ANNALS OF MATHEMATICAL LOGIC – Volume 1, No. 4 (1970) pp. 423–468

GENERALIZED INTERPOLATION AND DEFINABILITY

David W.KUEKER
University of Michigan

Received 1 August 1969

In this paper we present various results on interpolation and definability which generalize the well-known theorems of Craig [5] and Beth [1]. The proofs are almost entirely model-theoretic, and rely heavily on the use of special models (see [15]). Statements of some known results we will refer to are included in §1. Our basic result is the Main Lemma 2.2 of §2 characterizing subsets of special models definable by infinite conjunctions. The Main Lemma is reformulated as Theorem 2.3 on Σ_1^1-definable subsets. These results are applied in §3 to yield interpolation-type theorems, such as Theorem 3.2, which concern certain second-order conditions whose only second-order quantifiers are universal. In §4 we obtain some definability results using interpolation theorems. In particular, by applying Theorem 3.2 we obtain Theorem 4.1, a generalization of Beth's theorem treating conditions intermediate between those in Beth's theorem and the theorem of Chang [2] and Makkai [13]. The Main Lemma is applied in §5 to intersections of elementary submodels, yielding a proof of some results of Park [16, 17]. In the last section we give a syntactical proof of Theorem 4.1.

Many of the results of this paper appeared in the author's doctoral dissertation [9] written under Professor C.C.Chang. Theorem

4.1 was also announced in [8]. The author wishes to express his appreciation to Professor Chang for many helpful comments and suggestions for improvements concerning the results of this paper.

§ 1. Preliminaries

We consider a finitary first-order predicate language L (with identity) which is fixed throughout the paper. Models for the language L will be denoted by $\mathfrak{A}$, $\mathfrak{B}$, ... We will follow the convention that the universe of $\mathfrak{A}$ is A, that of $\mathfrak{B}$ is B, etc. We assume familiarity with the basic concepts of model theory, and also the notion of special model (from [15]). For the most part we employ standard terminology and notations; for example, we use $\cong$ for isomorphism, $\equiv$ for elementary equivalence, $\prec$ for elementary submodel, and $|X|$ for the cardinality of a set X. We use $\models$ both for the relation of satisfiability in a model ($\mathfrak{A} \models \phi(a_1, ..., a_k)$ where $a_1, ..., a_k \in A$) and for the (semantic) relation of consequence ($T \models \phi$ where T is a theory). In the rest of this section we explain some other notations and conventions, give some facts about special models, and state some known results on definability and interpolation.

If **R** is a k-place predicate symbol not belonging to L, then L(**R**) is the new language formed by adding **R** to L. Models for L(**R**) will be written as $(\mathfrak{A}, R)$, where R is a k-place relation on A. Similarly, if we are given a sequence $\mathbf{R}_0, \mathbf{R}_1, ...$ of new predicates, we form the new language $L(\mathbf{R}_0, \mathbf{R}_1, ...)$, whose models are written $(\mathfrak{A}, R_0, R_1, ...)$.

We will assume throughout that **P**, **Q**, **R**, and **S** (sometimes with subscripts) are distinct predicate symbols which do not occur in L. In addition, **P** and **Q** are assumed to be unary.

In writing formulas of these expanded languages we will sometimes exhibit the new predicate symbols. For example, a formula ϕ of L(**P**) may be written as $\phi(\mathbf{P})$. $\phi(\mathbf{Q})$ would then be the formula of L(**Q**) obtained by substituting **Q** everywhere in ϕ for **P**. Added predicates will also be treated at times as second-order variables, and we will form second-order sentences such as $\exists \mathbf{P}\, \phi(\mathbf{P})$.

We also sometimes exhibit the free (individual) variables of a formula, writing $\phi(v_1, ..., v_k)$ for ϕ. Using the notation $\phi(v_1, ..., v_k)$

will imply that the only free variables of ϕ are $v_1, ..., v_k$ (but $v_1, ..., v_k$ need not all occur free in ϕ). We also use x, y, and z for variables.

Given formulas $\phi_1, ..., \phi_n$ we will write $\bigwedge_{1\leq i\leq n} \phi_i$ and $\bigvee_{1\leq i\leq n} \phi_i$ for $\phi_1 \wedge ... \wedge \phi_n$ and $\phi_1 \vee ... \vee \phi_n$ respectively. More generally, if $\{\phi_i : i \in I\}$ is any family of formulas, we use $\bigwedge_{i\in I} \phi_i$ and $\bigvee_{i\in I} \phi_i$ for the (possibly infinite) conjunction and disjunction of all the formulas ϕ_i. The satisfaction of these infinitary formulas in a model is defined by the obvious extension of the usual definition for finitary formulas.

We use $\exists^{\leq n} x\, \phi$ as an abbreviation for an expression meaning "there are at most n x such that ϕ". Similarly, $\exists ! x\, \phi$ means "there is exactly one x such that ϕ". We always use n, k, and m to denote natural numbers, that is, elements of ω. The empty set is denoted by 0.

If T is a theory of $\mathrm{L}(\mathbf{P})$ and $\mathfrak{A}$ is any model (that is, model for L) then we define

$$M_T(\mathfrak{A}) = \{ P \subseteq A : (\mathfrak{A}, P) \models T \} .$$

More generally, if T is a theory of some language containing $\mathrm{L}(\mathbf{P})$, say $\mathrm{L}(\mathbf{P}, \mathbf{R}, ...)$, then by $M_T(\mathfrak{A})$ we mean the set of all $P \subseteq A$ such that $(\mathfrak{A}, P)$ can be expanded to some model $(\mathfrak{A}, P, R, ...)$ of T.

We assume throughout that T is a theory in $\mathrm{L}(\mathbf{P})$ or in some language containing $\mathrm{L}(\mathbf{P})$. Hence for any T the set $M_T(\mathfrak{A})$ is unambiguously defined for every $\mathfrak{A}$.

If $(\mathfrak{A}, P)$ is any model for $\mathrm{L}(\mathbf{P})$ then we define

$$M(\mathfrak{A}, P) = \{ P' \subseteq A : (\mathfrak{A}, P) \cong (\mathcal{A}, P') \} .$$

Therefore $P' \in M(\mathfrak{A}, P)$ if and only if there is an automorphism of $\mathfrak{A}$ mapping P onto P'. Notice that P is always an element of

$M(\mathfrak{A}, P)$, but that $M_T(\mathfrak{A})$ may be empty. But if $(\mathfrak{A}, P)$ is a model of T, then $M(\mathfrak{A}, P) \subseteq M_T(\mathfrak{A})$.

We will freely use some facts from [15] about special models without explicit mention. Among them are the facts that every consistent theory has special models ("existence"), that elementarily equivalent special models of the same power are isomorphic ("uniqueness"), and that if $\mathfrak{A}$ is special and $a \in A$ then $(\mathfrak{A}, a)$ is also special. We also use the fact that special models are "relation-universal", that is:

(U) If $\mathfrak{A}$ is special and $\mathfrak{A} \equiv \mathfrak{A}'$, then for any relation R' on A' there is some relation R on A such that $(\mathfrak{A}, R) \equiv (\mathfrak{A}', R')$.

Now, actually it is only proved in [15] that special models of certain cardinalities are relation-universal (namely those cardinalities in which any theory of the language has a special model, in which case it is immediate by uniqueness). This would be enough for most of our purposes, only making it necessary to add this cardinality restriction to the Main Lemma 2.2 and Theorem 2.3. However, (U) is true in general, so we here indicate the proof.

Let $\mathfrak{A}$ be special, and assume for simplicity that $\kappa = |A| > |L| \cup \omega$. Let $\mathfrak{A}' \equiv \mathfrak{A}$ and R' be given. By the downward Löwenheim-Skolem theorem we may assume that $|A'| < |A|$. We construct a sequence $\{(\mathfrak{A}_\nu, R_\nu)\}$ of models and a sequence $\{f_\nu\}$ of functions, each of length κ (or cofinality of κ if κ is singular) such that the following hold:

$$(\mathfrak{A}_0, R_0) = (\mathfrak{A}', R'),$$

$$(\mathfrak{A}_\nu, R_\nu) \prec (\mathfrak{A}_\mu, R_\mu) \text{ for } \nu < \mu,$$

$$|A_\nu| < \kappa,$$

f_ν is an elementary map of $\mathfrak{A}_\nu$ into $\mathfrak{A}$,

$f_\nu \subseteq f_\mu$ for $\nu < \mu$,

and every $a \in A$ is in the range of some f_ν.

The construction of these sequences offers no serious difficulties, using, say, the characterization of special models given by remark (5) in [15].

Now let

$$\mathfrak{A}'' = \bigcup_\nu \mathfrak{A}_\nu , \qquad R'' = \bigcup_\nu R_\nu , \qquad g = \bigcup_\nu f_\nu .$$

Then $(\mathfrak{A}', R') \prec (\mathfrak{A}'', R'')$ and g is an isomorphism of $\mathfrak{A}''$ onto $\mathfrak{A}$. Letting R be the image of R'' under g, we have $(\mathfrak{A}, R) \cong (\mathfrak{A}'', R'')$, and so $(\mathfrak{A}, R) \equiv (\mathfrak{A}', R')$ as desired.

The case $|A| = |\mathrm{L}| \cup \omega$ offers a few more complications but is not essentially different.

The case $\omega \leq |A| < |\mathrm{L}|$ can be reduced to the previous cases by finding a language $\mathrm{L}' \subseteq \mathrm{L}$ such that $|\mathrm{L}'| \cup \omega \leq |A|$ and every symbol of $\mathrm{L} - \mathrm{L}'$ is definable in $\mathscr{A}$ by those of L' alone. This reduction depends on the fact that $\mathfrak{A}$, being special, must realize every type in finitely many variables which is consistent with it. Hence, there can be at most $|A|$ maximal types, and therefore by compactness at most $|A|$ non-equivalent predicates in L. (Recall that a *type* in $v_1, ..., v_k$ is a set Φ of formulas with just $v_1, ..., v_k$ free, and that the type Φ is *realized* in $\mathfrak{A}$ if there are $a_1, ..., a_k \in A$ such that $\mathfrak{A} \models \phi(a_1, ..., a_k)$ for every $\phi \in \Phi$.)

If $|A| < \omega$, of course, (U) is obvious.

In our notation we may state Beth's definability theorem as follows.

Theorem 1.1 (Beth [1]). *For any T the following conditions are equivalent:*

(i) *For every* $\mathfrak{A}$, $|M_T(\mathfrak{A})| \leq 1$;

(ii) *There is a formula* $\phi(x)$ *of* L *such that*

$$T \models \forall x[\mathbf{P}(x) \leftrightarrow \phi(x)] \ .$$

Beth proved this theorem only for theories T in the language L(**P**). The extension to theories T in any language containing L(**P**) is due to Craig [5], and follows naturally from his proof of the theorem using his interpolation theorem.

Theorem 1.2 (Craig [5]). *Let* $\psi(\mathbf{R})$ *be a formula of* L(**R**) *and* $\chi(\mathbf{S})$ *a formula of* L(**S**). *Then the following are equivalent:*
(i) $\models \psi(\mathbf{R}) \to \chi(\mathbf{S})$;
(ii) *There is some formula* ϕ *of* L *such that*

$$\models \psi(\mathbf{R}) \to \phi \ \text{ and } \ \models \phi \to \chi(\mathbf{S}) \ .$$

Beth's theorem is true also for predicates of any (finite) number of places, not just unary predicates. The same is true for all the other definability results in this paper. We have stated them just for unary predicates solely for ease of presentation. Similarly, in Craig's theorem **R** and **S** could be replaced by sequences of new and all different predicates (also, of course, functions and individual constants). This comment also will apply to the interpolation results we will give later.

