Math. Struct. in Comp. Science (2000), *vol.* 10, *pp.* 595–615. Printed in the United Kingdom

On the number of fixed points of a combinator in lambda calculus

B. INTRIGILA[†] and E. BIASONE

Dipartimento di Matematica Pura e Applicata, Universitá Di L'Aquila,
Via Vetoio, 67010 Coppito, L'Aquila, Italy

Email: intrigil@univaq.it and e.biasone@reply.it

Received 13 September 1998; revised 19 November 1999

Dedicated to Roger Hindley, on the occasion of his 60th birthday.

We consider the problem of determining the number of fixed points of a combinator (a closed λ-term). This appears to be a still unsolved problem. We give a partial answer by showing that if there is a fixed point *in normal form*, then this fixed point is unique or there are infinitely many fixed points.

1. Introduction

One of the more striking characteristics of untyped λ-calculus is the fixed point theorem, which asserts that every combinator (that is a closed lambda-term) has at least one fixed point.

This paper is concerned with the following question:

Question. How many fixed points can a combinator have? In particular, is there a combinator that has a finite number (greater than one) of fixed points?

This is a very natural question, which, strangely enough, has remained unnoticed until now (as far as we know). For further references, let us call it Question NFP (Number of Fixed Points). Observe that we are always considering *closed terms*, so that we are asking for the number of fixed points that, in addition, are closed terms.

This is to be compared with the analogous question concerning the number of values of a combinator, which is the subject of the celebrated Range Theorem due to Barendregt and Myhill. The *range* $Ra(T)$ of a combinator T is the set $\{N \mid \exists M \; TM = N,\, N \text{ closed}\}$, where $=$ is the convertibility relation. The Range Theorem (see Barendregt (1984, Theorem 17.1.16)) asserts that every combinator has the *range property*, that is, its range is either infinite or a singleton (obviously, up to convertibility).

[†] Work partially supported by Progetto Cofinanziato MURST 'Tecniche formali per la specifica, l'analisi, la verifica, la sintesi e la trasformazione di sistemi software'.

Here the situation is well understood, as shown in Barendregt (1993). The reader is urged to consult this beautiful paper both for itself and because its conceptual framework is useful for analysing our present problem.

Question NFP may be seen as a particular case of the range property. In particular, we may think that only a constant combinator can have an unique fixed point. So, we start with the following puzzle.

Puzzle (solution at the end of this Introduction) Find a combinator F, such that F has infinite range, whilst the equation $FX = X$ has exactly one solution.

One may think of a more subtle reduction as follows. Let a combinator F be given, and let P_1,P_2 be *two* fixed points of F. If we can find a fixed point combinator Hx for F (that is $HU = F(HU)$ for every U), such that $P_1,P_2 \in Ra(H)$, then everything would follow from the range property, since H has infinite range. However, every direct construction (at least, every construction we were able to do) of such an H gives no assurance that both P_1 and P_2 are in the range of H. On the other hand, for special classes of terms this kind of approach may be successful. See Appendix A for an example.

Since such a direct attempt does not work in general, we can resort to the *recursion theoretic* approach that was so successful in proving the Range Theorem. Moreover, quoting from Barendregt (1993), 'the range property is really a result in recursion theory'. (This point of view is substantiated in Barendregt (1993), by results from Barendregt and Statman, which give a general formulation of the range property in the Ershov-enumerations setting.)

However, from an analysis of the recursion theoretic proof of the Range Theorem, it seems doubtful that this kind of argument can be useful in the present case. See Appendix B (Part 1) for details.

Moreover, one may also argue that for this problem a general recursion theoretic argument cannot exist by a comparison with analogous problems in Recursion Theory. In Appendix B (Part 2), we recall a few points about the number of fixed points of recursive functions with respect to the Recursion Theorem framework.

To summarise our previous discussion, we consider Question NFP a well-distinguished problem, different from the range property. In the following we give a partial answer by showing that if there is a fixed point *in normal form*, then this fixed point is unique or there are infinitely many fixed points. (It is perhaps worth noting that for a term F to have a fixed point in normal form by no means implies that F itself has a normal form.) We leave the general question open.

