AN INTERPOLATION THEOREM IN THE PREDICATE CALCULUS

Roger C. Lyndon

1. Introduction. In studying the formal structure of sentences whose validity is preserved under passage from an algebraic system to a homomorphic image of the system, we have had occasion to use a lemma from formal logic. A proof of this lemma, our Interpolation Theorem, can be given within the theory of deductive inference, as formalized by Gentzen. Gentzen's theory is rather complicated and perhaps not generally well known. Moreover, the use of any formalized system of deductive logic seems to an extent alien to the primarily algebraic nature of our intended application. Therefore we give here a proof of the Interpolation Theorem that lies entirely within the theory of models: our arguments are as far as possible in the spirit of abstract algebra, and, in particular, borrow nothing from formal logic beyond an understanding of the intended meaning, herein precisely defined, of the conventional symbolism.

The Interpolation Theorem deals with sentences of the Predicate Calculus. Roughly, these are sentences that can be build up using the usual logical connectives, symbols denoting operations (or functions), symbols denoting relations (or predicates), and variables whose range is individual elements of the systems under consideration, but no variables ranging over operations, relations, or sets. The theorem takes the same form whether or not we admit a predicate denoting identity, with suitable axioms, to the predicate calculus. For technical reasons we admit as sentential connectives only the signs for negation, conjunction and disjunction (regarding "if $\cdots$ then" as a defined concept), together with signs 0 and 1 for truth and falsehood. For each occurrence of a relation symbol in a sentence S, there is a unique maximal chain of well formed formulas, all containing the given occurrence and each occurring as a proper part of the next. The given occurrence of the relation symbol will be called *positive* if the number of formulas in this chain that begin with the negation sign is even, and *negative* if this number is odd. If S is in prenex disjunctive form, this criterion takes the simpler form that an occurrence is negative if and only if it is preceded by the negation sign.

INTERPOLATION THEOREM, *Let S and T be sentences such that S implies T. Then there exists a sentence M such that S implies M and $M*

implies T, and that a relation symbol has positive occurrences in M only if it has positive occurrences in both S and T, and has negative occurrences in M only if it has negative occurrences in both S and T.

This theorem is a generalization of a result of W. Craig [3, 4]; Craig's lemma is obtained from it by suppressing the distinction between positive and negative sentences. As indicated, our first proof of the Interpolation Theorem used the Gentzen calculus; it did not differ essentially from Craig's proof, at that time unpublished, of his lemma.

The leading idea of the present proof is to interpret S implies T to mean that T holds in every model for which S holds; we express this relation by writing $S \Rightarrow T$. By Gödel's Completeness Theorem [6], this semantic interpretation is equivalent to the interpretation $S \vdash T$, that T is a formal consequence of S in a deductive axiomatization of the predicate calculus. The crucial point in our argument is the Main Theorem, which serves as a substitute, under this interpretation, for results in the theory of proof due to Herbrand [8] and to Gentzen [5].

A theorem of the theory of proof may be taken, in general, as saying that if there exists any derivation of one set $\varDelta$ of formulas from a set $\varGamma$, then there exists a derivation with certain special properties. A semantic counterpart of such a theorem will take the form of an 'interpolation theorem': if $\varGamma \Rightarrow \varDelta$, then there exists a chain $\varGamma = \varGamma^1$, $\varGamma^2, \cdots, \varGamma^n = \varDelta$ of sets of formulas, with certain special properties, such that $\varGamma^1 \Rightarrow \varGamma^2, \cdots, \varGamma^{n-1} \Rightarrow \varGamma^n$. Theorems of this sort will ordinarily require the occurrence in the $\varGamma^k$ of additional symbols (for the 'Skolem functions') that do not appear in $\varGamma$ or $\varDelta$, although this is not true of the Interpolation Theorem. Our arguments abjure any formal use of the concept of deductive derivablilty, hence of the Completeness Theorem. In various special cases, where $\varGamma \vdash \varDelta$ would be immediate, that $\varGamma \Rightarrow \varDelta$ follows directly from our definitions. The more difficult half of the Completeness Theorem, that if $\varGamma \Rightarrow \varDelta$ then $\varGamma \vdash \varDelta$, is implicit in the Main Theorem, which guarantees the existence of a chain $\varGamma = \varGamma^1, \cdots, \varGamma^n = \varDelta$ such that at each step the relation $\varGamma^k \vdash \varGamma^{k+1}$ is immediately evident.

I have profited much from discussions related to the present topic with A. Tarski and L. Henkin[2]; in particular, Tarski has emphasized the desirability of establishing the Interpolation Theorem by methods independent of the theory of proof. The idea of providing semantic proofs of results from the theory of proof is not new: a proof by E. Beth [1, 2], in a quite different formalism, of Craig's Lemma would certainly serve as well to prove the Interpolation Theorem; and A. Robinson has likewise provided semantic proofs of closely related results [10]. Unpublished results similar to those presented here have recently been

---

[2] In particular, while the author was visiting at the University of California, Berkeley.

obtained by A. Grzegorczyk, A. Mostowski and C. Ryll-Nardzewski, and by R. Vaught.

2. Basic concepts.[3] A *language* L is determined by an ordered quadruple, V, W, R, ρ, where V, W, R are disjoint sets, V infinite, and ρ is a function from $W \cup R$ to the natural numbers. The elements of V will be called *variables*, those of W *operation symbols*, and those of R *relation symbols*; for w in W, r in R, $\rho(w)$ is the *rank* of w and $\rho(r)$ the *rank* of r. The *logical symbols* are $0, 1, \sim, \wedge \vee, \forall, \exists$. The expressions of L will be made up of these symbols together with parentheses and commas. A *term* is, recursively, any variable, and any expression $w(t_1, \cdots, t_{\rho(w)})$ where w is an operation symbol and $t_1, \cdots, t_{\rho(w)}$ are terms. An *atomic formula* is any expression $r(t_1, \cdots, t_{\rho(r)})$ where r is a relation symbol and $t_1, \cdots, t_{\rho(r)}$ are terms. A *formula* is, recursively, any atomic formula, and any expression $0, 1, \sim F, (F \wedge G), (F \vee G), \forall x F, \exists x F$ where F and G are formulas and x is a variable. Formally, we define L to be the set of its symbols, terms and formulas.