Notice that the strongest condition on the interpolating formula ϕ of Theorem 1.2 is obtained by taking L to be the language containing only the non-logical constants (that is, the predicate, function, and individual constant symbols) which occur in both ψ and χ. Thus, ϕ contains only the non-logical constants common to ψ and χ.

The following theorem is similar to Theorem 1.1 but concerns $M(\mathfrak{A}, P)$; condition (i) says that P is left fixed by the automorphisms of $\mathfrak{A}$.

Theorem 1.3 (Svenonius [22]). *For any T the following are equivalent:*

(i) *For every model* $(\mathfrak{A}, P)$ *of* T, $|M(\mathfrak{A}, P)| = 1$;

(ii) *There are formulas* $\phi_1(x), ..., \phi_n(x)$ *of* L *such that*

$$T \models \bigvee_{1 \leq i \leq n} \forall x [\mathbf{P}(x) \leftrightarrow \phi_i(x)] .$$

Svenonius' theorem also holds for theories T in any language containing L(**P**), it being understood in this case that the requirement in (i) that $(\mathfrak{A}, P)$ is a model of T means that $P \in M_T(\mathfrak{A})$. Theorem 1.3 may be derived from Beth's theorem by showing that if T satisfies condition (i) of 1.3 then every complete extension of T satisfies condition (i) of 1.1; this follows from a simple special model argument (essentially remark (7) of [15]).

Finally there is the following theorem, due independently to Chang and Makkai, which is an infinite generalization of both Theorems 1.1 and 1.3.

Theorem 1.4 (Chang [2], Makkai [13]). *For any T the following are equivalent:*

(i) *For every infinite* $\mathfrak{A}$, $|M_T(\mathfrak{A})| \leq |A|$;

(ii) *For every infinite* $\mathfrak{A}$, $|M_T(\mathfrak{A})| < 2^{|A|}$;

(iii) *For every infinite model* $(\mathfrak{A}, P)$ *of* T, $|M(\mathfrak{A}, P)| \leq |A|$;

(iv) *For every infinite model* $(\mathfrak{A}, P)$ *of* T, $|M(\mathfrak{A}, P)| < 2^{|A|}$;

(v) *There are formulas* $\phi_i(x, v_1, ..., v_k)$, $1 \leq i \leq n$, *of* L *such that*

$$T \models \bigvee_{1 \leq i \leq n} \exists v_1, ..., v_k \, \forall x \, [\mathbf{P}(x) \leftrightarrow \phi_i] .$$

Among other results concerning definability we mention just Robinson's consistency lemma ([18, 20]), which is equivalent to Craig's interpolation theorem, and thus also suffices to yield Beth's theorem. There are also several results improving Craig's theorem (e.g. [6, 12]), which have, however, little direct connection with the results we will give.

§2. Subsets of special models

Before proceeding to the main results of this section, we first note the following.

Lemma 2.1. *Let $\mathfrak{A}$ be special and let $P \subseteq A$. Then the following are equivalent:*
(i) $|M(\mathfrak{A}, P)| = 1$;
(ii) *There are formulas $\phi_{i,j}(x)$ of* L *($i \in I$, $j \in J$) such that*

$$(\mathfrak{A}, P) \models \forall x\, [\mathbf{P}(x) \leftrightarrow \bigvee_{i \in I} \bigwedge_{j \in J} \phi_{i,j}(x)]\ .$$

Proof. It is enough to show that (i) implies (ii). Let $a, b \in A$ and assume that $(\mathfrak{A}, a) \equiv (\mathfrak{A}, b)$. Then $(\mathfrak{A}, a) \cong (\mathfrak{A}, b)$ since $\mathfrak{A}$ is special, and hence $a \in P$ if and only if $b \in P$ by (i). So, let $\{a_i : i \in I\}$ enumerate the elements of P. For each $i \in I$ let $\{\phi_{i,j}(x) : j \in J\}$ enumerate all the formulas $\phi(x)$ of L satisfied by a_i in $\mathfrak{A}$. Then $b \in P$ if and only if $(\mathscr{A}, b) \equiv (\mathfrak{A}, a_i)$ for some $i \in I$, that is, if and only if

$$\mathfrak{A} \models \bigwedge_{j \in J} \phi_{i,j}(b) \text{ for some } i \in I\ .$$

Hence (ii) holds.

If $(\mathfrak{A}, P)$ is also special then a compactness argument can be used to show that in (ii) a single formula $\phi(x)$ suffices to define P. From this fact, as is well-known, one can derive Beth's and Svenonius' theorems.

Main Lemma 2.2. *For any T there are formulas $\phi_i(x)$ of* L *($i \in I$) such that for every special model $\mathfrak{A}$, if $Q = \bigcup M_T(\mathscr{A})$ then*

$$(\mathfrak{A}, Q) \models \forall x [\mathbf{Q}(x) \leftrightarrow \bigwedge_{i \in I} \phi_i(x)]\ .$$

Conversely, given Q so definable there is such a theory T.

Proof. Let $\{\phi_i(x) : i \in I\}$ enumerate all the formulas $\phi(x)$ of L such that

$$T \models \forall x[\mathrm{P}(x) \to \phi(x)] .$$

Let $\mathfrak{A}$ be special and let $Q = \bigcup M_T(\mathfrak{A})$. Then we have

$$(\mathfrak{A}, Q) \models \forall x[\mathrm{Q}(x) \to \bigwedge_{i \in I} \phi_i(x)]$$

since that implication is true of every $P \in M_T(\mathfrak{A})$. So we will be through once we show

(1) $$(\mathfrak{A}, Q) \models \forall x[\neg \mathrm{Q}(x) \to \bigvee_{i \in I} \neg \phi_i(x)] .$$

Assume $a \notin Q$. Let $\{\chi_j(x) : j \in J\}$ enumerate all the formulas of L satisfied by a in $\mathfrak{A}$. Notice $|M(\mathfrak{A}, Q)| = 1$. Hence by the proof of the previous lemma we know

(2) $$\text{if } \mathfrak{A} \models \bigwedge_{j \in J} \chi_j(b) \text{ then } b \notin Q \text{ and so}$$

$$\text{if } (\mathfrak{A}, P) \models T, \qquad b \notin P .$$

Let T_0 be the complete theory consisting of all sentences of L true on $\mathfrak{A}$. We first show

(3) $$T_0 \cup T \cup \{\chi_j(x) \wedge \mathrm{P}(x) : j \in J\} \text{ is inconsistent.}$$

If not, there would be some model $(\mathfrak{A}', P')$ of $T_0 \cup T$ and some $b' \in P'$ satisfying all χ_j in $\mathfrak{A}'$. But then $\mathfrak{A} \equiv \mathfrak{A}'$ and therefore, since

$\mathfrak{A}$ is relation-universal, we could find $P \subseteq A$ and $b \in P$ such that $(\mathfrak{A}, P) \models T$ and $\mathscr{A} \models \bigwedge_{j \in J} \chi_j(b)$, thus contradicting (2).

Applying compactness to (3) we obtain some $j \in J$ and some sentence σ of L such that

$$T_0 \models \sigma \text{ and } T \models \neg [\sigma \wedge \chi_j(x) \wedge \mathbf{P}(x)] ,$$

hence

$$T \models \forall x(\mathbf{P}(x) \to [\sigma \to \neg \chi_j(x)]) .$$

It follows that $\sigma \to \neg \chi_j$ is some ϕ_i. But $\mathfrak{A} \models \sigma \wedge \chi_j(a)$, hence $\mathfrak{A} \models \neg \phi_i(c)$ and so we have shown (1).

The converse is clear, since given the definition of Q by the ϕ_i, we could take

$$T = \{ \forall x[\mathbf{P}(x) \to \phi_i(x)] : i \in I\} .$$

A few comments are in order. First, **P** need not be a unary predicate, and T may be a theory in any language containing L(**P**), provided we add the restriction that $|A|$ is at least as large as the number of new symbols added. These improvements are both clear from the proof. Also, the dual of the Lemma, obtained by replacing $\bigcup$ by $\bigcap$ and $\bigwedge$ by $\bigvee$, clearly holds. Notice that the hypothesis on Q is that it be defined by the second-order (and in general infinitary) formula $\exists \mathbf{P}[T(\mathbf{P}) \wedge \mathbf{P}(x)]$. The following theorem is a reformulation of the Main Lemma treating the case of an arbitrary Σ^1_1 formula.

Theorem 2.3. *Let* $\theta(\mathbf{R}, y)$ *be a formula of* L(**R**). *Then there are formulas* $\phi_i(y)$ *of* L $(i \in I)$ *such that for every special model* $\mathfrak{A}$

$$\mathfrak{A} \models \forall y[\exists \mathbf{R}\, \theta(\mathbf{R}, y) \leftrightarrow \bigwedge_{i \in I} \phi_i(y)] .$$

Proof. Define a theory T in $L(\mathbf{P}, \mathbf{R})$ by $T = \{\forall x[\mathbf{P}(x) \leftrightarrow \theta(\mathbf{R}, x)]\}$ Applying Lemma 2.2, in view of the comments following it, we obtain formulas $\phi_i(y)$ of L $(i \in I)$ such that for every special $\mathfrak{A}$ the set $Q = \bigcup M_T(\mathfrak{A})$ is definable by $\bigvee_{i \in I} \phi_i(y)$. But Q is also definable by

$$\exists \mathbf{P}\, \exists \mathbf{R}\, (\mathbf{P}(y) \wedge \forall x[\mathbf{P}(x) \leftrightarrow \theta(\mathbf{R}, x)])\,,$$

that is, by $\exists \mathbf{R}\, \theta(\mathbf{R}, y)$. Hence the conclusion holds.

Remarks. (a) Instead of $\mathbf{R}$ we could have any sequence of new predicates, and θ could have, instead of the single free variable y, any number of free variables. Also, the dual of Theorem 2.3 holds. These facts are clear from the corresponding comments about Lemma 2.2.

(b) Notice that the formulas ϕ_i depend only on the symbols of L which actually occur in θ, and the identity. Therefore we can take $I = \omega$.

(c) In place of θ we could have an infinite conjunction of formulas θ_j, $j \in J$. In this case we take T to be the theory

$$\{\forall x[\mathbf{P}(x) \rightarrow \theta_j(\mathbf{R}, x)] : j \in J\}\,.$$

(d) Chang and Moschovakis have recently found a proof of Theorem 2.3 very different from what we have given here. Their proof also yields an improved form, announced in [4], in which the set $\{\phi_i(x) : i \in I\}$ of defining formulas may be taken to be primitive recursive.

In this connection the author wishes to mention that the formulations given here of the Main Lemma and Theorem 2.3 have been much improved by comments and suggestions of C.C.Chang.

