Preprint of a paper to appear in the Proceedings of the 1992 International Workshop on Higher Order Logic Theorem Proving and its Applications, Leuven, September 1992, edited by L. J. M. Claesen and M. J. C. Gordon (North-Holland, 1992).

The HOL Logic Extended with Quantification over Type Variables

Thomas F. Melham

University of Cambridge Computer Laboratory, New Museums Site, Pembroke Street, Cambridge, CB2 3QG, England.

Abstract

This paper discusses a proposal to extend the primitive basis of the HOL logic with a very simple form of quantification over types. It is shown how certain practical problems with using the definitional mechanisms of HOL would be solved by the additional expressive power gained by making this extension.

Keyword Codes: F.4.1; I.2.3
Keywords: Mathematical Logic; Deduction and Theorem Proving.

1 Introduction

The version of higher order logic mechanized by the HOL system is essentially Church's formulation of simple type theory [2] extended with explicit rules of definition and with object-language polymorphism of the kind developed by Milner for the LCF logic PPλ [4]. In this paper a further extension to the logic is proposed—namely, the addition of a very limited form of object-language quantification over types. The motivation for this extension comes from a particular technical problem that arises when using the definitional mechanisms provided by the HOL logic. This paper explains this problem and shows how the proposed extension solves it.

It is assumed that the reader is familiar with the details of the HOL logic given in the DESCRIPTION volume of the HOL documentation [5].

2 Types and polymorphism in HOL

Type expressions in the HOL logic have the following syntax:

$$\sigma \quad ::= \quad c \quad | \quad \alpha \quad | \quad \sigma_1 \to \sigma_2 \quad | \quad (\sigma_1, \ldots, \sigma_n)op$$

where σ, σ_1, $\ldots$, σ_n range over types, c ranges over type constants, α ranges over type variables, and op ranges over n-ary type operators (for $n \geq 1$). It is the inclusion of type variables in this syntax of types that makes the HOL logic polymorphic. Typing of terms takes place within the context of an assignment of generic types to constants, and a constant is well-typed at any substitution instance of its generic type.

Theorems that contain polymorphic types are also true for any substitution instance of them, so there is a limited form of implicit universal quantification over types in the HOL logic. More precisely, there is an implicit universal quantification over type variables at the level of *sequents*. A sequent $\Gamma \vdash \mathrm{P}$ means that P is provable by natural deduction from the hypotheses Γ. All occurrences of a type variable in such a sequent are identified by the semantics, and there is an implicit universal quantification over the value of this variable whose scope is the entire sequent. This is reflected in the primitive rule of type instantiation shown below.

$$\frac{\Gamma \vdash \mathrm{P}}{\Gamma \vdash \mathrm{P}[\sigma_1, \ldots, \sigma_n \,/\, \alpha_1, \ldots, \alpha_n]} \quad \alpha_1,\, \ldots,\, \alpha_n \text{ not in } \Gamma.$$

where $\mathrm{P}[\sigma_1, \ldots, \sigma_n \,/\, \alpha_1, \ldots, \alpha_n]$ means the result of simultaneously substituting the type σ_i for the type variable α_i for $1 \leq i \leq n$ at every occurrence of α_i in the term P. The side condition on this rule is a consequence of the fact that if a type variable occurs in both the hypotheses and the conclusion of a sequent, then both occurrences are assumed to denote the same set. In fact, from this primitive rule one can derive the more general instantiation rule

$$\frac{\Gamma \vdash \mathrm{P}}{\Gamma[\sigma_1, \ldots, \sigma_n \,/\, \alpha_1, \ldots, \alpha_n] \vdash \mathrm{P}[\sigma_1, \ldots, \sigma_n \,/\, \alpha_1, \ldots, \alpha_n]}$$

which allows one to substitute types for type variables throughout an entire sequent.

There is an additional side condition on both of the rules shown above, namely the condition that no two distinct term variables in P (and, for the second rule, Γ) may become identified as a result of doing the substitution. The HOL system deals with this side condition by renaming variables as required.

3 Motivation

The proposal of this paper is to extend the very simple form of polymorphism described above with limited object-language quantification over types. In particular, we wish to add primitive terms to the HOL logic of the forms $\forall\alpha.\,\mathrm{P}$ and $\exists\alpha.\,\mathrm{P}$, where α is a type variable and P is a boolean term. Informally, the intended interpretation is that $\mathrm{P}[\sigma/\alpha]$ is true for *all* types σ and for *some* type σ, respectively. Note that we are not proposing an extension to the type language of HOL—the quantifications $\forall\alpha.\,\mathrm{P}$ and $\exists\alpha.\,\mathrm{P}$ are new term constructs, and not type constructs of the kind found (for example) in Girard's system F [3]. The extended logic proposed here resembles system Q, a transfinite type theory due to Andrews [1]. It is, however, still much weaker than Andrews' system.

The motivation for this extension originally arose in connection with work on new derived rules of definition for HOL—particularly in work on derived rules for defining abstract data types. The following sections explain this motivational background. Readers primarily interested in the semantics and other details of the proposed extension may wish to skip to section 4.

3.1 Type specifications

In the latest edition of the HOL documentation [5], the following *type specification* rule for the HOL logic is introduced:

$$\frac{\Gamma_1 \vdash \exists x.\, \mathrm{P}\ x, \quad \Gamma_2 \vdash (\alpha \approx \mathrm{P}{:}\sigma{\to}bool) \supset \mathrm{Q}}{\Gamma_1 \cup \Gamma_2 \vdash \mathrm{Q}[\nu/\alpha]} \qquad \alpha \text{ not in } \sigma \text{ or } \mathrm{P}.$$

where the notation '$\alpha \approx \mathrm{P}$' is an abbreviation for the assertion that there is a bijection between the values of type α and the set of values that satisfy P:

$$\exists f{:}\alpha{\to}\sigma.\, (\forall a_1\, a_2.\, (f\ a_1 = f\ a_2) \supset (a_1 = a_2)) \wedge (\forall r.\, \mathrm{P}\ r = (\exists a.\, r = f\ a))$$

This (as yet unimplemented) principle of definition is intended to allow new types to be introduced by loose specifications of their properties. The idea is that one finds a non-empty subset P of an existing type σ to represent the values of desired type. One then proves that if the type α is in bijection with this set, then the desired property Q holds of α. If this can be proved, then it is consistent to extend the logic with a new type ν for which $\mathrm{Q}[\nu/\alpha]$ holds.

