The Undecidability of λ-Definability

Ralph Loader, Oxford University*

Introduction

In this article, we shall show that the Plotkin-Statman conjecture [**?**, **?**] is false. The conjecture was that, in a model of the simply typed λ-calculus with only finitely many elements at each type, definability (by a closed term of the calculus) is decidable. This conjecture had been shown to imply many things, for example, Statman [**?**] (see also Wolfram's book [**?**]) has shown it implies the decidability of pure higher order pattern matching (a problem that remains open at the time of writing) and is equivalent to higher order pattern matching with δ-functions. The proof of undecidability given here uses encodings of semi-Thue systems as definability problems. It had been thought that λ-definability might be characterised by invariance under logical relations, which would imply the Plotkin-Statman conjecture. We give a relatively simple counterexample to this, using our encoding of word problems.

1 Preliminaries

We shall consider the simply typed λ-calculus with a single ground type o. All models M considered here will be finite (in the sense of finitely many elements at each type) and full (for any types T and U, $M_{T \to U}$ is the set of all functions from M_T to M_U). Terms of the simply typed λ-calculus shall be assumed to be in the Church form—*i.e.,* all variables are typed, but we shall usually omit the types for typographical convenience.

We shall consider terms in β-normal, η-long form. β-conversion is generated by the following rules:

$$(\lambda \mathrm{x}.\, s[\mathrm{x}])(t) \quad \longrightarrow_\beta \quad s[t]$$

*This research was supported by the Commonwealth Scholarship Commission in the United Kingdom. I would like to thank Allen Stoughton, Lincoln Wallen (my Doctoral supervisor), and a referee, for useful conversations and advice.

$$
\begin{array}{rcl}
s \longrightarrow_\beta s' & \Rightarrow & s(t) \longrightarrow_\beta s'(t) \\
t \longrightarrow_\beta t' & \Rightarrow & s(t) \longrightarrow_\beta s(t') \\
s \longrightarrow_\beta s' & \Rightarrow & \lambda \mathtt{x}.\, s \longrightarrow_\beta \lambda \mathtt{x}.\, s'
\end{array}
$$

where bound variables may have to be renamed appropriately. η-expansion is the opposite of the usual η-contraction;

$$
\begin{array}{rcll}
t^{A \to B} & \longrightarrow_\eta & \lambda \mathtt{x}^A.\, t(\mathtt{x}) & (1) \\
s \longrightarrow_\eta s' & \Rightarrow & s(t) \longrightarrow_\eta s'(t) & (2) \\
t \longrightarrow_\eta t' & \Rightarrow & s(t) \longrightarrow_\eta s(t') & \\
s \longrightarrow_\eta s' & \Rightarrow & \lambda \mathtt{x}.\, s \longrightarrow_\eta \lambda \mathtt{x}.\, s' &
\end{array}
$$

where, in (1), $\mathtt{x}$ is not free in t. To retain normalisation, we make the restriction that η-expansion cannot create new β-redexes; in other words, (1) may not be applied if the term t is already a λ-abstraction $\lambda \mathtt{x}.\, t'$, and (2) may not be applied if $s \longrightarrow_\eta s'$ is obtained as an instance of (1).

Given a simply typed λ-term t and a valuation v defined on the free variables of t, the valuation of t with respect to v will be denoted by $[\![t]\!]_v$. The subscript v may be omitted if the intended valuation is obvious from the context or if t is closed. If $t[\bar{x}]$ is a term, and $\bar{a} \in M$ have the appropriate types, then we shall write $[\![t[\bar{a}]]\!]$ or just $t[\bar{a}]$ to denote the valuation of t w.r.t. the valuation sending $\bar{\mathtt{x}}$ to $\bar{a}$. The valuation of a term is invariant under both β-conversion and η-expansion.

A closed term t is said to define $a \in M$ if $[\![t]\!] = a$. A term $t[\bar{\mathtt{x}}]$ is said to define b w.r.t. $\bar{a}$ if $[\![t[\bar{a}]]\!] = b$

The **(absolute) $\boldsymbol{\lambda}$-definability problem** is: Given a full and finite model M and a member $a \in M_T$ for some type T, is there a closed term that defines a?

The **relative $\boldsymbol{\lambda}$-definability problem** is: Given a model M, again full and finite, $\bar{a}, b \in M$, is there a term $t[\bar{\mathtt{x}}]$ that defines b w.r.t. $\bar{a}$?