We introduce the abbreviations $F \supset G$ for $(\sim F \vee G)$, $\bigwedge_1^n F_i$ for $F_1 \wedge \cdots \wedge F_n$ with the convention $\bigwedge_1^0 F_i = 1$, and $\bigvee_1^n F_i$ for $F_1 \vee \cdots \vee F_n$ with $\bigvee_1^0 F_i = 0$, and write $\forall x_1 \cdots x_n$ for $\forall x_1 \cdots \forall x_n$. A *matrix* is a formula that does not contain $\forall$ or $\exists$. A *normal matrix* is a matrix of the form $\bigvee_{i=1}^m \bigwedge_{j=1}^{n_m} F_{ij}$ where each F_{ij} is either A_{ij} or $\sim A_{ij}$, for A_{ij} an atomic formula. A *prenex* formula is one of the form $Q_1 x_1 \cdots Q_q x_q M$ where each Q_i is $\forall$ or $\exists$, each x_i is a variable, and M is a matrix; the formula is *normal* if the matrix M is normal. An occurrence of a variable x in a formula F is *free* in the formula F if it is not part of a subformula of the forms $\forall x G$ or $\exists x G$. A *sentence* is a formula without free occurrences of variables.

It is easily shown by induction that if G is any part of a formula F, then there is a smallest part of F that is a formula and contains G. It follows that there is a unique maximal chains of formulas $H_1, \cdots, H_n = F$, each a proper part of the next, and all containing G. The part G is *positive* in F if the number of $H_{i+1} = \sim H_i$ is even, and *negative* if it is odd. In what follows, G will always be an occurrence of a relation symbol in F.

An *interpretation* of a language L is determined by a set A and a function μ, defined on $V \cup W \cup R$, such that $\mu x \in A$ for $x \in V$, $\mu w \in A^{A^{\rho(w)}}$ for $w \in W$, and $\mu r \in 2^{A^{\rho(r)}}$ for $r \in R$. We regard 2 as the two element Boolean algebra with elements $0, 1$ and operations $\sim, \wedge, \vee$, so that μr is a function with values $(\mu r)(a_1, \cdots, a_{\rho(r)})$ equal to 0 or 1; but in practice we indulge in the harmless ambiguity of treating μw as a subset of $A^{\rho(w)+1}$

and μr of $A^{\rho(r)}$, and accordingly using such notation as $\mu w \subseteqq \mu w'$, $\mu r \subseteqq \mu r'$. Putting aside the trivial case that L contains no relation symbols of positive rank, μ unambiguously determines its *domain* A.

The function μ determines a unique extension mapping all terms of L into A, by the recursive definition

$$\mu[w(t, \cdots, t_{\rho(w)})] = (\mu w)(\mu t, \cdots, \mu t_{\rho(w)}) \, .$$

A further extension mapping all formulas of L into 2 is determined by the conditions.

$$(1) \qquad \mu 0 = 0, \mu 1 = 1, \quad \mu(\sim F) = \, \sim \mu F, \quad \mu(F \wedge G) = \mu F \wedge \mu G,$$
$$\mu(F \vee G) = \mu F \vee \mu G \, ,$$

and

$$(2) \qquad \begin{aligned} \mu(\forall x F) &= 1 \quad \text{if and only if} \quad \lambda F = 1 \quad \text{for all } \lambda \\ \mu(\exists x F) &= 1 \quad \text{if and only if} \quad \lambda F = 1 \quad \text{for some } \lambda \end{aligned} \Big\} \text{ such that}$$

$\lambda z = \mu z$ for all z in $V \cup W \cup R - \{x\}$. Formally, we define an interpretation to be a function μ thus extended ; in practice we shall say that μ and λ agree except on x when we mean that μ and λ agree for all z in $V \cup W \cup R - \{x\}$.

A *model* of L is the restriction $\mathfrak{A}$ of an interpretation μ to the operation and relation symbols of L. The model $\mathfrak{A}$ may be regarded as a 'relational system'[4] consisting of a set A, its domain, together with a set of *operations* $\mathfrak{A}w$ indexed by the operation symbols w of L, and a set of *relations* $\mathfrak{A}r$ indexed by the relation symbols r of L. If $\mathfrak{A}$ is the restriction of μ, we call μ an interpretation in the model $\mathfrak{A}$. If $\mu F = 1$, we say that F *holds* for the interpretation μ. Evidently μF depends only on the domain A of μ, the values of μ on the operation and relation symbols that occur in F, and the values of μ on the variables that occur free in F. In particular, if S is a sentence, μS depends only on the model $\mathfrak{A}$ to which μ belongs, and if $\mu S = 1$ we say that S holds in the model $\mathfrak{A}$.

If Γ and Δ are sets of formulas of L, we say that Γ *implies* Δ *in* L if $\mu\Delta = \{1\}$ for all interpretations of L such that $\mu\Gamma = \{1\}$. This interpretation is evidently independent of L, provided only that Γ and Δ belong to L ; we say simply that Γ *implies* Δ, and write $\Gamma \Rightarrow \Delta$. We write $\mu\Gamma = 1$ for $\mu\Gamma = \{1\}$, and employ such notation as $\Gamma_1, \Gamma_2 \Rightarrow F$ with the obvious meaning. If $\Gamma \Rightarrow \Delta$ and $\Delta \Rightarrow \Gamma$, then Γ and Δ are *equivalent* and we write $\Gamma \Longleftrightarrow \Delta$. That $1 \Rightarrow F$ expresses that F is a *theorem*. A set Γ is called *consistent* if there exists an interpretation μ such that $\mu\Gamma = 1$; thus $\Gamma \Rightarrow 0$ expresses that the set Γ is *inconsistent*.

---

[4] See [11], [12].

3. **Preliminary propositions.** The set $\Phi = \Phi(L)$ of all formulas of L constitutes, in an obvious sense, an algebraic system with operations $0, 1, \sim, \wedge, \vee$; in fact it is a ' word algebra ', a free algebra without axioms. The relation $F \Leftrightarrow G$ is a congruence on Φ, and the quotient system $\overline{\Phi}$ is a Boolean algebra, the Lindenbaum algebra of L. If κ is the canonical map of Φ onto $\overline{\Phi}$, then every interpretation μ of L, when restricted to Φ, can be factored uniquely in the form $\mu = \overline{\mu}\kappa$ where $\overline{\mu}$ is a homomorphism of $\overline{\Phi}$ onto 2.