(e) Let $\theta(\mathbf{S})$ be a formula containing a k-place predicate $\mathbf{S}$. Let $\mathbf{S}^*$ be a new $(k+1)$-place predicate. Recall the well-known equivalence

$$\models \forall z\, \exists \mathbf{S}\, \theta(\mathbf{S}) \leftrightarrow \exists \mathbf{S}^* \forall z\, \theta^*(\mathbf{S}^*)\,,$$

where θ^* is the formula resulting from θ by replacing $\mathbf{S}(t_1, ..., t_k)$ everywhere by $\mathbf{S}^*(z, t_1, ..., t_k)$, for any terms $t_1, ..., t_k$, assuming z is not bound in θ. The effect of this is to show that if $\theta(\mathbf{R}_1, ..., \mathbf{R}_k, y, z_1, ..., z_m)$ is a formula of $L(\mathbf{R}_1, ..., \mathbf{R}_k)$, then any second-order formula obtained from θ by quantifying $\mathbf{R}_1, ..., \mathbf{R}_k$, $z_1, ..., z_m$ in *any* order, provided the $\mathbf{R}_i$ are quantified existentially, is equivalent to a formula $\exists \mathbf{R}_1^* ... \exists \mathbf{R}_k^*\, \theta^*$, where θ^* is a (first-order) formula of $L(\mathbf{R}_1^*, ..., \mathbf{R}_k^*)$ with just y free. Therefore, by Remark (a), Theorem 2.3 applies also to such more general second-order formulas. This remark will be used in the next section in deriving Lemma 3.1*.

(f) Notice that the equivalence in the Main Lemma or Theorem 2.3 does not hold in general for models which are not special. Also, in Theorem 2.3 we may not allow universal second-order quantifiers to occur in addition to the existential ones. In this connection we refer the reader to Svenonius [23] for a reduction of Σ_1^1 sentences on countable models. Also, it is interesting to note that, assuming we have only predicate symbols, any second-order sentence which can be written in prenex form with only universal first-order quantifiers (but arbitrary second-order quantifiers) is logically equivalent to an infinite conjunction of universal first-order sentences (see [7] page 141). This fact depends essentially on the well-known special properties of universal first-order formulas.

§ 3. Interpolation

The interpolation theorems we give here generalize Craig's Theorem 1.2 in that we require that the implication from ψ to χ hold only for some choices of individuals, not for all. The resulting conditions are therefore second-order, but our results show that they are reducible to finitary first-order statements.

Before proceeding to the results which actually generalize Craig's theorem, we require the following, a sort of "one-sided" interpolation theorem.

Lemma 3.1. *Let* $\theta(\mathbf{R}, y)$ *be a formula of* $\mathrm{L}(\mathbf{R})$. *Then the following are equivalent:*

(i) $\models \exists y \, \forall \mathbf{R} \, \theta(\mathbf{R}, y)$;

(ii) *There is a formula* $\sigma(y)$ *of* L *such that*

$$\models \exists y \, \sigma(y) \quad \text{and} \quad \models \forall y [\sigma(y) \rightarrow \theta(\mathbf{R}, y)] \ .$$

Proof. From (ii) to (i) is obvious – any y satisfying σ works in (i) since σ does not contain R. The other direction is an easy consequence of the dual of Theorem 2.3. We know there are formulas $\phi_i(y)$ of L ($i \in I$) such that for every special model $\mathfrak{A}$

$$\mathfrak{A} \models \forall y [\forall \mathbf{R} \, \theta(\mathbf{R}, y) \leftrightarrow \bigvee_{i \in I} \phi_i(y)] \ .$$

By (i) $\mathfrak{A} \models \exists y \, \forall \mathbf{R} \, \theta$, and so there is some $i \in I$ such that $\mathfrak{A} \models \exists y \, \phi_i$. Hence every model yields $\exists y \, \phi_i$ for some $i \in I$, and so by compactness there are finitely many ϕ_i, say $\phi_1, ..., \phi_n$, such that

$$\models \exists y \, \phi_1 \vee ... \vee \exists y \, \phi_n \ .$$

So, defining $\sigma(y)$ as $\phi_1(y) \vee ... \vee \phi_n(y)$ the conditions of (ii) are satisfied.

As usual, we have simplified the statement of this result. The analogues of the remarks at the end of §2 all hold here too. In particular, the single quantifiers $\exists y$ and $\forall \mathbf{R}$ could each be replaced by sequences of similar quantifiers (Remark (a)), and θ could be an infinite disjunction (Remark (c)). The latter fact implies that if T_0 is any theory of L then the equivalence of (i) and (ii) continues to hold if we require the statements to be consequences of T_0 rather than universally valid. Although similar remarks apply to the other results of this section and will be used in applications in the next section, we will no longer explicitly mention them.

The next result is our basic generalization of Craig's interpolation theorem.

Theorem 3.2. *Let $\psi(\mathbf{R}, x, y)$ and $\chi(\mathbf{S}, x, y)$ be formulas of* L(**R**) *and* L(**S**) *respectively. Then the following are equivalent:*

(i) $\models \exists y\, \forall \mathbf{R}, \mathbf{S}\, \forall x[\psi \to \chi]$;

(ii) *There are formulas $\sigma(y)$ and $\phi(x, y)$ of* L *such that*

$$\models \exists y \sigma \quad \textit{and} \quad \models \forall x, y([\sigma \wedge \psi \to \phi] \wedge [\phi \to \chi]);$$

(iii) *There is a formula $\phi(x, y)$ of* L *such that*

$$\models \exists y\, \forall \mathbf{R}, \mathbf{S}\, \forall x([\psi \to \phi] \wedge [\phi \to \chi]).$$

Proof. (ii) implies (iii) and (iii) implies (i) are clear. To show that (i) implies (ii) we first apply Lemma 3.1, where θ is $\forall x[\psi \to \chi]$, to get a formula $\sigma(y)$ of L such that

$$\models \exists y \sigma \quad \text{and} \quad \models \forall y(\sigma \to \forall x[\psi \to \chi]).$$

The last sentence may be rewritten as

$$\models \forall x, y[\sigma \wedge \psi \to \chi],$$

and then an application of Theorem 1.2 yields the desired formula $\phi(x, y)$ of L.

Instead of appealing here to Craig's theorem we could derive it from Theorem 2.3 as follows. Assume that

$$\models \psi(\mathrm{R}, x) \to \chi(\mathrm{S}, x) .$$

This is the same as

$$\models \exists \mathrm{R}\, \psi(\mathrm{R}, x) \to \forall \mathrm{S} \chi(\mathrm{S}, x) .$$

By Theorem 2.3 there are formulas $\phi_i(x)$ of L ($i \in I$) whose conjunction is equivalent to $\exists \mathrm{R}\, \psi(\mathrm{R}, x)$ on every special $\mathfrak{A}$, and hence

$$\mathfrak{A} \models [\exists \mathrm{R}\, \psi \to \bigwedge_{i \in I} \phi_i] \wedge [\bigwedge_{i \in I} \phi_i \to \forall \mathrm{S} \chi] .$$

We may drop the superfluous second-order quantifiers, and then a compactness argument shows that some finite conjunction of the ϕ_i will interpolate between ψ and χ.

Notice that (ii) is stronger than (iii) since it yields not only an interpolating formula ϕ but also an L-definable set of y's for which the implications hold. Also, (ii) implies that $\phi \to \chi$ is true for all y. Because of this, one can find examples of an interpolating formula ϕ which will work in (iii), but which will not work in (ii) for any σ.

Also notice that the following is true: if in (i) the single implication $\psi \to \chi$ is replaced by a finite conjunction of implications $\psi_i \to \chi_i$ ($i = 1, ..., n$) then in (ii) we can find interpolating formulas ϕ_i ($i = 1, ..., n$) which all work with the same formula σ. It is this slight generalization of Theorem 3.2 which is actually used in the next section.

There are situations which vary somewhat from the one in Theorem 3.2 in which we also may interpolate. For example, there is the following, which will also be applied in the next section. (We omit the conditions corresponding to (iii).)

Theorem 3.3. *Let* $\psi(\mathbf{R}, x, y)$ *and* $\chi(\mathbf{S}, x, y)$ *be formulas of* $L(\mathbf{R})$ *and* $L(\mathbf{S})$ *respectively.* Then:

(a) *The following two conditions are equivalent:*

(i) $\models \forall \mathbf{R}\, \exists y\, \forall \mathbf{S}\, \forall x [\psi \to \chi]$;

(ii) *There are formulas* $\sigma(\mathbf{R}, y)$ *of* $L(\mathbf{R})$ *and* $\phi(x, y)$ *of* L *such that*

$$\models \exists y \sigma \quad \textit{and}$$

$$\models \forall x, y([\sigma \wedge \psi \to \phi] \wedge [\phi \to \chi]) .$$

(b) *The following two conditions are equivalent:*

(i) $\models \forall \mathbf{R}\, \exists y\, \forall \mathbf{S}\, \forall x [\chi \to \psi]$;

(ii) *There are formulas* $\sigma(\mathbf{R}, y)$ *of* $L(\mathbf{R})$ *and* $\phi(x, y)$ *of* L *such that*

$$\models \exists y\, \sigma \quad \textit{and}$$

$$\models \forall x, y([\chi \to \phi] \wedge [\sigma \wedge \phi \to \psi]) .$$

Proof. (a) is proved just like Theorem 3.2. Thus, in (i) we drop the superfluous outer quantifier and apply Lemma 3.1 to get a formula $\sigma(\mathbf{R}, y)$ of $L(\mathbf{R})$ such that

$$\models \exists y\, \sigma \quad \textit{and}$$

$$\models \forall x, y[\sigma \wedge \psi \to \chi] .$$

Applying Theorem 1.2 to this implication then yields a formula $\phi(x, y)$ of L as interpolant. (b) is the same, except that Theorem 1.2 is applied to $\chi \to [\sigma \to \psi]$.

Chang has proved a generalization of the equivalence of (i) and

(iii) in Theorem 3.2. In this generalization, announced in [3], the prefix $\exists y$ is replaced by an arbitrary quantifier prefix, allowing also second-order quantifiers provided they are universal. We subsequently succeeded in generalizing Theorem 3.2, including condition (ii), to this situation. In fact, the resulting theorem, Theorem 3.4, also generalizes Theorem 3.3 above. Our proof, which is quite different from Chang's, depends upon first giving a corresponding generalization of Lemma 3.1.