These two problems are related as follows

Lemma 1 *The λ-definability problem is decidable if and only if the relative λ-definability problem is decidable.*[1]

Proof Suppose that the absolute problem is decidable. We reduce the relative problem to this as follows: Given $a_i \in M_{T_i}$, $i = 1 \cdots n$ and $b \in M_T$, look for $f \in M_{T_1 \to \cdots \to T_n \to T}$ such that

[1] I assume a sensible coding of finite λ-models, encoding functions as Gödel numbers of their graphs. Note that this lemma applies both to the problems in a fixed model, or the problems ranging over all (full and finite) models.

1. f is λ-definable by a closed term ϕ.

2. $f(a_1)\cdots(a_n) = b$.

If such a f is found, then the term $\phi(\mathtt{x}_1)\cdots(\mathtt{x}_n)$ defines b relative to the a_i. If a term $t[\mathtt{x}_1\cdots\mathtt{x}_n]$ defines b relative to $a_1\cdots a_n$, then 1 and 2 above are satisfied with $\phi = \lambda\mathtt{x}_1\cdots\mathtt{x}_n.\, t$, and $f = [\![\phi]\!]$.

Further, the search for f satisfying 1 and 2 is effective (by our assumption that the absolute problem is decidable), and terminating as we can list the finitely many objects of a given type in M. Thus the relative problem is decidable if the absolute problem is.

The converse implication is trivial as the absolute problem is a special case of the relative problem. $\square$

The statement that the λ-definability problem is decidable is known as the Plotkin-Statman conjecture. We will show that, in fact, these problems are undecidable, by encoding word problems as relative λ-definability problems.

2 Encoding Word Problems

We will encode word problems (specifically, semi-Thue systems) over the alphabet $\{A, B\}$ as members of the model M with (at least) seven elements $\{A, B, L, R, Y, N, *\}$ at ground type.

Given a word W_0 and rules $C_i \longrightarrow D_i$, $i = 1\cdots N$, where the C_i and the D_i are words, a derivation is a sequence of words $W_0, \ldots, W_n$ such that each W_{j+1} is obtained from W_j by replacing a sub-word C_i by D_i. The word W is derivable if it is the last word of some derivation.

The word problems we consider are as follows: Given an initial word W_0, and some rules $C_i \longrightarrow D_i$, $i = 1\cdots n$, can we derive a word W?

Proposition 2 *There is a word W_0 and rules $C_i \longrightarrow D_i$ such that the problem of determining if a word W can be derived is undecidable.*

Proof In [**?**], a universal Turing machine is encoded as a semi-Thue system so as to give a set of rules F_i and a word W_0 such that the problem "Given a word W, can W_0 be derived from W" is undecidable. By reversing all the rules of this problem, we get the proposition, except with a larger alphabet. It is easy to encode such a problem in the alphabet $\{A, B\}$. $\square$

We now encode words and rules as objects in M. A word W of length n will be coded by an object $\ulcorner W \urcorner$ of type $\underbrace{o \to \cdots \to o}_{n} \to o$. If words C and D

have lengths m and n respectively, then the rule $C \longrightarrow D$ will be encoded by an object $\ulcorner C \longrightarrow D \urcorner$ of type $(\underbrace{o \to \cdots \to o}_{m} \to o) \to (\underbrace{o \to \cdots \to o}_{n} \to o)$.

The encoding of a word W is defined by the following clauses:

W1 If W has $c \in \{A, B\}$ in the ith position ($i = 1 \cdots n$), then

$$\ulcorner W \urcorner \underbrace{(*) \cdots (*)}_{i-1} (c) \underbrace{(*) \cdots (*)}_{n-i} = Y.$$

W2 For $i = 1 \cdots n-1$,

$$\ulcorner W \urcorner \underbrace{(*) \cdots (*)}_{i-1} (L)(R) \underbrace{(*) \cdots (*)}_{n-i-1} = Y.$$

W3 $\ulcorner W \urcorner (x_1) \cdots (x_n) = N$ unless stated otherwise in W1 or W2.

It may help the reader's understanding to consider $\ulcorner W \urcorner$ as answering questions about the word W, with the ith parameter corresponding to the ith letter of W. W2 may then be thought of as encoding the ordering of the letters in W; without this axiom, any permutation of W would have its encoding λ-definable from $\ulcorner W \urcorner$, which is undesirable for our purposes.