With the usual HOL rule of type definition, a new type is defined by an axiom that explicitly relates its values to those of a particular subset of an existing type. By contrast, all that is known about a type introduced by the above rule is that it has the properties given by the formula Q. Explicit information about representation is not part of the definition of a type, so types can be only loosely specified by this rule.

A very natural intended application of this type specification rule is in introducing types whose specifications are given by a collection of constants with particular properties—for example, abstract data types with equational specifications. The property by which the new type is specified takes the following general form:

$$\exists v_1 \ \ldots \ v_n.\mathrm{Q}[v_1, \ldots, v_n]$$

That is, one shows that there are values $v_1, \ldots, v_n$ involving the new type (for example, operations over it) for which $\mathrm{Q}[v_1, \ldots, v_n]$ holds. A type having this property is then introduced by a type specification, and the constant specification rule can subsequently be used to introduce constants $\mathsf{c}_1, \ldots, \mathsf{c}_n$ that denote these values. The result is a data type characterised by the theorem $\vdash \mathrm{Q}[\mathsf{c}_1, \ldots, \mathsf{c}_n]$ and defined in such a way that consequences of this defining property are all that can be proved about the type.

An example application of this approach is in defining a type *two*, about which all that is known is that it has *at least* two distinct elements. This type can be defined by first proving the theorems

$$\vdash \exists x.\, (\lambda x{:}bool.\mathsf{T})\ x \quad \text{and} \quad \vdash (\alpha \approx (\lambda x{:}bool.\mathsf{T})) \supset \exists x{:}\alpha.\, \exists y{:}\alpha.\neg(x = y)$$

One can then use the type specification rule to introduce the type *two* satisfying

$$\vdash \exists x{:}two.\, \exists y{:}two.\neg(x = y)$$

Finally, two constants one and two can be introduced for the two values asserted to exist by this theorem, giving the characterising theorem $\vdash \neg(\mathsf{one} = \mathsf{two})$. The result is a type about which we know only that it has at least two distinct values, namely one and two.

3.2 The problem

The problem that motivates the present proposal to extend the HOL logic is illustrated by the following attempt to use the type definition method described above. Suppose we wish to define in HOL an abstract data type of non-empty bit-vectors, specified algebraically using two unit vectors t and f and an associative concatenation operation c. Suppose further that we wish to use the methodology sketched above—that is, we wish first to prove that the desired characterizing property holds of some appropriate representing set of values and then to introduce the type with this property using a type specification. In this case, we would proceed by first proving

$$
\begin{array}{l}
\vdash (\alpha \approx (\lambda l{:}(\mathit{bool})\mathit{list}.\, \neg(l = [\,]))) \;\supset \\
\quad \exists t{:}\alpha.\, \exists f{:}\alpha.\, \exists c{:}\alpha\to\alpha\to\alpha. \\
\quad\; (\forall b_1\, b_2\, b_3.\, c\ b_1\ (c\ b_2\ b_3) = c\ (c\ b_1\ b_2)\ b_3) \wedge \\
\quad\; (\forall x{:}\beta.\, \forall y{:}\beta.\, \forall op{:}\beta\to\beta\to\beta. \\
\quad\quad (\forall x\, y\, z.\, op\ x\ (op\ y\ z) = op\ (op\ x\ y)\ z) \;\supset \\
\quad\quad \exists!\mathit{fn}{:}\alpha\to\beta.\, (\mathit{fn}\ t = x) \wedge (\mathit{fn}\ f = y) \wedge \\
\qquad\qquad\qquad (\forall b_1\, b_2.\, \mathit{fn}\ (c\ b_1\ b_2) = op\ (\mathit{fn}\ b_1)\ (\mathit{fn}\ b_2)))
\end{array}
$$

This theorem says that if α is in bijection with the set of non-empty lists of booleans, then we have an abstract characterization of α as an initial algebra with two elements t and f and an associative concatenation operation c. In particular, we have that there is a unique homomorphism from α to any other type β with two elements x and y and an associative operation op.

Given this theorem, one can use the type specification rule shown in the previous section to introduce a type constant *vect* with the following defining property.

$$
\begin{array}{l}
\vdash \exists t{:}\mathit{vect}.\, \exists f{:}\mathit{vect}.\, \exists c{:}\mathit{vect}\to\mathit{vect}\to\mathit{vect}. \\
\quad (\forall b_1\, b_2\, b_3.\, c\ b_1\ (c\ b_2\ b_3) = c\ (c\ b_1\ b_2)\ b_3) \wedge \\
\quad (\forall x{:}\beta.\, \forall y{:}\beta.\, \forall op{:}\beta\to\beta\to\beta. \\
\quad\; (\forall x\, y\, z.\, op\ x\ (op\ y\ z) = op\ (op\ x\ y)\ z) \;\supset \\
\quad\; \exists!\mathit{fn}{:}\mathit{vect}\to\beta.\, (\mathit{fn}\ t = x) \wedge (\mathit{fn}\ f = y) \wedge \\
\qquad\qquad\qquad (\forall b_1\, b_2.\, \mathit{fn}\ (c\ b_1\ b_2) = op\ (\mathit{fn}\ b_1)\ (\mathit{fn}\ b_2)))
\end{array}
\tag{1}
$$

Having introduced the type *vect* with the above specification, it is natural to wish to use the constant specification rule to give names to the values t, f and c. This would yield the theorem

$$
\begin{array}{l}
\vdash (\forall b_1\, b_2\, b_3.\, \mathsf{c}\ b_1\ (\mathsf{c}\ b_2\ b_3) = \mathsf{c}\ (\mathsf{c}\ b_1\ b_2)\ b_3) \wedge \\
\quad (\forall x{:}\beta.\, \forall y{:}\beta.\, \forall op{:}\beta\to\beta\to\beta. \\
\quad\; (\forall x\, y\, z.\, op\ x\ (op\ y\ z) = op\ (op\ x\ y)\ z) \;\supset \\
\quad\; \exists!\mathit{fn}{:}\mathit{vect}\to\beta.\, (\mathit{fn}\ \mathsf{t} = x) \wedge (\mathit{fn}\ \mathsf{f} = y) \wedge \\
\qquad\qquad\qquad (\forall b_1\, b_2.\, \mathit{fn}\ (\mathsf{c}\ b_1\ b_2) = op\ (\mathit{fn}\ b_1)\ (\mathit{fn}\ b_2)))
\end{array}
\tag{2}
$$

where t and f are constants of type *vect*, and the concatenation operator c is a constant of type *vect*→*vect*→*vect*. This final theorem is the desired algebraic characterization of bit-vectors. The general approach to defining other equationally-specified abstract data types would be similar.

