Harmonious Logic: Craig's Interpolation Theorem and its Descendants

Solomon Feferman Stanford University http://math.stanford.edu/~feferman

. . . . . . . . .

Interpolations Conference in Honor of William Craig 13 May 2007

. . . . . . . . .

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.

1

- Early history
- Subsequent generalizations and applications, especially of many-sorted interpolation theorems
- A rare interaction between proof theory and model theory
- Interpolation and the quest for "reasonable" stronger logics.

# Craig's Interpolation Theorem ("Lemma")

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

$$\vdash \phi(\underline{R}, \underline{S}) \rightarrow \theta(\underline{S}) \text{ and } \models \theta(\underline{S}) \rightarrow \psi(\underline{S}, \underline{T}).$$

Here  $\mid$  is validity in the first order predicate calculus with equality (FOL) and  $\varphi$ ,  $\psi$ ,  $\theta$  are sentences.

W. Craig (1957a), "Linear reasoning. A new form of the Herbrand-Gentzen theorem"

\_\_\_\_\_(1957b), "Three uses of the Herbrand-Gentzen theorem in relating model theory and proof theory"

Let  $Rel(\varphi)$  = the set of relation symbols in  $\varphi$ . General statement:

Suppose  $\varphi$ ,  $\psi$  are sentences with  $\vdash \varphi \rightarrow \psi$ . Then (i)  $\operatorname{Rel}(\varphi) \cap \operatorname{Rel}(\psi) \neq \emptyset \Rightarrow \exists a \text{ sentence } \theta \text{ s.t.}$   $\vdash \varphi \rightarrow \theta \text{ and } \vdash \theta \rightarrow \psi \text{ and}$   $\operatorname{Rel}(\theta) \subseteq \operatorname{Rel}(\varphi) \cap \operatorname{Rel}(\psi).$ (ii)  $\operatorname{Rel}(\varphi) \cap \operatorname{Rel}(\psi) = \emptyset \Rightarrow \vdash \neg \varphi \text{ or } \vdash \psi.$ 

<u>N.B.</u> In the following, assumptions like (i) are implicit and we ignore boundary cases like (ii).

# **Proof theory and model theory**

 Proof theory concerns the structure and transformation of proofs in formal axiomatic systems. (Hilbert, Herbrand, Gentzen, ...)

2. Model theory concerns the relation of satisfaction between formulas from a formal language and structures. M  $\models \varphi$  if M is a model of  $\varphi$ .

 $\Sigma \models \phi$  if every model of  $\Sigma$  is a model of  $\phi$ .

(Skolem, Tarski, A. Robinson, ...)

3. Gödel's completeness theorem relates provability in FOL to validity:

 $\varphi$  is provable from  $\sum$  iff  $\sum \vdash \varphi$ . So  $\sum$  has a model iff  $\sum$  is consistent, iff every finite subset of  $\sum$  has a model (the Compactness Theorem).

4. Gentzen showed proofs in FOL can be transformed into *direct proofs* ("cut-free" with the "subformula property").

5. The Herbrand-Gentzen mid-sequent theorem for prenex formulas. Craig's version.

# <u>Craig's first application of the Interpolation</u> <u>Theorem: Beth's Definability Theorem</u>

A. Padoa (1900), "Logical introduction to any deductive theory" (English translation in *From Frege* to Gödel.) Padoa's claim: To prove that a basic symbol S is independent of the other basic symbols in a system of axioms  $\Sigma$ , it is n.a.s. that there are two interpretations of  $\Sigma$  which agree on all the basic

symbols other than S and which differ at S.

First justification for FOL:

E. W. Beth (1953), "On Padoa's method in the theory of definition".

<u>Beth's Theorem</u> ("Implicit definability implies explicit definability"). Suppose  $\varphi(\underline{R}, S) \land \varphi(\underline{R}, S') \longmapsto S(\underline{x}) \Leftrightarrow S'(\underline{x})$ . Then there is a  $\theta(\underline{R}, \underline{x})$  such that  $\varphi(\underline{R}, S) \longmapsto S(\underline{x}) \Leftrightarrow \theta(\underline{R}, \underline{x})$ .

Beth's first proof; Tarski's reaction. Beth's published proof.

Craig's simple proof: Apply interpolation to  $\vdash \phi(\underline{R}, S) \land S(\underline{x}) \rightarrow (\phi(\underline{R}, S') \rightarrow S'(\underline{x})).$ 

Robinson's earlier proof of Beth's theorem: A. Robinson (1956), "A result on consistency and its application to the theory of definition"

Craig's theorem implies Robinson's theorem (easy).

