the quantifiers need not be relativized in an interpolant. This issue was revisited by Martin Otto (2000) in a way that will be discussed below. My proof of Theorem 8 was carried out by a proof-theoretical argument in a many-sorted version of LK.

One application of Theorem 8 is to characterize up to equivalence the sentences preserved under extensions. In the single-sorted case, we have the well-known characterization independently due to Los and Tarski (1955) by a model-theoretic argument, namely that a sentence is preserved under extensions iff it is equivalent to an existential sentence. This can be generalized to the many-sorted case as follows. For $M = (\langle A_j \rangle_{j \in J}, \ldots)$ and $M' = (\langle A'_j \rangle_{j \in J}, \ldots)$ and $I \subseteq J$ we write $M \subseteq_I M'$ and M' is called an I-*stationary extension* of M if M is a substructure of M' with $A_i = A'_i$ for each $i \in I$. A sentence φ is said to be *preserved under* I-*stationary extensions rel. to* $\sum$ if whenever M, M' are models of $\sum$ and $M \models \varphi$ and $M \subseteq_I M'$ then $M' \models \varphi$. A sentence θ is said to be *existential outside of* I if $\mathrm{Un}(\theta) \subseteq I$.

THEOREM 9. φ *is preserved under* I-*stationary extensions rel. to* $\sum$ *iff for some* θ *that is existential outside of* I, $\sum \vdash \varphi \leftrightarrow \theta$.

Proof. For each sort of variable $x_j, \ldots$ with $j \in J - I$, adjoin a new sort $x_j', \ldots$, and associate with each relation symbol R of L (other than =) a new symbol R′. Let φ' be the copy of φ, leaving the variables of sort $i \in I$ unchanged. Let Ext_I = the conjunction of $\forall x_j \exists x_j' (x_j = x_j')$ for each $j \in J - I$ together with $\forall \underline{x} [R(\underline{x}) \leftrightarrow R'(\underline{x})]$ for each R. Then φ is preserved under I-stationary extensions iff $\sum \cup \sum' \vdash Ext_I \wedge \varphi \rightarrow \varphi'$, where $\sum'$ is the result of replacing each sentence σ of $\sum$ by σ'. By compactness there is a finite conjunction ψ of sentences of $\sum$ such that $(\psi \wedge Ext_I \wedge \varphi) \rightarrow (\psi' \rightarrow \varphi')$ is valid. By Theorem 8, an interpolant for this implication can be chosen to be of the form θ', with $Un(\theta')$ contained in $Un(Ext_I)$; but there are no variables of the new sort in J–I, so $Un(\theta') \subseteq I$. Collapsing the new sorts to the old ones gives the desired result.

Looked at model-theoretically, what the proof does is combine two many-sorted structures M and M′ into a new one considered with the liberal interpretation, even when the original structures are taken with a strict interpretation. The Los-Tarski theorem is the special case of this preservation theorem for J a singleton and I empty.

The technique of proof for Theorem 9 was modified in Feferman (1968a) to characterize the sentences preserved under end-extensions, where for single-sorted structures M and M′ with distinguished binary relations $<$ and $<'$, $M' = (A', <', \ldots)$ is called an *end-extension* of $M = (A, <, \ldots)$ it is an extension such that for each $a \in A$ and $b \in A'$, $b <' a \Rightarrow b \in A$. A sentence φ is said to be *preserved under end-extensions relative to* $\sum$ if whenever M and M′ are models of $\sum$ and $M \models \varphi$ and M′ is an end-extension of M, then $M' \models \varphi$. For the characterization, one adjoins *bounded quantifiers* $(\forall y < x)(\ldots)$ and $(\exists y < x)(\ldots)$ as basic logical operators; then a formula is called *essentially existential* if its nnf in the expanded language is existential when one ignores the bounded quantifiers.

THEOREM 10. (Feferman 1968a) φ *is preserved under end extensions rel. to* $\sum$ *iff it is equivalent in* $\sum$ *to an essentially existential sentence.*

When < is taken to be the membership relation and $\sum$ is an axiomatic theory of sets, Theorem 10 yields a characterization of the (provably) *absolute properties rel. to* $\sum$, as just those which are equivalent to both an essentially existential and an essentially universal formula. The notion of end-extension can be generalized to many-sorted structures with a given set of stationary sorts, and then this result can be generalized in a way analogous to Theorem 9.

Jacques Stern obtained the following variant of Theorem 8 and showed how that could be used to prove Theorem 9 while hewing to the strict interpretation.