The trouble with the construction sketched above is that the primitive rule of constant specification in HOL does not admit the direct inference of theorem (2) from theorem (1). The rule of constant specification is:

$$\frac{\vdash \exists x_1{:}\sigma_1 \; \ldots \; x_n{:}\sigma_n.\, \mathrm{P}}{\vdash \mathrm{P}[\mathsf{c}_1, \ldots, \mathsf{c}_n / x_1, \ldots, x_n]} \quad \mathrm{tyvars}(\sigma_i) = \mathrm{tyvars}(\mathrm{P}) \text{ for } 1 \leq i \leq n.$$

The side condition on this rule requires all the type variables in the defining property P to be included in the types of the constants to be introduced. But in the defining property given by theorem (1) there is a type variable β which does not appear in the type of the existentially quantified variables t, f and c, so the constant specification rule cannot in this case be applied.

The problem illustrated by this example is closely related to the implicit quantification over type variables mentioned in the previous section. Consider a theorem of the form $\vdash \exists x{:}\sigma.\, \mathrm{P}$ where the type variable α appears in P but not in the type σ. Informally, what is meant by such a theorem is:

'for any type τ, we have $\vdash \exists x{:}\sigma.\, \mathrm{P}[\tau/\alpha]$'.

That is, for any type τ, there is a value x such that $\mathrm{P}[\tau/\alpha]$ comes out true. In general, of course, the choice of value x that makes this term true may depend on the type τ. And if the type variable α does not appear in the type of x, then this dependence will be lost if a constant is introduced to denote this value—hence the side-condition on the rule for constant specifications. For example, in the theorem

$$\vdash \exists b.\, b = \forall x{:}\alpha.\, \forall y{:}\alpha.\, x = y$$

the value b must be either true or false, depending on whether the type α denotes a set with exactly one element or not. It is therefore unsound to admit a constant specification

$$\vdash \mathsf{b} = \forall x{:}\alpha.\, \forall y{:}\alpha.\, x = y$$

in which a fixed name b is given to this value, since then by instantiating α in two different ways one can obtain from this theorem both $\vdash \mathsf{b} = \mathsf{T}$ and $\vdash \mathsf{b} = \mathsf{F}$.

Returning to the example, consider again the defining property for the type *vect* given by theorem (1). This contains a type variable β that does not appear in the types of the variables t, f and c, and so the desired constant specification cannot be made. In this particular case, however, the values t, f and c that make the specification true do not, in fact, depend on the set denoted by type variable β; there are three fixed values t, f and c that make this specification hold for *any* type β. This suggests that the defining property for *vect* should be something more like:

$$\begin{array}{l}
\vdash \exists t{:}\mathit{vect}.\, \exists f{:}\mathit{vect}.\, \exists c{:}\mathit{vect}{\rightarrow}\mathit{vect}{\rightarrow}\mathit{vect}. \\
\quad (\forall b_1\, b_2\, b_3.\, c\ b_1\ (c\ b_2\ b_3) = c\ (c\ b_1\ b_2)\ b_3) \wedge \\
\quad (\forall \boldsymbol{\beta}.\, \forall x{:}\beta.\, \forall y{:}\beta.\, \forall op{:}\beta{\rightarrow}\beta{\rightarrow}\beta. \\
\quad\quad (\forall x\, y\, z.\, op\ x\ (op\ y\ z) = op\ (op\ x\ y)\ z) \supset \\
\quad\quad \exists!\, fn{:}\mathit{vect}{\rightarrow}\beta.\, (fn\ t = x) \wedge (fn\ f = y) \wedge \\
\quad\quad\quad\quad\quad (\forall b_1\, b_2.\, fn\ (c\ b_1\ b_2) = op\ (fn\ b_1)\ (fn\ b_2)))
\end{array} \qquad (3)$$

where there is a universal quantification over the type variable β (shown here in boldface) within the scope of the existential quantification over t, f and c.

This new theorem more accurately expresses the initiality property of the type *vect* than theorem (1). Moreover, if the quantification over β is taken to bind the type variable β, and if the function tyvars in the constant specification rule shown above is reinterpreted as returning the *free* type variables in a term, then it would be valid to infer from this theorem the desired constant specification

$$\begin{array}{l}
\vdash (\forall b_1\, b_2\, b_3.\, \mathsf{c}\ b_1\ (\mathsf{c}\ b_2\ b_3) = \mathsf{c}\ (\mathsf{c}\ b_1\ b_2)\ b_3) \wedge \\
\quad (\forall \beta.\, \forall x{:}\beta.\, \forall y{:}\beta.\, \forall op{:}\beta{\to}\beta{\to}\beta. \\
\qquad (\forall x\, y\, z.\, op\ x\ (op\ y\ z) = op\ (op\ x\ y)\ z) \supset \\
\qquad \exists!\, fn{:}vect{\to}\beta.\, (fn\ \mathsf{t} = x) \wedge (fn\ \mathsf{f} = y) \wedge \\
\qquad\qquad\qquad\qquad (\forall b_1\, b_2.\, fn\ (\mathsf{c}\ b_1\ b_2) = op\ (fn\ b_1)\ (fn\ b_2)))
\end{array}$$

This is the correct form of algebraic specification for the type *vect*.

4 Universal quantification over types

The proposal of this paper is to extend the HOL logic with universal quantification over types, so that properties like that given by theorem (3) in the previous section can be expressed directly in the logic. With this extension to HOL, the syntax of untyped terms becomes:

$$\mathrm{M} \quad ::= \quad c \quad | \quad v \quad | \quad \mathrm{M\,N} \quad | \quad \lambda v.\, \mathrm{M} \quad | \quad \forall \alpha.\, \mathrm{M}$$

That is, the term language of HOL (i.e. the simply-typed λ-calculus) is extended with terms of the form $\forall \alpha.\, \mathrm{M}$, where α is a type variable and M is a (boolean) term. The intended meaning of such a term is that the proposition represented by M is true no matter what set is denoted by the type variable α.