The set Φ_0 of all matrices of L constitutes a subalgebra of Φ, and its image $\overline{\Phi_0} = \kappa\Phi_0$ is a subalgebra of $\overline{\Phi}$. Every homomorphism θ of $\overline{\Phi_0}$ onto 2 can be extended to a homomorphism θ' of $\overline{\Phi}$ onto 2 such that $\theta'\kappa$ is an interpretation. To prove this we construct the special interpretation μ *induced* by θ. For the domain A of μ we take the set of all terms of L. For a variable x, define $\mu x = x$. For an operation symbol w and terms $t_1, \cdots, t_{\rho(w)}$, we define μw by assigning to $(\mu w)(t_1, \cdots, t_{\rho(w)})$ as value the term $w(t_1, \cdots, t_{\rho(w)})$. For a relation symbol r and terms $t_1, \cdots, t_{\rho(r)}$ we define μr by assigning to $(\mu r)(t_1, \cdots, t_{\rho(r)})$ the value $\theta\kappa[r(t_1, \cdots, t_{\rho(r)})]$ in 2. By virtue of the last definition, $\mu F = \theta\kappa F$ for all atomic formulas F. Since the images κF of the atomic formulas F generate $\overline{\Phi_0}$, and $\overline{\mu}\kappa F = \theta\kappa F$ for atomic F, it follows that $\overline{\mu} = \theta$ on $\overline{\Phi_0}$ and $\overline{\mu}$ is an extension of θ.

PROPOSITION 1. *If Γ is a set of matrices, and J the dual ideal in the Boolean algebra $\overline{\Phi_0}$ generated by $\kappa\Gamma$, then $\Gamma \Rightarrow 0$ if and only if $0 \in J$.*

Proof. Assume $0 \in J$. Then $0 = \kappa F_1 \wedge \cdots \wedge \kappa F_n$ for some $F_1, \cdots, F_n$ in Γ. If μ is an interpretation such that $\mu\Gamma = 1$, then each $\overline{\mu}\kappa F_i = \mu F_i = 1$, whence $1 = \overline{\mu}\kappa \bigwedge F_i = \overline{\mu} \bigwedge \kappa F_i = \overline{\mu}0 = 0$, a contradiction. Assume $0 \notin J$. Then $J \neq \Phi_0$ and $J \subseteq K$ for some maximal dual ideal K in $\overline{\Phi_0}$ If θ is the canonical map of Φ_0 onto 2 with kernel the maximal ideal $\overline{\Phi_0} - K$ complementary to the dual ideal K, then $\theta\kappa\Gamma \subseteq \theta J \subseteq \theta\kappa = 1$. If μ is the special interpretation of L induced by the homomorphism θ, then $\mu\Gamma = \overline{\mu}k\Gamma = \theta\kappa\Gamma = 1$, whence Γ is consistent.

COROLLARY 1.1. *If Γ is a set of matrices, then $\Gamma \Rightarrow 0$ if and only if $\Gamma_0 \Rightarrow 0$ for some finite subset Γ_0 of Γ.*

Every map σ of the atomic formulas of L, as free generators of Φ_0, into Φ_0, extends to an endomorphism of Φ_0, which in turn induces an endomorphism $\overline{\rho}$ of $\overline{\Phi_0}$. It follows that if $\Gamma \Rightarrow 0$ then $\sigma\Gamma \Rightarrow 0$. Every map σ of the variables of L into terms of L extends in an obvious way to a map of the terms of L into terms of L, hence of formulas of L into formulas of L ; a transformation induced in this fashion will be

called a *substitution*.

PROPOSITION 2. *Let Γ be set of sentences S of the form $\forall x_1 \cdots x_n M$ where the M are matrices, and Γ'' the set of all formulas σM where σ is a substitution and M is the matrix of some sentence S in Γ. Then $\Gamma \Rightarrow 0$ if and only if $\Gamma'' \Rightarrow 0$.*

Proof. Suppose that Γ'' is consistent. Then $\lambda \Gamma'' = \bar\lambda \kappa \Gamma'' = 1$ for some interpretation λ. Let μ be the special interpretation induced by the homomorphism λ of Φ_0 onto 2. Let $F = \forall x_1 \cdots x_n M$ be in Γ', and ν be an interpretation that agrees with μ except on $x_1, \cdots, x_n$. Since the values νx for variables x are terms, we may define a substitution by setting $\sigma x = \nu x$. Since $\mu x = x$ for all variables x, $\nu M = \mu \sigma M = \bar\lambda \sigma M = 1$. This establishes that $\mu F = 1$. Suppose Γ' is inconsistent. Then for all interpretations μ there is some $F = \forall x_1 \cdots x_n M$ in Γ and some substitution σ such that $\mu \sigma M = 0$. Then setting $\lambda x_i = \mu \sigma x_i$, $i = 1, \cdots, n$ defines an interpretation λ that agrees with μ except on $x_1, \cdots, x_n$, and such that $\lambda M = 0$. It follows that $\mu F = 0$.

COROLLARY 2.1. *If Γ is a set of universal sentences, of the form $F = \forall x_1 \cdots x_n M$, where M is a matrix, then $\Gamma \Rightarrow 0$ if and only if $\Gamma_0 \Rightarrow 0$ for some finite subset Γ_0 of Γ.*

A prenex sentence S of the language L may be written in the form

$$S = \forall x_{11} \cdots x_{1m_1} \exists y_1 \cdots \forall x_{n1} \ldots x_{nm_n} \exists y_n M$$

where $n, m_1, \cdots, m_n$ are natural numbers, the x_{pq} and y_r are variables, and M is a matrix. The *Skolem matrix* of S is the result σM of substituting $\sigma y_r = s_r(x_{11}, \cdots, x_{rm_r})$ and $\sigma z = z$ for all other variables z; here the $s_1, \cdots, s_n$ are new and distinct operation symbols which we may suppose uniquely associated with the pair consisting of S and L. The *Skolem form* of S is the sentence $\forall x_{11} \cdots x_{nm_n} \sigma M$. The Skolem form belongs to the language L' obtained by adjoining the symbols $s_1, \cdots, s_n$ to L.