Lemma 3.1*. *Let* $\theta(\mathbf{R}_0, ..., \mathbf{R}_k, y_0, ..., y_k)$ *be a formula of* $\mathrm{L}(\mathbf{R}_0, ..., \mathbf{R}_k)$. *Then the following are equivalent:*

(i) $\models \exists y_0 \, \forall \mathbf{R}_0 \, ... \, \exists y_k \, \forall \mathbf{R}_k \, \theta$;

(ii) *There are formulas* $\sigma_i(\mathbf{R}_0, ..., \mathbf{R}_{i-1}, y_0, ..., y_i)$ *of* $\mathrm{L}(\mathbf{R}_0, ..., \mathbf{R}_{i-1})$, $i = 0, ..., k$, *such that*

$$\models \exists y_0 \, \sigma_0 \, ,$$

$$\models \forall y_0, ..., y_i [\sigma_i \rightarrow \exists y_{i+1} \, \sigma_{i+1}] \, , \qquad i = 0, ..., k-1 \, ,$$

$$\textit{and} \quad \models \forall y_0, ..., y_k [\sigma_k \rightarrow \theta] \, .$$

Proof. As before, from (ii) to (i) is easy. The other direction is proved by repeated application of Lemma 3.1. The essential point in the proof is to notice that, by Remark (e) of § 2, we may apply Lemma 3.1 to the situation in which instead of having a single block of universal second-order quantifiers, they are broken up by first-order quantifiers. So, applying this to (i) we obtain a formula $\sigma_0(y_0)$ of L such that

$$\models \exists y_0 \, \sigma_0$$

and

$$\models \forall y_0 [\sigma_0 \rightarrow \exists y_1 \, \forall \mathbf{R}_1 \, ... \, \exists y_k \, \forall \mathbf{R}_k \, \theta] \, .$$

Since $y_1, \ldots, y_k, \mathbf{R}_1, \ldots, \mathbf{R}_k$ do not appear in σ_0 this last consequence is equivalent to

$$\models \exists y_1 \, \forall \mathbf{R}_1 \ldots \exists y_k \, \forall \mathbf{R}_k \, [\sigma_0 \to \theta] \, .$$

Applying 3.1 to this, treating y_0 as a constant, we get a formula $\sigma'(\mathbf{R}_0, y_0, y_1)$ of $\mathrm{L}(\mathbf{R}_0)$ such that

$$\models \exists y_1 \, \sigma'$$

and

$$\models \forall y_1 (\sigma' \to \exists y_2 \, \forall \mathbf{R}_2 \ldots \exists y_k \, \forall \mathbf{R}_k [\sigma_0 \to \theta]) \, ,$$

that is

$$\models \forall y_1 [\sigma_0 \wedge \sigma' \to \exists y_2 \, \forall \mathbf{R}_2 \ldots \exists y_k \, \forall \mathbf{R}_k \, \theta] \, .$$

So, defining σ_1 as $\sigma_0 \wedge \sigma'$ we have

$$\models \exists y_0 \sigma_0 \, , \qquad \models \forall y_0 [\sigma_0 \to \exists y_1 \, \sigma_1] \, ,$$

and

$$\models \forall y_0, y_1 [\sigma_1 \to \exists y_2 \, \forall \mathbf{R}_2 \ldots \exists y_k \, \forall \mathbf{R}_k \, \theta] \, .$$

Continuing in this fashion we get $\sigma_0, \sigma_1, \ldots, \sigma_k$ satisfying (ii).

From this the desired generalization of Theorem 3.2 follows easily. Chang's theorem of [3] is the equivalence of (i) and (iii) (without the further condition on ϕ).

Theorem 3.4. *Let* $\psi(y_0, \ldots, y_k, x)$ *and* $\chi(y_0, \ldots, y_k, x)$ *be formulas of* $\mathrm{L}(\mathbf{R}_0, \ldots, \mathbf{R}_k, \mathbf{S}_1)$ *and* $\mathrm{L}(\mathbf{R}_0, \ldots, \mathbf{R}_k, \mathbf{S}_2)$ *respectively. Then the following are equivalent:*

(i) $\models \exists y_0 \, \forall \mathbf{R}_0 \ldots \exists y_k \, \forall \mathbf{R}_k \, \forall \mathbf{S}_1, \mathbf{S}_2 \, \forall x [\psi \to \chi]$;

(ii) *There are formulas* $\sigma_i(y_0, \ldots, y_i)$ *of* $\mathrm{L}(\mathbf{R}_0, \ldots, \mathbf{R}_{i-1})$,

$i = 0, ..., k$, and a formula $\phi(y_0, ..., y_k, x)$ of $\mathrm{L}(\mathbf{R}_0, ..., \mathbf{R}_k)$ such that

$$\models \exists y_0\, \sigma_0 ,$$

$$\models \forall y_0, ..., y_i [\sigma_i \to \exists y_{i+1}\, \sigma_{i+1}] , \qquad i = 0, ..., k-1 ,$$

and

$$\models \forall y_0, ..., y_k\, \forall x([\sigma_k \wedge \psi \to \phi] \wedge [\phi \to \chi]) ;$$

(iii) *There is a formula $\phi(y_0, ..., y_k, x)$ of $\mathrm{L}(\mathbf{R}_0, ..., \mathbf{R}_k)$ such that*

$$\models \exists y_0\, \forall \mathbf{R}_0 ... \exists y_k\, \forall \mathbf{R}_k\, \forall \mathbf{S}_1, \mathbf{S}_2\, \forall x([\psi \to \phi] \wedge [\phi \to \chi])$$

Furthermore, in (ii) *and* (iii) *ϕ contains $\mathbf{R}_i$ only if χ does ($i = 0, ..., k$).*

Proof. The only direction requiring proof is from (i) to (ii). First apply Lemma 3.1* to get $\sigma_0, ..., \sigma_k$ as in (ii). Then σ_k is such that

$$\models \sigma_k \to \forall x[\psi \to \chi] , \quad \text{i.e.,} \quad \models \sigma_k \wedge \psi \to \chi .$$

We apply Craig's Theorem 1.2 to this implication to get the interpolating formula ϕ. Then certainly ϕ belongs to $\mathrm{L}(\mathbf{R}_0, ..., \mathbf{R}_k)$, and moreover ϕ contains only the predicate symbols occurring in both $\sigma_k \wedge \psi$ and χ. In particular, then, ϕ contains $\mathbf{R}_i$ only if χ does.

Remarks

(a) One can also give a version of Theorem 3.4 in which one instead requires that ϕ contains $\mathbf{R}_i$ only if ψ does. Conditions (i) and (iii) remain the same, and (ii) is altered by attaching σ_k to ϕ rather than ψ in the last assertion. Theorem 3.3(a) is an instance of the original version of Theorem 3.4, and Theorem 3.3(b) is an instance of this other version.

(b) The simplifications made in the statement of Theorem 3.4, to make it easier to state, may have made it harder to see what it does say in many instances. Especially note that, in general, universal first-order quantifiers could appear in the prefix (which we rendered without them as $\exists y_0 \, \forall \mathbf{R}_0 \ldots \exists y_k \, \forall \mathbf{R}_k$); we have simply imagined them as collapsed with the universal second-order quantifiers. In these cases it may not be obvious what the best corresponding condition (ii) is. The following example, however, should make the procedure clear.

If $\psi(\mathbf{S}_1, y_0, y_1, z, x)$ and $\chi(\mathbf{S}_2, y_0, y_1, z, x)$ are formulas of $\mathrm{L}(\mathbf{S}_1)$ and $\mathrm{L}(\mathbf{S}_2)$ respectively, then the following are equivalent:

(i) $\models \exists y_0 \, \forall z \, \exists y_1 \, \forall \mathbf{S}_1, \mathbf{S}_2 \, \forall x [\psi \to \chi]$;

(ii) There are formulas $\sigma(y_0, y_1, z)$ and $\phi(y_0, y_1, z, x)$ of L such that

$$\models \exists y_0 \, \forall z \, \exists y_1 \sigma$$

and

$$\models \forall y_0, z, y_1, x([\sigma \wedge \psi \to \phi] \wedge [\phi \to \chi]) .$$

That (ii) implies (i) is, as usual, obvious. We see that (i) implies (ii) by applying Theorem 3.4 while treating z as a constant. This yields a formula $\sigma_0(y_0)$ of L (not containing z) and formulas $\sigma_1(y_0, y_1, z)$ and $\phi(y_0, y_1, z, x)$ of L (but also containing z) such that

$$\models \exists y_0 \, \sigma_0 , \qquad \models \forall y_0 [\sigma_0 \to \exists y_1 \, \sigma_1] ,$$

and

$$\models \forall y_0, y_1([\sigma_1 \wedge \psi \to \phi] \wedge [\phi \to \chi]) .$$

Because z does not occur in σ_0 the second statement leads to

$$\models \forall y_0 [\sigma_0 \to \forall z \, \exists y_1 \, \sigma_1] .$$

Hence, if we define $\sigma(y_0, y_1, z)$ to be $\sigma_0 \wedge \sigma_1$, condition (ii) above is satisfied.

The remaining remarks concern possible improvements and extensions of the results of this section.

(c) One question concerning Theorem 3.4 which arises immediately is whether one can require (at least in (iii)) that the interpolant ϕ contains only the non-logical constants occurring in *both* ψ and χ. The following example shows that in general this improvement is not possible.

Let L have only identity. Let θ be the sentence of L saying that the universe has exactly three elements, let $\tau_0(\mathbf{P})$ be the sentence of L(**P**) syaing that **P** has exactly one element, and let $\tau_1(\mathbf{Q})$ be the sentence of L(**Q**) saying that **Q** has exactly two elements. Then the following holds:

$$\models \forall \mathbf{P}, \mathbf{Q}\, \exists y\, \forall x (\theta \wedge \tau_0(\mathbf{P}) \wedge [\mathbf{P}(y) \leftrightarrow \mathbf{P}(x)]$$

$$\rightarrow [\tau_1(\mathbf{Q}) \rightarrow \mathbf{Q}(x)]) \, .$$

In fact, given $(\mathcal{A}, P, Q) \models \theta \wedge \tau_0(\mathbf{P}) \wedge \tau_1(\mathbf{Q})$ we may choose $y \in P$ if $P \subseteq Q$, $y \notin P$ otherwise (in which case $A - P \subseteq Q$); and these are the *only* choices of y satisfying the implication for all x.

If the above conjecture were true, we could interpolate a formula $\phi(x, y)$ of L. Up to equivalence with respect to θ there are only four possibilities: ϕ is either $x = y$, $x \neq y$, logically true, or logically false. But, given $(\mathfrak{A}, P, Q) \models \theta \wedge \tau_0(\mathbf{P}) \wedge \tau_1(\mathbf{Q})$ and choosing y such that the above holds, the consequent of the implication is false for some x, the antecedent is true when $x = y$, and also the antecedent is true for some x different from y (in the case $y \notin P$). Therefore no such formula ϕ can interpolate in the above implication.

(d) A second question which arises is whether one can allow existential second-order quantifiers in the prefix and still get a first-order interpolant ϕ (at least in (iii)). The following example, which looks forward to the definability applications of the next section, shows that this also is generally not possible.

Let L have just identity, and let **R** be a binary predicate symbol. Let $\theta(\mathbf{R})$ be the sentence of $L(\mathbf{R})$ saying that **R** is a discrete linear order of the universe. Let $\sigma(x, y)$ be the formula of $L(\mathbf{R})$ saying that x is the immediate predecessor of y in the ordering **R**. Let $\tau(\mathbf{R}, \mathbf{P})$ be the following sentence of $L(\mathbf{R}, \mathbf{P})$:

$$\forall x[\mathbf{P}(x) \leftrightarrow (\neg \exists z\, \sigma(z, x)$$

$$\vee \exists z, y[\mathbf{P}(z) \wedge \sigma(z, y) \wedge \sigma(y, x)])] .$$

Thus, $\tau(\mathbf{R}, \mathbf{P})$ says that **P** is the set of "even-numbered" elements in the ordering **R**. For any $(\mathfrak{A}, R) \models \theta(\mathbf{R})$ there is some $P \subseteq A$ such that $(\mathfrak{A}, R, P) \models \tau(\mathbf{R}, \mathbf{P})$; but if R is, for example, a well-order then there is exactly one P such that $(\mathfrak{A}, R, P) \models \tau(\mathbf{R}, \mathbf{P})$. Hence the following holds:

$$\models \exists \mathbf{R} \forall \mathbf{P}, \mathbf{Q} \forall x([\theta(\mathbf{R}) \rightarrow \tau(\mathbf{R}, \mathbf{P}) \wedge \mathbf{P}(x)]$$

$$\rightarrow [\tau(\mathbf{R}, \mathbf{Q}) \rightarrow \mathbf{Q}(x)]) .$$

More precisely, this implication holds for R if and only if R satisfies $\theta(\mathbf{R})$ and there is no infinite descending sequence $\{a_k : k \in \omega\}$ such that a_{k+1} is the immediate predecessor of a_k. Now, if we could get an interpolating formula $\phi(\mathbf{R}, x)$ of $L(\mathbf{R})$, this would define the unique P satisfying $\tau(\mathbf{R}, \mathbf{P})$ for some R for which such a P is unique. But whenever A is infinite and $(\mathfrak{A}, R, P) \models \theta(\mathbf{R}) \wedge \tau(\mathbf{R}, \mathbf{P})$ we can find $(\mathfrak{A}', R', P')$ such that $(\mathfrak{A}, R, P) \equiv (\mathfrak{A}', R', P')$, but

$$|\{P'' \subseteq A' : (\mathfrak{A}', R', P') \cong (\mathfrak{A}', R', P'')\}| = 2^{|A'|}$$

Therefore no such P can be first-order definable, and so interpolation fails.