More precisely, we take the syntax of types for the extended logic to be

$$\sigma \quad ::= \quad bool \quad | \quad c \quad | \quad v \quad | \quad \sigma_1 \to \sigma_2 \quad | \quad (\sigma_1, \ldots, \sigma_n)op$$

That is, we assume right from the start that the type expressions include the special type constant *bool*. This is necessary because we wish to form quantifications $\forall \alpha.\, \mathrm{M}$ only for *boolean* terms M. The extended syntax of typed terms is then given by

$$\mathrm{M} \quad ::= \quad c_\sigma \quad | \quad v_\sigma \quad | \quad (\mathrm{M}_{\sigma_1 \to \sigma_2}\ \mathrm{N}_{\sigma_1})_{\sigma_2} \quad | \quad (\lambda v_{\sigma_1}.\, \mathrm{M}_{\sigma_2})_{\sigma_1 \to \sigma_2} \quad | \quad (\forall \alpha.\, \mathrm{M}_{bool})_{bool}$$

where (as usual) a constant c_σ is a well-formed term only if σ is a substitution instance of the generic type of c, and where a quantification $\forall \alpha.\, \mathrm{M}$ is well-formed only if for every occurrence of a variable v_σ in M, either the type variable α does not appear in the type σ or the occurrence of v_σ is bound in M by a textually enclosing lambda-abstraction λv_σ. In what follows, it is assumed that all terms are well-formed according to these formation rules.

A quantification $\forall \alpha.\, \mathrm{M}$ binds occurrences of the type variable α in M, so this extended syntax of terms comes with a notion of free and bound type variable, in addition to the usual notion of free and bound (term) variables. Furthermore, the standard notion of a free occurrence of a term in another term must be modified to take into account binding by type quantifiers. Following the definition given by Andrews for his system Q, we say that an occurrence of a term M_σ is *free for* the variable v_σ in a term N if it does not occur after the dot in a subterm of N that has the form $\lambda v_\sigma.\,\mathrm{P}$ and if it does not occur after the dot in any subterm of the form $\forall\,\alpha.\,\mathrm{P}$ where α occurs in σ. We define a free occurrence of a variable v_σ in a term N as an occurrence which is free for itself in N. In the presence of the well-formedness condition stated above for type quantifications, this notion of free variable is equivalent to the usual HOL definition: a variable v_σ is free in N if it does not appear after the dot in a subterm of the form $\lambda v_\sigma.\,\mathrm{P}$.

In what follows, we assume suitably-defined notions of term and type substitution, with renaming of bound term and type variables as appropriate to avoid variable capture. In particular, we assume for type substitution that if $\sigma s = \sigma_1, \ldots, \sigma_n$ are types and $\alpha s = \alpha_1, \ldots, \alpha_n$ are distinct type variables, then the metalinguistic notation $\mathrm{M}[\sigma s/\alpha s]$ stands for the result of simultaneously substituting σ_i for α_i for $1 \leq i \leq n$ at every free occurrence of α_i in the term M, with the condition that no type variable in any σ_i and no free term variable in M becomes bound in the result of the substitution. In practice, an implementation will rename type and term variables in order to satisfy this condition. So, for example,

$$\begin{aligned}(\forall\beta.\,(x{:}\alpha = x{:}\alpha))[\beta/\alpha] &= \forall\beta'.\,(x{:}\beta = x{:}\beta)\\ (\forall x{:}\alpha.\,x{:}bool)[bool/\alpha] &= \forall x'{:}bool.\,x{:}bool\end{aligned}$$

Similar renamings are also carried out in the current HOL implementation; see [5].

4.1 Semantics

Chapters 9 and 10 of the DESCRIPTION volume of the HOL manual [5] contain a set-theoretic semantics of the HOL logic due to Andy Pitts. In this section, we very briefly sketch this semantics and show how it can be extended to cover the quantification over type variables described above.

The semantics of HOL is defined in terms of particular set $\mathcal{U}$ called the *universe*, the elements of which are the sets denoted by the (monomorphic) type expressions. Among other things, the universe is closed under the function space operation and contains the distinguished two-element set $2 = \{1, 0\}$ which is the meaning of the boolean type *bool*.

The semantics of types is given relative to a model M which assigns to each type constant an element of $\mathcal{U}$ and to each n-ary type operator a function $\mathcal{U}^n \to \mathcal{U}$. The notion of a *type-in-context* is also used in defining the semantics of types. A *type context* αs is just a finite list of distinct type variables, and a type-in-context $\alpha s.\sigma$ is a type σ together with a type context αs which contains (at least) all the type variables in σ. The meaning of a type in context $\alpha s.\sigma$, where the context αs is of length n, is then given by a function

$$[\![\alpha s.\sigma]\!]_M : \mathcal{U}^n \to \mathcal{U}$$

which is defined so that for any assignment of sets $Xs = (X_1, \ldots, X_n) \in \mathcal{U}^n$ to the type variables in αs (and hence to the type variables in σ), the element $[\![\alpha s.\sigma]\!]_M(Xs)$ of $\mathcal{U}$ is the corresponding set denoted by σ. The details of the definition of $[\![\_]\!]_M$, which proceeds by induction on the structure of types, can be found in [5].

The notion of a context is also employed in defining the meaning of terms. A *term-in-context* is written 'αs,xs.M' and consists of a term M together with a type context αs and a finite list of variables xs called a *variable context.* The variable context xs of a term-in-context αs,xs.M contains all the variables that occur free in M, and the type context αs contains all the type variables that occur in xs and M. It is evident that this notion of terms in context is easily adapted for the extended term language described above; we simply make the condition that the type context for a term-in-context αs,xs.M contains all the type variables that appear in xs or appear *free* in M. Following [5], we assume that xs contains no variable that appears bound in M—bound term variables may always be renamed to satisfy this condition. Likewise, we assume that the type context αs contains no type variables that are bound in M. This assumption is valid because the well-formedness condition on terms of the form $\forall\alpha.\,$M ensures that M contains no *free* variable v_σ where α occurs in σ. Hence bound type variables can always be renamed so as to be distinct from all type variables in αs.