The **characteristic** $\boldsymbol{\chi S} \in \boldsymbol{M_{o \to \cdots \to o \to o}}$ of a set $S \subset M_o \times \cdots \times M_o$ is defined by

- $\chi S \, (x_1) \cdots (x_n) = Y$ whenever $(x_1, \ldots, x_n) \in S$
- $\chi S \, (x_1) \cdots (x_n) = N$ whenever $(x_1, \ldots, x_n) \notin S$

Thus W1–3 above may be stated as: The encoding of a word $c_1 \cdots c_n$ is the characteristic of the union of the two sets

$$\{ \, (\underbrace{* \cdots *}_{i-1}, c, \underbrace{* \cdots *}_{n-i}) \mid i = 1 \cdots n \, \}$$

and

$$\{ \, (\underbrace{* \cdots *}_{i-1}, L, R, \underbrace{* \cdots *}_{n-i-1}) \mid i = 1 \cdots n-1 \, \}.$$

A rule $F = C \longrightarrow D$, where C and D have length m and n, is encoded as follows:

R1 $\ulcorner F \urcorner (\chi \{ (\underbrace{*, \ldots, *}_{m}) \}) = \chi \{ (\underbrace{*, \ldots, *}_{n}) \}$

R2 $\ulcorner F \urcorner(\chi\{(R, \underbrace{*, \ldots, *}_{m-1})\}) = \chi\{(R, \underbrace{*, \ldots, *}_{n-1})\}$

R3 $\ulcorner F \urcorner(\chi\{(\underbrace{*, \ldots, *}_{m-1}, L)\}) = \chi\{(\underbrace{*, \ldots, *}_{n-1}, L)\}$

R4 $\ulcorner F \urcorner(\ulcorner C \urcorner) = \ulcorner D \urcorner$

R5 $\ulcorner F \urcorner(g) = \chi(\emptyset)$ in all cases not covered above.

The encoding of rules is designed so that:

Lemma 3 *If $W_1 = g_1 \cdots g_k \, C \, h_1 \cdots h_l$ and $W_2 = g_1 \cdots g_k \, D \, h_1 \cdots h_l$ are words, and F is the rule $C \longrightarrow D$, then $\ulcorner W_2 \urcorner$ is given by*

$$f =_{def} \lambda \bar{\mathtt{x}}\bar{\mathtt{y}}\bar{\mathtt{z}}. \ulcorner F \urcorner(\lambda \bar{\mathtt{u}}. \ulcorner W_1 \urcorner \bar{\mathtt{x}}\bar{\mathtt{u}}\bar{\mathtt{z}})(\bar{\mathtt{y}})$$

where $\bar{\mathtt{x}}$, $\bar{\mathtt{y}}$, $\bar{\mathtt{z}}$ and $\bar{\mathtt{u}}$ are vectors of k, n, l and m distinct variables respectively.

Proof An easy but tedious verification. For example, by W1 and W3,

$$\lambda \bar{u}. \ulcorner W_1 \urcorner(g_1) \underbrace{(*) \cdots (*)}_{k-1}(u_1) \cdots (u_m) \underbrace{(*) \cdots (*)}_{l}$$

is the characteristic of the set $\{(*, \ldots, *)\}$ and thus by R1,

$$f(g_1) \underbrace{(*) \cdots (*)}_{k+n+l-1} = Y. \qquad \Box$$

Corollary 4 *If a word W is derivable from W_0 by rules F_i, then $\ulcorner W \urcorner$ is λ-definable from $\ulcorner W_0 \urcorner$ and the $\ulcorner F_i \urcorner$.*

Proof By an induction over the derivation of W, using the lemma. $\Box$

3 Faithfulness of the Encoding

In this section we shall fix a word W_0 and rules $F_i = C_i \longrightarrow D_i$. We show that if $\ulcorner W \urcorner$ is λ-definable relative to $\ulcorner W_0 \urcorner$ and the $\ulcorner F_i \urcorner$, then W is derivable from W_0 by the rules F_i, and thus conclude that the Plotkin-Statman conjecture fails.