THEOREM 11. (Stern 1975) *If* $\vdash \varphi \to \psi$ *then it has an interpolant* θ *w.r.t.* Rel^{+}, Rel^{-} *and* Sort, *for which*

(††) $\mathrm{Un}(\theta) \subseteq \mathrm{Un}(\psi)$ *and* $\mathrm{Ex}(\theta) \subseteq \mathrm{Ex}(\varphi)$.

Stern's proof of this used the model-theoretic method of forcing. Note that in his version, θ is not required to satisfy that it is an interpolant for the Free (variables) function. My proof of Theorem 8 using LK can also be modified so as to drop that condition. In doing so, one sees the reason for the switch from (†) to (††), by keeping track of when one must introduce quantifiers in the build-up of the interpolant. By imposing the free variables condition on a subset of the sorts, Stern obtained a common generalization of the two many-sorted interpolation theorems 8 and 11. The following strengthened form of Herbrand's theorem is a nice immediate consequence of the latter.

Corollary. If φ and ψ are formulas with φ universal and ψ existential and if $\varphi \to \psi$ is valid then it has a quantifier-free interpolant θ w.r.t. Rel^{+} and Rel^{-}.

Note that it is essential for this that θ is not required to be an interpolant w.r.t. Free.
In Feferman (1974) I applied Theorem 8 and this corollary to establish a simple model-theoretic necessary and sufficient condition for eliminability of quantifiers for axiom sets

$\sum$ that are model-consistent relative to some subset of their universal consequences, in a way that generalizes to certain infinitary language.

Returning to the question of dealing with many-sorted interpolation via unification of domains and relativization of quantifiers, Martin Otto (2000) showed how to deal with U-*relativized formulas*, where $\underline{U} = \langle U_i \rangle_{i \in I}$ is a sequence of unary predicates, and every quantifier is relativized to some U_i. Otto obtains there a generalization of Lyndon's Interpolation Theorem in which an interpolant for U-relativized formulas is also required to be U-relativized; in addition, his result implies the many-sorted interpolation theorems of Feferman and Stern above. Otto's proof makes use of ω-saturated structures, but a proof via LK should also be possible.

5. Beyond first-order logic: interpolation properties and abstract model theory.

Many semantically specified logics stronger than FOL have been studied in the last 50 years; the following are some prominent examples:

1. ω-logic
2. 2^{nd} order logic
3. Logic with the cardinality quantifier Q_α (i.e., $\exists \geq \aleph_\alpha$)
4. $L_{\kappa,\lambda}$, logic with conjunctions of length $< \kappa$ and quantifier strings of length $< \lambda$ (κ, λ infinite cardinals)
5. L_A for A admissible (conjunctions over sets in A, ordinary 1^{st} order quantification).

FOL can be identified with $L_{\omega,\omega}$ or with L_{HF}, where HF is the collection of hereditarily finite sets. For HC = the hereditarily countable sets and $A \subseteq HC$, $L_A \subseteq L_{\kappa,\omega}$ with $\kappa = \omega_1$.

The subject of *abstract model theory* arose through the study of these and other such languages from a general perspective; a comprehensive source for material on it is provided by the volume Barwise and Feferman (1985). Abstract model theory deals with properties of *model-theoretic logics* L, specified by an abstract syntax—i.e. a set of "sentences" satisfying suitable closure conditions—and "satisfaction" relation $M \models \varphi$ for φ a sentence of L. With each such L is associated its collection of Elementary Classes, EC_L, and from that its collection of Projective Classes, PC_L in the way explained above. $L \subseteq L^*$ is defined to hold if $EC_L \subseteq EC_{L^*}$. Using these notions we can formulate various properties of model-theoretic logics and examine specific logics such as 1-5 in terms of them. Before getting into interpolation and related properties, we begin with the following:

1° *Countable compactness property.*
2° *Löwenheim-Skolem (L-S) property.*
3° *R.e. completeness property.*

By 1° is meant that if ∑ is any countable set of L sentences every finite subset of which has a model then ∑ has a model. By 2° is meant that if an L-sentence φ has an infinite model then it has a countable model. By 3° is meant that the set of valid sentences is recursively enumerable (usually simply referred to as the completeness property).

Example: Other than $L_{\omega,\omega}$ only its extension by the uncountability quantifier (Q_1) among the specific examples 1-5 has countable compactness and r.e. completeness (Keisler 1970); obviously L-S fails. None of the others has either property.