For the semantics of terms, a model consists of a type model (as described above) together with a function that assigns to each constant c with generic type σ an element of the set of functions

$$\prod_{Xs \in \mathcal{U}^n} [\![\alpha s.\sigma]\!]_M(Xs)$$

where n is the length of the type context αs. For a given model M, the meaning of a term-in-context αs,xs.M, where αs has length n, xs has length m, and M has type τ, is given by a function $[\![\_]\!]_M$ defined by induction on terms such that:

$$[\![\alpha s,xs.\mathrm{M}]\!]_M \in \prod_{Xs \in \mathcal{U}^n} \left(\prod_{j=1}^{m} [\![\alpha s.\sigma_j]\!]_M(Xs) \right) \to [\![\alpha s.\tau]\!]_M(Xs).$$

where $xs = x_1, \ldots, x_m$ and σ_i is the type of the corresponding variable x_i. The idea is that given an assignment of sets

$$Xs = (X_1, \ldots, X_n) \in \mathcal{U}^n$$

to the type variables in αs (and hence to the free type variables in M) and given an assignment of elements

$$ys = (y_1, \ldots, y_m) \in [\![\alpha s.\sigma_1]\!]_M(Xs) \times \cdots \times [\![\alpha s.\sigma_m]\!]_M(Xs)$$

to the variables in xs (and hence to the variables that occur free in the term M), the result of $[\![\alpha s,xs.\mathrm{M}]\!]_M(Xs)(ys)$ will be an appropriate element of the set $[\![\alpha s.\tau]\!]_M(Xs)$ denoted by the type of M. Again, full details of the definition of $[\![\_]\!]_M$ can be found in [5].

A sequent with hypotheses $\Gamma = \{\mathrm{P}_1, \ldots, \mathrm{P}_p\}$ and conclusion P is *satisfied* by a model M if any assignment of values to free variables that makes all the hypotheses true in M also makes the conclusion true in M. In particular, M satisfies the sequent if for all $Xs \in \mathcal{U}^n$ and all $ys \in [\![\alpha s.\sigma_1]\!]_M(Xs) \times \cdots \times [\![\alpha s.\sigma_m]\!]_M(Xs)$,

$$[\![\alpha s,xs.\mathrm{P}_1]\!]_M(Xs)(ys) = 1, \quad \ldots, \quad [\![\alpha s,xs.\mathrm{P}_p]\!]_M(Xs)(ys) = 1$$

imply that

$$[\![\alpha s,xs.\mathrm{P}]\!]_M(Xs)(ys) = 1,$$

where $\alpha s, xs$ is any valid context for each of P, P_1, ..., P_p with αs of length n, $xs = x_1, \ldots, x_m$, and σ_i the type of the corresponding variable x_i. We write $\Gamma \models_M \mathrm{P}$ to mean that M satisfies the sequent with hypotheses Γ and conclusion P.

It is completely straightforward to extend the definition of $[\![\_]\!]_M$, and hence the notion of satisfaction, to cover type quantifications $\forall\alpha.\,\mathrm{M}$. Suppose $\alpha s,xs.\forall\alpha.\,\mathrm{M}$ is a term-in-context, with αs of length n, and $xs = x_1, \ldots, x_m$ with σ_i the type of x_i. Then, since the term M has type *bool* we have

$$[\![\alpha s,\alpha,xs.\mathrm{M}]\!]_M \in \prod_{Xs \in \mathcal{U}^{n+1}} \left(\prod_{j=1}^{m} [\![\alpha s,\alpha.\sigma_j]\!]_M(Xs) \right) \to 2$$

Note that α is distinct from all the type variables in αs, by the assumption that this type context contains no type variable bound in $\forall\alpha.\,\mathrm{M}$. This ensures that $\alpha s,\alpha,xs.\mathrm{M}$ is in fact a well-formed term-in-context. Now define

$$[\![\alpha s,xs.\forall\alpha.\,\mathrm{M}]\!]_M \in \prod_{Xs \in \mathcal{U}^{n}} \left(\prod_{j=1}^{m} [\![\alpha s.\sigma_j]\!]_M(Xs) \right) \to 2$$

by taking $[\![\alpha s,xs.\forall\alpha.\,\mathrm{M}]\!]_M(Xs)(ys)$ to be 1 if and only if the function in $\mathcal{U} \to 2$ given by $X \mapsto [\![\alpha s,\alpha,xs.\mathrm{M}]\!](Xs, X)(ys)$ yields 1 for all $X \in \mathcal{U}$.

4.2 Primitive inference rules

Having added propositions $\alpha.\,\mathrm{M}$ as primitive terms to HOL, the next step is to extend the logic with new primitive inference rules for reasoning with them. The required rules are the introduction and elimination rules for universal quantification over types shown in figure 1 below. Note that in both rules, it is assumed that $\forall\alpha.\,\mathrm{P}$ is a well formed term—that is, that P contains no free variable v_σ with α occurring in σ.

$$\forall\text{-E:}\quad \frac{\Gamma \vdash \forall\alpha.\,\mathrm{P}}{\Gamma \vdash \mathrm{P}[\sigma/\alpha]} \qquad\qquad \forall\text{-I:}\quad \frac{\Gamma \vdash \mathrm{P}{:}bool}{\Gamma \vdash \forall\alpha.\,\mathrm{P}} \quad \alpha \text{ not free in } \Gamma.$$

Figure 1: Rules for universal quantification over types.

In what follows, we sketch the proof that the elimination rule $\forall$-E rule for universal type quantification is sound. That is, we show that if a model satisfies the hypothesis of the rule, then it also satisfies the conclusion. The proof depends on the series of lemmas given below.