Our proofs will be by induction over terms in the following form:

Definition 5 *A* **preword-term** *is a term t of type o, containing constants amoung $\ulcorner W_0 \urcorner$ and $\ulcorner F_1 \urcorner \cdots \ulcorner F_N \urcorner$ such that*

- *t is in β-normal, η-long form.*
- *t is not a variable.*
- *All free variables in t are of type o.*

If a preword-term s has a closure $\lambda \mathtt{x}_1 \cdots \mathtt{x}_n.\, s$ that encodes a word, then $\mathtt{x}_1 \cdots \mathtt{x}_n$ must all occur free in s (as by W1 and W3 a function encoding a word of length n must depend on all of its n arguments) and the ordering of $\mathtt{x}_1 \cdots \mathtt{x}_n$ (and thus the word encoded) is unique (by W2 and W3). Note that a preword-term is in one of the forms $\ulcorner W_0 \urcorner(t_1) \cdots (t_n)$ or $\ulcorner F_i \urcorner(\lambda \bar{\mathtt{u}}.\, t)(t_1) \cdots (t_n)$.

In preparation for the main proof, we prove two technical lemmas about preword-terms. Although lemmas 6 and 7, and proposition 9, are stated about preword-terms, they are really used to give information about their closures, as we wish to identify which closures of preword-terms encode words. It is smoother formally, however, to prove things by induction over preword-terms.

Lemma 6 *If s is a preword-term and v any valuation, then $[\![s]\!]_v$ is either Y or N.*

Proof As s is in normal form and not a variable, s must be in the form $K(t_1) \cdots (t_j)$, where K is $\ulcorner W_0 \urcorner$ or one of the $\ulcorner F_i \urcorner$. That $[\![s]\!] \in \{Y, N\}$ is now obvious from W1–3 and R1–5. □

Lemma 7 *Let s be a preword-term*

1. *If $\mathtt{x}$ is free in s, and v is a valuation with $v(\mathtt{x}) \in \{A, B\}$ such that $[\![s]\!]_v = Y$, then $v(\mathtt{y}) = *$ for all $\mathtt{y}$ other than $\mathtt{x}$ free in s.*
2. *If $[\![\lambda \mathtt{u}_1 \cdots \mathtt{u}_n.\, s]\!]_v$ encodes a word, then $v(\mathtt{x}) = *$ for every variable $\mathtt{x}$ free in $\lambda \bar{\mathtt{u}}.\, s$.*

Proof We prove the lemma by induction over the preword-term s. We first derive 1 from the induction hypothesis. Consider the case when s is $\ulcorner W_0 \urcorner(t_1) \cdots (t_n)$.

- If one of the t_i is not a variable, then $[\![t_i]\!]_v \in \{Y, N\}$ by the previous lemma, so that $[\![s]\!]_v = N$ by W3, and hence there is nothing to prove.
- If all of the t_i are variables, then 1 is immediate by W1–3.

Suppose s is $\ulcorner F_i \urcorner(\lambda \bar{\mathtt{u}}.\, t)(t_1) \cdots (t_n)$. For any valuation v such that $[\![s]\!]_v = Y$,

i. Each of the t_i is a variable, as else $[\![t_i]\!]_v \in \{Y, N\}$ and $[\![s]\!]_v = N$.

ii. By R5, t cannot be a variable, and thus t is a preword-term.

iii. By R5, $[\![\lambda\bar{\mathrm{u}}.\, t]\!]_v \neq \chi(\emptyset)$.

We can now verify 1 for this s:

- If x is one of the variables t_j, then examining R1–5, the only way we can have $v(x) \in \{A, B\}$, $[\![s]\!]_v = Y$, is if $[\![\lambda\bar{\mathrm{u}}.\, t]\!]_v = \ulcorner C_i \urcorner$ (so that by R4, $\ulcorner F_i \urcorner([\![\lambda\bar{\mathrm{u}}.\, t]\!]_v) = \ulcorner D_i \urcorner$), $v(x)$ is the jth letter of the word D_i, and $v(t_k) = *$ for $k \neq j$. Now by the induction hypothesis 2 for t, $v(z) = *$ for all variables z free in $\lambda\bar{\mathrm{u}}.\, t$ also.

- Suppose the variable x is free in $\lambda\bar{\mathrm{u}}.\, t$, and $v(x) \in \{A, B\}$, $[\![s]\!]_v = Y$. By iii there is a valuation v' with $v'(y) = v(y)$ for all y free in $\lambda\bar{\mathrm{u}}.\, t$ and $[\![t]\!]_{v'} = Y$. Applying the induction hypothesis to t and v', we see that $v(y) = v'(y) = *$ for each $y \neq x$ free in $\lambda\bar{\mathrm{u}}.\, t$, and that $[\![\lambda\bar{\mathrm{u}}.\, t]\!]_v = \chi\{(* \cdots *)\}$, so by R1, we must have $v(t_k) = *$ for each t_k also.