### **Craig's second application: Projective classes**

A class K of models is a *projective class* (PC) if it is the set of M = (A, <u>S</u>) satisfying  $\exists \underline{R} \varphi(\underline{R}, \underline{S})$  for some  $\varphi$  of FOL. K is an *elementary class* (EC) if it it consists of the models of a sentence  $\theta(\underline{S})$  of FOL. In these terms Craig's theorem is reexpressed as:

<u>Interpolation Theorem</u> Any two disjoint projective classes can be separated by an elementary class. <u>Proof.</u> Use  $\vdash - \varphi(\underline{R}, \underline{S}) \rightarrow -\psi(\underline{T}, \underline{S})$  from  $\vdash - [\exists \underline{R} \varphi(\underline{R}, \underline{S}) \land \exists \underline{T} \psi(\underline{T}, \underline{S})].$ 

Corollary ( $\Delta$ -Interpolation Theorem). *If* K and its complement are both in PC then K is in EC.

<u>Note</u>:  $\Delta$ -Interpolation is analogous to the Souslin-Kleene theorem in (Effective) Descriptive Set Theory and Post's theorem in Recursion Theory.

### Lyndon's theorems

R. Lyndon (1959a), "An interpolation theorem in the predicate calculus"

\_\_\_\_\_ (1959b), "Properties preserved under homomorphism"

Let F be a map from formulas to sets.  $\theta$  is called an interpolant for  $\mid -\phi \rightarrow \psi$  w.r.t. F if  $\mid -\phi \rightarrow \theta$  and  $\mid -\theta \rightarrow \psi$  and  $F(\theta) \subseteq F(\phi) \cap F(\psi)$ .

Let Rel<sup>+</sup>( $\phi$ ) (Rel<sup>-</sup>( $\phi$ )) be the set of relation symbols of  $\phi$  with at least one positive (negative) occurrence in  $\phi^{\sim}$ , the *negation normal form* (n.n.f.) of  $\phi$ .

Lyndon's interpolation theorem If  $\mid -\phi \rightarrow \psi$  then it has an interpolant w.r.t. Rel<sup>+</sup> and Rel<sup>-</sup>.

Given, say, M = (A, R) and M' = (A', R') with R, R' binary, a map h:  $A \rightarrow A'$  is said to be a homomorphism of M onto M' if h is onto and for any  $x, y \in A, R(x, y) \Rightarrow R'(h(x), h(y))$ . (When R, R' are functions, this is the usual notion of homomorphism.) A sentence  $\varphi$  is said to be *preserved under homomorphisms* if whenever M  $\models \phi$  and M' is a homomorphic image of M then M'  $\models \varphi$ . Replace x = y in  $\varphi$  by E(x, y) and write  $\varphi(R, E)$  for  $\varphi$ . Let Cong(R, E) express that E is an equivalence relation, together with  $\forall x_1 \forall x_2 \forall y_1 \forall y_2 [ E(x_1, y_1) \land E(x_2, y_2) \land R(x_1, x_2) \rightarrow$ 

 $R(y_1, y_2)$ ].

<u>Lemma</u>  $\varphi$  *is preserved under homomorphisms iff*  ⊢ Cong(R, E) ∧ Cong(R', E') ∧ ∧R ⊆ R' ∧ E ⊆ E' ∧  $\varphi$ (R, E) →  $\varphi$ (R', E'). A sentence  $\varphi$  is said to be *positive* if Rel<sup>-</sup>( $\varphi$ ) is empty.

Lyndon's characterization theorem  $\varphi$  is preserved under homomorphisms iff it is equivalent to a positive sentence.

<u>Proof</u> Apply Lyndon's interpolation theorem to

 $Cong(R, E) \land R \subseteq R' \land E \subseteq E' \land \varphi(R, E) \rightarrow$ 

 $[\operatorname{Cong}(\mathsf{R}',\mathsf{E}') \rightarrow \varphi(\mathsf{R}',\mathsf{E}')].$ 

R' and E' have no negative occurrences in the hypothesis.

Lyndon's first proof of his interpolation theorem; Tarski's reaction.

# **Many-sorted interpolation theorems and their** <u>uses</u>

S. Feferman (1968a), "Lectures on proof theory" \_\_\_\_\_ (1974), "Applications of many-sorted interpolation theorems" (in Proc. of the 1971 Tarski Symposium)

Many-sorted structures  $M = (\langle A_j \rangle_{j \in J},...)$ . Language L with variables  $x_j, y_j, z_j,...$  for each  $j \in J$ . <u>Example</u>: Two-sorted, with variables x, y, z, ..., and x', y', z',... *Liberal equality* (x = x') vs. *strict equality* (x = y, x' = y') atomic formulas. We allow liberal equality.