The following are the famous theorems of Per Lindström characterizing FOL in terms of these properties:[5]

THEOREM 13 (Lindström 1969)
(i) $L_{\omega,\omega}$ is the largest logic having the countable compactness and L-S properties.
(ii) $L_{\omega,\omega}$ is the largest logic having the r.e. completeness and L-S properties.

We now turn to abstract formulations of interpolation and related properties for any model-theoretic language L:

4° *Interpolation property.*

5° *Δ-Interpolation property.*

6° *Beth property.*

7° *Weak Beth property.*

8° *Weak projective Beth property.*

By 4° is meant that any two disjoint classes K in PC_L can be separated by an EC_L, while 5° means that if K and its complement are both PC_L then K is in EC_L. The Beth property 6° is that for $K \in EC_L$, if each M has *at most one* expansion [M, S] ∈ K then S is uniformly definable over M, while 7° gives the same conclusion when each M has *exactly one* expansion [M, S] ∈ K. Finally, by 8° is meant that the weak Beth property holds for $K \in PC_L$.

The following is a ready consequence of these explanations.

Lemma.
(i) Interpolation ⇒ Δ-interpolation ⇒ Beth ⇒ weak Beth.
(ii) Δ-interpolation ⇔ weak projective Beth.

It was shown by Lopez-Escobar (1965) that L_{HC} has the interpolation property and more generally by Barwise (1969) that the same holds for every L_A with A a countable admissible subset of HC. Barwise also generalized the r.e. completeness property to L_A when 'r.e.' is replaced by 'A-r.e.', i.e. $\sum_1$-definable over A in the language of set theory. Even more, one has:

THEOREM 14 (Feferman 1968, 1968a). *All of the results stated in sections 1-4 above generalize to* L_A *for* A *a countable admissible subset of* HC *and* ∑ *an* A-*r.e. set of sentences.*

My proof of this made use of the completeness of an extension of Gentzen's calculus LK to such L_A, together with a cut-elimination theorem thereof. Model-theoretic proofs of overlapping results have been given by Keisler (1971), using the so-called technique of

Consistency Properties. As remarked by Bienvenido Nebres (1972), p. 464, these are dual to Validity Properties, i.e. closure conditions on the sequents in cut-free LK derivations. To the best of my knowledge, no other model-theoretic methods extend to these L_A for the results comprehended by Theorem 14.

None of the other logics in 1-5 has even the weak Beth property, as shown by Craig (1965), Mostowski (1968), and Friedman (1973). One way to show that a specific logic L does not satisfy the weak Beth property is to show that the satisfaction (and hence truth) predicate for L over any model M is implicitly definable in L uniformly in M, while it is not explicitly definable over suitable M. However, it is often the case that the satisfaction predicate is uniformly definable in L up to any formula of L, via its subformulas. This led me in my article (1974) to introduce the notion, L *is adequate to truth in* L*, when the syntax of L* is represented in a transitive set A, roughly speaking if the satisfaction predicate $Sat_{L^*}(m, a, \underline{x})$, which holds iff $m \models a[\underline{x}]$ in L*, is uniformly implicitly definable in L up to any $a \in A$ that represents an L* formula. Then L is said to be *truth maximal* if whenever it is adequate to truth in L*, $L^* \subseteq L$. It is *truth complete* if it is both truth maximal and adequate to truth in itself.

THEOREM 15 (Feferman 1974)
(i) L has the Δ-interpolation property iff it is truth maximal.
(ii) L_A is truth complete for each admissible $A \subseteq HC$.

Jouko Väänänen (1985) has used the notion of truth adequacy to characterize logics whose satisfaction relation is absolute relative to certain systems of axiomatic set theory.

Now—drawing to a close—I want to say something about work that has been done on a problem that I raised in the 1970s in various informal venues, and more formally in my paper (1975), sec. 4:

QUESTION. *Does there exists a logic* L *properly extending* $L_{\omega,\omega}$ *satisfying the countable compactness property and the interpolation property, or at least one of its related consequences such as the* Δ*-interpolation property or one of the Beth definability properties*?

Re Δ-interpolation: as has been pointed out above, that fails for the countably compact logic $L_{\omega,\omega}(Q_1)$ (with a counter-example due to Keisler). However, one can form the Δ-*closure* Δ(L) of any logic L so as to satisfy Δ-interpolation, and similarly the *Beth-closure* of L. But if L is compact, the resulting closures need not be compact. Note that by the Lindström theorem 13(i) characterizing $L_{\omega,\omega}$, any logic L of the kind that might exist in answer to this question must violate the L-S property, such as, for example, extensions of $L_{\omega,\omega}(Q_1)$.

