**Centrum voor Wiskunde en Informatica**
Centre for Mathematics and Computer Science

P.H. Rodenburg, R.J. van Glabbeek

An interpolation theorem in equational logic

Computer Science/Department of Software Technology Report CS-R8838 October

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

# An Interpolation Theorem in Equational Logic

P.H. Rodenburg
*Programming Research Group, University of Amsterdam*
*P.O. Box 41882, 1009 DB Amsterdam, The Netherlands*

R.J. van Glabbeek
*Centre for Mathematics and Computer Science*
*P.O. Box 4079, 1009 AB Amsterdam, The Netherlands*

In a natural formulation, Craig's interpolation theorem is shown to hold for equational logic. We also discuss the prevalent claims that equational logic does *not* have the interpolation property.

*1980 Mathematics Subject Classification* (Zentralblatt für Mathematik): 03C40, 03F99, 08B05, 68B10.
*1987 CR Classification scheme* (Computing Reviews): D.3.3 (Modules), F.4.1 (Proof theory).
*Key Words & Phrases:* Interpolation, Craig's lemma, Equational logic, Similarity, Module algebra, Splitting interpolation.
*Note:* During most of the time that this paper was being written, the first author was attached to the department of Philosophy of the Rijksuniversiteit at Utrecht, and partially supported by the Dutch government through the SPIN project PRISMA. The second author was partially supported by the European Community through the Esprit project Meteor.

## Introduction

The following proposition about first order logic is known as CRAIG'S LEMMA:

(1) If $\varphi$ and $\psi$ are sentences of some first order language $\mathcal{L}$, and $\varphi\rightarrow\psi$ is provable in the first order calculus (notation: $\vdash\varphi\rightarrow\psi$), then we can find a sentence $\theta$ such that the nonlogical symbols of $\theta$ occur in both $\varphi$ and $\psi$, and $\vdash\varphi\rightarrow\theta$ and $\vdash\theta\rightarrow\psi$ .

(Cf. [C].) The sentence $\theta$ is called an *interpolant* for $\varphi$ and $\psi$, and accordingly the property of first order logic expressed in (1) is sometimes referred to as the *interpolation property*, and (1) as the *interpolation theorem*. It should be noted that the equality symbol = is counted among the logical symbols.

One may ask of other logical systems, such as second order logic or equational logic, whether they have the interpolation property. It need not be immediately clear, however, what such a question means; especially when the system under consideration is weaker than first order logic. In particular, if we want to pronounce on the behavior of equational logic with respect to interpolation, we must first find a suitable statement of Craig's Lemma.

Report CS-R8838
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

The problem is essentially that equational logic is much less expressive than first order logic. It is not really weaker. Indeed, first order logic is a conservative extension of equational logic: if an equation $\beta$ can be derived, in the first order calculus, from the universal closures of a set A of equations, then $\beta$ can be derived from A in equational logic. Thus, on the assumption that deduction in the first order calculus from assumptions containing free variables is defined in such a way that a formula $\varphi$ can be deduced from a set $\Phi$ of formulas if and only if $\varphi$ can be deduced from the universal closures of the formulas in $\Phi$, our indiscriminate use, in the sequel, of the turnstile $\vdash$ for deducibility in both equational logic and first order logic is quite justified.

First, then, the statement (1) depends on a notion of implication in the formal language, which is lacking in equational logic. This can be overcome by passing to the equivalent statement — equivalent for first order logic —

(2) If $\varphi \vdash \psi$, then we can find a sentence $\theta$ such that the nonlogical symbols of $\theta$ occur in both $\varphi$ and $\psi$, and $\varphi \vdash \theta$ and $\theta \vdash \psi$.

This (with 'equation' substituted for 'sentence') is still not quite what we want. Deductions from a single equation are rather special, and not particularly interesting. It could in fact be maintained that the interest of (2), even in the first order case, is a consequence of the presence of conjunction. Thus we pass to the next reformulation:

(3) If $\varphi_1,\ldots,\varphi_n \vdash \psi$, then we can find a finite list of sentences $\theta_1,\ldots,\theta_k$ such that the nonlogical symbols of each $\theta_i$ ($1 \leq i \leq k$) occur both in $\psi$ and in at least one of $\varphi_1,\ldots,\varphi_n$, and $\varphi_1,\ldots,\varphi_n \vdash \theta_i$, for all $i$ from 1 to $k$, and $\theta_1,\ldots,\theta_k \vdash \psi$.

Statement (3) is still equivalent to (1), for first order logic; and *mutato mutando* we have an interesting claim about equational logic. Avoiding subscripts, and taking account of sorts as part of the nonlogical endowment of equations, the interpolation theorem for equational logic now runs as follows:

(4) If $\beta$ is an equation, and A a set of equations, such that $A \vdash \beta$, then we can find a finite set I of equations the signature of which is contained in that of A and that of $\beta$, such that $A \vdash I$ and $I \vdash \beta$.

This is what we will prove in the sequel. With some self-explanatory notation, and dropping the claim of finiteness (which was hardly worth making anyway; finiteness of I in (4) is a simple consequence of the finitary nature of equational logic), (4) can be condensed to

(4') If A and B are sets of equations such that A⊢B, then we can find a set I of equations such that $\Sigma(I) \subseteq \Sigma(A) \cap \Sigma(B)$, A⊢I and I⊢B .