(e) One sort of improvement in interpolation results which we have not considered is that of relating the syntactical form of the interpolant ϕ to that of the formulas ψ and χ between which it interpolates. However, by combining the known results of that kind with ours, we also obtain some improvements in that line.

For example, Lyndon [12] proved that Theorem 1.2 could be improved by adding to (ii) that

(1) a predicate which occurs positively (negatively) in ϕ also occurs positively (negatively) in both ψ and χ.

(See [12] for the definition of positive and negative.) Using this we can obtain, for example, an improvement of Theorem 3.2 by adding to (ii)

(2) a predicate which occurs positively (negatively) in ϕ also occurs positively (negatively) in both $\sigma \wedge \psi$ and χ; a predicate which occurs positively (negatively) in σ either occurs negatively (positively) in ψ or positively (negatively) in χ.

In fact, if σ and ϕ satisfy (ii) of Theorem 3.2, then the implications $\sigma \to [\psi \to \chi]$, $\sigma \wedge \psi \to \chi$, and $\phi \to \chi$ are all valid. Therefore if σ and ϕ do not satisfy (2) we could apply Lyndon's theorem (1) to get formulas which do.

We do not know whether any better, or basically different, refinements of this sort are possible.

(f) Notice that in the example given in Remark (d) we can interpolate the following infinitary formula $\phi(\mathbf{R}, x)$:

$$\neg \exists z\, \sigma(z, x) \vee \bigvee_{n \in \omega} \exists y_0, \ldots, y_n, z_0, \ldots, z_n \, (\neg \exists z\, \sigma(z, y_0)$$

$$\wedge \bigwedge_{0 \leq i \leq n} [\sigma(y_i, z_i) \wedge \sigma(z_i, y_{i+1})] \wedge \sigma(z_n, x)) \, .$$

Can this always be done? That is, given $\psi(\mathbf{R}, \mathbf{S}_1, x)$ and $\chi(\mathbf{R}, \mathbf{S}_2, x)$, finitary first-order formulas such that

$$\models \exists \mathbf{R} \forall \mathbf{S}_1, \mathbf{S}_2 \forall x [\psi \to \chi] ,$$

can we find some infinitary (first-order) formula $\phi(\mathbf{R}, x)$ (not containing $\mathbf{S}_1, \mathbf{S}_2$) such that

$$\models \exists \mathbf{R} \forall \mathbf{S}_1, \mathbf{S}_2 \forall x([\psi \to \phi] \wedge [\phi \to \chi])?$$

By "infinitary" we intend primarily one of the classical languages $L_{\kappa\lambda}$ or something similar; in the above example ϕ belongs to $L_{\omega_1\omega}$. (For information and further references on infinitary languages see [21] and the volume in which [10] appears.) We should remark, however, that the analogous generalization of Lemma 3.1 (that is, replacing $\exists y$ by a second-order existential quantifier and finding an infinitary σ) is false.

§ 4. Definability

The main result of this section is the following generalization of Beth's theorem, which will be derived as an application of Theorem 3.2.

Theorem 4.1. *For any T and any positive $n \in \omega$ the following are equivalent:*

$(\mathrm{i})_n$ *For every* $\mathfrak{A}$, $|M_T(\mathfrak{A})| \leq n$;

$(\mathrm{ii})_n$ *There are formulas* $\sigma(v_1, ..., v_k)$ *and* $\phi_i(x, v_1, ..., v_k)$, $1 \leq i \leq n$, *of* L *such that*

$$T \models \exists v_1, ..., v_k\, \sigma$$

and

$$T \models \forall v_1, ..., v_k\, (\sigma \to \bigvee_{1 \leq i \leq n} \forall x [\mathbf{P}(x) \leftrightarrow \phi_i])\,.$$

Proof. It is easy to see that $(\mathrm{ii})_n$ implies $(\mathrm{i})_n$ – for any $\mathfrak{A}$ pick $a_1, ..., a_k \in A$ satisfying σ; then any $P \in M_T(\mathfrak{A})$ must be one of sets defined by $\phi_i(x, a_1, ..., a_k)$, $i = 1, ..., n$, and so $|M_T(\mathfrak{A})| \leq n$. To prove that $(\mathrm{i})_n$ implies $(\mathrm{ii})_n$ we first apply compactness and so assume that T is given by a single sentence $\tau(\mathbf{P})$ of $\mathrm{L}(\mathbf{P})$. (Actually this is not necessary if one is willing to use the infinitary form of the interpolation results.) What we will do is show, assuming $(\mathrm{i})_n$ holds, that

(*) there are formulas $\psi_i(\mathbf{P}, v_1, ..., v_k)$ of $\mathrm{L}(\mathbf{P})$, $1 \leq i \leq n$, such that

$$\models \exists v_1, ..., v_k\, \forall \mathbf{P}, \mathbf{Q}\, \forall x ([\tau(\mathbf{P}) \to \bigvee_{1 \leq i \leq n} \psi_i(\mathbf{P})]$$

$$\wedge \bigwedge_{1 \leq i \leq n} ([\tau(\mathbf{P}) \wedge \psi_i(\mathbf{P}) \wedge \mathbf{P}(x)]$$

$$\to [\tau(\mathbf{Q}) \wedge \psi_i(\mathbf{Q}) \to \mathbf{Q}(x)]))\,.$$

(*) says that for any $\mathfrak{A}$ there are points $a_1, ..., a_k \in A$ which divide the sets in $M_T(\mathfrak{A})$ into n parts by means of $\psi_1, ..., \psi_n$, and each such part has at most one member.

Once we have (*) we apply Theorem 3.2 to obtain formulas $\sigma(v_1, ..., v_k)$ and $\phi_i(x, v_1, ..., v_k)$ of L, $1 \leq i \leq n$, such that

$$\models \exists v_1, ..., v_k\, \sigma\,, \qquad T \models \forall v_1, ..., v_k [\sigma \to \bigvee_{1 \leq i \leq n} \psi_i(\mathbf{P})]\ ,$$

$$T \models \forall v_1, ..., v_k\, \forall x [\sigma \wedge \psi_i(\mathbf{P}) \wedge \mathbf{P}(x) \to \phi_i]\ , \quad i = 1, ..., n\,,$$

and

$$T \models \forall v_1, ..., v_k\, \forall x(\phi_i \to [\psi_i(\mathbf{P}) \to \mathbf{P}(x)])\,, \quad i = 1, ..., n,$$

changing $\mathbf{Q}$ to $\mathbf{P}$ in the last line. The last two lines may be combined to yield

$$T \models \forall v_1, ..., v_k(\sigma \wedge \psi_i(\mathbf{P}) \to \forall x [\mathbf{P}(x) \longleftrightarrow \phi_i])\,,$$

$$i = 1, ..., n\ ,$$

which together with the second line gives us

$$T \models \forall v_1, ..., v_k(\sigma \to \bigvee_{1 \leq i \leq n} \forall x [\mathbf{P}(x) \longleftrightarrow \phi_i])\,.$$

Thus the theorem is established once we have (*).

We show (*) as follows. For any $\mathfrak{A}$, $|M_T(\mathfrak{A})| \leq n$, so let $P_1, ..., P_n$ be a list including all the sets in $M_T(\mathfrak{A})$. We may assume that

$$P_n \not\subseteq P_i \text{ for each } i = 1, ..., n-1\ .$$

Let $a_i \in P_n - P_i$ for $i = 1, ..., n-1$ and let $\psi_1(\mathbf{P}, v_1, ..., v_{n-1})$ be

$\mathbf{P}(v_1) \wedge \ldots \wedge \mathbf{P}(v_{n-1})$. Then P_n is the only set $P \in M_T(\mathfrak{A})$ such that

$$(\mathfrak{A}, P) \models \psi_1(\mathbf{P}, a_1, \ldots, a_{n-1}) .$$

Similarly we may assume that

$$P_{n-1} \not\subseteq P_i \text{ for each } i = 1, \ldots, n-2 ,$$

and find $a_{n-1+i} \in P_{n-1} - P_i$ for $i = 1, \ldots, n-2$. Then, defining $\psi_2(\mathbf{P}, v_1, \ldots, v_{2n-3})$ to be $\neg\psi_1 \wedge \mathbf{P}(v_n) \wedge \ldots \wedge \mathbf{P}(v_{2n-3})$, we know that P_{n-1} is the only set $P \in M_T(\mathcal{A})$ such that

$$(\mathfrak{A}, P) \models \psi_2(\mathbf{P}, a_1, \ldots, a_{2n-3}) .$$

Continuing in this fashion we obtain all ψ_i, $i = 1, \ldots, n$. It is then clear from their definitions that (*) holds. Therefore the theorem is proved.

Note that for $n = 1$ this does give Beth's Theorem 1.1, and also that the proof works for T a theory in any language containing $L(\mathbf{P})$.

The problem of finding an equivalent to $(\mathrm{i})_n$ was first raised by Craig in [5]. Later he and Daigneault considered this question and formulated a condition, similar to our $(\mathrm{ii})_n$ but much more complicated, which they proved equivalent to $(\mathrm{i})_n$ for theories T in $L(\mathbf{P})$. Their methods, which did not work for theories T in arbitrary languages containing $L(\mathbf{P})$, are much different and much more involved than those used here or in §6 below. The author is grateful to Professor Craig for sending him an account of their (unpublished) work.

Just as Svenonius' theorem may be derived from Beth's, a corresponding generalization of Svenonius' theorem may be derived from Theorem 4.1. The semantic condition concerns $M(\mathfrak{A}, P)$ rather than $M_T(\mathfrak{A})$, and the syntactical condition states that some finite disjunction of the corresponding conditions of Theorem 4.1

is a consequence of T. Combining these two results we obtain the following, which, like the Chang-Makkai theorem, is a generalization of both Beth's and Svenonius' theorems.

Theorem 4.2. *For any T the following are equivalent:*
(i) *For every* $\mathfrak{A}$, $|M_T(\mathfrak{A})| < \omega$;
(ii) *For every model* $(\mathfrak{A}, P)$ *of* T, $|M(\mathfrak{A}, P)| < \omega$;
(iii) *For some* n, *condition* (ii)$_n$ *of Theorem 4.1 holds.*

A natural question to ask is whether the individual parameters $v_1, \ldots, v_k$ are necessary in (ii)$_n$ of Theorem 4.1, or whether a disjunction of explicit definitions (as in Svenonius' theorem) would suffice. Equivalently, this asks whether the condition that $|M_T(\mathfrak{A})| < \omega$ for all $\mathfrak{A}$ implies that $|M(\mathfrak{A}, P)| = 1$ for every model $(\mathfrak{A}, P)$ of T. In general, the answer to this question is no, as is shown by the following example.