The question above led to considerable research, especially by Johann Makowsky and Saharon Shelah and with various collaborators, of which the following is a sample; I will revisit the motivations behind the question itself below. See also Makowsky and Shelah (1979 and 1981), Makowsky, Shelah and Stavi (1976) and Makowsky (1985), sec. 4, the latter for a discussion and survey (to that date) of this and related questions.

• Makowsky and Shelah (1983), "Positive results in abstract model theory" and Mundici (1982), "Compactness, interpolation and Friedman's third problem" independently prove that for extensions of $L_{\omega,\omega}$ with finitely many generalized quantifiers, the Robinson consistency property is equivalent to the interpolation property plus compactness.
• Shelah (1985), "Remarks in abstract model theory", proves there is a (fully) compact proper extension of $L_{\omega,\omega}$ with the Beth definability property, using the Δ-closure of the quantifier "the cofinality of < is at most the cardinality of the continuum". This logic does not satisfy interpolation; also, it does not extend $L_{\omega,\omega}(Q_1)$.
• Mekler and Shelah (1985), "Stationary logic and its friends I", proves that it is consistent for $L_{\omega,\omega}(Q_1)$ to have the weak Beth property.
• Hodges and Shelah (1991), "There are reasonably nice logics", proves that $L_{\omega,\omega}(Q_\alpha)$ is a countably compact logic with the interpolation property, for $\aleph_\alpha$ an uncountable strongly compact cardinal with at least one larger strongly compact cardinal.
• Shelah and Väänänen (n.d.), "New infinitary languages with interpolation", proves that there exists a logic beween $L_{\infty,\omega}$ and $L_{\infty,\infty}$ with the interpolation property, but it does not satisfy the countable compactness property.

To conclude, I have emphasized the harmonious relation between syntax and semantics due to the interpolation property for first-order logic and its immediate generalizations to the countable admissible logics. But the completeness of FOL—even more its strong completeness—is of course the fundamental property of that character. To formulate this in abstract terms, we consider only logics L represented in HF. (More abstract formulations with generalized notions of finiteness were proposed in Feferman (1975), sec. 4.) By an *axiomatic basis* B *for* L is meant a pair B = ⟨A, R⟩ consisting of a set A of "axioms" and a set R of finitary "rules of inference" such that for every set ∑ of L-sentences and any L-sentence φ, ∑ |= φ in L iff there is a derivation of φ from ∑ together with A using the rules of inference R. L is said to satisfy the *strong completeness property* if it has an axiomatic basis, and it is said to satisfy the *strong r.e. completeness property* if it has an axiomatic basis for which A and R are recursively enumerable. That of course is a property of FOL by Gödel's proof of his completeness theorem. The following relates these notions to 1° and 3° above :

<u>Lemma</u>.

(i) Strong completeness implies countable compactness.

(ii) Strong r.e. completeness property implies r.e. completeness.

This suggests the following new Lindström type characterization theorem of $L_{\omega,\omega}$:

CONJECTURE. There does not exist a proper extension of $L_{\omega,\omega}$ with the strong r.e. completeness and interpolation properties.

If this conjecture is correct, it may well be that one has still stronger characterizations of $L_{\omega,\omega}$, obtained by dropping 'r.e.' and/or replacing the interpolation property by one of its consequences 5°-8° above.

Acknowledgments. I wish to thank Wilfrid Hodges, Jerry Keisler, Janós Makowsky, Gabriel Sabbagh, Saharon Shelah, and Jouko Väänänen for their helpful comments on the material of this paper.

References

Barwise, J. (1969). Infinitary logic and admissible sets. *J. Symbolic Logic* 34, 226-252.

Barwise, J. and Feferman, S., Eds. (1985). *Model-Theoretic Logics*. (New York: Springer-Verlag).

Beth, E. W. (1953). On Padoa's method in the theory of definition. *Indagationes Mathematicae* 15, 330-339.

Chang, C.-C. and Keisler, H. J. (1973). *Model Theory*. (Amsterdam: North-Holland).

Craig, W. (1957). Linear reasoning. A new form of the Herbrand-Gentzen theorem. *J. Symbolic Logic* 22, 250-268.