Our investigation was motivated by [BHK]. In that report, BERGSTRA, HEERING and KLINT present a 'module algebra': a system of axioms for import and export. Interpreted for first order logic, one of these axioms (E4, to be precise) is shown to be a restatement of Craig's Lemma. The same axiom is shown to fail for equational logic. On the basis of a similar observation, MAIBAUM and SADLER [MS] claim that equational logic 'does not satisfy the Craig interpolation property'. Below, we shall discuss the situation in the light of our result.

**Preliminaries**

A *signature* consists of a set of sorts and a set of function symbols. It may help to assume that all sorts and all function symbols are taken from some pre-existing class. Every function symbol comes with a fixed *arity*: a specification of the number of its arguments, their sorts in order, and the sort of its value. If a signature contains a function symbol, then it must also contain the sorts specified in its arity. To communicate that a function symbol $f$ takes $n$ arguments of respective sorts $s_1,\ldots,s_n$, and values of sort $s$, we may write $f{:}s_1\times\ldots\times s_n\rightarrow s$. From function symbols and variables, terms are constructed as usual. If the function symbols in a term $t$ are all from some signature $\Sigma$, we say $t$ is a term *over* $\Sigma$. For a term, equation or system of equations $X$, $\Sigma(X)$, the signature of $X$, consists of all the sorts involved in $X$ and all function symbols occurring in $X$. In fact, sorts do not play an essential part in this note. We always assume that the terms and equations we are dealing with are wellformed. The symbol ≡ will be used for identity by definition and identity of syntactic objects.

A *context* is a term or equation with holes; with each such hole a sort is associated, the sort of terms that would fit into it. We shall mark holes by the names of their associated sorts. A precise definition of contexts then runs as follows:

(i) a term of sort $s$, and the sort $s$ itself, are contexts of sort $s$;

(ii) if $f{:}s_1\times\ldots\times s_n\rightarrow s$, and $c_1,\ldots,c_n$ are contexts of sorts $s_1,\ldots,s_n$ respectively, then $f(c_1,\ldots,c_n)$ is a context of sort $s$;

(iii) if $c_1$ and $c_2$ are contexts of the same sort, then $c_1{=}c_2$ is a context.

When it is necessary to distinguish contexts as formed by (i) and (ii) above from the contexts formed by (iii), we shall call the first kind *term contexts* and the second *equation contexts*. If the function symbols and sorts in context $c$ are all from signature $\Sigma$, we say $c$ is a context *over* $\Sigma$. If $c$ contains holes marked $s_1,\ldots,s_n$ respectively, then we may write $c[s_1,\ldots,s_n]$ for $c$; then if $t_1,\ldots,t_n$ are terms of suitable sorts, $c[t_1,\ldots,t_n]$ is the term or equation one gets by inserting each $t_i$ in the $i$-th slot ($1\leq i\leq n$). Note that we assume the holes marked by writing $c[s_1,\ldots,s_n]$ to be

real. In particular, insertion instances $c[s_1,\ldots,s_n]$ and $c[t_1,\ldots,t_n]$ are identical only if $s_i \equiv t_i$ for all $i$, $1 \leq i \leq n$. If $c \equiv (c_1 = c_2)$ is an equation context, with $c_1 \equiv c_1[s_1,\ldots,s_m]$ and $c_2 \equiv c_2[s_{m+1},\ldots,s_n]$, we may write $c_1[s_1,\ldots,s_m]=c_2[s_{m+1},\ldots,s_n]$ instead of $c[s_1,\ldots,s_n]$.

By a common convention, a term $t$ may be introduced as $t(y_1,\ldots,y_n)$, with the implication that $t$ contains no other variables than $y_1,\ldots,y_n$. In the sequel, a rather different convention will be used much more often, which we shall explain now. We shall introduce equations $\gamma \equiv (s=t)$ as $\gamma[y_1,\ldots,y_n]$, or $s[y_1,\ldots,y_m]=t[y_{m+1},\ldots,y_n]$ — an insertion instance of some equation context $c \equiv \gamma[s_1,\ldots,s_n] \equiv (s[s_1,\ldots,s_m]=t[s_{m+1},\ldots,s_n])$ — with the understanding that $s$ and $t$ contain no occurrences of variables other than the indicated occurrences of $y_1,\ldots,y_n$; so in particular, the context $c$ does not contain any variables. As with the parentheses notation, we may go on to discuss some equation $s[r_1,\ldots,r_m]=t[r_{m+1},\ldots,r_n]$ (or terms $s[r_1,\ldots,r_m]$, $t[r_{m+1},\ldots,r_n]$); but these will count as *insertion* instances of $\gamma[s_1,\ldots,s_n]$ (or $s[s_1,\ldots,s_m]$, $t[s_{m+1},\ldots,s_n]$), so that it is not required that $r_i \equiv r_j$ if $y_i$ and $y_j$ happen to be the same variable and $i \neq j$.

Equational logic consists of an identity axiom $x=x$ (one for every sort) and the rules of symmetry (conclude $t=s$ from $s=t$), transitivity (conclude $r=t$ from $r=s$ and $s=t$), substitution (conclude $s(s_1,\ldots,s_n)=t(s_1,\ldots,s_n)$ from $s(y_1,\ldots,y_n)=t(y_1,\ldots,y_n)$) and congruence (for any function symbol $f$, conclude $f(s_1,\ldots,s_n)=f(t_1,\ldots,t_n)$ from $s_1=t_1,\ldots,s_n=t_n$). $A \vdash \gamma$ means that the equation $\gamma$ can be derived by these rules from the identity axioms and the equations in A. For later reference we note the following simple fact.

