Subject: Lambda-definability and logical relations
Author: G.D. Plotkin

1. Introduction

In [4] we showed that in every model, $\mathbb{D}_\omega$, of the $\lambda$-calculus as constructed in [8] the strict ordering, $\preceq$, is first-order definable using only application. Here we look at the, perhaps more pertinent, question of definability by pure $\lambda$-terms of such lattice-theoretic entities as $\bot$, $\top$, $\cup$, $\cap$ and $\gamma$, the least fixed-point operator.

The main method will be to construct certain, so-called, logical relations which are satisfied by all (constant vectors of) $\lambda$-definable elements and yet are not satisfied by the lattice-theoretic entity under discussion. The definition of logical is derived from a corresponding one of M. Gordon for the typed $\lambda$-calculus. This in turn generalised the idea of an invariant functional [2]. R. Milne [3] has independently developed analogues of the logical relations for use in equivalence proofs about programming languages.

It is not known whether logical relations also provide sufficient conditions for definability. In the second half of this memorandum we discuss this question for the typed case, obtaining necessary and sufficient conditions by using the more inclusive concept of an I-logical relation.

This memorandum is by no means self-contained. The reader should have some knowledge of both the typed and untyped $\lambda K$-calculi and be fairly familiar with Scott's models of the untyped $\lambda K$-calculus.
2. Pure definability in $D_{\omega}$

A structure $<D,K,S,[\cdot]>\text{ is called a (non-trivial) model of the }\lambda\text{-calculus if } K \text{ and } S\text{ have the usual properties and extensionality holds (and } |D| > 1). \text{ Such structures give a denotational semantics for the }\lambda\text{-}\lambda K\text{-calculus which we will use informally, confusing use and mention. Generally we will consider only the models }D_{\omega}\text{, from [8], given by a Park retraction, }\psi_0 = \lambda f: D_1.f(t),\text{ where } t \text{ is an isolated element of } D_0. \text{ We will often use facts about such models, accompanied by a reference to the proof for the case } t = \perp. \text{ The general proof is always similar. Also needed is the fact that if } Y_{\lambda}\text{ is the paradoxical combinator, }\lambda f(\lambda x.f(xx))(\lambda x.f(xx)), \text{ then, in } D_{\omega}, Y_{\lambda}[f]_{n\geq 0}f^n[f[t] \equiv t \rightarrow t, \perp](f \in D_{\omega}) \text{ [5].}

A relation $R \subseteq D^\kappa$, ($\kappa$ an ordinal) on such a structure is logical iff:

$$\forall \vec{p} \in D^\kappa: (R(\vec{p})) \equiv (\forall \vec{x} \in D^\kappa: (R(\vec{x}) \rightarrow R(\vec{p}[\vec{x}])))$$

Here $\kappa$ is any ordinal and application of vectors is defined pointwise. An element $x \in D$ satisfies $R$ iff $R(\vec{x})$ is true, where $\vec{x} \in D^\kappa$ is the constant vector such that $(\vec{x})_\lambda = x(\lambda \lessdot \kappa)$.

An element $x \in D$ is $\lambda$-definable if $x=M$, for some closed $\lambda$-term $M$; it is $\lambda$-definable from $X \subseteq D$ iff there is a closed term $M$ and $x_1, \ldots, x_n$ in $X$ such that $x=Mx_1 \cdots x_n$.

**Theorem 1**

1. Any closed $\lambda$-term satisfies any logical relation.
2. If $x$ is $\lambda$-definable from $X \subseteq D$, and each element in $X$ satisfies the logical relation $R$, then so does $x$.

**Proof** Clearly, if $x$ and $y$ satisfy a logical relation $R$, so does $x[y]$.

So to finish the proof we need only show that $K$ and $S$ satisfy any such relation. Suppose $R$ is logical. To show $K$ satisfies $R$, assuming $R(\vec{x})$ we must show that $R(\vec{K}[\vec{x}])$. This, in turn, follows if $R(\vec{K}[\vec{x}][\vec{y}])$ when $R(\vec{y})$. But this holds as $\vec{K}[\vec{x}][\vec{y}] = \vec{x}$.

In/
In the same way we see that \( S \) satisfies \( R \) if \( R(\hat{x}^>)[\hat{y}^>][\hat{z}^>] \)
when \( R(\hat{x}^>) \), \( R(\hat{y}^>) \), and \( R(\hat{z}^>) \). But then we have successively, by the
remark made at the beginning of the proof that \( R(\hat{x}^>)[\hat{z}^>] \), \( R(\hat{y}^>)[\hat{z}^>] \) and
\( R(x[\hat{z}^>][\hat{y}^>][\hat{z}^>] \)), concluding the proof.

Nothing is known about the converse of theorem 1. However it will
be very useful for particular cases of undefinability. Here is a way
of constructing logical relations \( R \subseteq D^2 \).

Suppose \( R_0 \subseteq D^2 \). Define \( R_n \subseteq D^2 \) by:

\[
\forall f, g \in D_{n+1}, (R_{n+1}(f, g) \equiv \forall x, y \in D_n (R_n(x, y) \rightarrow R_n(fx, gy))).
\]

Define \( R_\infty \subseteq D^2 \) by:

\[
\forall d, e \in D_\infty (R_\infty(d, e) \equiv \forall n R_n(d_n, e_n)).
\]

**Theorem 2.** Suppose that \( R_0(t, t) \), that \( R_0(d, e) \) implies \( R_1(\varphi^1_d, \varphi^1_e) \),
for any \( d, e \) in \( D_0 \), and that \( R_0 \) is closed under unions of increasing
sequences. Then:

1. \( R_\infty \) is logical.
2. \( R_\infty \) is closed under increasing sequences.
3. If \( R_0 \) is closed under \( \bigcup (\Pi) \) so is \( R_\infty \); if \( R_0(\bot, \bot) \)
   \( (R_0(T_0, T_0)) \) then \( R_\infty(\bot, \bot) \) \( (R_\infty(T_0, T_0)) \).

The construction also works for any \( R_0 \subseteq D^K \) under the corresponding
conditions, and the theorem analogous to theorem 2 can be proved; this
extension will be assumed.