LEMMA 3. *Let S be a sentence of the form*

$$S = \forall x_{11} \cdots x_{1m_1} \exists y_1 \cdots x_{n1} \forall \cdots x_{nm_n} \exists y F,$$

where the x_{pq} and y_r are distinct variables and F is a formula in which all occurrences of these variables are free. Let F' result from F by substituting for each y_r a term σy_r that contains no variables other than $x_{11}, \cdots, x_{rm_r}$. Let S' be the sentence

$$S' = \forall x_{11} \cdots x_{1m_1} x_{21} \cdots x_{nm_n} F'.$$

Then $S' \Rightarrow S$.

Proof[5]. We proceed by induction. For $n = 0$ the assertion is trivial. For $n = 1$ it suffices to observe that if μ is an interpretation such that $\mu F' = 1$, then defining an interpretation λ to agree with μ except on y_1, and setting $\lambda y_1 = \mu \sigma y_1$, gives $\lambda F = \mu F''$, hence $\lambda F = 1$. For $n > 1$, form F''' from F by substituting σy_r for y_r, all y_r except y_n, and let $S'' = \forall x_{11} \cdots x_{nm_n} \exists y F'''$. Then the case $n = 1$ applies to give $S' \Rightarrow S''$, and the case $n - 1$ to give $S'' \Rightarrow S$.

PROPOSITION 4. *Let Γ be a set of prenex sentences of a language L, and Γ', in an extended language L', the set of all Skolem forms of the sentences in Γ. Then Γ holds in a model $\mathfrak{A}$ of L if and only if Γ' holds in some extension $\mathfrak{A}'$ of $\mathfrak{A}$ to a model of L'.*

Proof. By an induction it evidently suffices to establish the conclusion under the assumption that Γ' results from Γ by replacing a single sentence S by its Skolem form S'. If Γ' holds in an extension $\mathfrak{A}'$ of $\mathfrak{A}$ to L', it follows by Lemma 3 that Γ holds in $\mathfrak{A}'$, and, since Γ belongs to L, that Γ holds in $\mathfrak{A}$. For the rest, by a second induction it suffices to establish the conclusion for $S = \forall x_1, \cdots, x_m \exists y F$, $S' = \forall x_1 \cdots x_m \sigma F$, F a formula, $\sigma x_i = x_i$, $i = 1, \cdots, m$, and $\sigma y = s(x_1, \cdots, x_m)$, where s does not belong to L and L' is obtained by adjoining s to L.

Assume now that Γ holds in $\mathfrak{A}$. For any $a_1, \cdots, a_m$ in the domain A of $\mathfrak{A}$, there exists an interpretation μ in $\mathfrak{A}$ such that $\mu x_i = a_i$, $i = 1, \cdots, m$. Since $\mu S = 1$, it follows that $\mu(\exists y F) = 1$, and there exists an interpretation λ that agrees with μ except on y such that $\lambda F = 1$. By the axiom of choice we may define a function f from A^m into A by choosing for all $a, \cdots, a_m$ interpretations μ and λ as above and setting $f(a_1, \cdots, a_m) = \lambda y$. Extend $\mathfrak{A}$ to $\mathfrak{A}'$ by defining $\mathfrak{A}'s = f$. If μ' is an interpretation in $\mathfrak{A}'$, then μ' agrees with some μ, λ as above on the variables $x_1, \cdots, x_m$. Moreover, $\mu' \sigma y = f(\mu' x_1, \cdots, \mu' x_m) = f(a_1, \cdots, a_m) = \lambda y$, whence $\mu' \sigma F = \lambda F = 1$. It follows that $\mu' S' \; 1$ for all interpretations μ' in $\mathfrak{A}'$, whence Γ' holds in $\mathfrak{A}'$.

COROLLARY 4.1. *If Γ is any set of prenex sentences, then $\Gamma \Rightarrow 0$ if and only if $\Gamma_0 \Rightarrow 0$ for some finite subset Γ_0 of Γ.*

Every sentence is equivalent to a prenex sentence, and, indeed, a normal sentence. This follows by induction from various immediate consequences of the definitions, of which $\sim(F \wedge G) \Leftrightarrow (\sim F \vee \sim G)$ and $\forall x(F \wedge G) \Leftrightarrow (\forall x F \wedge \forall x G)$ are typical. In fact, it is easily seen that

⁵ C. C. Chang pointed out to me a gap in an earlier version of this proof.

every sentence S is equivalent to a normal sentence S' such that a relation symbol occurs positively (negatively) in S' only if it occurs positively (negatively) in S.

In view of this, Corollary 4.1 yields the Compactness Theorem.

PROPOSITION 5. *If Γ is any set of sentences, then $\Gamma \Rightarrow 0$ if and only if $\Gamma_0 \Rightarrow 0$ for some finite subset Γ_0 of Γ.*

4. The main theorem. Let S be a prenex sentence, of the form

$$S = \forall x_{11} \ldots \forall_{1m_1} \exists y_1 \cdots \forall x_{n1} \ldots x_{nm_n} \exists y_n M .$$

A second sentence S_0 will be said to arise from S by *duplication* if

(i) $\pi_1, \cdots, \pi$ are substitutions such that all $\pi_i x_{pq} = x_{pq}^i$, $\pi_i y_r = y_r^i$, where the x_{pq}^i and y_r^i are distinct variables; and

(ii) S_0 results from $\pi_1 M \wedge \cdots \wedge \pi_a M$ by prefixing quantifiers $\forall x_{pq}^i$ and $\exists y_r^i$ in some order such that, for $p \leq r$, $\forall x_{pq}^i$ precedes $\exists y_r^i$.

PROPOSITION 6. *If S_0 arises from S by duplication, then $S \Rightarrow S_0$.*

Proof. Let S have Skolem matrix σM, in the language L', where $\sigma x_{pq} = x_{pq}$ and $\sigma y_r = s_r(x_{11}, \cdots x_{rm_r})$. By Proposition 4, if S holds in any model $\mathfrak{A}$, then its Skolem form S' holds in some extension $\mathfrak{A}'$ of $\mathfrak{A}$ to L'. If μ is an interpretation of L' in $\mathfrak{A}'$, then every substitution instance of σM holds in μ; in particular, all $\pi_i \sigma M$ hold in μ, whence $\bigwedge \pi_i \sigma M$ holds in μ. But $\bigwedge \pi_i \sigma M$ results from $\bigwedge \pi_i M$ by substituting $s_r(x_{11}^i, \cdots, x_{rm_r}^i)$ for each y_r^i, whence, by Lemma 3, S_0 holds in $\mathfrak{A}'$, and therefore in $\mathfrak{A}$.