**Lemma 1.** If $A \vdash \gamma[y_1,\ldots,y_k]$, and for certain terms $r_1,\ldots,r_k$, $A \vdash r_i=r_j$ whenever $y_i \equiv y_j$, then $A \vdash \gamma[r_1,\ldots,r_k]$.

## Similarity

Let A be a system of equations. The relation of *similarity over* A is the least relation $\approx$ on the class of all terms such that, for any terms $s$ and $t$, $s \approx t$ if

(i) $s \equiv t$; or

(ii) for some function symbol $f$ not in $\Sigma(A)$, and terms $s_1,\ldots,s_n,t_1,\ldots,t_n$ such that for $1 \leq i \leq n$, $s_i \approx t_i$, $s \equiv f(s_1,\ldots,s_n)$ and $t \equiv f(t_1,\ldots,t_n)$, or

(iii) there are terms $s' \equiv s'[y_1,\ldots,y_m]$ and $t' \equiv t'[y_{m+1},\ldots,y_n]$ over $\Sigma(A)$, and terms $r_1,\ldots,r_n$ such that for $1 \leq i < j \leq n$, $r_i \approx r_j$ if $y_i \equiv y_j$, with $s \equiv s'[r_1,\ldots,r_m]$ and $t \equiv t'[r_{m+1},\ldots,r_n]$, and $A \vdash s'=t'$.

If $s$ and $t$ are in this relation, we say that $s$ and $t$ are *similar over* A, notation: $s \approx_A t$ (and we shall often drop the subscript when confusion is unlikely). If two terms are similar over A, there must be a *derivation* of this similarity, consisting of successive applications of clauses (i)-(iii) above to intermediate similarities. By *the set of theorems associated with* some given derivation $D$ of similarity we shall mean the set of all A-derivable equations $s'=t'$ over $\Sigma(A)$ used in ap-

plications of clause (iii) in $D$. In proving properties of similarity we will sometimes take recourse to induction on the length of derivations.

By clause (i), the relation of similarity, over any given A, is reflexive. It is also symmetric, as may be seen by a trivial induction on the lengths of derivations.

**Examples.**
1. Let A be a system of equations axiomatizing the theory of abelian groups, with a single sort $s$, a binary function symbol + (written between its arguments), a unary function symbol –, and a nullary function symbol 0. Writing $x - y$ for $x + (-y)$, as is natural, we have $A \vdash x-x=0$. Since $x \approx x$ by (i), $x-x \approx_A 0$ by clause (iii). Let $f: s \times s \to s$ be a new function symbol. Since $y \approx y$ by (i), $f(x-x,y) \approx f(0,y)$ by clause (ii). Again by clause (iii), $A \vdash x-x=0$ justifies $f(x-x,y)-f(0,y) \approx 0$.
2. Let $A = \{x=y\}$, for distinct variables $x$ and $y$. Then $\Sigma(A)$ is empty (but for a single sort). Let $s$ and $t$ be arbitrary terms (of that sort). Then replacing $x$ by $s$ and $y$ by $t$, we find $s \approx_A t$.

**Similarity implies provable equality**

We now prove a sequence of lemmas connecting provable equality from axioms A with similarity over A. One direction is easy.

**Lemma 2.** If $s \approx_A t$, then $A \vdash s = t$.

**Proof.** By induction on the length of the derivation of $s \approx t$. If $s \approx t$ by clause (i), $s=t$ is a substitution instance of an identity axiom. If the last step is by clause (ii), $A \vdash s=t$ follows from the induction hypothesis by the rule of congruence. If the last step is by clause (iii), $A \vdash s=t$ is an immediate consequence of the induction hypothesis by lemma 1. ¶

The converse is considerably harder. We shall not be able to avoid juggling with the contexts involved in clause (iii) of the definition of similarity. As it stands, this clause makes for easy proofs of substitutivity and congruence (lemmas 6 and 7 below); but some more insight will be required to show that similarity is an equivalence relation on the class of all terms. For this reason, we introduce a seemingly more restricted notion of similarity.

**Strict similarity**

**Definition.** A *strict* derivation of similarity over A is a derivation of similarity over A obeying the following *strictness condition*: there are no applications of clause (iii) in which either

(a) $s' \equiv y_1 \equiv t'$, or

(b) any one of the terms $r_1, \ldots, r_n$ begins with a function symbol belonging to $\Sigma(A)$.

If $s \approx_A t$ by a strict derivation, we say $s$ and $t$ are *strictly similar* over A, and write $s \cong_A t$.

Intuitively, the alternatives (a) and (b) represent ways in which a derivation might be longer than necessary. If $s' \equiv y_1 \equiv t'$, one of the premises needed to conclude $s \approx t$ is $s \approx t$ itself; and if $r_i$ begins with a function symbol from $\Sigma(A)$, presumably some part of $r_i$ could be added to $s'[\ldots]$ and/or $t'[\ldots]$, and the subderivations $D_{ij}$ involving $r_i$ replaced by subderivations of $D_{ij}$. We will justify this intuition only to the extent of proving that all similar terms are strictly similar (lemma 4 below). Observe that for any term $r$, there is exactly one context $c$ that could figure in applications of clause (iii) to obtain a similarity with $r$ that satisfy strictness: with $r \equiv c[q_1,\ldots,q_n]$, where $q_1,\ldots,q_n$ are the maximal subterms of $r$ that do not begin with a function symbol from $\Sigma(A)$.

**Lemma 3.** Strict similarity over A is an equivalence relation on the class of all terms.

**Proof.** Reflexivity and symmetry are trivial, as with ordinary similarity; we assume they hold, and concentrate on transitivity.