We now derive 2 from 1. If $[\![\lambda\bar{\mathrm{u}}.\, s]\!]_v$ encodes a word, then either for $c = A$ or for $c = B$ we have

$$[\![\lambda\bar{\mathrm{u}}.\, s]\!]_v(c)(*)\cdots(*) = Y.$$

Define a valuation v' by

- $v'(y) = v(y)$ for y free in $\lambda\bar{\mathrm{u}}.\, s$.

- $v'(u_1) = c$ and $v'(u_i) = *$ for $i = 2 \cdots n$.

so that $[\![s]\!]_{v'} = [\![\lambda\bar{\mathrm{u}}.\, s]\!]_v(c)(*)\cdots(*) = Y$, and thus by 1 we have $v(x) = v'(x) = *$ for any variable x free in $\lambda\bar{\mathrm{u}}.\, s$. □

Corollary 8 *If a preword-term s has a closure $\lambda\bar{\mathrm{u}}.\, s$ that satisfies, for some word W,*

$$[\![\lambda\bar{\mathrm{u}}.\, s]\!](x_1)\cdots(x_n) = Y$$

whenever

$$\ulcorner W \urcorner(x_1)\cdots(x_n) = Y$$

then each of the variables $\bar{\mathrm{u}}$ occurs free at most once in s.

Proof Suppose that a variable u_i occurs more than once in s, so that we can find a preword-term $t[y, z]$ such that both y and z occur free in $t[y, z]$, and s is the term $t[u_i, u_i]$. Now by part 1 of the lemma above, $[\![t]\!]_v = N$ whenever $v(y) = v(z) \in \{A, B\}$, so that $[\![s]\!]_v = N$ whenever $v(u_i) \in \{A, B\}$. But by the hypothesis of the lemma, and W1, we must have a valuation v with $v(u_i) \in \{A, B\}$ and $v(s) = Y$, a contradiction. □

Proposition 9 *If a closure $\lambda\bar{\mathtt{u}}.\, s$ of a preword-term s satisfies, for some word W,*

$$[\![\lambda\bar{\mathtt{u}}.\, s]\!](x_1)\cdots(x_n) = Y$$

whenever

$$\ulcorner W \urcorner(x_1)\cdots(x_n) = Y$$

then the word W is derivable, and $[\![\lambda\bar{\mathtt{u}}.\, s]\!] = \ulcorner W \urcorner$.[2]

Proof The proof is by induction on s. If s is $\ulcorner W_0 \urcorner(t_1)\cdots(t_n)$, then the t_i must be distinct variables by W1–3, and $\lambda\bar{\mathtt{u}}.\, s$ must encode the word W_0.

If s is $\ulcorner F_i \urcorner(\lambda x_1, \ldots, x_m.\, t)(t_1)\cdots(t_n)$ then by R4, R5 and W1, all the t_i must be distinct variables. Then by W2 and R1-5, the closure of s that satisfies the condition of the proposition must be

$$s' = \lambda\bar{\mathtt{u}}, t_1, \ldots, t_n, \bar{\mathtt{v}}.\, s$$

for some sequences $\bar{\mathtt{u}}$ and $\bar{\mathtt{v}}$ of variables. Let the lengths of $\bar{\mathtt{u}}$ and $\bar{\mathtt{v}}$ be k and l respectively. By corollary 8, none of the t_i occurs free in $\lambda\bar{\mathtt{x}}.\, t$, so that $t' = \lambda\bar{\mathtt{u}}, \bar{\mathtt{x}}, \bar{\mathtt{v}}.\, t$ is a closed term. The induction predicate for s will follow from the induction hypothesis and lemma 3 if we show that

$$[\![t']\!](x_1)\cdots(x_n) = Y \text{ whenever } \ulcorner W' \urcorner(x_1)\cdots(x_n) = Y \tag{3}$$

for some word W' whose sub-word at positions $k{+}1\cdots k{+}m$ is C_i.