For S as before, a second sentence S_0 will be said to arise from S by *specialization* if

(iii) θ is a substitution such that $\theta y_r = y_r$, while each θx_{pq} is a term in certain new variables $u_1 \cdots, u_a$ together with the y_r for $r < p$; and

(iv) S_0 results from θM by prefixing quantifiers $\forall u_h$ and $\exists y_r$ in some order such that $\forall u_h$ precedes $\exists y_r$ if u_h occurs in any θx_{pq} for $p \leq r$, and $\exists y_s$ precedes $\exists y_r$ if y_s occurs in any θx_{pq} for $p \leq r$.

PROPOSITION 7. *If S_0 arises from S by specialization, then $S \Rightarrow S_0$.*

Proof. Let S have Skolem matrix σM in L' as before. Define a substitution ρ by setting $\rho z = z$ for all variables z other than the y_r, and, by recursion on the order of quantification of the y_r in S_0, defining $\rho y_r = \rho \theta \sigma y_r = s_r(\rho \theta x_{11}, \cdots, \rho \theta x_{rm_r})$. Since all y_s that occur in $\theta \sigma y_r$ occur in some θx_{pq} for $p \leq r$, all such y_s precede y_r in S_0, and the recursion

if legitimate. Since $\theta y_r = y_r$, $\rho\theta y_r = \rho y_r = \rho\theta\sigma y_r$ by the above definition, while for all other variables z, $\sigma z = z$ and again $\rho\theta z = \rho\theta\sigma z$. Suppose now that S holds in a model $\mathfrak{A}$ of L, and hence, by Proposition 4, that the Skolem form S' of S holds in an extension $\mathfrak{A}'$ of $\mathfrak{A}$ to L'. Then, for every interpretation μ in $\mathfrak{A}'$, all instances of σM hold, and, in particular, $\rho\theta\sigma M$ holds. Since $\rho\theta\sigma = \rho\theta$, $\rho\theta\sigma M = \rho\theta M$. Now $\rho\theta M$ results from θM by the substitution ρ, and $\rho u_h = u_h$, while ρy_r contains only those u_h that occur in the $\rho\theta x_{pq}$ for $p \leqq r$; by induction, using (iii), these are among the u_h that occur in θx_{pq} for $p \leqq r$, and hence among the u_h that precede y_r in S_0. Therefore Lemma 3 applies to establish that S_0 holds in $\mathfrak{A}'$ and thus in $\mathfrak{A}$.

Let S^1, S^2 be prenex sentences of the form, for $\delta = 1, 2$,

$$S^\delta = \forall x_{11}^\delta \cdots x_{1m_1}^\delta \exists y_1^\delta \cdots \forall x_{n1}^\delta \cdots x_{nm_n}^\delta \exists y_n^\delta M^\delta$$

with Skolem matrices σM^δ in a language L', where $\sigma x_{pq}^\delta = x_{pq}^\delta$, $\sigma y_r^\delta = s_r^\delta(x_{11}^\delta, \cdots x_{rm_r}^\delta)$. Then S^1 and S^2 will be called *propositionally inconsistent* if there exists a substitution η in L' that is one-to-one on all atomic formulas of each σM^δ such that $\eta\sigma M^1, \eta\sigma M^2 \Rightarrow 0$.

PROPOSITION 8. *If S^1, S^2 are propositionally inconsistent, then $S^1 S^2 \Rightarrow 0$.*

Proof. Suppose S^1, S^2 were consistent, hence both held in some model $\mathfrak{A}$ of L. Using Proposition 4, all instances of σM^1 and σM^2 would hold for all interpretations in a certain extension $\mathfrak{A}'$ of $\mathfrak{A}$ to a model of L'. Then $\eta\sigma M^1$ and $\eta\sigma M^2$ would hold for all such μ, and $\mu 0 = 1$, a contradiction.

In propositions 6, 7 and 8 we have attempted to isolate the chief ideas that underly the Main theorem; the proof of this theorem can now be accomplished by easier and more natural stages, although at the cost of a small amount of repetition.

MAIN THEOREM. *Let S^1 and S^2 be prenex sentences such that $S^1, S^2 \Rightarrow 0$. Then there exist prenex sentences T^1, T^2, U^1 and U^2 such that (1) T^1 arises from S^1, and T^2 from S^2, by duplication; (2) U^1 arises from T^1, and U^2 from T^2, by specialization; and (3) U^1 and U^2 are propositionally inconsistent.*

Proof. Let $S^1, S^2, M^1, M^2, \sigma$ and L, L' be as above. (There is clearly no loss of generality in taking common values of n and the m_r, and a common substitution σ, for S^1 and S^2.) By Proposition 4, $S^1, S^2 \Rightarrow 0$ implies that their Skolem forms are inconsistent. By Proposition 2, the set of all instances of σM^1 and σM^2 is consistent. By Corollary 1.1 some finite set of these instances is inconsistent. Therefore there exist substitutions $\eta_1, \cdots, \eta_a$ in the language L' such that

$$\eta_1 \sigma M^1, \cdots, \eta_a \sigma M^1, \eta_1 \sigma M^2, \cdots, \eta_a \sigma M^2 \Rightarrow 0 \ .$$

Define substitutions $\pi_1, \cdots, \pi_a$ such that all $\pi_i x_{pq}^\delta = x_{pq}^{\delta i}$ and $\pi_i y_r^\delta = y_r^{\delta i}$, where the $x_{pq}^{\delta i}$ and $y_r^{\delta i}$ are new and distinct variables. Define σ' such that $\sigma' x_{pq}^{\delta i} = x_{pq}^{\delta i}$ and $\sigma' y_r^{\delta i} = s_r^\delta(x_{11}^{\delta i}, \cdots, x_{rm_r}^{\delta i})$; thus $\sigma' \pi_i M^\delta = \pi_i \sigma M^\delta$ for all π_i. Define η such that $\eta x_{pq}^{\delta i} = \eta_i x_{pq}^\delta$; then $\eta \sigma' \pi_i M^\delta = \eta \pi_i \sigma M^\delta = \eta_i \sigma M^\delta$. Define $M_0^\delta = \bigwedge \pi_i M$; then $\eta \sigma' M_0^\delta = \bigwedge \pi_i \sigma M^\delta$, and $\eta \sigma' M_0^1, \eta \sigma' M_0^2 \Rightarrow 0$.