Let for any term $t$, lh$t$ be the length of $t$. We order three-element multisets of terms, to be called triples for short, as follows. (Recall that multisets are something between sequences and sets: the order of elements is neglected, but not the number of times they occur.) Suppose we are given multisets $\{r_1,s_1,t_1\}$ and $\{r_2,s_2,t_2\}$, with $\mathrm{lh}r_i \geq \mathrm{lh}s_i \geq \mathrm{lh}t_i$, for $i$=1,2. Then $\{r_1,s_1,t_1\} > \{r_2,s_2,t_2\}$ iff

$\mathrm{lh}r_1 > \mathrm{lh}r_2$, or

$\mathrm{lh}r_1 = \mathrm{lh}r_2$ and $\mathrm{lh}s_1 > \mathrm{lh}s_2$, or

$\mathrm{lh}r_1 = \mathrm{lh}r_2$ and $\mathrm{lh}s_1 = \mathrm{lh}s_2$ and $\mathrm{lh}t_1 > \mathrm{lh}t_2$.

This ordering of triples is clearly inductive. We shall prove that $\simeq_A$ is transitive by induction over triples.

Suppose we have three terms $r$, $s$ and $t$, with $r \simeq_A s$ and $s \simeq_A t$; and whenever $r' \simeq_A s'$ and $s' \simeq_A t'$ with $\{r',s',t'\} < \{r,s,t\}$, $r' \simeq_A t'$. We must show that $r \simeq_A t$.

We first exclude a simple pathological case. It is easy to see that, if $A \vdash x=y$ for distinct variables $x$ and $y$ of some sort $s$, $r \simeq_A t$ for all terms $r$ and $t$ of sort $s$. Hence for the remainder of this proof we may assume that $A \vdash x=y$, for variables $x$ and $y$ of the sort of our terms $r$, $s$ and $t$, holds only if $x \equiv y$. Now since we obviously may ignore applications of clause (i), there are four cases left.

(a) There is a function symbol $f \notin \Sigma(A)$ such that for certain terms $r_1,\ldots,r_n$, $s_1,\ldots,s_n$ and $t_1,\ldots,t_n$, $r \equiv f(r_1,\ldots,r_n)$, $s \equiv f(s_1,\ldots,s_n)$ and $t \equiv f(t_1,\ldots,t_n)$, with $r_i \simeq s_i \simeq t_i$, $1 \leq i \leq n$. For each $i$, $\{r_i,s_i,t_i\} < \{r,s,t\}$, so $r_i \simeq t_i$ by induction hypothesis. It follows that $r \simeq t$ by an application of clause (ii) of the definition of similarity.

(b) In both derivations, the last step is an application of clause (iii). Then there are terms $r' \equiv r'[y_1,\ldots,y_k]$, $s' \equiv s'[y_{k+1},\ldots,y_l]$, $s'' \equiv s''[z_1,\ldots,z_{k'}]$ and $t' \equiv t'[z_{k'+1},\ldots,z_{l'}]$ over $\Sigma(A)$ such that

$A\vdash r'=s'$ and $A\vdash s''=t'$, and for certain terms $q_1,\ldots,q_l, r_1,\ldots,r_{l'}$ with $q_i\simeq q_j$ if $y_i\equiv y_j$ and $1\le i<j\le l$, and $r_i\simeq r_j$ if $z_i\equiv z_j$ and $1\le i<j\le l'$, $r\equiv r'[q_1,\ldots,q_k]$, $s\equiv s'[q_{k+1},\ldots,q_l]\equiv s''[r_1,\ldots,r_{k'}]$ and $t\equiv t'[r_{k'+1},\ldots,r_{l'}]$. Observe that by the requirement of strictness the contexts $s'[s_{k+1},\ldots,s_l]$ and $s''[t_1,\ldots,t_{k'}]$ (where $s_i$ is the sort of $y_i$, and $t_j$ the sort of $z_j$) are identical; in particular, $l-k = k'$; and it follows that $q_{k+i}\equiv r_i$ for $1\le i\le k'$. We assume without loss of generality that the sequences $y_1,\ldots,y_l$ and $z_1,\ldots,z_{l'}$ are disjoint.

We have $\mathrm{lh}q_i<\mathrm{lh}r$, for $1\le i\le k$, unless $r'\equiv y_1$, and so on; by our assumptions (strictness and nontriviality of the sort under consideration) we may be sure that at most one of $r'$ and $s'$ and of $s''$ and $t'$ is a variable. Then for any triple $\{p_1,p_2,p_3\}$ of distinct elements of $\{q_1,\ldots,q_l, r_1,\ldots,r_{l'}\}$, $\{p_1,p_2,p_3\} < \{r,s,t\}$, since $p_1,p_2,p_3$ are subterms of $r,s,t$, and one at least is a proper subterm; so by reflexivity and symmetry and the induction hypothesis, $\simeq$ is an equivalence on $\{q_1,\ldots,q_l, r_1,\ldots,r_{l'}\}$.