Lemma 4.1 *Suppose α does not occur in $\alpha s, xs$ and α does not occur in σ, and suppose that $\forall\alpha.\, \mathrm{P}$ is well-formed. Then if $\alpha s, xs$ is a valid context for each of $\mathrm{P}[\sigma/\alpha]$, $\mathrm{P}_1, \ldots, \mathrm{P}_p$, then $\alpha s, xs$ is also a valid context for each of $\forall\alpha.\, \mathrm{P}$, $\mathrm{P}_1, \ldots, \mathrm{P}_p$.*

Proof. Obvious, by the definitions of well-formedness and valid term contexts.[1] $\square$

Lemma 4.2 *If $\alpha s, \alpha.\sigma$ is a type-in-context where α does not occur in σ, then*

$$[\![\alpha s, \alpha.\sigma]\!]_M(Xs, X) = [\![\alpha s.\sigma]\!]_M(Xs)$$

for all $X \in \mathcal{U}$ and $Xs \in \mathcal{U}^n$ where n is the length of αs.

Proof. By structural induction on σ. $\square$

Lemma 4.3 *Suppose $\forall\alpha.\, \mathrm{P}$ is a well-formed term (and hence P contains no free variable v_τ where α occurs in τ). If $\alpha s, xs.\mathrm{P}[\sigma/\alpha]$ is a term-in-context where α does not occur in $\alpha s, xs$ and α does not occur in σ, then for all $Xs \in \mathcal{U}^n$, where n is the length of αs*

$$[\![\alpha s, xs.\mathrm{P}[\sigma/\alpha]]\!]_M(Xs) = [\![\alpha s, \alpha, xs.\mathrm{P}]\!]_M(Xs, [\![\alpha s.\sigma]\!](Xs))$$

Proof. By structural induction on P, using lemma 4.2 to establish that the left- and right-hand sides are functions with the same domain. $\square$

Theorem 4.4 *Rule $\forall$-E is sound. I.e. for any model M and type σ, if $\Gamma \models_M \forall\alpha.\, \mathrm{P}$, then $\Gamma \models_M \mathrm{P}[\sigma/\alpha]$.*

Proof. We may suppose that α appears nowhere in Γ and that σ does not contain α, as the bound type variable α can be renamed if necessary without changing the semantics. Now suppose that $\Gamma \models_M \forall\alpha.\, \mathrm{P}$ with $\Gamma = \{\mathrm{P}_1, \ldots, \mathrm{P}_p\}$. Then for any $\alpha s, xs$ which is a context of each of $\forall\alpha.\, \mathrm{P}$, $\mathrm{P}_1, \ldots, \mathrm{P}_p$ where the length of αs is n, $xs = x_1, \ldots, x_m$, and x_i has type σ_i, we have that for all $Xs \in \mathcal{U}^n$ and $ys \in [\![\alpha s.\sigma_1]\!]_M(Xs) \times \cdots \times [\![\alpha s.\sigma_m]\!]_M(Xs)$, if

$$[\![\alpha s, xs.\mathrm{P}_k]\!]_M(Xs)(ys) = 1 \quad \text{for } 1 \leq k \leq p$$

then $[\![\alpha s, xs.\forall\alpha.\, \mathrm{P}]\!]_M(Xs)(ys) = 1$, which by the semantics of $\forall\alpha.\, \mathrm{P}$ given above means that for all $X \in \mathcal{U}$ we have

$$[\![\alpha s, \alpha, xs.\mathrm{P}]\!]_M(Xs, X)(ys) = 1.$$

Now, suppose that $\alpha s', xs'$ is a context for each of $\mathrm{P}[\sigma/\alpha]$, $\mathrm{P}_1, \ldots, \mathrm{P}_p$. We may assume that $\alpha s', xs'$ does not contain α, since α does not appear in any of $\mathrm{P}[\sigma/\alpha]$, $\mathrm{P}_1, \ldots, \mathrm{P}_p$. Let n' be the length of $\alpha s'$, $xs' = x'_1, \ldots, x'_{m'}$, and σ'_i be the type of the corresponding variable x'_i. Assume that for some $Xs' \in \mathcal{U}^{n'}$ and $ys' \in [\![\alpha s.\sigma'_1]\!]_M(Xs') \times \cdots \times [\![\alpha s'.\sigma'_{m'}]\!]_M(Xs')$ we have for the hypotheses that

$$[\![\alpha s', xs'.\mathrm{P}_k]\!]_M(Xs')(ys') = 1 \quad \text{for } 1 \leq k \leq p.$$

[1] Note that we have assumed that no bound type variables in P have been renamed in $\mathrm{P}[\sigma/\alpha]$. One can always contrive for this to be the case by doing appropriate renamings beforehand.

By lemma 4.1 we have that $\alpha s', xs'$ is a valid context for each of $\forall\alpha.\,\mathrm{P}$, $\mathrm{P}_1, \ldots, \mathrm{P}_p$, so the above assumption implies that for all $X \in \mathcal{U}$

$$[\![\alpha s', \alpha, xs'.\mathrm{P}]\!]_M(Xs', X)(ys') = 1 \tag{4}$$

One then has by lemma 4.3 that

$$[\![\alpha s', xs'.\mathrm{P}[\sigma/\alpha]]\!]_M(Xs')(ys') = [\![\alpha s', \alpha, xs'.\mathrm{P}]\!]_M(Xs', [\![\alpha s'.\sigma]\!](Xs'))(ys')$$

and hence by (4) that $[\![\alpha s', xs'.\mathrm{P}[\sigma/\alpha]]\!]_M(Xs')(ys') = 1$, as required. □

The proof of soundness for the introduction rule ∀-I is left as an exercise for the reader.

4.3 Derived inference rules

Various higher-level inference rules for universal quantification over types are derivable from the primitive introduction and elimination rules given in the previous section. We give the derivations of some of the more important rules below.

The significance of the following congruence rule is that it is needed to extend the implementation of HOL's conversion-based rewriting; see [5] for the details.

Theorem 4.5 *Universal quantification over types satisfies the following congruence rule*

$$\forall\text{-EQ}:\quad \frac{\Gamma \vdash (\mathrm{M}{:}bool) = (\mathrm{N}{:}bool)}{\Gamma \vdash (\forall\alpha.\,\mathrm{M}) = (\forall\alpha.\,\mathrm{N})} \quad \alpha \text{ not free in } \Gamma.$$

Proof. By the following derivation.