Let S_0^σ be the sentence obtained from M_0^δ by prefixing quantifiers $\forall x_{pq}^{\delta i}$ and $\exists y_r^{\delta i}$ in an order such that, if z and z' are two of these variables and the term $\eta \sigma' z$ is shorter than the term $\eta \sigma' z'$, then the quantification of z precedes that of z'. If $p \leq r$, the term $\eta \sigma' x_{pq}^{\delta i} = \eta x_{pq}^{\delta i}$ is a proper part of the term $\eta \sigma' y_r^{\delta i} = s_r^\delta(\eta x_{11}^{\delta i}, \cdots, \eta x_{rm_r}^{\delta i})$, whence $\forall x_{pq}^{\delta i}$ precedes $\exists y_t^{\delta i}$ in S_0^δ. Thus S_0^δ arises from S^δ by duplication.

Let S_0^δ have Skolem matrix $\sigma_0 M_0$ where $\sigma_0 x_{pq}^{\delta i} = x_{pq}^{\delta i}$ and $\sigma_0 y_r = s_r^{\delta i}(\cdots, x_{Lq}^{\delta j}, \cdots)$, the arguments ranging, in order of occurrence in S_0^δ, over all $x_{pq}^{\delta j}$ that precede y_r in S_0^δ. One has $\eta \sigma' x_{pq}^{\delta i} = \eta \sigma_0 x_{pq}^{\delta i}$, but $\eta \sigma' y_r = s_r^\delta(\eta x_{11}^{\delta i}, \cdots, \eta x_{r m_r}^{\delta i})$ while the term $\eta \sigma_0 y_r^{\delta i} = s_r^{\delta i}(\cdots, \eta x_{pq}^{\delta j}, \cdots)$ begins with a different operation symbol and contains additional arguments. To bring these into agreement, define a transformation χ on terms as follows :

(1) $\chi z = z$ for a variable z ;

(2) $\chi \eta \sigma' y_r^{\delta i} = \chi \eta \sigma_0 y_r^{\delta i}$;

(3) for any term $t = w(t_1, \cdots, t_{\rho(w)})$ not of the form $\eta \sigma' y_r^{\delta i}$,
   $$\chi t = w(\chi t_1, \cdots, \chi t_{\rho(w)}) \ .$$

The clause (2) if legitimate, by an induction on length of $\eta \sigma' y_r$. For $\chi \eta \sigma_0 y_r^{\delta i} = s_r^{\delta i}(\cdots, \chi \eta x_{pq}^{\delta j}, \cdots)$ contains $\chi \eta \sigma' y_s^{\delta k}$ only for those $\chi \eta \sigma' y_r^{\delta k}$ that occur in some $\chi \eta x_{pq}^{\delta j}$ for $p \leq r$, and it follows by an induction that for all of these $s < p$. Let L_0 be the language obtained from L by adjoining the symbols $s_r^{\delta i}$. Although neither χ nor $\chi \eta$ is in general a substitution, when applied to terms of L_0, which do not contain symbols s_r^δ, the clause (2) is never invoked ; consequently the restriction η_0 of $\chi \eta$ to L_0 is a substitution.

Since $\eta \sigma' M_0^1, \eta \sigma' M_0^2 \Rightarrow 0$, and χ induces a transformation on terms, it follows that $\chi \eta \sigma' M_0^1, \chi \eta \sigma' M_0^2 \Rightarrow 0$. Now $\chi \eta \sigma' y_r^{\delta i} = \chi \eta \sigma_0 y_r^{\delta i}$ by definition, while $\sigma' x_{pq}^{\delta i} = x_{pq}^{\delta i} = \sigma_0 x_{pq}^{\delta i}$ implies that $\chi \eta \sigma' x_{pq}^{\delta i} = \chi \eta \sigma_0 x_{pq}^{\delta i}$; it follows that $\chi \eta \sigma' M_0^\delta = \chi \eta \sigma_0 M_0^\delta = \eta_0 \sigma_0 M_0^\delta$, the last since $\sigma_0 M_0^\delta$ belongs to L_0. Hence, $\eta_0 \sigma^0 M_0^1, \eta_0 \sigma_0 M_0^2 \Rightarrow 0$.

Dropping the subscripts on S_0^δ, we now have the situation at the beginning of the proof, but with $a = 1$, that is with a single substitution η such that $\eta \sigma M^1, \eta \sigma M^2 \Rightarrow 0$. From the set of all terms that occur in $\eta \sigma M^\delta$ obtain a set B^δ by deleting successively any term that is expressible, by means of the operation symbols of L, in terms of the rest. Since each $\eta \sigma y_r^\delta = s_r^\delta(\eta x_{11}^\delta, \cdots, \eta x_{rm_r}^\delta)$ where s_r^δ does not belong to L, we

can suppose that all the $\eta \sigma y_r^\delta$ belong to B^δ. Let $b_1^\delta, \cdots, b_a^\delta$ be the remaining elements of B^δ. Then for each x_{pq}^δ (that occurs in M^δ) ηx_{pq}^δ is expressible in terms of the $\eta \sigma y_r^\delta$ and b_h^δ. More precisely, if $u_1^\delta, \cdots, u_a^\delta$ are new and distinct variables, and τ a substitution such that $\tau y_r^\delta = \eta \sigma y_r^\delta$, $\tau u_h^\delta = b_h^\delta$. then there exists in L a term θx_{pq}^δ in the variables y_r^δ and b_h^δ such that $\tau \theta x_{pq}^\delta = \eta x_{pq}^\delta$. We extend θ to a substitution by setting $\theta z = z$ for all z other than the x_{pq}^1, x_{pq}^2.