**Lemma 1.1** Suppose that \( R_0(t, t) \) and \( R_0(d, e) \) implies \( R_1(\varphi^1_d, \varphi^1_e) \)
for any \( d, e \) in \( D_0 \). Then,

\[
\forall n \forall f, g \in D_n (R_n(f, g) \rightarrow R_{n+1}(\varphi_n(f), \varphi_n(g))) \text{ and}
\forall n \forall f, g \in D_{n+1} (R_{n+1}(f, g) \rightarrow R_n(\varphi_n(f), \varphi_n(g)));
\]

1.2 If \( R_0 \) is closed under increasing sequences so is each \( R_n \).

**Proof/**
Proof 1.1 By induction on \( n \).

For \( n = 0 \), note that if \( R_1(f, g) \) then \( R_0(ft, gt) \) from the definition of \( R_1 \) and the fact that \( R_0(t, t) \).

For \( n + 1 \), suppose \( R_{n+1}(f, g) \) and suppose \( R_{n+1}(f', g') \).

By induction hypothesis \( R_n(\psi_n, \psi_n) \). Therefore

\[
R_n(f(\psi_n, \psi_n, \psi_n)), \quad \text{and by the induction hypothesis,}
\]

\[
R_{n+1}(\phi \circ f \circ \psi_n(f'), \phi \circ f \circ \psi_n(g')), \quad \text{which shows that}
\]

\[
R_{n+2}(\phi \circ \chi_n, \phi \circ \chi_n). \quad \text{The other half is similar.}
\]

1.2 By induction on \( n \). For \( n + 1 \), let \( \langle f^m, g^m \rangle_{m=0}^\infty \) be an (infinite) increasing sequence in \( R_{n+1} \) and suppose \( R_n(x, y) \). Then

\[
\langle f^m, g^m \rangle_{m=0}^\infty \text{ is an increasing sequence in } R_n \text{ and so}
\]

\[
\langle (\bigcup_m^x), (\bigcup_m^y) \rangle \text{ is in } R_n \text{ by induction hypothesis and the complete additivity of application in its first argument. This concludes the proof.}
\]

Proof of theorem 2 1 First suppose that \( R_\infty(f, g) \) and \( R_\infty(x, y) \). We will show that \( R_\infty(fx, gy) \).

Now \( (fx) = \bigcup_{n=0}^\infty \psi_n(f_{m=0}^n, x) \) and similarly for \( (gy) \).

Since \( R_n(f_{m=0}^n, x, g_{m=0}^n, y) \) is true for any \( m \geq n, R_n(\psi_n(f_{m=0}^n, x), \psi_n(g_{m=0}^n, y)) \) holds by \( m - n \) applications of lemma 1.1, and then we see that \( R_n((fx), (gy)) \) by lemma 1.2 and the above formulae for \( (fx) \) and \( (gy) \).

Conversely, suppose that whenever \( R_\infty(x, y) \) then \( R_\infty(fx, gy) \) and yet for some \( R_n(f_n, g_n) \) is false. By lemma 1.1 we can assume that \( n \neq 0 \), and so for some \( (x_{n-1}, y_{n-1}) \in R_{n-1}(fx_{n-1}, gy_{n-1}) \) is false. Let \( x = f_{n-1} \) and define \( y \) similarly. By lemma 1.1, \( R_\infty(x, y) \) is true and so therefore is \( R_\infty(fx, gy) \) and, consequently,

\[
R_{n-1}((fx)_{n-1}, (gy)_{n-1}).
\]

But \( (fx)_{n-1}(f_{n-1}x_{n-1}) \) and similarly for \( (gy)_{n-1} \) (cf. the laws of application in [9]) and so \( R_{n-1}(f_{n-1}x_{n-1}, g_{n-1}y_{n-1}) \), a contradiction.

2 Suppose \( \langle x^m, y^m \rangle_{m=0}^\infty \) is an (infinite) increasing sequence in \( R_\infty \).

Now...
Now, \( \biguplus_{m=0}^{n} x_{m} \uparrow_{n} \) and similarly for the y's (cf. [7]).

Then one sees, successively that, \( R_{n}((x_{m})_{n},(y_{m})_{n}) \) for all m and n, 
\( R_{n}(\bigcup_{m=0}^{n} x_{m},\bigcup_{m=0}^{n} y_{m}) \), for all n, by lemma 1.2, \( R_{n}(\psi_{n},(\bigcup_{m} x_{m})_{n}) \),
\( \psi_{n},(\bigcup_{m} y_{m})_{n} \) for \( n' \geq n \), by lemma 1.1 and finally \( R_{n}(\bigcup_{m} x_{m},(\bigcup_{m} y_{m})_{n}) \) by lemma 1.2.

A straightforward inductive argument shows that if \( R_{0} \) is closed under \( \bigcup \) so is each \( R_{n} \). Then, clearly, \( R_{\infty}(\lambda x \lambda y(x_{n} \bigcup y_{n})) \),
\( \lambda x \lambda y(x_{n} \bigcup y_{n}) \) and so \( R_{\infty} \) is closed under \( \bigcup \) as \( \bigcup = \bigcup_{n=0}^{\infty} (\lambda x \lambda y(x_{n} \bigcup y_{n})) \) expresses \( \bigcup \) as an increasing sequence in \( D_{\infty} \).

The argument for \( D_{\infty} \) is similar; it uses the fact that if \( f, g : X \rightarrow Y \)
where \( X \) and \( Y \) are continuous lattices then \( (f \sqcap g)x = (fx) \sqcap (gx) \).

If \( R_{0}(\bot, \bot) \) then \( R_{n}(\bot, \bot) \), for any \( n \), by lemma 1.1.

If \( R_{0}(T_{0}, T_{0}) \) an easy inductive argument shows that \( R_{n}(T_{n}, T_{n}) \) for all \( n \), concluding the proof.

As an example, let \( P_{0} = \{ t \} \). Then by the assumed extension of) theorem 2, \( P_{\infty} \) is logical and so the 0th component of any closed \( \lambda \)-term is \( t \). Therefore if \( T \neq \bot \) neither \( \bot \) nor, since \( \bot = Y \), \( Y \) is \( \lambda \)-definable; this is a result of Park [6].