For each $j = 1\cdots k$, there is $a_j \in \{A, B\}$ such that

$$s' \underbrace{(*)\cdots(*)}_{j-1}(a_j)\underbrace{(*)\cdots(*)}_{k-j+n+l} = Y$$

and so by R1 and R5 we must have

$$t' \underbrace{(*)\cdots(*)}_{j-1}(a_j)\underbrace{(*)\cdots(*)}_{k-j+m+l} = Y \tag{4}$$

also. Similarly, for $j = 1\cdots l$, there is $b_j \in \{A, B\}$ such that

$$s' \underbrace{(*)\cdots(*)}_{k+n+j-1}(b_j)\underbrace{(*)\cdots(*)}_{l-j} = Y$$

[2]The reader will no doubt notice that this proposition is stronger than needed. This is a good example of saving work by choosing an induction predicate carefully — the author's original proof of what is actually needed ("if $[\![\lambda\bar{\mathtt{u}}.\, s]\!] = \ulcorner W \urcorner$ then W is derivable") required several pages of tedious technical lemmas similar to lemma 7 in order to make a proof, virtually identical to the one here, work.

so by R1 and R5 we have

$$t' \underbrace{(*)\cdots(*)}_{k+m+j-1}(b_j)\underbrace{(*)\cdots(*)}_{l-j} = Y. \tag{5}$$

We will show that (3) is satisfied when W' is the word $a_1\cdots a_k\, C_i\, b_1\cdots b_l$. For one of $c = A$ or $c = B$, we must have

$$s' \underbrace{(*)\cdots(*)}_{k}(c)\underbrace{(*)\cdots(*)}_{n-1+l} = Y$$

Inspecting R1–5, this can only occur if $\lambda\bar{\mathtt{x}}.\, t[\bar{\mathtt{u}}, \bar{\mathtt{v}} := *]$ encodes the word C_i, and c is the first letter of D_i, so that R4 applies. Thus for $j = 1\cdots m$, we have

$$t' \underbrace{(*)\cdots(*)}_{k+j-1}(c_j)\underbrace{(*)\cdots(*)}_{m-j+l} = Y \tag{6}$$

where c_j is the jth letter of the word C_i. Similarly to the above, for $j = 1\cdots l{+}m{+}k{-}1$,

$$t' \underbrace{(*)\cdots(*)}_{j-1}(L)(R)\underbrace{(*)\cdots(*)}_{l+m+k-j-1} = Y.$$

Together with (4), (5) and (6), this gives (3) with $W' = a_1\cdots a_k\, C_i\, b_1\cdots b_l$ as required. □

Theorem 10 *The λ-definability problem is undecidable.*

Proof By lemma 1, it suffices to show that the relative problem is undecidable. Let W_0 and F_i be a word and set of rules such that the problem, of whether or not a given word is derivable, is undecidable, as given by proposition 2. As our encoding is effective, it now suffices to show that $\ulcorner W \urcorner$ is definable relative to $\ulcorner W_0 \urcorner$ and the $\ulcorner F_i \urcorner$ if and only if W is derivable from W_0 by the rules F_i. One direction is provided by corollary 4. For the other direction, suppose that $\ulcorner W \urcorner = t[\ulcorner W_0 \urcorner, \ulcorner F_1 \urcorner, \ldots, \ulcorner F_n \urcorner]$. By taking the β-normal, η-long form, we may assume that t is the closure of a preword-term, and hence apply the previous proposition. □

The proof given is not optimal in the sense that the size of the model used may be reduced. It is however known that definability of objects of rank two or less is decidable, a decision procedure being given by logical relations [**?**], as is definability in the model with just one element at each type (this latter case is just the decidability of intuitionistic logic). In particular, the proofs given are optimal with respect to the rank of the objects concerned. Recently Jung and Tiuryn [**?**] proved a partial decidability result by restricting the number of bound variables in a λ-term, by using modifications of the notion of logical relation.

Appendix: The Failure of Logical Relations

One of the main reasons for formulating the notion of logical relations was to use them to characterise λ-definablility. For example, in [?], a Kripke logical relation is used to characterise definability in infinite models.

However, as will be shown in this section, logical relations must fail to give such a characterisation in finite models, as otherwise this would yield a decision procedure for definability. As our proof of undecidability is constructive, we could in principle extract an example of this failure from the proof, by diagonalising the appropriate algorithm. Instead we shall construct a reasonably simple example of how logical relations fail to characterise relative definability.