Sort( $\varphi$ ) = { $j \in J \mid a \text{ variable of sort } j \text{ occurs in } \varphi$ } Free( $\varphi$ ) = the set of free variables of  $\varphi$ Un( $\varphi$ ) = { $j \in J \mid \text{there is } a \forall x_j \text{ in } \varphi^{\sim}$ } Ex( $\varphi$ ) = { $j \in J \mid \text{there is an } \exists x_j \text{ in } \varphi^{\sim}$ } <u>Many-sorted interpolation theorem</u>. *If*  $\mid -\phi \rightarrow \psi$ *then it has an interpolant*  $\theta$  *w.r.t.* Rel<sup>+</sup>, Rel<sup>-</sup>, Sort, *and* Free, *for which* 

(\*) 
$$Un(\theta) \subseteq Un(\phi) \text{ and } Ex(\theta) \subseteq Ex(\psi).$$

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

For  $M = (\langle A_j \rangle_{j \in J}, ...)$  and  $M' = (\langle A'_j \rangle_{j \in J}, ...)$  and  $I \subseteq J$ ,  $M \subseteq_I M'$  if M is a substructure of M' with  $A_i = A'_i$  for each  $i \in I$ .

 $\varphi$  is *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'  $\models \varphi$ .

<u>Theorem</u>  $\varphi$  *is preserved under* I*-stationary extensions rel. to*  $\sum$  *iff for some*  $\theta$  *that is existential outside of* I,  $\sum \vdash \varphi \Leftrightarrow \theta$ . <u>Proof</u>. For each sort of variable  $x_j,...$  with  $j \in J - I$ , adjoin a new sort  $x'_j,...$ , and associate with each relation symbol R of L (other than =) a new symbol R'. Let  $\varphi'$  be the copy of  $\varphi$ , 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  $\varphi$  is preserved under I-stationary extensions iff  $\sum \cup \sum' \models Ext_I \land \varphi \rightarrow \varphi'$ . Finally, apply compactness and many-sorted interpolation.

<u>Note</u>: The Los-Tarski theorem (1955) is the special case of this for J a singleton and I empty.

To avoid use of liberal equality between sorts, the following was proved by

J. Stern (1975), "A new look at the interpolation problem":

<u>Many-sorted interpolation theorem (Stern version)</u>. If  $\mid - \phi \rightarrow \psi$  then it has an interpolant  $\theta$  w.r.t. Rel<sup>+</sup>, Rel<sup>-</sup> and Sort, for which

(\*\*)  $\operatorname{Un}(\theta) \subseteq \operatorname{Un}(\psi) \text{ and } \operatorname{Ex}(\theta) \subseteq \operatorname{Ex}(\varphi).$ 

<u>N.B.</u> The interpolant may have free variables not in both  $\phi$  and  $\psi$ .

<u>Corollary</u> (Herbrand) If  $\varphi$  is universal and  $\psi$  is existential and  $\mid -\varphi \rightarrow \psi$  then it has a quantifier-free interpolant.

The Herbrand theorem is combined with a use of basic many-sorted interpolation in Feferman (1974) to establish a simple model-theoretic n.a.s.c. for eliminability of quantifiers for  $\Sigma$  that are model-consistent relative to some subset of their universal consequences. This holds, e.g., for real closed and algebraically closed fields.

#### Preservation under end-extensions

For (possibly many-sorted languages) with a binary relation symbol < (on one of the sorts), we can introduce *bounded quantifiers*  $(\forall y < x)(...)$  and  $(\exists y < x)(...)$ , and then *essentially existential* and *essentially universal formulas*.

 $M' = (A', <', ...) \text{ is an$ *end-extension* $of}$ M = (A, <, ...) if it is an extension such that for each

 $a \in A \text{ and } b \in A', b <' a \Rightarrow b \in A.$ 

S. Feferman (1968b) "Persistent and invariant formulas for outer extensions" uses a modification of the methods of the (1968a) article to prove:

<u>Theorem</u>  $\varphi$  is preserved under end extensions rel. to  $\sum$  iff it is equivalent in  $\sum$  to an essentially existential sentence. Similarly with I-stationary sorts. When < is taken to be the membership relation and  $\sum$  is an axiomatic theory of sets, this yields a characterization of the (provably) *absolute properties rel. to*  $\sum$ .

# **Beyond First Order Logic**

Many logics stronger than FOL have been studied in the last 50 years.

Examples:

- 1. ω-logic
- 2. 2<sup>nd</sup> order logic
- 3. Logic with cardinality quantifiers  $Q_{\alpha} (= \exists \geq \omega_{\alpha})$

4.  $L_{\kappa,\lambda}$ , logic with conjunctions of length <  $\kappa$  and quantifier strings of length <  $\lambda$  ( $\kappa$ ,  $\lambda$  inf. cards.)