The next example establishes all the definabilities among \( \bot, T, \bigcup \),
\( \Pi \) and \( Y \) for all possible values of \( t \).

**Theorem 3**

1. If \( t = \bot \), \( Y = Y \lambda \) and \( \bot = Y \lambda \).

2. In general, \( \bot = Y \) and \( Y = Y \lambda f(Y \lambda (\lambda g \lambda x f(g \bot)) \).

3. If \( t = T_{0} \), then \( T = Y \lambda \).

4. If \( t = T_{0} \) and \( D_{0} = \{ \bot, T \} \) then
\( \Pi = Y \lambda (\lambda g \lambda x \lambda y \lambda z g(xz)(yz)) \).

2.
2. The only definabilities among $\bot, T, U, \Pi$ and $Y$ are those implied by 1.

**Proof**

1.1 This result is known - see [9].

1.2 $\bot = Y \Pi$ is obvious.

Suppose $f \in D_{\infty}$ and let $T^\alpha = \lambda x \lambda y f(x \bot)$. As $T^\alpha \bot = T^\alpha \bot = \lambda x f(x \bot)$, we see that $\lambda \lambda \Pi = \bigcup_{n \geq 0} T^n(\lambda x f(x \bot))$. By induction on $n$, $T^n(\lambda x f(x \bot)) = \lambda x f^{n+1}(\bot)$, giving $\lambda \lambda \Pi = \lambda x (Yf)$, and the result follows.

1.3 As $Kt \equiv t$, $\lambda \lambda K = \bigcup_{n \geq 0} K^nt$. As $K^nt \equiv T_n$ ($t=T_0$ here), for all $n$, $\lambda \lambda K = T$.

1.4 Let $T^\alpha = \lambda g \lambda x \lambda y \lambda z g(xz)(yz)$. Since, in this $D_{\infty}$, $x \equiv t$ iff $xt \equiv t$, one sees that $T^\alpha \bot \equiv t$. Now, $t = \lambda x \lambda y x_0 \Pi y_0$ is true in this lattice and then $\Pi = Y \lambda \Pi$ follows by the usual inductive argument.

2. As $\bot$ and $Y$ are interdefinable, only definabilities among $\bot, T, U$ and $\Pi$ need be considered.

We must show that if $t \neq \bot$, then $\bot$ is not $\lambda$-definable from $\{T, U, \Pi\}$; that it $t \neq T_0$, $T$ is not $\lambda$-definable from $\{\bot, U, \Pi\}$; that $U$ is not $\lambda$-definable from $\{\bot, T, \Pi\}$ in all cases; and that if $t \neq T_0$ or $D_{\infty} \not \equiv 0$ then $\Pi$ is not $\lambda$-definable from $\{\bot, T, U\}$.

To show that $\bot$ is not $\lambda$-definable from $\{T, U, \Pi\}$, when $t \neq \bot$, let $R_0 = \{t, T_0\}$. The conditions of theorem 2 are easily checked and so $P_0$ is logical. It also follows from theorem 2 that $U$, $\Pi$ and $T$ satisfy $R_0$. Clearly $\bot$ does not. The conclusion then follows from theorem 1.2.

In the rest of the proof we shall first display an appropriate $R_0$ and leave the (admittedly tedious) details to the reader.

To show that if $t \neq T_0$, $T$ is not $\lambda$-definable from $\{\bot, \Pi, U\}$ take $R_0 = \{\bot, t\}$.

To/
The conditional, $\triangleright$, is in $T_\infty \to T_\infty$, and is regarded as being in $T_\infty$, in the usual way. It is defined by:

$$ (z \triangleright x, y) = \begin{cases} x \uplus y & \text{(if } z = T) \\ x & \text{(if } tt \in z \neq T) \\ y & \text{(if } ff \in z \neq T) \\ \bot & \text{(otherwise)} \end{cases} $$

It is known that $T$ is $\lambda$-definable from $\{tt, ff, \uplus\}$, $\uplus$ is $\lambda$-definable from $\{\triangleright, T\}$ and $\bot$ can be defined from $\{tt, ff, \uplus, \triangleright\}$. We will show that there are no more $\lambda$-definabilities of $tt, ff, T, \uplus$ or $\bot$ other than those implied by the above ones; the situation for $\bot$ has only been partly clarified.

First, $\triangleright$ is not definable from $\{tt, ff, \uplus, \bot, T\}$. Take $R_0 = \{\bot, ff, tt, ff, tt, \triangleright, tt, \bot, ff, tt, \triangleright, tt\}$ and note that $R_0(\triangleright tt ff tt, \triangleright ff ff tt)$ is false. Here and later theorem 2 is used implicitly.

- $tt$ is not definable from $\{ff, \uplus, \triangleright, T\}$; take $R_0 = \{\bot, ff, T\}$.
- $ff$ is not definable from $\{tt, \uplus, \triangleright, \bot\}$; take $R_0 = \{\bot, tt, T\}$.
- $T$ is not definable from any one of $\{tt, ff, \triangleright, \bot\}$, $\{tt, \uplus, \triangleright, \bot\}$ or $\{ff, \uplus, \triangleright, \bot\}$; take $R_0 = \{\bot, tt, ff\}, \{\bot, tt\}$ or $\{\bot, ff\}$ respectively.

$\uplus$ is not definable from either one of $\{tt, ff, \bot, \bot\}$ or $\{tt, ff, \bot, \triangleright\}$; take/
take $R_0 = \{ <tt,tt>, <ff,ff>, <L,L>, <T,T>, <tt,ff>, <tt,L>, <L,ff> \}

and note that $R_0(\cup tt tt, \cup tt ff)$ is false in the first case and

and take $R_0 = \{ L, tt, ff \}$ in the second case.

In the case of $\pi$, we would like to show that $\pi$ is not definable

from any of the sets $\{ tt, ff, \cup , T \}$, $\{ ff, \cup , \pi , T \}$ or

$\{ tt, \cup , \pi , T \}$. For the first of these take $R_0 = \{ <tt,tt>, <ff,ff>, <T,T>, <L,L>, <tt,ff>, <tt,T>, <T,ff> \}$ and note that $R_0(\pi ff \cup \pi \pi ff)$ is false.