Definition 11 *Given a model M of the λ-calculus, an* ***n*-ary logical relation** *is a collection of relations $R_A \subset M_A^n$ for each type A, such that $(f_1, \ldots, f_n) \in R_{A \to B}$ if and only if*

$$(f_1(x_1), \ldots, f_n(x_n)) \in R_A \text{ whenever } (x_1, \ldots, x_n) \in R_B.$$

*Given a Kripke frame K (*i.e.*, a partial order $(K, \leq)$), an* ***n*-ary Kripke logical relation over *K*** *is a collection of relations $R_{A,i} \subset M_A^n$ for $i \in K$ and A a type, such that*

- *$R_{A,i} \subset R_{A,j}$ whenever $i \leq j$.*
- *$(f_1, \ldots, f_n) \in R_{A \to B,i}$ if and only if $(f_1(x_1), \ldots, f_n(x_n)) \in R_{B,j}$ whenever $i \leq j$ and $(x_1, \ldots, x_n) \in R_{A,j}$.*

Note that a (Kripke) logical relation is completely determined by its behaviour at ground type. We say that $a \in M_A$ is **invariant** *in a (Kripke) logical relation R if $(a, \ldots, a) \in R_A$ ($(a, \ldots, a) \in R_{A,i}$ for all $i \in K$).*

The following lemma is originally from [?], and is easily proved by induction over terms.

Lemma 12 *Let R be a logical relation, t be a term and v_i ($i = 1 \cdots n$) be valuations such that $(v_1(\mathtt{x}), \ldots, v_n(\mathtt{x})) \in R$ for each variable $\mathtt{x}$ free in t. Then $([\![t]\!]_{v_1}, \ldots, [\![t]\!]_{v_n}) \in R$ also.* $\square$

Corollary 13 *If a is λ-definable (relative to $b_1 \cdots b_m$) then a is invariant in every logical relation R (such that the b_i are invariant).* $\square$

These also hold using Kripke logical relations. For classical rather than Kripke logical relations, the following is straightforward.

Proposition 14 *As a predicate of a, A, M and n, "$a \in M_A$ is invariant in every n-ary Kripke logical relation" is decidable. There is a member of our seven-element model which is invariant under every Kripke logical relation, but not λ-definable.*

Proof Given M and n, define K^1 to be

$$\left\{ R_i \,\middle|\, \begin{array}{l} K \text{ is a Kripke frame, } i \in K \text{ and } R \text{ is an} \\ n\text{-ary Kripke logical relation over } K \end{array} \right\}$$

where R_i means the collection $\{R_{B,i}\}_{B:\text{Type}}$, and set $R \leq R'$ iff $R_B \subset R'_B$ for all types B. Now we can define a logical relation R^1 over K^1 by $\bar{x} \in R^1_{B,R}$ iff $\bar{x} \in R_B$. Obviously, $a \in M_A$ is invariant in R^1 if and only if a is invariant in every n-ary Kripke logical relation.

Fixing $a \in M_A$, we now reduce K^1 to a finite Kripke frame by a method similar to that of filtrations in modal logic (see [**?**]). Define an equivalence relation $\equiv$ on K^1 by $R \equiv R'$ iff $R_B = R'_B$ for all sub-types B of A. Let K^A be the quotient $K^1/\equiv$, and for $i, i' \in K^A$ put $i \leq i'$ iff, for $R \in i$ and $R' \in i'$, $R_B \subset R'_B$ for all sub-types B of A. Now let R^A be the unique n-ary Kripke logical relation such that $R^A_{o,R/\equiv} = R_o$ for each equivalence class $(R/\equiv) \in K^A$. It is now easy to verify by induction that for any sub-type B of A, and $(R/\equiv) \in K^A$, we have $R^A_{B,R/\equiv} = R_B$ so that $a \in M_A$ is invariant in R^A iff a is invariant in R^1 iff a is invariant in every n-ary Kripke logical relation.

Further, K^A is finite, as the equivalence class of $R \in K^1$ is determined by the collection $(R_B : B \text{ is a sub-type of } A)$, and thus the size $|K^A|$ is bounded by $N = \prod_B 2^{|M_B|^n}$ where B ranges over the sub-types of A, which is recursive in A, M and n. Thus a is invariant in every n-ary Kripke logical relation if and only if it is invariant in every n-ary Kripke logical relation over a Kripke frame of size at most N. This shows that the given predicate is recursive.