1. $\Gamma, \forall\alpha.\,\mathrm{M} \vdash \forall\alpha.\,\mathrm{M}$	[`ASSUME`, `ADD_ASSUM`]
2. $\Gamma, \forall\alpha.\,\mathrm{M} \vdash \mathrm{M}$	[∀-E rule: 1]
3. $\Gamma, \forall\alpha.\,\mathrm{M} \vdash \mathrm{N}$	[`EQ_MP`: hypothesis, 2]
4. $\Gamma, \forall\alpha.\,\mathrm{M} \vdash \forall\alpha.\,\mathrm{N}$	[∀-I rule: 3]
5. $\Gamma \vdash \forall\alpha.\,\mathrm{M} \supset \forall\alpha.\,\mathrm{N}$	[`DISCH`: 4]
6. $\Gamma \vdash \forall\alpha.\,\mathrm{N} \supset \forall\alpha.\,\mathrm{M}$	[similar to steps 1–5]
7. $\Gamma \vdash \forall\alpha.\,\mathrm{M} = \forall\alpha.\,\mathrm{N}$	[`EQ_IMP_RULE`: 5, 6]

Note that the side condition on the ∀-EQ rule is necessary for the steps in this proof that use the introduction rule ∀-I to be valid. □

The following theorem shows that the existing HOL type instantiation rule, described above in section 2, can be derived in the extended logic.

Theorem 4.6 *The following type instantiation rule is derivable.*

$$\forall\text{-INST}:\quad \frac{\Gamma \vdash \mathrm{P}}{\Gamma \vdash \mathrm{P}[\sigma/\alpha]} \quad \alpha \text{ not free in } \Gamma.$$

where there is the additional side condition that no two distinct term variables in P *become identified as a result of doing the substitution (see section 2).*

Proof. Let $x_1, \ldots, x_n$ be all the free variables in P whose types contain α.

1.	$\Gamma \vdash \mathrm{P}$	[hypothesis]
2.	$\Gamma \vdash \forall x_1. \cdots \forall x_n. \mathrm{P}$	[`GENL` 1]
3.	$\Gamma \vdash \forall \alpha. \forall x_1. \cdots \forall x_n. \mathrm{P}$	[$\forall$-I rule: 2]
4.	$\Gamma \vdash (\forall x_1. \cdots \forall x_n. \mathrm{P})[\sigma/\alpha]$	[$\forall$-E rule: 3]
5.	$\Gamma \vdash \forall x_1[\sigma/\alpha]. \cdots \forall x_n[\sigma/\alpha]. (\mathrm{P}[\sigma/\alpha])$	[defn. of type substitution: 4]
6.	$\Gamma \vdash \mathrm{P}[\sigma/\alpha]$	[`SPECL`: 5]

Note that step 2 is valid because the side condition that α is not free in Γ ensures that none of $x_1, \ldots, x_n$, each of whose types contains α, is free in Γ. Step 3 is valid because the generalizations in step 2 ensure that '$\forall \alpha. \forall x_1. \cdots \forall x_n. \mathrm{P}$' will be well-formed. Step 5 is valid because the additional side condition mentioned above ensures that no x_i must be renamed to avoid capture by type substitution of other variables free in P. □

It is straightforward use this $\forall$-INST rule to derive the full rule for simultaneous type instantiation shown in section 2. This means that a primitive type instantiation rule, as present in the existing HOL logic, is redundant in the extended system—it can instead be a derived rule. The rule shown in section 2 for instantiation throughout a sequent is also derivable.

Theorem 4.7 *The following alpha-conversion rule holds.*

$$\forall\text{-ACONV:}\quad \frac{}{\vdash (\forall \alpha. \mathrm{M}) = (\forall \beta. \mathrm{M}[\beta/\alpha])}\quad \beta \text{ not free in } \forall \alpha. \mathrm{M}.$$

Proof. By the following derivation.

1.	$\forall \alpha. \mathrm{M} \vdash \forall \alpha. \mathrm{M}$	[`ASSUME`]
2.	$\forall \alpha. \mathrm{M} \vdash \mathrm{M}[\beta/\alpha]$	[$\forall$-E rule: 1]
3.	$\forall \alpha. \mathrm{M} \vdash \forall \beta. \mathrm{M}[\beta/\alpha]$	[$\forall$-I rule: 2]
4.	$\vdash \forall \alpha. \mathrm{M} \supset \forall \beta. \mathrm{M}[\beta/\alpha]$	[`DISCH`: 3]
5.	$\vdash (\forall \beta. \mathrm{M}[\beta/\alpha]) \supset \forall \alpha. \mathrm{M}$	[similar to steps 1–4]
6.	$\vdash \forall \alpha. \mathrm{M} = (\forall \beta. \mathrm{M}[\beta/\alpha])$	[`EQ_IMP_RULE`: 4, 5]

The side condition that β is not free in $\forall \alpha. \mathrm{M}$ is necessary for step 3 to be valid. It also ensures that $\mathrm{M}[\beta/\alpha][\alpha/\beta] = \mathrm{M}$, which is used in step 5. □

5 Existential quantification over types

Adding universal quantification over types to HOL immediately suggests that existential quantification may be possible as well. The simplest approach is to make a term of the form '$\exists \alpha. \mathrm{P}$' an abbreviation for '$\neg \forall \alpha. \neg \mathrm{P}$'. This cannot, of course, be done as an object-language abbreviation using the existing HOL definitional mechanisms. It could, however, be implemented in HOL as a parser-supported metalinguistic abbreviation. Alternatively, the term language could be further extended to

$$\mathrm{M} \quad ::= \quad c \quad | \quad v \quad | \quad \mathrm{M\,N} \quad | \quad \lambda v. \mathrm{M} \quad | \quad \forall \alpha. \mathrm{M} \quad | \quad \exists \alpha. \mathrm{M}$$

with formation and typing rules similar to those given in the previous section for universal

$$\exists\text{-I:} \quad \frac{\Gamma \vdash \mathrm{P}[\sigma/\alpha]{:}bool}{\Gamma \vdash \exists\alpha.\ \mathrm{P}}$$

$$\exists\text{-E:} \quad \frac{\Gamma \vdash \exists\alpha.\ \mathrm{P}, \quad \Gamma, \mathrm{P}[\beta/\alpha] \vdash \mathrm{Q}}{\Gamma \vdash \mathrm{Q}} \quad \beta \text{ not free in } \Gamma \text{ or } \mathrm{Q}.$$

Figure 2: Derived rules for existential quantification over types.

quantification. Meaning could then be given to terms of the form '$\exists\alpha.\,\mathrm{M}$' by the addition of the following primitive axiom-scheme (conversion):

$$\exists\text{-CONV:} \quad \frac{}{\vdash (\exists\alpha.\,\mathrm{P}) = \neg\forall\alpha.\neg\mathrm{P}} \quad \beta \text{ not free in } \forall\alpha.\,\mathrm{P}$$

This conversion is the only primitive rule required for existential quantification over types. Appropriate introduction and elimination rules for $\exists\alpha.\,\mathrm{P}$ are easily derived using it.