The trouble with the others is that if $R_\infty (\pi, \pi)$ then $R_\infty (\pi, \pi)$, for the $R\,'s$ considered here. For if $x, y \in T_0$ then

$(x \pi (y \pi L), (y \pi L, y)) = x \pi y$, and so one can define from $\pi$ terms

$M_n (n \geq 0)$ such that $M_n xy = x \pi y$ if $x$ and $y$ are in $T_n$. Therefore if

$R_\infty (\pi, \pi)$ then $R_\infty (\lambda x \lambda y x_n \pi y_n)$ for any $n$ and so $R_\infty (\pi, \pi)$.

On the other hand, $\pi$ is, in fact, not $\lambda$-definable from $\pi$.

For suppose $\pi = M \pi$ for some closed $\lambda$-term $M$. If $M$ is unsolvable then

$\pi = L$, a contradiction; therefore $M$ has the form $\lambda x_1 \ldots \lambda x_n x_j M_1 \ldots M_k$

where $n \geq 0$ and $1 \leq j \leq n$. One can assume that $j \leq 3$ since one can always apply

the identity $\pi = (M \pi) (M \pi)$. If $j = 1$, then $\pi x_1 \ldots x_n = \pi M_1 \ldots M_k$

where $M_r = (\lambda x_i M_r) (1 \leq r \leq k)$. Taking the $x_i = 1$ and $R_0 = \{ L \}$ we see that

$(M'_1)_0 = L$ and so $\pi 1 \ldots 1 = L$, a contradiction. If $j \neq 1$ then since

$\pi xy = \pi yx$, for any $x, y$ in $T_0$, we have $x M'_1 \ldots M'_k \pi xy = y M'_1 \ldots M'_k$

where the $M'_i$ and $M'_i$ are $\lambda$-definable from $x, y, \pi$ and $\pi$. Since $x, y$

are arbitrary members of $T_0$, this is a contradiction.

Perhaps an extension of Wadsworth's ideas to LAMBDA [9], would

settle these questions.
3. \( \lambda \)-definability in the full type hierarchy

For the sake of clarity, we will be a little more formal than in the last section.

The set of type symbols is the least set containing \( \mathfrak{c} \) and containing \( (\sigma \rightarrow \tau) \) whenever it contains \( \sigma \) and \( \tau \); \( \sigma \) and \( \tau \) are metavariables, possibly suffixed, ranging over type symbols and \( (\sigma_1, \ldots, \sigma_n, \tau) \) abbreviates \( (\sigma_1 \rightarrow (\sigma_2 \rightarrow \ldots (\sigma_n \rightarrow \tau) \ldots )) \) \( (n \geq 0) \).

The language of the typed \( \lambda \)-calculus has denumerably many variables \( \alpha_i^{\tau} (i \geq 0) \) of each type \( \tau \). We will use \( \alpha \) and \( \beta \), with or without various decorations as metavariables over variables. The language has a set of terms which is given by:

1. \( \alpha_i^{\tau} \) is a term of type \( \tau \), \( (i \geq 0) \),
2. if \( M \) and \( N \) are terms of type \( (\sigma \rightarrow \tau) \) and \( \sigma \) respectively then \( (MN) \) is a term of type \( \tau \),
3. if \( M \) is a term of type \( \tau \) then \( (\lambda \alpha_i^{\tau} M) \) is a term of type \( (\sigma \rightarrow \tau) \), \( (i \geq 0) \);

\( M \) and \( N \), possibly with suffixes, will be used as metavariables over terms. The reader is assumed to know what a \( \beta \cdot \gamma \)-normal form of a term is and the elementary properties of normal forms; \( M \cong N \) means that \( M \) and \( N \) have identical \( \beta \cdot \gamma \)-normal forms. By the Church-Rosser theorem this is an equivalence relation. Suppose that \( K_{\sigma \tau} = (\lambda \alpha \sigma (\lambda \alpha \tau \alpha_0^{\sigma})) \) and \( S_{\sigma_1, \sigma_2, \sigma_3} = (\lambda \alpha (\sigma_1, \sigma_2, \sigma_3)(\lambda \alpha (\sigma_1, \sigma_2) (\lambda \alpha (\sigma_1, \sigma_2, \sigma_3)) \alpha (\sigma_1, \sigma_2, \sigma_3))) \) are closed terms under application, to within \( \cong \). The type subscripts in \( K \) and \( S \) will often be omitted, as will be as many other type symbols as is convenient; the resulting propositions are to be understood as being asserted for every consistent way of putting the symbols back in.

Our language also has a semantics based on the full type hierarchy \( \{D_\sigma\} \) defined by:

\[
D/
\[ D(\sigma \rightarrow \tau) = D(\sigma \rightarrow D_{\tau}) \] (the set of all functions from \(D_{\sigma}\) to \(D_{\tau}\)),

where \(D\) is some given set.

The semantics is a function \(\llbracket \cdot \rrbracket\): \(\text{Terms} \rightarrow \bigcup_{\sigma \in D_{\sigma}} D_{\sigma}\) where \(\text{Env}\), the set of environments, is the set of type respecting functions from the set of variables to \(\bigcup D_{\tau}\), and is ranged over by \(\rho\). Then, \(\llbracket \cdot \rrbracket\) is the unique function of that type such that:

1. \(\llbracket \alpha_{i}^{\tau} \rrbracket(\rho) = \rho(\alpha_{i}^{\tau})(i \geq 0)\)
2. \(\llbracket (MN) \rrbracket(\rho) = [M] \llbracket N \rrbracket(\rho)\)
3. \(\llbracket (\lambda \alpha_{i}^{\tau}.M) \rrbracket(\rho)(x) = [M]([\rho[x/\alpha_{i}^{\tau}]])(i \geq 0, x \in D_{\tau})\)

where \(\rho[x/\alpha_{i}^{\tau}]\) is the environment \(\rho'\) such that

\[
\rho'(\alpha_{i}^{\tau'}) = \begin{cases} 
    x & (\alpha_{i}^{\tau'} = \alpha_{i}^{\tau}) \\
    \rho(\alpha_{i}^{\tau'}) & \text{(otherwise)}
\end{cases}
\]

Note that if \(M\) has type \(\sigma\), \(\llbracket M \rrbracket(\rho) \in D_{\sigma}\). If \(M\) is closed then \(\llbracket M \rrbracket(\rho) = \llbracket M \rrbracket(\rho')\) for any \(\rho\) and \(\rho'\) - so we often drop the reference to \(\rho\) for closed \(M\). If \(M \approx M'\) then \(\llbracket M \rrbracket(\rho) = \llbracket M' \rrbracket(\rho)\) for any \(\rho\); we will give a converse later.