Let S_0^δ be the sentence obtained from θM^δ by prefixing the quantifiers $\forall u_h^\delta$ and $\exists y_r^\delta$ in an order such that if z and z' are two of these variables, and τz is shorter than $\tau z'$, then z precedes z' in S_0^δ. To verify that S_0^δ arises from S^δ by specialization, we observe that, for (iii), if y_r^δ occurs in θx_{pq} then $\tau y_r^\delta = \eta \sigma y_r^\delta$ is a proper part of $\tau \theta x_{pq}^\delta = \eta x_{pq}^\delta$ whence $r < p$; and, for (iv), if z is any y_s^δ or u_h^δ and z occurs in θx_{pq}^δ for $p \leqq r$, then τz is a part of ηx_{pq}^δ which is in turn a proper part of $\eta \sigma y_r^\delta = \tau y_r^\delta$, whence z precedes y_r^δ in S_0^δ.

Let S_0^δ have Skolem matrix $\sigma_0 \theta M^\delta$, where $\sigma_0 z_{pq} = z$ for all variables z other than the y_r^δ and $\sigma_0 y_r^\delta = s_{0r}^\delta(\cdots, u_h^\delta, \cdots)$, the arguments ranging in order over all u_h^δ that precede y_r^δ in S_0^δ. From $\eta \sigma M^1, \eta \sigma M^2 \Rightarrow 0$ it remains to construct η_0, one-to-one on the atomic formulas of $\sigma_0 \theta M^1$, $\sigma_0 \theta M^2$, such that $\eta_0 \sigma_0 \theta M^1, \eta_0 \sigma_0 \theta M^2 \Rightarrow 0$. For this define a transformation χ on terms as follows :

(1) $\chi z = z$ for a variable z ;
(2) $\chi \theta \tau \sigma y_r^\delta = \chi \tau \sigma_0 y_r^\delta$;
(3) for any term $t = w(t_1, \cdots, t_{\rho(w)})$ not of the form $\tau \theta \sigma y_r^\delta$,
      $\chi t = w(\chi t_1, \cdots, \chi t_{\rho(w)})$.

As in an earlier situation, this definition is legitimate, and the restriction η_0 of $\chi \tau$ to the language L_0 obtained from L by adjoining the symbols s_{0r}^δ is a substitution. As before we conclude from $\eta \sigma M^1, \eta \sigma M^2 \Rightarrow 0$ that $\chi \tau \theta \sigma M^1, \chi \tau \theta \sigma M^2 \Rightarrow 0$,
Now

$$\chi \tau \theta \sigma y_r^\delta = \chi \tau \sigma_0 y_r^\delta = \chi \tau \sigma_0 \theta y_r^\delta = \eta_0 \sigma_0 \theta y_r^\delta ,$$

and

$$\chi \tau \theta \sigma x_{pq}^\delta = \chi \tau \theta x_{pq}^\delta = \chi \tau \sigma_0 x_{pq}^\delta = \chi \tau \sigma_0 \theta x_{pq}^\delta = \eta_0 \sigma_0 \theta x_{pq}^\delta .$$

It follows that $\chi \tau \theta \sigma M^\delta = \eta_0 \sigma_0 \theta M^\delta$, whence

$$\eta_0 \sigma_0 \theta M^1, \eta_0 \sigma_0 \theta M^2 \Rightarrow 0 .$$

It remains to show that $\eta_0 \sigma_0 = \chi \tau \sigma_0$ is one-to-one on the terms of each θM^δ. We show first that $\tau \theta \sigma$ is one-to-one on such terms. These terms are terms in the variables u_h^δ and y_r^δ, containing only the operation symbols of L. Note that $\tau \theta \sigma u_h^\delta = \tau \theta u_h^\delta = \tau u_h^\delta = b_h^\delta$ and $\tau \theta \sigma y_r^\delta = \eta \sigma y_r^\delta$.

From the construction of B, it follows that, for two such terms t and t', $\tau\theta\sigma t = \tau\theta\sigma t'$ cannot hold for one of t, t' a variable unless $t = t'$. Suppose now that $t = w(t_1, \cdots, t_{\rho(w)})$ and $t' = w'(t'_1, \cdots, t'_{\rho(w')})$. Comparing the first symbols we conclude from $\tau\theta\sigma t = \tau\theta\sigma t'$ that $w = w'$, and the arguments agree:

$$\tau\theta\sigma t_i = \tau\theta\sigma t'_i \qquad\qquad i = 1, \cdots, \rho(w) = \rho(w') .$$

By induction on the length of the shorter of t, t' we conclude that each $t_i = t'_i$, whence $t = t'$.

Finally, $\chi\tau\sigma_0 y_r^\delta = \chi\tau\theta\sigma y_r^\delta$ by definition, and $\chi\tau\sigma_0 u_h^\delta = \chi\tau u_h^\delta = \chi\tau\theta\sigma u_h^\delta$. Hence $\chi\tau\sigma_0 = \chi\tau\theta\sigma$ on terms of θM^δ. But χ is evidently one-to-one on terms that do not contain the symbols s_{0r}^δ. Hence, for terms t and t' of θM^δ, $\sigma\tau\sigma_0 t = \chi\tau\sigma_0 t^1$ implies $\chi\tau\theta\sigma t = \chi\tau\theta\sigma t'$, hence $\tau\theta\sigma t = \tau\theta\sigma t'$, and, by the property of $\tau\theta\sigma$ established above, $t = t'$. This completes the proof of the Main Theorem.

5. The Interpolation theorem. *Let S and T be sentences of a language L such that $S \Rightarrow T$. Then there exists a sentence S^0 of the language L such that $S \Rightarrow S^0$, $S^0 \Rightarrow T$, and that a relation symbol occurs positively in S^0 only if it occurs positively in both S and T, and occurs negatively in S^0 only if it occurs negatively in both S and T.*

Proof. S is equivalent to a prenex sentence S^1 such that a relation symbol occurs positively (negatively) in S^1 only if it occurs positively (negatively) in S. And $\sim T$ is equivalent to a prenex sentence S^2 such that a relation symbol occurs positively (negatively) in S^2 only if it occurs negatively (positively) in T. Since $S^1, S^2 \Rightarrow 0$, by the Main Theorem there exist prenex sentencs U^1 and U^2 such that $S^1 \Rightarrow U^1$, $S^2 \Rightarrow U^2$, that U^1 contains the same kinds of occurrences of relation symbols as S^1 and U^2 as S^2, and that $\eta\sigma M^1$, $\eta\sigma M^2 \Rightarrow 0$ where σM^1, σM^2 are the Skolem matrices of U^1, U^2, and η is a substitution that is one-to-one on the atomic formulas of each of σM^1, σM^2. All this is not altered if we modify U^1, U^2 by reducing M^1, M^2 to normal form.