_______ (1957a). Three uses of the Herbrand-Gentzen theorem in relating model theory and proof theory. *J. Symbolic Logic* 22, 269-285.

_______ (1965). Satisfaction for *n*-th order languages defined in *n*-th order languages. *J. Symbolic Logic* 30, 13-25.

Ebbinghaus, H.-D. and Flum, J. (1999). *Finite Model Theory,* 2nd edn.. (Berlin: Springer-Verlag).

Feferman, S. (1968). Lectures on proof theory. In M. H. Löb (Ed.), *Proc. Summer School in Logic, Leeds, 1967,* 1-107.

__________ (1968a). Persistent and invariant formulas for outer extensions. *Compositio Mathematicae* 20, 29-52.

__________ (1974). Applications of many-sorted interpolation theorems. In L. Henkin, et al. (Eds.), *Proceedings of the Tarski Symposium*. (Providence: American Mathematical Society).

__________ (1975). Two notes on abstract model theory, II: Languages for which the set of valid sentences is s.i.i.d. *Fundamenta Mathematicae* 89, 111-130.

Flum, J. (1985). Characterizing logics. In Barwise and Feferman (1985), 77-120.

Friedman, H. (1973). Beth's theorem in cardinality logics. *Israel J. of Mathematics* 14, 205-212.

Gentzen, G. (1934). Untersuchungen über das logische Schliessen. *Mathematische Zeitschrift* 39, 176-210, 405-431; English translation in Gentzen (1969) 86-131.

_________ (1969). *The Collected Papers of Gerhard Gentzen* (M. E. Szabo, ed.) (Amsterdam: North-Holland).

Henkin, L. (1963). An extension of the Craig-Lyndon interpolation theorem. *J. Symbolic Logic* 28, 201-216.

Herbrand, J. (1930). *Recherches sur la théorie de la démonstration*. Thesis at the University of Paris; English translation in van Heijenoort (1967), 525-581.

Hodges, W. (1993). *Model Theory*. (Cambridge: Cambridge University Press).

_________ (1998). The laws of distribution for syllogisms. *Notre Dame J. of Formal Logic* 39, 221–230.

_________ (2008). Tarski's theory of definition. In D. Patterson (Ed.), *New Essays on Tarski and Philosophy.* (New York: Oxford University Press). (In press)

Hodges, W. and Shelah, S. (1991). There are reasonably nice logics. *J. Symbolic Logic* 56, 300-322.

Keisler, H. J. (1959). A mathematical characterization of elementary equivalence. *Notices Amer. Math. Soc.* 6 (abstract).

___________ (1960). Theory of models with generalized atomic formulas. *J. Symbolic Logic* 25, 1-26.

___________ (1970). Logic with the quantifier "there exist uncountably many". *Annals of Mathematical Logic* 1, 1-93.

___________ (1971). *Model Theory for Infinitary Logic: Logic with countable conjunctions and finite quantifiers*. (Amsterdam: North-Holland).

Kochen, S. (1959). An algebraic characterization of arithmetical classes. *Notices Amer. Math. Soc.* 6 (abstract).

Lindström, P. (1969). On extensions of elementary logic. *Theoria* 35, 1-11.

López-Escobar, E.G.K. (1965). An interpolation theorem for denumerably long formulas. *Fundamenta Mathematicae* 57, 253-272.

Lyndon, R. (1959). An interpolation theorem in the predicate calculus. *Pacific J. of Mathematics* 9, 129-142.

_________ (1959a). Properties preserved under homomorphism. *Pacific J. of Mathematics* 9, 143-154.

Makowsky, J.A. (1985). Compactness, embeddings and definability. In Barwise and Feferman (1985), 645-715.

Makowsky, J.A. and Shelah, S. (1979). The theorems of Beth and Craig in abstract model theory, I: The abstract setting. *Transactions Amer. Math. Soc.* 256, 215-239.

_________________________ (1981). The theorems of Beth and Craig in abstract model theory, II: Compact logics. *Archiv für Mathematische Logik und Grundlagen der Mathematik* 21, 13-35.

_________________________ (1983). Positive results in abstract model theory: A theory of compact logics. *Annals of Pure and Applied Logic* 25, 263-300.

Makowsky, J.A., Shelah, S. and Stavi, J. (1976). Δ-logics and generalized quantifiers. *Annals of Mathematical Logic* 10, 155-192.

Mekler, A. H. and Shelah, S. (1985). Stationary logic and its friends, I. *Notre Dame J. of Formal Logic* 26.