Suppose \(\pi \in D(\sigma \rightarrow \lambda)\) is a permutation. Permutations \(\pi_{\sigma}\) in any \(D(\sigma \rightarrow \sigma)\) can be defined by:

\[ \pi(\sigma \rightarrow \tau)(f) = \pi_{\tau} \circ f \circ \pi_{\sigma}^{-1} \quad (f \in D(\sigma \rightarrow \tau)) \]

If \(M\) is closed term then \(\pi([M]) = [M]\) (see [2]). However this does not characterise \(\lambda\) - definability.

For example ground equality, \(=_{\lambda}\), is permutation-invariant, but is certainly not \(\lambda\) - definable. Explicitly let \(0\) abbreviate \((\lambda, \lambda, \lambda)\) and let \(tt\) and \(ff\) be \(\lambda \alpha_{0}^{\lambda} \lambda \alpha_{0}^{\lambda} \) and \(\lambda \alpha_{0}^{\lambda} \lambda \alpha_{1}^{\lambda}\) respectively. Then \(=_{\lambda}\) is defined by:

\[ =_{\lambda} xy \begin{cases} 
    tt & \text{(if } x = y) \\
    ff & \text{(if } x \neq y) \end{cases} \quad (x, y \in D_{\lambda}) \]

But/
But the only $\lambda$-definable functionals of type $(\lambda, \lambda, 0)$ are
$\lambda_{\alpha_0} \lambda\alpha_1 \lambda\alpha_2 \lambda\alpha_3 \lambda\alpha_j$ for $0 \leq j \leq 3$ none of which are $\equiv_\lambda$ if $|D_\lambda| > 1$.

M. Gordon proposed, as a possible remedy, that relations $R_\lambda \subseteq D_\lambda^2$ should be extended — not just permutations. Starting with such an $R_\lambda$, the $R_\sigma$s are defined by:

$$R_{(\sigma \rightarrow \tau)}(f, g) \equiv \forall x, y \in D_\sigma (R_\sigma(x, y) \rightarrow R_\lambda(fx, gy)).$$

When $R_\lambda$ is a permutation $\Pi_{\lambda} R_\sigma = \Pi_{\sigma}$ for all $\sigma$. The definition generalises, in the obvious way, if one starts with $R_\lambda \subseteq D^K$, for any ordinal $K$. If $R_\sigma \subseteq D^K_\sigma$ is obtained from an $R_\lambda$ in that way it is called $K$-logical; $f \in D_\sigma$ satisfies it iff $R_\sigma(f)$ holds. With the obvious definitions of $\lambda$-definability and $\lambda$-definability from a set $X \subseteq \cup D_\sigma$, one shows that any $\lambda$-definable functional satisfies any $K$-logical relation, of the right type and that if $\{R_\sigma\}$ is the system of relations obtained from some $R_\lambda$, and each member of $X$ satisfies the appropriate $R_\sigma$ and $f$ is $\lambda$-definable from $X$, then $x$ satisfies the appropriate $R_\sigma$.

The proof is like that of theorem 1.1.

One can now see why $\equiv_\lambda$ is not $\lambda$-definable if $|D_\lambda| > 1$.

Let $0, 1$ be distinct elements of $D_\lambda$. Let $R_\lambda$ = $\{<0, 0>, <0, 1>, <1, 0>\}$. Then $R(tt, ff)$ is false for $R(1, 0)$ and $R(0, 1)$ but not $R(tt10, ff01)$. Therefore $R(\equiv_\lambda, \equiv_\lambda)$ is false for $R(0, 0)$ and $R(0, 1)$ but not $R(\equiv_\lambda(0, 0), \equiv_\lambda(0, 1))$.

As an example of non-relative definability, consider the universal quantifier $\forall_\lambda$ of type $(\lambda \rightarrow 0 \rightarrow 0)$ defined by:

$$\forall_\lambda(f) = \begin{cases} tt & \text{(if } fx = tt \text{ for all } x \text{ in } D_\lambda) \\ ff & \text{(otherwise).} \end{cases}$$

Now $\forall_\lambda$ is permutation-invariant; however if $|D_\lambda| > 3$ it is not $\lambda$-definable from $\equiv_\lambda$. To see this let $R_\lambda$ = $\{<0, 0>, <1, 1>\}$ where $0, 1$ are distinct elements of $D_\lambda$. $R(\equiv_\lambda, \equiv_\lambda)$ is true, but if $f, g \in D(\lambda \rightarrow 0)$ are such that $f(x)$ is always $tt$ but $g(x)$ is $tt$ iff $x$ is $0$ or $1$ then $R/$
R(f, g) but not R(∀f(∀g)). Incidentally, if |D| ≤ 3, ∀λ is λ-definable from =λ.

We can only characterise definability using logical relations, for types of level ≤ 2.

**Theorem 1** Suppose r has the form (τ₁, ..., τₙ, ℓ) where each τᵢ has the form (τ, ..., τ, ℓ). Then if |D| ≥ λ₀ and f ∈ D r satisfies every 2-logical relation, it is λ-definable.

**Proof** We will just give two cases since this should give the idea without overmuch detail.

Suppose r = (τ, τ, ℓ). Let x, y, O, 1 be elements of D, with 0 and 1 distinct and take Rₜ = {<x, 1>, <y, O>}. Then R(fxy, f⁽1₀⁾). So for every x, y ∈ D₁ either fxy = x and f⁽1₀⁾ = 1 or else fxy = y and f⁽1₀⁾ = 0. Therefore either f⁽1₀⁾ = 1 or f⁽1₀⁾ = 0 and so, since 1 ≠ 0 either f = [[λα₀⁺¹₀ λα₁⁺¹₁]].