It follows that "$a \in M_A$ is invariant in every Kripke logical relation" is co-r.e. in a, A and M. λ-definability is obviously r.e., and so cannot be co-r.e. as this would contradict theorem 10. Since λ-definability implies invariance, the converse implication must fail, so that there is a member of the seven element model that is invariant but not definable. $\Box$

We finish by constructing an example where relative invariance holds but relative λ-definability fails. The example is given by the encoding of a semi-Thue system in our seven element model. For $n \geq 1$, let W_n be the word consisting of n letter A's. Let F_1 be the rule $A \longrightarrow AAA$ and F_2 be the rule $AAA \longrightarrow A$, so that W_n is derivable from W_1 if and only if n is odd. However, we will see that (the encoding of) W_2 (in fact any W_k) is invariant in any logical relation such that F_1, F_2 and W_1 are.

Proposition 15 *Let R be an n-ary (Kripke) logical relation.*

1. *If $k > 3n$ and $\ulcorner W_{k-1} \urcorner$ is invariant in R, then so is $\ulcorner W_k \urcorner$.*
2. *If $\ulcorner F_1 \urcorner$, $\ulcorner F_2 \urcorner$ and $\ulcorner W_1 \urcorner$ are invariant in R, then so is $\ulcorner W_2 \urcorner$. But $\ulcorner W_2 \urcorner$ is not λ-definable from $\ulcorner F_1 \urcorner$, $\ulcorner F_2 \urcorner$ and $\ulcorner W_1 \urcorner$.*

Proof We shall assume R is a classical logical relation, although the proof works for Kripke relations also.

For the first part, suppose that $k > 3n$, and $\ulcorner W_{k-1} \urcorner$ is invariant in R. We show that $\ulcorner W_k \urcorner$ is invariant in R also. Consider x_i^j such that $(x_i^1, \ldots, x_i^n) \in R_o$ for $i = 1 \cdots k$. We must show that also

$$(\ulcorner W_k \urcorner (x_1^1) \cdots (x_k^1), \ldots, \ulcorner W_k \urcorner (x_1^n) \cdots (x_k^n)) \in R_o. \tag{7}$$

For each $j = 1 \cdots n$, we shall choose a set $C_j \subset \{1, \ldots, k\}$ with at most three members as follows:

- If there is $i_0 \in \{1, \ldots, n-2\}$ such that $x_{i_0}^j = L$, $x_{i_0+2} = R$ and $x_i^j = *$ for all other $i = 1 \cdots n$, then let $C_j = \{i_0, i_0+1, i_0+2\}$.
- If there are more than three $i \in \{1, \ldots, n\}$ such that $x_i^j \neq *$ then let C_j be the set containing the first three such i.
- Otherwise let $C_j = \{\, i \mid x_i^j \neq * \,\}$.

The C_j have been chosen so that if $i \in \{1, \ldots, k\} - C_j$, then

$$\ulcorner W_{k-1} \urcorner (x_1^j) \cdots (x_{i-1}^j)(x_{i+1}^j) \cdots (x_k^j) = \ulcorner W_k \urcorner (x_1^j) \cdots (x_k^j). \tag{8}$$

Since $k > 3n$, and the C_j have at most three members, there is $i \in \{1, \ldots, k\}$ that is not in any of the C_j. Now

$$(\ulcorner W_{k-1} \urcorner (x_1^j) \cdots (x_{i-1}^j)(x_{i+1}^j) \cdots (x_k^j))_{1 \leq j \leq k} \in R$$

as $\ulcorner W_{k-1} \urcorner$ is invariant in R, and so by (8) we see that (7) holds also.

For the second part of the proposition, suppose that $\ulcorner F_1 \urcorner$, $\ulcorner F_2 \urcorner$ and $\ulcorner W_1 \urcorner$ are invariant in R. Take an even number $k > 3n$. Now $\ulcorner W_{k-1} \urcorner$ is derivable as $k-1$ is odd, so that by corollary 4, $\ulcorner W_{k-1} \urcorner$ is definable, and thus invariant by corollary 13. Now by the first part $\ulcorner W_k \urcorner$ is invariant also. $\ulcorner W_2 \urcorner$ is derivable from $\ulcorner W_k \urcorner$ by the rule $\ulcorner F_2 \urcorner$, so that similarly $\ulcorner W_2 \urcorner$ is invariant. That $\ulcorner W_2 \urcorner$ is not definable is obvious from proposition 9. $\square$