Theorem 5.1 *The introduction and elimination rules in figure 2 are derivable.*

Proof. For the introduction rule, the derivation is

1. $\Gamma \vdash \mathrm{P}[\sigma/\alpha]$ [hypothesis]
2. $\forall\alpha.\,\neg\mathrm{P} \vdash \forall\alpha.\,\neg\mathrm{P}$ [`ASSUME`]
3. $\forall\alpha.\,\neg\mathrm{P} \vdash \neg\mathrm{P}[\sigma/\alpha]$ [$\forall$-E rule: 2]
4. $\Gamma, \forall\alpha.\,\neg\mathrm{P} \vdash \mathsf{F}$ [`MP`: 3, 1]
5. $\Gamma \vdash \neg\forall\alpha.\,\neg\mathrm{P}$ [`DISCH`, `NOT_INTRO`: 4]
6. $\Gamma \vdash \exists\alpha.\,\mathrm{P}$ [$\exists$-CONV]

For the elimination rule, the derivation is

1. $\Gamma \vdash \neg\forall\alpha.\,\neg\mathrm{P}$ [$\exists$-CONV: hypothesis]
2. $\Gamma \vdash \mathrm{P}[\beta/\alpha] \supset \mathrm{Q}$ [`DISCH`: hypothesis]
3. $\Gamma, \neg\mathrm{Q} \vdash \neg\mathrm{P}[\beta/\alpha]$ [`CONTRAPOS`, `UNDISCH`: 2]
4. $\Gamma, \neg\mathrm{Q} \vdash \forall\beta.\,\neg\mathrm{P}[\beta/\alpha]$ [$\forall$-I: 3]
5. $\Gamma, \neg\mathrm{Q} \vdash \forall\alpha.\,\neg\mathrm{P}$ [$\forall$-E, $\forall$-I: 4]
6. $\Gamma, \neg\mathrm{Q} \vdash \mathsf{F}$ [`MP`: 1, 5]
7. $\Gamma \vdash \mathrm{Q}$ [`DISCH`, simplify: 6]

The side condition that β is not free in Γ or Q is needed for the validity of steps 4 and 5 in this second derivation. □

The derivations of congruence and alpha-conversion rules for existential quantification over types are straightforward and will therefore not be given here.

We conclude this section with some preliminary discussion of the idea that extending HOL with existential quantification over types may make it possible to simplify the type specification rule discussed above in section 3.1. In particular, by using existential type quantification the type specification rule may be expressed as

$$\frac{\Gamma \vdash \exists\alpha.\,\mathrm{Q}}{\Gamma \vdash \mathrm{Q}[\nu/\alpha]}$$

That is, if one can prove that there exists a type α with the property given by Q, then it is consistent to extend the logic with a new type expression ν for which $\mathrm{Q}[\nu/\alpha]$ holds. This new formulation of the type specification rule seems somewhat more direct than the one shown in section 3.1, as it does not need the notion of a bijection between sets of values.

With this new formulation, however, the question arises as to whether one can always prove the hypothesis $\Gamma \vdash \exists\alpha.\,\mathrm{Q}$ required for any desired application of the rule. The existential introduction rule $\exists$-I may be used to prove propositions of this form, but this rule requires a witness for the type α asserted to exist. Moreover, this witness must already be expressible in the type language of HOL. It may, therefore, be necessary to add the following additional primitive rule

$$\frac{\Gamma \vdash \exists x:\sigma.\,\mathrm{P}\;x}{\Gamma \vdash \exists\alpha.\,\alpha \approx \mathrm{P}}$$

in order to make the new type specification rule shown above actually useful. This rule states that for any non-empty set of values, there exists a type whose elements are in bijection with this set. There need not, of course, be an actual type expression σ that denotes this set.

6 Implementation

The extensions proposed in the preceding sections have not yet been implemented. Adding terms of the forms $\forall\alpha.\,\mathrm{P}$ and $\exists\alpha.\,\mathrm{P}$ is an extension of the primitive term language of HOL so an implementation based on the existing HOL88 or HOL90 systems would require fairly extensive modifications. The type checker[2] would have to be extended, as would the parser and pretty-printer. The definitions of many of the syntax functions would have to be modified—for example the functions dealing with alpha-conversion, term substitution, type substitution, and free and bound variables. All this should be fairly straightforward, though reimplementing operations like substitution is likely to be error-prone. It is hoped that a prototype system can be be built in the near future, most likely based on HOL90, so that practical experiments can be carried out.

[2]Type checking and type inference for the extended system are still decidable.

Acknowledgements

I am grateful to Andy Pitts for discussions about extending the set-theoretic semantics of HOL, and to the reviewers for helpful comments on the presentation. I have found Andrews' book [1] invaluable in working out many of the technical details of the extension to HOL proposed here. This work was carried out at the University of Cambridge under SERC grant number GR/F 36675.

References

[1] P. B. Andrews, *A Transfinite Type Theory with Type Variables*, Studies in Logic and the Foundations of Mathematics series (North-Holland, 1965).

[2] A. Church, 'A Formulation of the Simple Theory of Types', *The Journal of Symbolic Logic*, Vol. 5 (1940), pp. 56–68.

[3] J.-Y. Girard, 'The System F of Variable Types, Fifteen Years Later', *Theoretical Computer Science*, vol. 45 (1986), pp. 159–192.

[4] M. J. Gordon, A. J. Milner, and C. P. Wadsworth, *Edinburgh LCF: A Mechanised Logic of Computation*, Lecture Notes in Computer Science, vol. 78 (Springer-Verlag, 1979).

[5] University of Cambridge Computer Laboratory, *The HOL System: DESCRIPTION*, revised edition (July 1991).