The other case we consider is r = (τ, λ, ℓ, 0). Identify the integers, with a subset of D₁ and let the restriction of s to the integers be the successor function. Given g in D(λ) and x ∈ D₁ let Rₜ = {<gⁿ(x), s⁻¹(0) | n ≥ 0}. Clearly R(x, 0) and R(g, s). Therefore R(f(g)(x), f(s)(0)) and so for every g ∈ D(λ) and x ∈ D₁ there is an n such that f(g)(x) = gⁿ(x) and f(s)(0) = s⁻¹(0). Since s⁻¹(0) = s⁻¹(0) iff n = n', n must be independent of g and x and so for some n, f = [[λα₀⁺¹₀ λα₁⁺¹₁ α₀⁺¹₀ α₀⁺¹₀(α₀⁺¹₀(α₀⁺¹₀(α₀⁺¹₀(α₀⁺¹₀(...))))... n times]]).

We believe the theorem holds without the restriction on D₁. The simplest type which has us baffled when |D₁| ≥ 2 is (((λ)→(λ)→)→(λ)).

Some characterisation of definability can be obtained by strengthening the implication in the definition of Rσ to an intuitionistic one, à la Kripke [1].

To this end, suppose we have a set W (of worlds) a reflexive, transitive binary relation ≤ which is a subset of W² (alternativeness) and a relation R₁ ≤ D³ × W such that:

∀/
∀ x, y, z ∈ D_ξ ∀ w' ∈ W(R(x, y, z, w) ∧ (w ≤ w')) → R(x, y, z, w').

Define relations R_σ ⊆ D^3_ξ × W by:

R.(σ→τ)(f, g, h, w) ≡ (∀ x, y, z ∈ D_ξ ∀ w' ∈ W((R(x, y, z, w') ∧ (w ≤ w'))

Any such R_σ is called 3-I-logical; f ∈ D_σ satisfies R_σ iff R_σ(f, f, f)

Theorem 2 If | D_ξ | ≥ χ_0, then f ∈ D_σ satisfies every 3-I-logical relation

We don't know if the restriction on D_ξ can be dropped, or if 3 can
be reduced to 2; it cannot be reduced to 1 because if D_ξ is the integers
and f ∈ D(<→→(→→)) is defined by:

f(g)(x) = g(x)(x) (g ∈ D(→→) x ∈ D_ξ), then f satisfies every
1-I-logical predicate but is not λ-definable.

Neither do we know anything about characterising relative
definability, even in interesting special cases, or what happens in
other models of the typed λ-calculus or, of course, what happens in
the case of the untyped λ-calculus. The rest of the memorandum is
devoted to the proof of theorem 2.

Lemma 1 Suppose R_σ ⊆ D^3 × W is 3-I-logical. Then, for the appropriate
ξ: ∀ f, g, h ∈ D_σ ∀ w, w' ∈ W((R_σ(f, g, h, w) ∧ w ≤ w') → R_σ(f, g, h, w')).

Proof By induction on σ.

Lemma 2 If f ∈ D_σ is λ-definable, it satisfies every 3-I-logical
relation.

Proof Let \{R_σ\} be the collection of 3-I-logical relations built up, as
above, from some W, ≤ and R_ξ ⊆ D^3 × W. Clearly, if f ∈ D(σ→τ)
satisfies R_σ and g ∈ D_σ satisfies R_σ then f(g) satisfies R_τ. So we need
only/
only show, by a remark made above, that \([k]\) and \([s]\) satisfy the appropriate R's.

For K, suppose that \(w' \leq w\), \(R(f,g,h,w')w' \leq w^w\) and \(R(x,y,z,w')\). Then \(R([k](f)(x),[k](g)(y),[k](h)(z),w^w)\) by the previous lemma, since \(R(f,g,h,w')\) and \(w' \leq w^w\).

For S, suppose that \(w' \leq w\), \(R(f,f',f'',w')w' \leq w^w\), \(R(g,g',g'',w')\), \(w' \leq w''\) and \(R(x,x',x'',w'')\). As \(w' \leq w''\), \(R(f(x),f'(x'),f''(x''),w'')\) and as \(w' \leq w''\), \(R(g(x),g'(x'),g''(x''),w'')\). Therefore, as \(w' \leq w''\), \(R(f(x),g(x),f'(x),g'(x'),f''(x''),g''(x''),w'')\), concluding the proof.

This establishes the "consistency" half of theorem 2. Of course, the lemma also holds if 3 is replaced by any \(\kappa\) and, further, an analogue of theorem 1.1.2 also holds in general.

To obtain the "completeness" half we use a special \(w, \leq, R\) which in turn requires us to give a "standard" environment, \(\rho_0\), which assigns to each \(\alpha_i^\gamma\) an element of \(D_\gamma\) which behaves like \(\alpha_i^\gamma\), in a way to be made clear by lemmas 3 and 4.

Now we suppose that \(|D_\gamma| \geq \kappa_0\) and ((( ))) is a map from the set of terms of type \(\zeta\) to \(D_\gamma\) such that:

\[(\langle M \rangle) = (\langle N \rangle) \iff M \cong N.\]

A vector, \(\alpha^\gamma\), of variables is non-repeating if no variable occurs twice in it. If \(M\) is a term and \(\bar{N}^\gamma = \langle N_1, \ldots, N_n \rangle\) \((n \geq 0)\) is a vector of terms, \(M \bar{N}^\gamma\) abbreviates \((\ldots((M N_1)N_2)\ldots N_n)\); similarly if \(\bar{x}^\gamma\) is a vector of elements of \(\bigcup D_\gamma\) then \(\bar{f}^{\bar{x}^\gamma}\) abbreviates \(f(x_1)\ldots(x_n)\), where \(\bar{x}^\gamma = \langle x_1, \ldots, x_n \rangle\) and \(f\) is a functional of the appropriate type. If \(\alpha^\gamma = \langle \alpha_1^\gamma, \ldots, \alpha_n^\gamma \rangle\) is a vector of variables, then \(\rho(\alpha^\gamma) = \langle \rho(\alpha_1^\gamma), \ldots, \rho(\alpha_n^\gamma) \rangle\) for any environment \(\rho\).