Assume that L has just a binary predicate **E**. Let T be the theory in L(**P**) which says that **E** is an equivalence relation which divides the universe into two infinite equivalence classes, and **P** is one of these equivalence classes. Then T is a complete theory satisfying the conditions in Theorem 4.1 with $n = 2$, but **P** cannot be defined without parameters (there are models $(\mathfrak{A}, P)$ of T such that $|M(\mathfrak{A}, P)| = 2$). A definition of **P** with parameters is given by

$$T \models \forall v_1 (\forall x [\mathbf{P}(x) \leftrightarrow \mathbf{E}(x, v_1)]$$
$$\vee \forall x [\mathbf{P}(x) \leftrightarrow \neg \mathbf{E}(x, v_1)]) .$$

However, there is a large class of theories T for which the parameters are not necessary. This is the case whenever T satisfies the following "choice" condition:

(C) For every formula $\theta(x)$ of L there is some formula $\psi(x)$ of L such that

$$T \models \exists x\, \theta \rightarrow \exists !x(\theta \wedge \psi)\,.$$

For, if T satisfies (C) and also $(\mathrm{ii})_n$ of Theorem 4.1, then repeated applications of (C) to σ yield a formula σ' of L which picks out exactly one k-tuple of elements satisfying σ. It follows that whenever $(\mathfrak{A}, P)$ is a model of T, P must be definable by one of the formulas $\exists v_1, ..., v_k\, [\sigma' \wedge \phi_i]$, $i = 1, ..., n$, which have no parameters. In this case, then, the conditions in Theorem 4.2 and Svenonius' theorem are all equivalent.

Theorem 4.3. *If T satisfies (C), then the conditions in Theorem 1.3 may be added to the list of equivalent conditions in Theorem 4.2*

It should be noted that whether or not T satisfies (C) depends only on the consequences of T in L. So if T_0 is any theory of L satisfying (C), then any extension of T_0 in any language also satisfies (C). This is the case, for example, when T_0 is Peano arithmetic.

Before going on to a few other applications, let us survey some of the kinds of model-theoretic condition concerning a predicate **P** which are equivalent to some condition stating the explicit definability of **P**. In what follows we drop our notation about $M_T(\mathfrak{A})$ and instead write the conditions as second-order sentences; τ here is a first-order formula which varies from condition to condition. The equivalent definability conditions have either been given previously or are easily obtainable from what we have done previously.

Thus, Beth's theorem gives a definability condition equivalent to

$$(1) \qquad \models \exists^{\leq 1} \mathbf{P} \tau\,.$$

The immediate effect of our interpolation Theorem 3.2 is to enable us to give a definability equivalent also to

(2) $\models \exists y\, \exists^{\leq 1}\, \mathbf{P}\tau$.

But we showed in Theorem 4.1 that using a variant of (2) we actually find a condition equivalent to

(3) $\models \exists^{\leq n}\, \mathbf{P}\tau$.

Using Theorem 3.4 instead of 3.2 we get a definability equivalent to

(4) $\models \exists y_0\, \forall \mathbf{R}_0 \ldots \exists y_k\, \forall \mathbf{R}_k\, \exists^{\leq n}\, \mathbf{P}\tau$.

Craig's improvement in Beth's theorem may be expressed here as generalizing (1) to

(1′) $\models \exists^{\leq 1}\, \mathbf{P}\, \exists \mathbf{S}\, \tau$,

and adding that the definition of **P** does not involve **S**. More generally, recalling Remark (e) from §2, we can replace $\exists \mathbf{S}$ by any sequence of quantifiers, provided the second-order quantifiers are all existential. In particular, then, (4) becomes

(4′) $$\models (\exists y_0\, \forall \mathbf{R}_0 \ldots \exists y_k\, \forall \mathbf{R}_k)\, \exists^{\leq n}\, \mathbf{P}$$
$$(\forall z_0\, \exists \mathbf{S}_0 \ldots \forall z_m\, \exists \mathbf{S}_m)\tau ,$$

and in the corresponding definability condition the definition does not involve $z_0, \ldots, z_m, \mathbf{S}_0, \ldots, \mathbf{S}_m$.

By combining these arguments with those of the Chang-Makkai theorem we obtain corresponding results for $\exists^{\leq |A|}\, \mathbf{P}$ ($\mathfrak{A}$ infinite) in place of $\exists^{\leq n}\, \mathbf{P}$. Also, all of these results have corresponding Svenonius forms.

Note the following:

(a) A standard sort of compactness argument shows that for any T, either $|M_T(\mathfrak{A})| < \omega$ for every $\mathfrak{A}$, or there is some infinite $\mathfrak{A}$ such that $|M_T(\mathfrak{A})| \geq |A|$; in the latter case, in fact, there is an infinite model $(\mathfrak{A}, P)$ of T such that $|M(\mathfrak{A}, P)| \geq |A|$. Therefore no cardinality restrictions on $M_T(\mathfrak{A})$ (or on $M(\mathfrak{A}, P)$ for $P \in M_T(\mathfrak{A})$) other than those in the previous theorems and the Chang-Makkai theorem can possibly hold. It follows that those are also the only possible cardinality restrictions on **P** in (1)–(4′) above. Hence they cannot be further generalized by altering the cardinality condition.

(b) The example in Remark (d) of §3 shows that $\models \exists \mathbf{R} \exists^{\leq 1} \mathbf{P}\, \tau$ need not imply that **P** is first-order definable (in terms of some **R**). Hence (4) and (4′) cannot be generalized by allowing existential second-order quantifiers to the left of **P**. One may similarly show that universal second-order quántifiers cannot be allowed to the right of **P**. However, the question of the infinitary definability of **P** in these cases, as raised in Remark (f) of the last section, is open.

In general it seems that obtaining further definability results requires looking at different types of conditions than those above. Indeed, very many different types of results follow simply from other applications of the interpolation theorems. Without attempting any comprehensive survey of such results, we give here two interesting and related examples of such applications.

The first result is an apparently new consequence of Craig's interpolation theorem.

Theorem 4.4. *For any T the following are equivalent:*
(i) *For every $\mathfrak{A}$, if $P, P' \in M_T(\mathfrak{A})$ and $P \neq P'$ then $P \cap P' = 0$;*
(ii) *There is a formula $\phi(x, y)$ of* L *such that*

$$T \models \forall y(\mathbf{P}(y) \to \forall x[\mathbf{P}(x) \leftrightarrow \phi(x, y)]) .$$

Proof. ϕ is obtained from Theorem 1.2 as the interpolating formula for the valid implication

$$\models [T(\mathbf{P}) \wedge \mathbf{P}(y) \wedge \mathbf{P}(x)] \to [T(\mathbf{Q}) \wedge \mathbf{Q}(y) \to \mathbf{Q}(x)] \ .$$

The other result, an application of Theorem 3.3, also deals with a particular way in which **P** can be definable with parameters.

Theorem 4.5. *For any T the following are equivalent:*
(i) *For every $\mathfrak{A}$, if $P \in M_T(\mathfrak{A})$ then there is some $a \in P$ such that P is the only set in $M_T(\mathfrak{A})$ containing a;*
(ii) *There is a formula $\phi(x, y)$ of* L *such that*

$$T \models \exists y (\mathbf{P}(y) \wedge \forall x [\mathbf{P}(x) \longleftrightarrow \phi(x, y)])$$

and

$$T \models \forall x, y [\mathbf{P}(x) \wedge \mathbf{P}(y) \to \phi(x, y)] \ .$$

Proof. We first show that (ii) implies (i). Assuming (ii), let $P \in M_T(\mathfrak{A})$ and let a be a point in P which defines P as in the first line of (ii). Assume that $a \in P' \in M_T(\mathfrak{A})$. Then by the second line of (ii) we have

$$(\mathfrak{A}, P') \models \forall x [\mathbf{P}(x) \to \phi(x, a)]$$

and so $P' \subseteq P$. Repeating the argument with a point $a' \in P'$ we find that also $P \subseteq P'$, and so $P = P'$, which shows (i).

To show that (i) implies (ii) notice that (i) implies

$$(1) \qquad \models \forall \mathbf{P}\, \exists y\, \forall \mathbf{Q}\, \forall x [[T(\mathbf{P}) \to \mathbf{P}(y)]$$
$$\wedge ([T(\mathbf{Q}) \wedge \mathbf{Q}(y) \wedge \mathbf{Q}(x)] \to [T(\mathbf{P}) \to \mathbf{P}(x)])] \ .$$

So applying Theorem 3.3(b) we get formulas $\sigma(\mathbf{P}, y)$ of $\mathrm{L}(\mathbf{P})$ and $\phi(x, y)$ of L such that

(2) (a) $T \models \exists y\, \sigma \wedge \forall y[\sigma \to \mathbf{P}(y)]$,

(b) $T \models \forall x, y[\mathbf{P}(y) \wedge \mathbf{P}(x) \to \phi(x, y)]$

(after changing $\mathbf{Q}$ to $\mathbf{P}$), and

(c) $T \models \forall x, y[\sigma(\mathbf{P}, y) \wedge \phi(x, y) \to \mathbf{P}(x)]$.

Combining (2) (a) and (c) we get

$$T \models \exists y(\mathbf{P}(y) \wedge \forall x[\phi(x, y) \to \mathbf{P}(x)]) ,$$

which because of (2) (b) yields

$$T \models \exists y(\mathbf{P}(y) \wedge \forall x[\phi(x, y) \leftrightarrow \mathbf{P}(x)]) ,$$

which completes the proof.

Finally, we also have the following result which, although not explicitly mentioning definability, is a consequence of Theorem 4.1.

Theorem 4.6. *Let T be a theory of* L(**P**). *Let $\mathfrak{A}$ be such that $M_T(\mathfrak{A}) = 0$ but for some $\mathfrak{A}' \equiv \mathfrak{A}$, $M_T(\mathfrak{A}') \neq 0$. Then there is some $\mathfrak{B} \equiv \mathfrak{A}$ such that*

$$|M_T(\mathfrak{B})| \geq |B| \geq \omega .$$

Proof. Let T_0 be the set of all sentences of L true on $\mathfrak{A}$. Let $\mathfrak{B}$ be a special model of T_0 (that is, a special model elementarily equivalent to $\mathfrak{A}$). If $\mathfrak{B}$ does not satisfy the conclusion of the theorem, then for every model $\mathfrak{A}'$ of T_0, $|M_T(\mathfrak{A}')| < \omega$. Therefore we can apply Theorem 4.1 to $T_0 \cup T$ to obtain formulas $\sigma(v_1, ..., v_k)$ and $\phi_i(x, v_1, ..., v_k)$ of L ($i = 1, ..., n$) such that

$$T_0 \cup T \models \exists v_1, \ldots, v_k\, \sigma$$

and

$$T_0 \cup T \models \forall v_1, \ldots, v_k \,(\sigma \rightarrow \bigvee_{1 \leq i \leq n} \forall x [\mathbf{P}(x) \leftrightarrow \phi_i])$$

The hypothesis that $M_T(\mathfrak{A}') \neq 0$ for some $\mathfrak{A}' \equiv \mathfrak{A}$ implies that $M_T(\mathfrak{B}) \neq 0$ (since $\mathfrak{B}$ is relation-universal). Let $P_0 \in M_T(\mathfrak{B})$, so that $(\mathfrak{B}, P_0) \models T_0 \cup T$. $\mathfrak{A} \models \exists v_1, \ldots, v_k \sigma$ since $\mathfrak{A} \equiv \mathfrak{B}$, so let $a_1, \ldots, a_k \in A$ satisfy σ in $\mathcal{A}$. Since $\mathfrak{B}$ is special there are $b_1, \ldots, b_k \in B$ such that

$$(+) \qquad (\mathfrak{B}, b_1, \ldots, b_k) \equiv (\mathfrak{A}, a_1, \ldots, a_k)\,.$$

In particular, $b_1, \ldots, b_k$ satisfy σ, so for some i_0

$$(\mathfrak{B}, P_0) \models \forall x [\mathbf{P}(x) \leftrightarrow \phi_{i0}(x, b_1, \ldots, b_k)]$$

Define $P = \{a \in A : \mathfrak{A} \models \phi_{i0}(a, a_1, \ldots, a_k)\}$. Then, by (+), $(\mathfrak{A}, P) \equiv (\mathfrak{B}, P_0)$ and therefore $(\mathfrak{A}, P) \models T$, contradicting our hypothesis that $M_T(\mathfrak{A}) = 0$. Therefore $\mathfrak{B}$ satisfies the conclusion of the theorem.