Let $\sim$ be the least equivalence relation on the set $\{y_1,\ldots,y_l,z_1,\ldots,z_{l'}\}$ such that $y_{k+i}\sim z_i$ for $1\le i\le k'$. Pick a representative for each equivalence class. List these representatives as $u_1,\ldots,u_l$, $\ldots,u_{k+l'}$, with $u_i\sim y_i$ if $1\le i\le l$, and $u_i\sim z_{i-k}$ if $i>k$. Since identical entries in the list $y_1,\ldots,y_l$, $z_1,\ldots,z_{l'}$ have identical representatives, we infer

$$A\vdash r'[u_1,\ldots,u_k]=s'[u_{k+1},\ldots,u_l] \text{ and } A\vdash s''[u_{k+1},\ldots,u_l]=t'[u_{l+1},\ldots,u_{k+l'}]$$

by lemma 1. Hence

$$(*)\ A\vdash r'[u_1,\ldots,u_k]=t'[u_{l+1},\ldots,u_{k+l'}]$$

by the rule of transitivity. Since $q_{k+i}\equiv r_i$ for $1\le i\le k'$, and $\sim$ is the *least* equivalence relation with $y_{k+i}\sim z_i$ for $1\le i\le k'$, by induction hypothesis we have $q_i\simeq q_j$ when $y_i\sim y_j$ $(1\le i,j\le l)$, $q_i\simeq r_j$ when $y_i\sim z_j$ $(1\le i\le l,\ 1\le j\le l')$, and $r_i\simeq r_j$ when $z_i\sim z_j$ $(1\le i,j\le l')$. Now if $r'$ and $t'$ are variables (which implies $r'[u_1]\equiv u_1\equiv u_{l+1}\equiv t'[u_{l+1}]$), $r\equiv q_1\simeq r_1\equiv t$; otherwise, replacing $u_i$ by $q_i$ in (*), for $1\le i\le k$, and $u_j$ by $r_{j-k}$ for $l<j\le k+l'$, $r\simeq t$ by an application of clause (iii) of the definition of (strict) similarity.
(c),(d) In one derivation the last step is by clause (ii), in the other it is by clause (iii). These cases are symmetric; we treat just one. Suppose $r\equiv f(r_1,\ldots,r_k)$, $s\equiv f(s_1,\ldots,s_k)$, and there are terms $s'\equiv s'[y_1,\ldots,y_m]$ and $t'\equiv t'[y_{m+1},\ldots,y_n]$ over $\Sigma(A)$ such that $A\vdash s'=t'$, and for certain terms $t_1,\ldots,t_n$ with $t_i\simeq t_j$ if $y_i\equiv y_j$ and $1\le i<j\le n$, $s\equiv s'[t_1,\ldots,t_m]$ and $t\equiv t'[t_{m+1},\ldots,t_n]$. Then $s'$ must be the variable $y_1$ — so $t'$ is not a variable. Thus whenever $y_1\equiv y_j$ for $1<j\le n$, we have $r\simeq s\equiv t_1\simeq t_j$, with $t_j$ a proper subterm of $t$, so that $\{r,s,t_j\} < \{r,s,t\}$. Hence $r\simeq t_j$ by induction hypothesis; so from $A\vdash y_1=t'$ we may conclude, by clause (iii), that $r\simeq t$. ¶

Next we show that the additional restriction in the definition of strict similarity does not lead to a smaller relation.

**Lemma 4.** If $s \approx_A t$, then $s \cong_A t$.

**Proof.** Induction on the length of the given derivation $D$ of $s \approx_A t$. If the last step in this derivation is not by clause (iii), or satisfies, if it is, part (b) of the strictness condition, then the argument is straightforward. So suppose the last step of $D$ consists in taking terms $s' \equiv s'[y_1,\ldots,y_m]$ and $t' \equiv t'[y_{m+1},\ldots,y_n]$ over $\Sigma(A)$, and terms $r_1, \ldots, r_n$ such that for $1 \le i < j \le n$, $r_i \approx r_j$ if $y_i \equiv y_j$, with $s \equiv s'[r_1,\ldots,r_m]$ and $t \equiv t'[r_{m+1},\ldots,r_n]$, and $A \vdash s' = t'$. By induction hypothesis, $r_i \cong r_j$, say by a strict derivation $D_{ij}$, whenever $1 \le i < j \le n$ and $y_i \equiv y_j$; by reflexivity and symmetry of $\cong$, we may drop the requirement that $i<j$.

For each $i$, $1 \le i \le n$, there is a unique variable-free context $c_i \equiv c_i[s_{i,1},\ldots,s_{i,k_i}]$ over $\Sigma(A)$ such that $r_i \equiv c_i[q_{i,1},\ldots,q_{i,k_i}]$ for certain terms $q_{i,1},\ldots,q_{i,k_i}$ not beginning with a function symbol from $\Sigma(A)$. If the last step of $D_{ij}$ is an application of clause (i), then $c_i \equiv c_j$, and $q_{i,l} \equiv q_{j,l}$ for $1 \le l \le k_i$ $(=k_j)$. Take distinct variables $u_{i,j,1},\ldots,u_{i,j,k_i}$, and set $v_{i,j,l} \equiv u_{i,j,l}$, for $1 \le l \le k_i$. Then we have $A \vdash c_i[u_{i,j,1},\ldots,u_{i,j,k_i}] = c_j[v_{i,j,1}, \ldots, v_{i,j,k_j}]$, and $q_{i,l} \cong q_{j,l}$ by clause (i). If the last step of $D_{ij}$ is an application of clause (ii), then $c_i \equiv s_{i,1} \equiv c_j$. Let $v_{i,j,1} \equiv u_{i,j,1}$ be some variable of sort $s_{i,1}$. Then we have $A \vdash c_i[u_{i,j,1}] = c_j[v_{i,j,1}]$, and $D_{ij}$ proves $q_{i,1} \cong q_{j,1}$. Finally, suppose the last step of $D_{ij}$ is an application of clause (iii). Then in this step, sequences $u_{i,j,1},\ldots,u_{i,j,k_i}$ and $v_{i,j,1}, \ldots, v_{i,j,k_j}$ are specified such that