The standard environment, \(\rho_0\), is defined by:

1. \(\rho_0(\alpha_i^\zeta) = (\alpha_i^\zeta)\) \((i \geq 0)\)
2. /
2. If there is a vector $\tilde{N}$ of terms of length $n$ such that for any $j (1 \leq j \leq n)$ and almost every* non-repeating $\vec{\alpha}_j$ containing no variable free in $(\tilde{N})_j$, $f_j \rho_0(\vec{\alpha}_j) = \left( (\tilde{N})_j \vec{\alpha}_j \right)$, then $\rho_0(\vec{\alpha}_1, \ldots, \vec{\alpha}_n, \vec{\alpha}) = \left( \vec{\alpha}_1, \ldots, \vec{\alpha}_n \vec{\alpha} \right)$, where $\tilde{N}$ is one such vector; if there is no such vector then

$$\rho_0(\vec{\alpha}_1, \ldots, \vec{\alpha}_n, \vec{\alpha})(f_1) \ldots (f_n) = (\vec{\alpha} \vec{\alpha}_0).$$

Notice that if $\tilde{N}_1$ and $\tilde{N}_2$ satisfy the conditions of the second clause then they are, componentwise, $\sim$. So $\rho_0$ is well-defined.

**Lemma 2** For every term $M$ and vector $\vec{\alpha}$ of variables of the appropriate type, $[M^\vec{\alpha}](\rho_0)=((M^\vec{\alpha})).$

**Proof** First we prove the lemma for variables, by induction on types. If $M=\vec{\alpha}$, the result is immediate from the definition of $\rho_0$. For $\vec{\alpha}_1, \ldots, \vec{\alpha}_n$, let $\vec{\beta}_j$ be any vector of variables such that $\vec{\alpha}_j, \vec{\beta}_j$ has type $\vec{\gamma}$. By induction hypothesis, $[\vec{\alpha}_j, \vec{\beta}_j](\rho_0)=((\vec{\alpha}_j, \vec{\beta}_j))$. Therefore $\vec{\alpha}$ satisfies the condition in clause 2 of the definition of $\rho_0$, when $\vec{\alpha}_1, \ldots, \vec{\alpha}_n \equiv \rho_0(\vec{\alpha})$. Therefore, $[M^\vec{\alpha}](\rho_0)=\rho_0(\vec{\alpha}_1, \ldots, \vec{\alpha}_n, \vec{\alpha})(\rho_0(\vec{\alpha})=((M^\vec{\alpha}))$.

Finally the lemma is proved for terms in $\vec{\beta}, \vec{\gamma}$-normal form by induction on their size. If $M$ is a variable, we are finished. Otherwise, if $M^\vec{\alpha} \approx (\ldots (\vec{\alpha} M_j) \ldots M_k)^{(k)}$ where $\alpha$ is some variable and the $M_j (1 \leq j \leq k)$ all are $\vec{\beta}, \vec{\gamma}$-normal forms of smaller size than $M$, the proof then proceeds by applying the induction hypothesis to the $M_j$, in a way similar to the above. The only other case is when $M \approx \lambda^{\beta_1} \ldots \lambda^{\beta_k} \alpha^\vec{\alpha}_i$ for some $i$, and this is very easy.

* Here and elsewhere assertions of the form "for almost every $\vec{\alpha}, \ldots"$ should be read as "there is a finite set of variables such that for every $\vec{\alpha}$, none of whose component variables are in the set, $\ldots" 

From now on, we drop references to $\rho_0$ in $[M]^\rho_0$. Lemma 3 implies that $[M] = [M']$ iff $M \equiv M'$. If $D_\xi$ were finite then there would be always a closed $M$, $M'$, even of type $((\epsilon \rightarrow \epsilon) \rightarrow (\epsilon \rightarrow \epsilon))$, such that $[M] = [M']$ but $M \not\equiv M'$. The above methods can be used to show that given closed $M$, $M'$, there is an integer $m$, such that if $|D_\xi| > m$ and $[M] = [M']$ then $M \simeq M'$. These remarks form the converse to the consistency of $\simeq$, mentioned above.

A term $M$ is of order zero iff it has the form $\alpha_1 M_1 \ldots M_k$.

**Lemma 4** Suppose $M$ has order-zero and $x \rho_0(\alpha^x) = y \rho_0(\alpha^y)$ for almost all non-repeating vectors $\alpha$ such that $x \rho_0(\alpha^x) \in D_\xi$. Then $[M]^\rho_0(x) = [M]^\rho_0(y)$.

**Proof** Immediate from the definition of $\rho_0$.

Now $W$, $\leq$ and $R_\xi$ can be defined. The worlds, $W$ are triples, $\langle \epsilon^\alpha, \alpha^\alpha, \beta^\alpha \rangle$, where $\beta^\alpha$ is a vector of members of $U D_\xi$, $\alpha^\alpha$, $\beta^\alpha$ are non-repeating and $\beta^\alpha$, $\alpha^\alpha$, $\beta^\alpha$ have the same length and corresponding members have the same type. If $w_i = \langle \alpha^\alpha_1, \alpha_1^\beta_1, \beta_1^\alpha \rangle$ $(i = 1, 2)$ then $w_1w_2 = \langle \alpha_1^\beta \alpha_2, \alpha_2^\beta, \beta_1^\alpha \beta_2 \rangle$, where the component vectors have been concatenated. Then $w_1 \leq w_2$ if $w_2 = w_1w_3$, for some world, $w_3$. Finally, $R_\xi$ is defined by:

$$R_\xi(x,y,z,\langle \epsilon^\alpha, \alpha^\alpha, \beta^\alpha \rangle) \text{ iff } \exists \text{ a closed term } M \text{ such that } x = [M]^\beta, y = [M]^\rho_0(\alpha^x) \text{ and } z = [M]^\rho_0(\beta^x).$$

Clearly $W$, $\leq$ and $R_\xi$ satisfy the required conditions for obtaining a system of 3-I-logical $R_\xi$'s.