Notice that this result clearly fails if T is allowed to be a theory in a language containing L(**P**). The theorem with the weaker conclusion that $|M_T(\mathfrak{B})| \geq 2$ is an immediate consequence of Beth's theorem. It is given and applied by K.L.de Bouvère in his book [24].

For other sorts of results the reader is also referred to Chang [2], which includes some theorems on the comparability of the sets in $M_T(\mathfrak{A})$. Some results of [2] are new consequences of Craig's theorem, and others (including the Chang-Makkai theorem) follow from Chang's Main Theorem, which may be considered as a sort of interpolation theorem.

§ 5. Intersection of elementary submodels

Let $\mathfrak{A}$ be a model for L and let $Y \subseteq A$. Then by $\mathfrak{A} \upharpoonright Y$ we mean the submodel of $\mathfrak{A}$ whose universe is Y (if our language has functions then we use submodel in the weak sense in which a submodel need not be closed under the functions of the language). We define $D(\mathfrak{A}, Y)$ to be the set of all $a \in A$ such that for some $\phi(x, v_1, ..., v_k)$ of L and some $b_1, ..., b_k \in Y$ we have, for some n,

$$\mathfrak{A} \models \phi(a, b_1, ..., b_k) \wedge \exists^{\leq n} x\, \phi(x, b_1, ..., b_k) .$$

Thus, $D(\mathfrak{A}, Y)$ is the set of all points of A definable in the sense of Theorem 4.1 from points of Y.

Notice that if $\mathfrak{A} \upharpoonright Y \prec \mathfrak{A}$ then $D(\mathfrak{A}, Y) = Y$, but that the converse fails in general. Those theories for whose models the converse holds are characterized in Theorem 5.3.

Let T be the theory of L(**P**) such that $(\mathfrak{A}, P) \models T$ if and only if $\mathfrak{A} \upharpoonright P \prec \mathfrak{A}$. Then, applying the Main Lemma to this theory T we obtain

Lemma 5.1. *Let $\mathfrak{A}$ be special. Then the intersection of all elementary submodels of $\mathfrak{A}$ has universe $D(\mathfrak{A}, 0)$.*

Proof. The universe of this intersection is precisely $\bigcap M_T(\mathfrak{A})$, by the definition of T. By the dual of the Main Lemma there are formulas $\phi_i(x)$ of L $(i \in I)$ such that

$$\bigcap M_T(\mathfrak{A}) = \{a \in A : \mathfrak{A} \models \bigvee_{i \in I} \phi_i(a)\} .$$

Now, it is easy to see that $D(\mathfrak{A}, 0) \subseteq \bigcap M_T(\mathfrak{A})$, so to get equality it will be enough to show that for each $i \in I$, $\mathfrak{A} \models \exists^{\leq n} x\, \phi_i(x)$ for some n.

Since the formulas ϕ_i do not depend on the particular special model in question we may assume that $|A| > |L| \cup \omega$. Then, by the downward Löwenheim-Skolem theorem we see that

$$|\bigcap M_T(\mathfrak{A})| < |A| ,$$

and so

$$|\{a \in A : \mathfrak{A} \models \phi_i(a)\}| < |A| \text{ for each } i \in I .$$

But a definable set in a special model either has the power of the model or is finite (since special models are universal; cf. [15] Theorem 3.7). Hence for each i there is some n such that $\mathfrak{A} \models \exists^{\leq n} x\, \phi_i(x)$, which proves the Lemma.

As an almost immediate consequence of this Lemma we derive a theorem of Park characterizing the sets $Y \subseteq A$ closed under the above notion of definability.

Theorem 5.2 (Park [16, 17]). *Let $Y \subseteq A$. The following are equivalent:*

(i) $D(\mathfrak{A}, Y) = Y$;

(ii) *There is some $\mathfrak{B}$ and some collection $\{\mathfrak{B}_j : j \in J\}$ of elementary submodels of $\mathfrak{B}$ such that*

$$\mathfrak{A} \prec \mathfrak{B} \quad \textit{and} \quad \mathfrak{B} \upharpoonright Y = \bigcap_{j \in J} \mathfrak{B}_j .$$

Proof. From (ii) to (i) is easy. To show the other direction we first expand the language L by adding an individual constant symbol for every element of Y. Call the resulting expansion of $\mathfrak{A}$, $\mathfrak{A}^*$. Then

$$D(\mathfrak{A}^*, 0) = D(\mathfrak{A}, Y) = Y ,$$

where $D(\mathfrak{A}^*, 0)$ refers to definability in the expanded language.

Let $\mathfrak{B}^*$ be a special elementary extension of $\mathfrak{A}^*$, and let $\{\mathfrak{B}_j^* : j \in J\}$ be the set of all elementary submodels of $\mathfrak{B}^*$. Clearly $D(\mathfrak{B}^*, 0) = D(\mathfrak{A}^*, 0)$, so by Lemma 5.1 (for models of the expanded language)

$$\mathfrak{B}^* \restriction Y = \bigcap_{j \in J} \mathfrak{B}_j^* .$$

Throwing away the added constants, $\{\mathfrak{B}_j : j \in J\}$ is the set of all elementary submodels of $\mathfrak{B}$ which contain Y, and $\mathfrak{B} \restriction Y = \bigcap_{j \in J} \mathfrak{B}_j$.

Using Theorem 5.2 one may derive another result of Park, characterizing those theories such that the intersection of any collection of elementary submodels of any model $\mathfrak{A}$ of the theory is again an elementary submodel of $\mathfrak{A}$.

Theorem 5.3 (Park [16, 17]). *Let T_0 be a theory of* L. *Then the following are equivalent:*

(i) *For every $\mathfrak{A} \models T_0$ and every set $\{\mathfrak{A}_j : j \in J\}$ of elementary submodels of $\mathfrak{A}$*

$$\bigcap_{j \in J} \mathfrak{A}_j \prec \mathfrak{A} ;$$

(ii) *For every $\mathfrak{A} \models T_0$ and every $Y \subseteq A$*

$$\mathfrak{A} \restriction D(\mathfrak{A}, Y) \prec \mathfrak{A} ;$$

(iii) *For every formula $\phi(x, v_1, ..., v_k)$ of* L *there is some formula $\psi(x, v_1, ..., v_k)$ of* L *such that for some n*

$$T_0 \models \forall v_1, ..., v_k [\exists x\, \phi \to \exists x(\phi \wedge \psi) \wedge \exists^{\leq n} x\, \psi]$$

Proof (see also [16]). The equivalence of (i) and (ii) is immediate from Theorem 5.2. The implication from (iii) to (ii) is straightforward (since $D(\mathfrak{A}, D(\mathfrak{A}, Y)) = D(\mathfrak{A}, Y)$). We give only a very brief indication of the proof of the remaining direction, from (ii) to (iii). Notice that, by compactness, to show (iii) it is sufficient to show

(iii′) For every formula $\phi(x, v_1, ..., v_k)$ of L, for every $\mathfrak{A} \models T_0$, and every $b_1, ..., b_k \in A$ such that $\mathfrak{A} \models \exists x\, \phi(x, b_1, ..., b_k)$, there is some $\psi(x, v_1, ..., v_k)$ of L such that

$$\mathfrak{A} \models \exists x [\phi(x, b_1, ..., b_k) \wedge \psi(x, b_1, ..., b_k)]$$

and

$$\mathfrak{A} \models \exists^{\leq n} x\, \psi(x, b_1, ..., b_k) \text{ for some } n.$$

But (iii′) follows easily from (ii). Thus, we know

$$\mathfrak{A} \upharpoonright D(\mathscr{A}, \{ b_1, ..., b_k \}) \prec \mathfrak{A}$$

and hence if $\mathfrak{A} \models \exists x\, \phi(x, b_1, ..., b_k)$ there is some $a \in D(\mathfrak{A}, \{ b_1, ..., b_k \})$ such that $\mathfrak{A} \models \phi(a, b_1, ..., b_k)$. This point a is then defined by some ψ which will work in (iii′).

Earlier results similar to Theorem 5.3 may be found in [19] and [14].

§ 6. A syntactical proof of Theorem 4.1

We give here a syntactical proof of the implication from $(\mathrm{i})_n$ to $(\mathrm{ii})_n$ of Theorem 4.1. It proceeds roughly as follows. We first apply compactness to reduce the theory T to a single sentence τ. Then we give a direct argument, using only Craig's interpolation theorem, to obtain the formulas σ and ϕ_i, $1 \leq i \leq n$, used in defining P. In particular, there are no further applications of compactness.

Aside from the usual interest of effective syntactical proofs, this proof is of interest because of its applicability to $\mathrm{L}_{\omega_1\omega}$, an infinitary language allowing countable conjunctions and disjunctions (see Scott [21]). Lopez-Escobar [11] has shown that Craig's theorem holds for $\mathrm{L}_{\omega_1\omega}$, and therefore so does Beth's theorem, for theories given by a single sentence of $\mathrm{L}_{\omega_1\omega}$. Thus this proof shows that Theorem 4.1 holds in $\mathrm{L}_{\omega_1\omega}$ for theories given by a single sentence. Other sorts of infinitary definability results are given in [21] and [10].

Actually the proof given is not purely syntactical, bu it could be rewritten as such a proof. This would, of course, involve stating $(\mathrm{i})_n$ syntactically.

We remark that we do not know to what extent the interpolation theorems in §3 have syntactical proofs, nor whether they hold for $\mathrm{L}_{\omega_1\omega}$.