$$(1) \qquad A \vdash c_i[u_{i,j,1},\ldots,u_{i,j,k_i}] = c_j[v_{i,j,1},\ldots,v_{i,j,k_j}],$$

and

$$(2) \qquad \begin{array}{l} q_{i,k} \cong q_{i,l} \text{ if } u_{i,j,k} \equiv u_{i,j,l} \ (1 \le k < l \le k_i), \\ q_{i,k} \cong q_{j,l} \text{ if } u_{i,j,k} \equiv v_{i,j,l} \ (1 \le k \le k_i,\ 1 \le l \le k_j), \text{ and} \\ q_{j,k} \cong q_{j,l} \text{ if } v_{i,j,k} \equiv v_{i,j,l} \ (1 \le k < l \le k_j). \end{array}$$

As set out above, we may assume (1) and (2) hold for *every* relevant pair $(i,j)$. We assume that all sets $U_{i,j} \equiv \{u_{i,j,1},\ldots,u_{i,j,k_i},v_{i,j,1},\ldots,v_{i,j,k_j}\}$ are disjoint. Let $\sim$ be the least equivalence relation on the union $\bigcup_{i,j} U_{i,j}$ such that $u_{i,j,l} \sim u_{i,j',l}$ for all $i,j,j'$ and $l$ with $1 \le i,j,j' \le n$ and $1 \le l \le k_i$, $u_{i,j,l} \sim v_{j,i,l}$ for all $i,j$ and $l$ with $1 \le i,j \le n$ and $1 \le l \le k_i$, and $v_{i,j,l} \sim v_{i',j,l}$ for all $i,i',j$ and $l$ with $1 \le i,i',j \le n$ and $1 \le l \le k_j$. Note that in constructing $r_i$ and $r_j$ from $c_i[u_{i,j,1},\ldots,u_{i,j,k_i}]$ and $c_j[v_{i,j,1},\ldots,v_{i,j,k_j}]$, $u_{i,j',l}$, $u_{i,j,l}$, $v_{j,i,l}$ and $v_{j'',i,l}$ are replaced by the same term $q_{i,l}$. Pick a representative for each equivalence class; say $w_{i,l}$ is the representative of $u_{i,j,l}$ and $v_{j',i,l}$, for all $j$ and $j'$. Since identical entries in the lists $u_{i,j,1},\ldots,u_{i,j,k_i}$, $v_{i,j,1},\ldots,v_{i,j,k_j}$ have identical representatives, we have

$$(3) \qquad A \vdash c_i[w_{i,1},\ldots,w_{i,k_i}] = c_j[w_{j,1},\ldots,w_{j,k_j}], \text{ for all } i,j \text{ with } y_i \equiv y_j,$$

from (1) by lemma 1. Moreover, $q_{i,k} \simeq q_{j,l}$ must hold if $w_{i,k} \equiv w_{j,l}$. For $w_{i,k} \equiv w_{j,l}$ implies $u_{i,j,k} \sim v_{i,j,l}$; since $\simeq$ is an equivalence relation by lemma 3, and $\sim$ is a *least* equivalence relation, $q_{i,k} \simeq q_{j,l}$ follows from what we just noted about the construction of $r_i$ and $r_j$.

By $A \vdash s'=t'$ and lemma 1, (3) implies

$$A \vdash s'[c_1[w_{1,1},\ldots,w_{1,k_1}],\ldots,c_m[w_{m,1},\ldots,w_{m,k_m}]] =$$

$$t'[c_{m+1}[w_{m+1,1},\ldots,w_{m+1,k_{m+1}}],\ldots,c_n[w_{n,1},\ldots,w_{n,k_n}]].$$

Since $r_i \equiv c_i[q_{i,1},\ldots,q_{i,k_i}]$, we conclude that $s \simeq t$ by clause (iii). ¶

**Corollary.** $s \approx_A t$ if and only if $s \simeq_A t$.

**Proof.** The direction from right to left is trivial. The reverse direction is the lemma we just proved. ¶

**Provable equality implies similarity**

As an immediate consequence of lemma 3 and the corollary to lemma 4 we have

**Lemma 5.** Similarity over A is an equivalence relation on the class of all terms.

We continue by proving that similarity over a set of axioms A is closed under the rules of substitution and congruence.

**Lemma 6.** Suppose $s \equiv s(x_1,\ldots, x_n)$, $t \equiv t(x_1,\ldots, x_n)$. If $s \approx_A t$ , then for any terms $r_1,\ldots,r_n$,

$$s(r_1,\ldots,r_n) \approx_A t(r_1,\ldots,r_n).$$

**Proof.** By induction on the length of the derivation of $s\approx t$. If $s\equiv t$, then $s(r_1,\ldots,r_n) \approx t(r_1,\ldots,r_n)$ by reflexivity of $\approx$. If the last step in the derivation is an application of clause (ii), $s(r_1,\ldots,r_n) \approx t(r_1,\ldots,r_n)$ as a straightforward consequence of the induction hypothesis. So suppose the last step is an application of clause (iii): there are terms $s'\equiv s'[y_1,\ldots,y_k]$ and $t'\equiv t'[y_{k+1},\ldots,y_l]$ with $A\vdash s'=t'$, and $s_1,\ldots,s_l$ such that $s_i\approx s_j$ if $y_i\equiv y_j$ $(1\le i<j\le l)$, $s\equiv s[s_1,\ldots,s_k]$ and $t\equiv t'[s_{k+1},\ldots,s_l]$. We may write $s_i\equiv s_i(x_1,\ldots, x_n)$ $(1\le i\le l)$; then by induction hypothesis $s_i(r_1,\ldots,r_n)\approx s_j(r_1,\ldots,r_n)$ if $y_i\equiv y_j$ $(1\le i<j\le l)$. Combining all, we have