It will suffice to find S^0 such that $U^1 \Rightarrow S^0$, and $S^0, U^2 \Rightarrow 0$, and a relation symbol occurs positively (negatively) in S^0 only if it occurs positively (negatively) in U^1 and negatively (positively) in U^2. Write $M^\delta = \bigvee M_i^\delta$, each $M_i^\delta = \bigwedge M_{ij}^\delta$, and each M_{ij}^δ either A_{ij}^δ or $\sim A_{ij}^\delta$ where A_{ij}^δ is an atomic formula. Define $M^0 = \bigvee M_i^0$ where $M_i^0 = 0$ if $M_i^1 \Rightarrow 0$, and otherwise M_i^0 results from M_i^1 by deleting all M_{ij}^1 such that $\sim\eta\sigma M_{ij}^1$ is not equivalent to some $\eta\sigma M_{hk}^2$. Let S^0 be the sentence obtained from U^1 by replacing its matrix M^1 by the matrix M^0. It is immediate that the occurences of relation symbols in S^0 are related to those in U^1 and U^2 in the required manner. Moreover. since $M^1 \Rightarrow M^0$ is immediate, it follows easily that $U^1 \Rightarrow S^0$.

It remains to show that $S^0, U^2 \Rightarrow 0$, and for this it will suffice to to show that $\eta\sigma M^0, \eta\sigma M^2 \Rightarrow 0$. Since $\eta\sigma M^1 \wedge \eta\sigma M^2 \Rightarrow 0$, then for all i, h, $\eta\sigma M_i^1 \wedge \eta\sigma M_h^2 \Rightarrow 0$. We want to conclude that for all i, h, $\eta\sigma M_i^0 \wedge \eta\sigma M_h^2 \Rightarrow 0$. Since σ is clearly one-to-one on the terms of M^1, so is $\eta\sigma$, and $\eta\sigma M_i^1 \Rightarrow 0$ implies $M_i^1 \Rightarrow 0$, whence by definition $M_i^0 = 0$, hence $\eta\sigma M_i^0 = 0$ and the conclusion follows. If $\eta\sigma M_h^2 \Rightarrow 0$ the conclusion is immediate. In the remaining case there exist j and k such that $\sim\eta\sigma M_{ij}^1 \Longleftrightarrow \eta\sigma M_{hk}^2$. But then, by definition, M_i^0 still contains the conjunct M_{ij}^1, and again $\eta\sigma M_i^0 \wedge \eta\sigma M_h^2 \Rightarrow 0$. Since $\eta\sigma M_i^0 \wedge \eta\sigma M_h^2 \Rightarrow 0$ for all i, h, it follows that $\eta\sigma M^0 \wedge \eta\sigma M^2 \Rightarrow 0$, completing the proof.

It was stated in the introduction that the Interpolation Theorem remains true for the predicate calculus with identity. Precisely, we restrict the definition of a language to apply only to those that contain a fixed relation symbol e of rank two, and the definition of interpretation to admit only those μ for which μe is the identity relation on the domain of μ. The relation $S \Rightarrow T$ then acquires a stronger meaning. Nonetheless, the Interpolation Theorem as stated remains true in this new sense. (It may be well to note that e is included among the relation symbols mentioned in the conclusion of the theorem.) In fact, all statements in this paper remain true in the new sense, apart from two modifications. First, Proposition 1 must be modified by enlarging J to contain (the coset of) each formula $e(t, t)$, t a term, and to contain any formula F' obtainable from a formula F in J by replacing an occurrence of a term t by a new term t', provided that $e(t, t')$ is in J. Second, in the proof of the Interpolation Theorem, the M_i^δ as described above must be similarly enlarged by adjoining to each the finite set of all M_{ij}^δ of the form A or $\sim A$, A atomic, such that $M_i^\delta \Rightarrow M_{ij}^\delta$ in the present sense.

The Interpolation Theorem can be refined in other ways. Conditions can be imposed on the internal structure of the atomic formulas $r(t_1, \cdots, t_{\rho(r)})$ containing the relation symbol r. For example, define an I-occurrence of r in S to be one in which each t_i, for $i \in I \subseteq \{1, \cdots \rho(r)\}$, is a variable universally quantified in S. Then it can be required that r have I-occurrences is S^0 only if it has I'-occurrences in S and I''-occurrences in T, where $I'' \subseteq I \subseteq I'$. Alternatively, stronger conditions can be imposed on the external context in which a relation symbol occurs. For example, suppose all positive occurrences in S of a relation symbol r are in formulas $A' \supset A$ where A and A' are atomic formulas, and that none of the relation symbols appearing in the parts A' of these formulas have positive occurrences in S, except possibly in parts A; then S^0 can be required to contain no positive occurrences of r. Such refinements of the Interpolation Theorem have proved useful in the study of homomorphisms and subdirect products of models, but because of their special nature it does not seem worthwhile to give separately formal statements and proofs of these results.

REFERENCES

1.   E. W. Beth, *On Padoa's method in the theory of definition,* Indag, Math. **15** (1953), 66.

2.   ———— , *The Foundations of Mathematics,* North Holland, Amsterdam, 1958.

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

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

5.   G. Gentzen, *Untersuchungen über das logische Schliessen,* Math. Z. **39** (1934–35), 176, 405.

6.   K. Gödel, *Die Vollständigkeit der Axiome des logischen Funktionenkalküls,* Monatsh. Math. Phys. **37** (1930), 349.

7.   L. Henkin, *The completeness of the first order functional calculus,* J. Symbolic Logic **14** (1949), 159.

8.   J. Herbrand, *Recherches sur la théorie de la démonstration,* Travaux Soc. Sci. et Let. Varsovie, Cl. III, No. 33, 1930.

9.   S. C. Kleene, *Introduction to Metamathematics,* Van Nostrand, New York, 1952.

10.   A. Robinson, *A result on consistency and its application to the theory of definition,* Indag. Math. **18** (1956), 47.

11.   A Tarski, *Some notions and methods on the borderline of algebra and metamathematics,* Proc. International Congress of Math. **1**, Cambridge, 1950, 705.

12.   A. Tarski, *Contributions to the theory of models, I, II, III,* Indag. Math. **16** (1954) 572, 582; **17** (1955), 55.

UNIVERSITY OF MICHIGAN