Our proof that $(\mathrm{i})_n$ implies $(\mathrm{ii})_n$ proceeds by induction on n. We assume the result is known for all $k < n$, for *any* choice of the language L. Let T be a theory such that $(\mathrm{i})_n$ holds; by compactness we may assume that T is given by a single sentence $\tau(\mathbf{P})$ of $\mathrm{L}(\mathbf{P})$. (Essentially the same proof also works if T, and hence τ, belong to any language containing $\mathrm{L}(\mathbf{P})$.) We first show

(1) *There is a formula $\chi(x, v_1, ..., v_{n-1})$ of* L *such that for any* $\mathfrak{A}$ *with* $|M_T(\mathfrak{A})| = n$, *there is some* $P \in M_T(\mathfrak{A})$ *such that*

$$(\mathfrak{A}, P) \models \exists v_1, ..., v_{n-1} \, \forall x [\mathbf{P}(x) \leftrightarrow \chi] \, .$$

Proof. Let $\mathbf{P}_1, \ldots, \mathbf{P}_n$ be new unary predicate symbols. Let θ be the sentence of $\mathrm{L}(\mathbf{P}_1, \ldots, \mathbf{P}_n)$ defined as

$$\tau(\mathbf{P}_1) \wedge \ldots \wedge \tau(\mathbf{P}_n) \wedge \bigwedge_{1 \leq i < j \leq n} \neg \forall x [\mathbf{P}_i(x) \leftrightarrow \mathbf{P}_j(x)]$$

Then the following implication is valid.

$$\models [\theta \wedge \neg \mathbf{P}_1(v_1) \wedge \ldots \wedge \neg \mathbf{P}_{n-1}(v_{n-1}) \wedge \mathbf{P}_n(x)]$$

$$\rightarrow [\tau(\mathbf{P}) \rightarrow \neg \mathbf{P}(v_1) \vee \ldots \vee \neg \mathbf{P}(v_{n-1}) \vee \mathbf{P}(x)]$$

Applying Theorem 1.2 we get a formula $\chi(x, v_1, \ldots, v_{n-1})$ of L interpolating between the antecedent and the consequent of this implication.

Let $\mathfrak{A}$ be such that $|M_T(\mathfrak{A})| = n$, say $M_T(\mathfrak{A}) = \{P_1, \ldots, P_n\}$. We may assume that P_n is maximal, and so find $a_i \in P_n - P_i$ for each $i = 1, \ldots, n-1$. Then

$$(\mathfrak{A}, P_n) \models \forall x [\mathbf{P}_n(x) \rightarrow \chi(x, a_1, \ldots, a_{n-1})] ,$$

and

$$(\mathfrak{A}, P) \models \forall x [\chi(x, a_1, \ldots, a_{n-1})$$

$$\rightarrow \neg \mathbf{P}(a_1) \vee \ldots \vee \neg \mathbf{P}(a_{n-1}) \vee \mathbf{P}(x)]$$

for each $P \in M_T(\mathfrak{A})$. For $P = P_n$ none of $\neg \mathbf{P}(a_1), \ldots, \neg \mathbf{P}(a_{n-1})$ can hold, and hence

$$(\mathfrak{A}, P_n) \models \forall x [\mathbf{P}(x) \leftrightarrow \chi(x, a_1, \ldots, a_{n-1})] ,$$

which proves (1).

We define $\tau_1(v_1, \ldots, v_{n-1})$ to be the formula obtained from $\tau(\mathbf{P})$

by replacing every occurrence in τ of $\mathbf{P}(t)$, for any term t, by $\chi(t, v_1, ..., v_{n-1})$, it being assumed that none of $v_1, ..., v_{n-1}$ occur in $\tau(\mathbf{P})$. Then for any $\mathfrak{A}$ and any $a_1, ..., a_{n-1} \in A$

$$\mathfrak{A} \models \tau_1(a_1, ..., a_{n-1}) \text{ iff } (\mathfrak{A}, P) \models \tau(\mathbf{P}),$$

where

$$P = \{a \in A : \mathfrak{A} \models \chi(a, a_1, ..., a_{n-1})\} .$$

Let ψ be the formula

$$\exists y_1, ..., y_{n-1}\, \tau_1(y_1, ..., y_{n-1}) \rightarrow \tau_1(v_1, ..., v_{n-1}) .$$

Note that τ_1 and ψ are both formulas of L and that $\models \exists v_1, ..., v_{n-1}\ \psi$.

Now we expand L to a language L^* by adding $n - 1$ new individual constant symbols, $\mathbf{c}_1, ..., \mathbf{c}_{n-1}$. We consider the theory T^* in the language $\mathrm{L}^*(\mathbf{P})$ given by

$$\tau(\mathbf{P}) \wedge \neg \forall x[\mathbf{P}(x) \leftrightarrow \chi(x, \mathbf{c}_1, ..., \mathbf{c}_{n-1})] \wedge \psi(\mathbf{c}_1, ..., \mathbf{c}_{n-1}) .$$

Then we have

(2) *Let* $\mathfrak{A}^* = (\mathfrak{A}, a_1, ..., a_{n-1})$ *be any model for* L^*. *Then:*

(a) *if* $\mathfrak{A} \models \psi(a_1, ..., a_{n-1})$ *then*

$$M_{T^*}(\mathfrak{A}^*) \subseteq M_T(\mathfrak{A}) \subseteq M_{T^*}(\mathfrak{A}^*) \cup \{P_0\},$$

where

$$P_0 = \{a \in A : \mathfrak{A} \models \chi(a, a_1, ..., a_{n-1})\} .$$

(b) $|M_{T^*}(\mathfrak{A}^*)| \leq n - 1$.

Proof. (a) is clear from the definition of T^*. (b) Assume that $M_{T^*}(\mathfrak{A}^*) \neq 0$. Then $a_1, ..., a_{n-1}$ satisfy ψ, and so (a) holds. But, by the definition of T^*, $P_0 \notin M_{T^*}(\mathfrak{A}^*)$; hence if $P_0 \in M_T(\mathfrak{A})$ then (a) implies that $M_{T^*}(\mathfrak{A}^*) = M_T(\mathfrak{A}) - \{P_0\}$, and so it can have at most $n-1$ elements. On the other hand, if $P_0 \notin M_T(\mathfrak{A})$ then no set in $M_T(\mathfrak{A})$ is definable by χ with any choice of individual parameters. Therefore by (1), $M_T(\mathfrak{A})$ cannot have n members, and so $|M_{T^*}(\mathfrak{A}^*)| \leq n-1$.

Now by (2)(b) we can apply our inductive hypothesis to T^* to get formulas $\sigma^*(v_1, ..., v_{n-1}, v_n, ..., v_k)$, $\phi_i(x, v_1, ..., v_k)$, $1 \leq i \leq n-1$, of L such that

(3) $$T^* \models \exists v_n, ..., v_k\, \sigma^*(\mathrm{c}_1, ..., \mathrm{c}_{n-1}, v_n, ..., v_k)$$

and

$$T^* \models \forall v_n, ..., v_k (\sigma^*(\mathrm{c}_1, ..., \mathrm{c}_{n-1}, v_n, ..., v_k)$$

$$\rightarrow \bigvee_{1\leq i\leq n-1} \forall x[\mathrm{P}(x) \leftrightarrow \phi_i(x, \mathrm{c}_1, ..., \mathrm{c}_{n-1}, v_n, ..., v_k)]).$$

Since $T^* \models \psi(\mathrm{c}_1, ..., \mathrm{c}_{n-1})$, (3) continues to hold when σ^* is replaced by the formula σ^{**} defined as

$$\sigma^*(v_1, ..., v_k) \wedge \psi(v_1, ..., v_{n-1}).$$

Now define ϕ_n to be $\chi(x, v_1, ..., v_{n-1})$. Then

(4) $$T \models \forall v_1, ..., v_k (\sigma^{**} \rightarrow \bigvee_{1\leq i\leq n} \forall x[\mathrm{P}(x) \leftrightarrow \phi_i]).$$

(4) is clear from the definition of σ^{**}, (2)(a), (3), and the definition of ϕ_n.

In general, however, $\exists v_1, ..., v_k \sigma^{**}$ may not follow from T. But if $\mathfrak{A} \models \neg \exists v_1, ..., v_k \sigma^{**}$ then it follows from (3) that for every $a_1, ..., a_{n-1} \in A$ we have $M_{T^*}((\mathfrak{A}, a_1, ..., a_{n-1})) = 0$. So if

$M_T(\mathfrak{A}) \neq 0$ it follows from (2)(a) that

$$M_T(\mathfrak{A}) = \{\{a \in A : \mathfrak{A} \models \chi(a, a_1, ..., a_{n-1})\}\}$$

for any $a_1, ..., a_{n-1}$ satisfying ψ. So, define $\sigma(v_1, ..., v_k)$ to be

$$\sigma^{**}(v_1, ..., v_k) \vee [\neg \exists v_1, ..., v_k \sigma^{**} \wedge \psi(v_1, ..., v_{n-1})]$$

Then what we have said above shows that $(\mathrm{ii})_n$ holds for this σ and the previously defined $\phi_1, ..., \phi_n$. Therefore the proof is completed.

References

[1] E.W.Beth, On Padoa's method in the theory of definition, Indag. Math. 15 (1953) 330–339.

[2] C.C.Chang, Some new results in definability, Bull. Am. Math. Soc. 70 (1964) 808–813.

[3] C.C.Chang, A generalization of the Craig interpolation theorem, Am. Math. Soc. Notices 15 (1968) 934.

[4] C.C.Chang and Y.N.Moschovakis, On Σ_1^1-relations on special models. Am. Math. Soc. Notices 15 (1968) 934.

[5] W.Craig, Three uses of the Herbrand-Gentzen theorem in relating model theory and proof theory, J. Symb. Logic 22 (1957) 269–285.

[6] L.Henkin, An extension of the Craig-Lyndon interpolation theorem, J. Symb. Logic 28 (1963) 201–216.

[7] G.Kreisel and J.L.Krivine, Elements of Mathematical Logic (Model Theory) (North-Holland, Amsterdam, 1967).

[8] D.W.Kueker, A generalization of Beth's theorem on definability, Am. Math. Soc. Notices 14 (1967) 718.

[9] D.W.Kueker, Some results on definability theory, Doctoral dissertation, University of California, Los Angeles, 1967.

[10] D.W.Kueker, Definability, automorphisms, and infinitary languages, in: J.Barwise (editor), Syntax and Semantics of Infinitary Languages, Springer Lecture Notes 72 (1968) 152–165.

[11] E.G.K.Lopez-Escobar, An interpolation theorem for denumerably long sentences, Fund. Math. 57 (1965) 253–272.

[12] R.C.Lyndon, An interpolation theorem in the predicate calculus, Pacific J. Math. 9 (1959) 129–142.

[13] M.Makkai, A generalization of a theorem of E.W.Beth, Acta Math. Acad. Sci. Hungar. 15 (1964) 227–235.

[14] R.M.Montague and R.L.Vaught, A note on theories with selectors, Fund. Math. 47 (1959) 243–247.

[15] M.Morley and R.L.Vaught, Homogeneous universal models, Math. Scand. 11 (1962) 37–57.

[16] D.M.R.Park, Set-theoretic constructions in model-theory, Doctoral dissertation, M.I.T., 1964.

[17] D.M.R.Park, Intersections of elementary substructures I, II, Am. Math. Soc. Notices 14 (1967) 250, 405.

[18] A.Robinson, A result on consistency and its application to the theory of definition, Indag. Math. 18 (1956) 47–58.

[19] A.Robinson, Model-theory and non-standard arithmetic, in: Infinitistic Methods (PWN, Warsaw, 1961) pp. 265–302.

[20] A.Robinson, Introduction to Model Theory and to the Metamathematics of Algebra (North-Holland, Amsterdam, 1963).

[21] D.Scott, Logic with denumerably long formulas and finite strings of quantifiers, in: J.W.Addison, L.Henkin and A.Tarski (editors), The Theory of Models (North-Holland, Amsterdam, 1965) pp. 329–341.

[22] L.Svenonius, A theorem on permutations in models, Theoria (Lund) 25 (1959) 173-178.
[23] L.Svenonius, On the denumerable models of theories with extra predicates, in: J.W.Addison, L.Henkin and A.Tarski (editors), The Theory of Models (North-Holland, Amsterdam, 1965) pp. 376–389.
[24] K.L.de Bouvère, A Method in Proofs of Undefinability (North-Holland, Amsterdam, 1959).