$$s(r_1,\ldots,r_n)\equiv s'[s_1(r_1,\ldots,r_n),\ldots,s_k(r_1,\ldots,r_n)],$$

$$t(r_1,\ldots,r_n)\equiv t'[s_{k+1}(r_1,\ldots,r_n),\ldots,s_l(r_1,\ldots,r_n)],$$

and $s(r_1,\ldots,r_n) \approx t(r_1,\ldots,r_n)$ by clause (iii) of the definition of similarity. ¶

**Lemma 7.** Let $f$ be a $k$-ary function symbol. If for $1\le l\le k$, $s_l\approx_A t_l$, then

$$f(s_1,\ldots,s_k) \approx_A f(t_1,\ldots,t_k).$$

**Proof.** Immediate: if $f\notin \Sigma(A)$, by clause (ii) of the definition of similarity; and otherwise by clause (iii), since $A\vdash f(x_1,\ldots,x_k) = f(x_1,\ldots,x_k)$. ¶

The last three lemmas may be epitomized as follows.

**Lemma 8.** If $A\vdash s=t$, then $s\approx_A t$.

## Interpolation

For similarity, interpolation is easy to prove. Suppose we have a derivation of $s\approx_A t$; let I be the associated set of theorems. Then $\Sigma(I) \subseteq \Sigma(A)\cap\Sigma(s=t)$; and obviously $s\approx_I t$. By the equivalence established between similarity and provable equality, ordinary interpolation has become easy too.

**Theorem.** If $A \vdash s=t$, then there exists a finite set I of equations over $\Sigma(A) \cap \Sigma(s=t)$ such that $A \vdash I$ and $I \vdash s=t$.

**Proof.** If $A \vdash s=t$, then $s \approx_A t$ by lemma 8. Take I as above. Then clearly $A \vdash I$, and $\Sigma(I) \subseteq \Sigma(A) \cap \Sigma(s=t)$. From $s \approx_I t$ we get $I \vdash s=t$ by lemma 2. ¶

**Module algebra**

In the notation of module algebra (see [BHK]), and equating modules with logically closed theories, the theorem above may be expressed by the equation

$$\text{(E3)} \qquad x \square (T(y)+Z) = T(x \cap y) + (x \square Z).$$

The apparent dependence of (E3) on interpolation was first observed by G. RENARDEL [R]. It is easy to prove (E3), interpreted in equational logic as in [BHK], from interpolation, and vice versa. We shall do this now, somewhat informally. By Cl(A) we shall mean the closure of A, i.e. the set of all equations over $\Sigma(A)$ that can be derived from A, in equational logic.
Interpolation implies (E3): It will suffice to prove $x \square (T(y)+Z) \subseteq T(x \cap y) + (x \square Z)$, since the reverse inclusion is trivial. Suppose $\sigma \in x \square (T(y)+Z)$. Then

$$\Sigma(\sigma) \subseteq x \cap (y \cup \Sigma(Z)) = (x \cap y) \cup (x \cap \Sigma(Z))$$

and $Z \vdash \sigma$. Take an interpolant I: then $Z \vdash I$ and $\Sigma(I) \subseteq \Sigma(Z) \cap \Sigma(\sigma) \subseteq x \cap \Sigma(Z)$, so $I \subseteq x \square Z$. Moreover $I \vdash \sigma$; and since $\Sigma(\sigma) \subseteq (x \cap y) \cup (x \cap \Sigma(Z)) = \Sigma(T(x \cap y)+(x \square Z))$, $\sigma \in T(x \cap y)+(x \square Z)$.
(E3) implies interpolation: Suppose $A \vdash \sigma$; then

$$\sigma \in \Sigma(\sigma) \square (T(\Sigma(\sigma))+Cl(A)) = T(\Sigma(\sigma)) + (\Sigma(\sigma) \square Cl(A)),$$

hence there must be $I \subseteq \Sigma(\sigma) \square Cl(A)$, i.e. such that $\Sigma(I) \subseteq \Sigma(\sigma) \cap \Sigma(A)$ and $A \vdash I$, with $I \vdash \sigma$.

The stronger axiom

$$\text{(E4)} \qquad \Sigma(Y) \cap \Sigma(Z) \subseteq x \Rightarrow x \square (Y+Z) = (x \square Y)+(x \square Z)$$

may be formulated for extensions of equational logic as a combined interpolation and *splitting* property:

(*) if $A \cup B \vdash \Gamma$, then there is a set I with $\Sigma(I) \subseteq \Sigma(A) \cap (\Sigma(B) \cup \Sigma(\Gamma))$ such that $A \vdash I$, and $B \cup I \vdash \Gamma$.

One gets interpolation from (*) by taking $B=\emptyset$.