Mostowski, A. (1968). Craig's interpolation theorem in some extended sysems of logic. In B. van Rootselaar and J.F. Staal (Eds.), *Logic, Methodology and Philosophy of Science III.* (Amsterdam: North-Holland), 87-103.

Mundici, D. (1982). Compactness, interpolation and Friedman's third problem. *Annals of Mathematical Logic* 22, 197-211.

Nebres, B. (1972). Infinitary formulas preserved under unions of models. *J. Symbolic Logic* 37, 449-465.

Otto, M. (2000). An interpolation theorem. *Bull. Symbolic Logic* 6, 447-462.

Padoa, A. (1901). Essai d'une théorie algébrique des nombres entiers, précédé d'une introduction logique à une théorie déductive quelconque. *Bibliothèque du Congrès international de philosophie, Paris, 1900,* vol. 3. (Paris: Armand Colin) 309-365; English translation in part in van Heijenoort (1967), 118-123.

Pohlers, W. (1989). *Proof Theory: An introduction.* (Berlin: Springer-Verlag).
Robinson, A. (1956). A result of consistency and its application to the theory of definition. *Indagationes Mathematicae* 18, 47-58.
___________ (1963). *Introduction to Model Theory and to the Metamathematics of Algebra.* (Amsterdam: North-Holland).

Rossman, B. (2005). Existential positive formulas and preservation under homomorphism. *Proc. 20th IEEE Symposium on Logic in Computer Science,* 467-476.
__________ (n.d.). Homomorphisms and first-order logic. (In press)
Schütte, K. (1977). *Proof Theory.* (Berlin: Springer-Verlag).
Shelah, S. (1985). Remarks in abstract model theory. *Annals of Pure and Applied Logic* 29, 255-288.
Shelah, S. and Väänänen, J. (n.d.). New infinitary languages with interpolation. *Annals of Pure and Applied Logic.* (In press).
Stern, J. (1975). A new look at the interpolation problem. *J. Symbolic Logic* 40, 1-13.
Tarski, A. (1935). Einige mehodologische Untersuchungen über die Definierbarkeit der Begriffe. *Erkenntnis* 5, 80-100; English translation in Tarski (1983), 296-319.
_______ (1955). Contributions to the theory of models, III. *Indagationes Mathematicae* 17, 56-64.
_______ (1983). *Logic, Semantics, Metamathematics. Papers from 1923 to 1938*, 2nd edn. (J. Corcoran, ed.). (Indianapolis: Hackett).
Tarski, A. and Givant, S. (1999). Tarski's system of geometry. *Bull. Symbolic Logic* 5, 175-214.
Tarski, A. and Lindenbaum, A. (1926). Sur l'indépendance des notions primitives dans les systèmes mathématiques. *Annales de la Société polonaise de mathématique* 5, 111-113.
Van Heijenoort, J. (1967). *From Frege to Gödel. A source book in mathematical logic 1879-1931,* Harvard University Press, Cambridge.
Väänänen, J. (1985). Set-theoretic definability of logics. In Barwise and Feferman (1985), 597-643.
Van Ulsen, P. (2000). *E. W. Beth als logicus.* (Amsterdam: Institute for Logic, Language and Computation).

---

[1] Padoa (1901), in van Heijenoort (1967), p. 122.
[2] H. Jerome Keisler said he heard it from one or more of the Berkeley logicians in the period 1957-1959 (personal communication). Henkin (1963) fn. 3 says that Robinson's theorem is "now known" to be equivalent to Craig's theorem, without giving credit to anyone. And Robinson (1963), pp. 137-138 writes that "[t]he fact that 5.1.6 [the Consistency Theorem] entails 5.1.8 [the Interpolation Theorem] has been pointed out to the author by several logicians from Warsaw and Berkeley."
[3] A short proof that Craig's theorem implies Robinson's can be found in Chang and Keisler (1973), pp. 88-89. The reverse direction is left there as a starred exercise.
[4] Interestingly, in a personal communication, Wilfrid Hodges brought to my attention that "[t]he Lyndon theorem has a claim to be the 20th century metatheorem with the longest pedigree, because it is a generalisation of the late medieval laws of distribution for

syllogisms. Obviously the theorem itself and any of its proofs would have been way beyond the understanding of any of the medievals, but it does seem to contain the correct formalisation of a number of medieval intuitions." For elaboration, see Hodges (1998).

[5] See also Flum (1985) for an exposition of this and related characterization theorems.