To give an idea of the method of proof used in the following sections, let us consider the fixed point equation $FX = X$ under the hypothesis that a solution N in normal form exists. Treating the unknown X as a free variable, we see that FX must have a *head normal form*, say $FX = \lambda x_1 \cdots x_n.\xi M_1 \cdots M_m$. If $\xi \equiv X$, an infinite number of solutions can be immediately obtained making use of (a family of) fixed points of the combinator $\mathbf{K}$. Indeed, if $\mathbf{K}^\infty$ is a fixed point of $\mathbf{K}$, then it is easily seen that

$$\mathbf{K}^\infty = \lambda x_1 \cdots x_n.\mathbf{K}^\infty M_1 \cdots M_m$$

using the equalities $\mathbf{K}^\infty x = \mathbf{K}^\infty$ and $\lambda x.\mathbf{K}^\infty = \mathbf{K}^\infty$.

https://doi.org/10.1017/S0960129500003091 Published online by Cambridge University Press

Otherwise ξ is one of the x_i, and *any* solution X must have the form:

$$\lambda x_1 \cdots x_n.x_i N_1 \cdots N_m, \text{ for some } N_1, \dots, N_m.$$

The idea is now to shift to *systems of fixed point equations* considering the terms N_i as new unknowns $X_1, \dots, X_m$. Observe that for each new unknown there is a natural fixed point equation that is, roughly speaking, $X_i = M_i$. This argument can be iterated, and since there is a solution in normal form, the process must eventually stop. At this point (simplifying matters) all the equations either have no unknowns or are in the form:

$$X_i = \lambda x_1 \cdots x_n.X_j M_1 \cdots M_m.$$

In the former case there is one solution only; in the latter we can again make use of fixed points of $\mathbf{K}$.

We end by remarking that our approach is completely *constructive*, that is we shall set up a semialgorithm that given a system of fixed point equations returns either the unique solution or a generic member of an infinite family of solutions; however, the semialgorithm may not terminate when there is no solution in normal form.

The paper is organised as follows:

- In Section 2 we introduce the notions needed to treat systems of fixed point equations; our formal setting is very similar to Böhm and Tronci (1991), Tronci (1996a) and Tronci (1996b), but such papers treat different kinds of equations. In fact, the right-hand side of their equations always has to start with a fresh free variable.
- In Section 3 we show how to transform systems of fixed point equation and prove our main result.
- Section 4 contains other related results we are working out.
- In Appendix A we show that if F has a suitable fixed point combinator, then F has either one or infinite fixed points, irrespective of whether or not it has a fixed point in normal form.
- In Appendix B, Part 1, we discuss the recursion theoretic proof of the Range Theorem with respect to the NFP Problem. In Part 2, we make some remarks on the NFP Question in Recursion Theory.

Solution of the puzzle. Take, for example, $F \equiv \lambda xy.y(x\mathbf{O})\mathbf{I}$.

2. Systems of fixed point equations

In general we follow Barendregt (1984) for notation and terminology. In particular, we make use of the following combinators:

$$\mathbf{I} \equiv \lambda x.x,$$
$$\mathbf{K} \equiv \lambda xy.x,$$
$$\mathbf{O} \equiv \lambda xy.y,$$
$$\omega \equiv \lambda x.xx,$$
$$\mathbf{\Omega} \equiv \omega\omega.$$

https://doi.org/10.1017/S0960129500003091 Published online by Cambridge University Press

It is preferable to treat systems of fixed point equations, rather than one single equation. To this end, we introduce the following notions.

Definition 2.1.

1 We consider the following alphabet:
$x_0, x_1, \dots$ variables,
$X_0, X_1, \dots$ unknowns,
λ abstractor,
(,) parentheses.

2 Given a finite set V of unknowns, the set of *terms* $\Gamma(V)$ on V is defined inductively as follows:

(a) $X \in \Gamma(V)$ for $X \in V$.

(b) $\Lambda \subseteq \Gamma(V)$, where Λ is the set of λ-terms.

(c) $M, N \in \Gamma(V) \Rightarrow MN \in \Gamma(V)$.

(d) $M \in \Gamma(V) \Rightarrow \lambda x.M \in \Gamma(V)$.

Remark 2.1. The notion of *β-reduction* and the related notion of *convertibility*, are treated in $\Gamma(V)$ as usual, considering the unknowns as free variables, and they are denoted by the same symbols.

Definition 2.2. Let $V = \{X_1, \dots, X_m\}$ be given, with $m \geqslant 0$ and $E = \{X_i = M_i \mid M_i \in \Gamma(V)\}_{0 \leqslant i \leqslant m}$.