(E4) implies (*): Suppose $A \cup B \vdash \Gamma$. Take $x = \Sigma(\Gamma) \cup (\Sigma(A) \cap \Sigma(B))$; then the condition of (E4) is satisfied for $Y = A$ with signature $\Sigma(A)$ and $Z = B$ with signature $\Sigma(B)$. Since $A \cup B \vdash \Gamma$, and $\Sigma(\Gamma) \subseteq x$, we must have $(x \square A) \cup (x \square B) \vdash \Gamma$. Take $I \equiv x \square A$. Then

$$\Sigma(I) \subseteq (\Sigma(\Gamma) \cup (\Sigma(A) \cap \Sigma(B))) \cap \Sigma(A) = \Sigma(A) \cap (\Sigma(B) \cup \Sigma(\Gamma)),$$

and clearly $I \cup B \vdash \Gamma$.

(*) implies (E4): Suppose $\Sigma(Y) \cap \Sigma(Z) \subseteq x$. We prove only $x \square (Y+Z) \subseteq (x \square Y)+(x \square Z)$, since the other inclusion is nothing special. If $\gamma \in x \square (Y+Z)$, then $\Sigma(\gamma) \subseteq x$. Take an interpolant $\Theta$ between $Y$ and $Z,\gamma$; and $I$ between $Z$ and $\Theta,\gamma$. Then

$$\begin{aligned}\Sigma(\Theta) &\subseteq \Sigma(Y) \cap (\Sigma(Z) \cup \Sigma(\gamma)) \subseteq \Sigma(Y) \cap (\Sigma(Z) \cup x) = (\Sigma(Y) \cap \Sigma(Z)) \cup (\Sigma(Y) \cap x) \\ &= ((\Sigma(Y) \cap \Sigma(Z)) \cap x) \cup (\Sigma(Y) \cap x) = \Sigma(Y) \cap x.\end{aligned}$$

Thus $\Theta \subseteq x \square Y$; and likewise $I \subseteq x \square Z$. Since $I \cup \Theta \vdash \gamma$, we see that $\gamma \in (x \square Y)+(x \square Z)$.

In first order logic, property (*) (*splitting interpolation*) easily follows from the interpolation property as formulated in (3) and (4) of the introduction. Splitting interpolation, and its 'algebraic' counterpart (E4), are more conspicuous in module algebra than (E3), and it may have been for this reason that in [BHK] only (E4) was considered in the context of equational logic. It fails, and this failure complicates the module algebra of equational logic. Essentially this had been noted earlier by Maibaum and Sadler [MS], who also formulated the splitting property (*).

We conclude by considering what it is about predicate logic that makes (E4) follow from (E3), or equivalently, why interpolation in predicate logic implies splitting interpolation. Suppose we have a finitary deduction system — finite formulas, rules involving finitely many formulas — extending equational logic. Note that in equational logic, every formula $\alpha$ is interpreted as its universal closure $\forall \alpha$; and that therefore the correct form of "$\alpha$ implies $\beta$" in the notation of predicate logic is $\forall \alpha \rightarrow \beta$. Suppose then that if $\alpha$ and $\beta$ are formulas (of our system), we also have a formula $(\forall \alpha \rightarrow \beta)$. The precise form of the symbols does not matter, of course, but we *will* use some wellknown conventions; in particular, we write $\alpha \rightarrow \beta \rightarrow \gamma$ instead of $\alpha \rightarrow (\beta \rightarrow \gamma)$. Assume modus ponens and the deduction theorem for formulas, in the following form:

$$A \cup \{\alpha\} \vdash \beta \text{ if and only if } A \vdash \forall \alpha \rightarrow \beta.$$

Suppose interpolation holds, and $A \cup B \vdash \Gamma$. Take any $\gamma \in \Gamma$. Then since our formal system is finitary, there is a finite $B_0 \subseteq B$, such that $A \cup B_0 \vdash \gamma$. Say $B_0 = \{\beta_1, \ldots, \beta_n\}$. Then by the deduction theorem,

$$A \vdash \forall\beta_1 \rightarrow \ldots \rightarrow \forall\beta_n \rightarrow \gamma.$$

By interpolation, there exists a set $I_\gamma$ of formulas with

$$\Sigma(I_\gamma) \subseteq \Sigma(A) \cap \Sigma(\forall\beta_1 \rightarrow \ldots \rightarrow \forall\beta_n \rightarrow \gamma) \subseteq \Sigma(A) \cap (\Sigma(B) \cup \Sigma(\Gamma))$$

such that $A \vdash I_\gamma$ and $I_\gamma \vdash \forall\beta_1 \rightarrow \ldots \rightarrow \forall\beta_n \rightarrow \gamma$. Now let $I \equiv \bigcup_{\gamma \in \Gamma} I_\gamma$; then I satisfies the requirements of splitting interpolation.

**Acknowledgements**

We thank C.P.J. Koymans, G.R. Renardel de Lavalette and J. Heering for their helpful comments.

REFERENCES

[BHK] J.A. Bergstra, J. Heering, P. Klint, *Module algebra*, Report CS-R8617, Department of Computer Science, Centre for Mathematics and Computer Science, Amsterdam (1986). Revised version to appear (1988).

[C] W. Craig, *Three uses of the Herbrand-Gentzen theorem in relating model theory and proof theory*, Journal of Symbolic Logic XXII (1957), 269-285.

[MS] T. Maibaum, M. Sadler, *Axiomatising specification theory*, in: H.-J. Kreowski ed., Recent trends in data type specification, Informatik-Fachberichte 116, Springer-Verlag (1985), 171-177.

[R] G.R. Renardel de Lavalette, *Modularisation, parameterisation, interpolation*, Logic Group Preprint Series 32, Department of Philosophy, University of Utrecht (1988).