**Lemma 5**

1. If $R_\xi(f, g, h, w)$ where $w = \langle \epsilon^\alpha, \alpha^\alpha, \beta^\alpha \rangle$ then there is a closed term $M$ such that whenever $\alpha^\alpha^+, \beta^\alpha^+$ are non-repeating vectors

$$f = [M]^\beta, g = [M|^\alpha \alpha^+] \text{ and } h = [M|^\beta \beta^+]$$

2. Suppose that $f = [M]^\beta, g = [M|^\alpha \alpha^+]$ and $h = [M|^\beta \beta^+]$ where $M$ is a closed term and $g$ and $h$ are denotations of order-zero terms and $w$/
$w = \langle \beta', \bar{\alpha}, \bar{\beta} \rangle \in W$. Then $R_{\sigma}(f, g, h, w)$ for the appropriate $\sigma$.

**Proof** Both parts are proved together by induction on $\sigma$.

1. For the result is immediate from the definition of $R_{\sigma}$.

   So suppose $f$ has type $(\sigma \rightarrow \mathcal{A})$.

   Suppose $R_{\sigma}(f, g, h, w)$ where $w = \langle \beta', \bar{\alpha}, \bar{\beta} \rangle$.

   Suppose next that $w = \langle x, \alpha_1, \beta_1 \rangle$, where $x$ has type $\sigma_f$ is in $W$.

   By induction hypothesis using 2 we see that, $R_{\sigma}(x, \rho_0(\alpha_1), \rho_0(\beta_1), w^+)$.

   Therefore $R_{\sigma}(fx, (g \rho_0(\alpha_1), \rho_0(\beta_1), w^+)$.

   By induction hypothesis, using 1, there is a closed term $M(x, \alpha_1, \beta_1)$ such that if $\bar{\alpha} \sim \alpha_1 \sim \alpha^{++}$ and $\bar{\beta} \sim \beta_1 \sim \beta^{++}$ are non-repeating vectors then $\langle M(x, \alpha_1, \beta_1) \rangle (g \rho_0(\alpha_1), \rho_0(\beta_1), w^+) = \langle \rho_0(\alpha_1), \rho_0(\beta_1), w^+ \rangle$ and $(h \rho_0(\beta_1)) \rho_0(\beta^{++}) = \langle \rho_0(\beta^{++}) \rangle$.

   The subscripts on $M(x, \alpha_1, \beta_1)$ indicate its dependence on $x, \alpha_1$ and $\beta_1$. Then, $\langle \rho_0(\alpha_1), \rho_0(\beta_1), w^+ \rangle$ for any $x, \alpha_1, \beta_1 \in \mathcal{D}_{\sigma}$ if $\alpha_1 \sim \alpha^{++}$ is non-repeating. Therefore, by lemma 3, $M(x, \alpha_1, \beta_1) \sim M(x', \alpha_1, \beta_1)$. A similar argument shows that $M(x, \alpha_1, \beta_1)$ is also independent of $\alpha_1$ and $\beta_1$, and the conclusion follows.

2. Suppose $f = \langle M \rangle \beta', g = \langle M \alpha \rangle$ and $h = \langle M \beta \rangle$ where $M$ is a closed term and $g$ and $h$ are denotations of order-zero terms, and $w = \langle \beta', \bar{\alpha}, \bar{\beta} \rangle \in W$.

   Let $w = \langle \beta', \bar{\alpha}, \bar{\beta} \rangle \in W$. Then by induction hypothesis for part 1, there is a closed term $M_1$ such that wherever $\bar{\alpha} \sim \alpha^{++}$ and $\bar{\beta} \sim \beta^{++}$ are non-repeating vectors,

   $x = \langle M_1 \rangle \beta \sim \alpha^{++}$, $y \rho_0(\alpha^{++}) = \langle M_1 \rangle \alpha \sim \alpha^{++}$ and $z \rho_0(\beta^{++}) = \langle M_1 \rangle \beta \sim \beta^{++}$.

   Then we have $fx = \langle (M \beta') \beta \rangle \langle (M \alpha) \alpha \rangle$ and $\rho_0(\beta^{++}) = \langle M_2 \rangle \beta \sim \beta^{++}$ for a certain closed term $M_2$.

   Since $y \rho_0(\alpha^{++}) = \langle M_1 \rangle \alpha \sim \alpha^{++} \rho_0(\alpha^{++})$ for almost all non-repeating vectors/
vectors $\tilde{\alpha}^{++}$ such that $y \rho_0(\tilde{\alpha}^{++})$ is in $D_\zeta$ and $g$ is the denotation of a term of order zero, lemma 4 applies and we have,

$$gy = g\left(\left[\begin{array}{c} M_1 \\
M_2 \end{array}\right] \tilde{\alpha} \sim \tilde{\alpha}^{++}\right)$$

$$= \left[\begin{array}{c} M_1 \\
M_2 \end{array}\right] \rho_0(\tilde{\alpha} \sim \tilde{\alpha}^{++})$$

$$= \left[\begin{array}{c} M_1 \\
M_2 \end{array}\right] \rho_0(\tilde{\alpha} \sim \tilde{\alpha}^{++})$$

where $M_2$ is the same term as above. In the same way,

$$hz = \left[\begin{array}{c} M_2 \end{array}\right] \rho_0(\tilde{\beta} \sim \tilde{\beta}^{++}).$$

It is now clear that $gy$ and $hz$ are themselves denotations of order zero terms, and so applying the induction hypothesis for part 2, we conclude that

$$R_{\sigma}(fx,gy,hz,w^+).$$

Therefore $R_{\sigma}(f,g,h,w)$, as was required, concluding the proof.

Lemma 5 gives us the second part of theorem 2. For suppose $f$ satisfies every 3-I-logical $R_{\sigma}$. Then in particular we have $R_{\sigma}(f,f,f,w_0)$ where $R_{\sigma}$ is the one to which lemma 5 applies and $w_0$ has all its components empty. By part 1 of lemma 5, there is a closed term $M$ such that $f = \left[\begin{array}{c} M \end{array}\right]$, which is just what was wanted.

Acknowledgements

I would like to thank Mike Gordon for his help and encouragement. The work was carried out with the aid of an S.R.C. research grant.

References


[5]