5.  $L_A$  for A admissible (conjunctions over sets in A, ordinary 1<sup>st</sup> 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  $\kappa = \omega_1$ .

#### Abstract model theory

S. Feferman and J. Barwise (eds.) (1985), *Model-Theoretic Logics*.

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  $M \models \phi$  for  $\phi$  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,  $PC_L$ .  $L \subseteq L^*$  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.

1° Countable compactness
2° Löwenheim-Skolem property
3° Löwenheim-Skolem-Tarski property
4° R.e. completeness.

By 4° is meant that the set of valid sentences is recursively enumerable.

<u>Example</u>: Other than  $L_{\omega,\omega}$  only the extension by the uncountablility quantifier (Q<sub>1</sub>) 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.

Lindström's theorems (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.

(iii)  $L_{\omega,\omega}$  is the largest logic having the L-S-T

property.

Interpolation related properties:

5° *Interpolation* (any two  $PC_L$  K's can be separated by an  $EC_L$ ).

6°  $\Delta$ -Interpolation (if K and its complement are both PC<sub>L</sub> then K is in EC<sub>L</sub>).

7° *Beth* (for  $K \in EC_L$ , if each M has *at most one* 

expansion  $[M, \underline{S}] \in K$  then  $\underline{S}$  is uniformly definable over M).

8° Weak Beth (...and each M has exactly one

expansion  $[M, \underline{S}] \in K \dots$ .

## 9° Weak projective Beth (for $K \in PC_L, ...$ ).

<u>Lemma</u>. Interpolation  $\Rightarrow \Delta$ -interpolation  $\Rightarrow$  Beth  $\Rightarrow$  weak Beth;  $\Delta$ -interpolation  $\Leftrightarrow$  weak projective Beth.

<u>Example</u>: Only  $L_A$  for  $A \subseteq HC$ , admissible, among the logics 1-5, has the interpolation property (Lopez-Escobar, Barwise).

All of the results above for many-sorted interpolation theorems and their applications to FOL carry over to these  $L_A$ . (Feferman 1968a, 1968b)

<u>Remark:</u> Most model-theoretic methods used in FOL to prove preservation theorems do not carry over to the  $L_A$  for  $A \subseteq HC$ , admissible.

*Consistency properties* do—but they are dual to use of cut-free sequents.

<u>Counter-examples to interpolation or even weak Beth</u> for other logics are due variously to Craig, Mostowski, Keisler, Friedman, etc.

W. Craig (1965), "Satisfaction for nth order languages defined in nth order languages."

E.g., for 2<sup>nd</sup> order logic, the truth predicate is implicitly but not explicitly definable.

<u>Truth adequacy and truth maximality</u>. These are notions introduced in my 1974 article. Roughly speaking, *L* is adequate to truth in *L*\*, when the syntax of L\* is represented in a transitive set A, if the truth predicate Sat(m, a) is uniformly implicitly definable up to any  $a \in A$ . *L* is truth maximal if whenever it is adequate to truth in L\*, L\*  $\subseteq$  L. It is truth complete if it is truth maximal and adequate to truth in itself. <u>Theorem</u>  $\Delta$ -interpolation is equivalent to truthmaximality.

# The quest for "reasonable" logics

It has been suggested that for a logic to be reasonable, it ought to satisfy countable compactness and  $\Delta$ -interpolation, or at least the Beth property.

•  $\Delta$ -interpolation fails for  $L_{\omega,\omega}(Q_1)$  (Keisler).

<u>Question</u>: are there any reasonable proper extensions of  $L_{\omega,\omega}$ ?

<u>Note</u>: One can form the  $\Delta$ -*closure*  $\Delta(L)$  of any logic L to satisfy  $\Delta$ -interpolation, but then the problem is to see if  $\Delta(L)$  has other reasonable properties.

#### Some results in the quest for reasonable logics

• S. Shelah (1985), "Remarks in abstract model theory" proves there is a compact proper extension of  $L_{\omega,\omega}$  with the Beth property, using the  $\Delta$ -closure of the quantifier "the cofinality of < is  $\leq 2^{\omega}$ . This logic does not satisfy interpolation.

A. Mekler and S. Shelah (1985), "Stationary logic and its friends I" proves that it is consistent for L<sub>ω,ω</sub>(Q<sub>1</sub>) to have the weak Beth property.

• W. Hodges and S. Shelah (1991), "There are reasonably nice logics" proves that  $L_{\omega,\omega}(Q_{\alpha})$  is a reasonable logic for  $\omega_{\alpha}$  an uncountable strongly compact cardinal with at least one larger strongly compact cardinal. <u>Question</u>: what if one adds r.e. completeness to the conditions for a reasonable logic?