1 The pair (E, V) is said to be a *system of (fixed point) equations* in the unknowns V.
2 Let $S = (E, V)$ be a system. The formula $M = N \in E$ is said to be an *equation* of S.
3 If $m = 0$, then S is called *the empty system.*
4 S is *closed* if every M_i contains no free variables (but possibly M_i contains unknowns).

Observe that S has m equations and m unknowns.

Notation 1.

1 $S, S_1, S_2, \dots$ denote systems of equations.
2 $e, e_1, e_2, \dots$ denote the equations of a system.
3 $E(S) =_{\text{def}}$ the set of equations of system S.
4 $V(S) =_{\text{def}}$ the set of unknowns of system S.

Definition 2.3. Let $S = (E, V)$ be a system with $E = \{X_i = M_i \mid M_i \in \Gamma(V)\}_{0 \leqslant i \leqslant m}$.

1 A *solution* of S is an m-tuple $(T_1, \dots, T_m)$, $T_i \in \Lambda$, such that $T_1 = \widetilde{M}_1, \dots, T_m = \widetilde{M}_m$, where $\widetilde{M}_i = M_i[T_1/X_1, \dots, T_m/X_m]$, $1 \leqslant i \leqslant m$. If $(T_1, \dots, T_m)$ is a solution of S, we say that S *has* $(T_1, \dots, T_m)$ as solution, and we call T_i the *i-th component* or the *X_i-component* of the solution.
2 A solution $(T_1, \dots, T_m)$ of S is *in normal form* if $\forall j$, $1 \leqslant j \leqslant m$, T_j has nf.
3 A solution $(T_1, \dots, T_m)$ of S is *closed*, if $\forall j$, $1 \leqslant j \leqslant m$, T_j is a closed term.

Convention 2.1. The empty set $\emptyset$, which we call the *empty solution*, is the only solution of the empty system and it is in normal form.

Notation 2. Let $S = (E, V)$ be a system.

1 $R(E) =_{\mathrm{def}} \{Q \mid P = Q \in E\}$.
2 If $P = Q \in E$ then $R(P = Q) =_{\mathrm{def}} Q$.

Definition 2.4.

1 Let $T \in \Gamma(V)$. T is said to be a *constant* if in T there are no occurrences of unknowns (that is, $T \in \Lambda$).
2 Let $S = (E, V)$ be a system and $X = T \in E$. Then X is called a *constant unknown* of S if T is a constant.

Definition 2.5. Let V be a set of unknowns.

1 A term $M \in \Gamma(V)$ is *in head normal form* (hnf) if M is of the form

$$M \equiv \lambda x_1 \cdots x_n.xM_1 \cdots M_m n, m \geqslant 0.$$

The *head variable* of this M is x. HNF is the set of hnf's.
2 M *has* a head normal form if $M = M_1$ for some $M_1 \in \mathrm{HNF}$.
3 A term $M \in \Gamma(V)$ is an *unknown-head normal form* (u-hnf) if M is of the form $M \equiv \lambda x_1 \cdots x_n.XM_1 \cdots M_m, n, m \geqslant 0$. The *unknown-head variable* of this M is X, for some $X \in V$. u-HNF is the set of u-hnf's.
4 M *has* a u-head normal form if $M = M_1$ for some $M_1 \in$ u-HNF.

Definition 2.6. Let $S = (E, V)$ be a system. S is *in standard form* if $\forall Q \in R(E)$, Q is not a constant and Q is either an hnf or a u-hnf.

Example 2.1. Let $S = (E, V)$ be a system with $V = \{X_1, X_2, X_3\}$ and

$$\begin{aligned} E = \{X_1 &= \lambda yz.z(X_3 y\mathbf{K}), \\ X_2 &= X_1\mathbf{IKIKK}, \\ X_3 &= \lambda y_2 y_3 y_4.y_4(X_2 y_2 \mathbf{I})\} \end{aligned}$$

S is in standard form.

Definition 2.7. Let $E = \{e_1, \ldots, e_m\}$ be a set of equations. Let $S = (E, V)$ be a system in standard form. We define the following sets:

— $hnf(S) = \{e \in E(S) : R(e) \text{ is an hnf}\}$
— $u\text{-}hnf(S) = \{e \in E(S) : e \notin hnf(S)\}$.

Example 2.2. Let $S = (E, V)$ be a system in standard form with

$$V = \{X_1, X_2, X_3, X_4, X_5, X_6\}$$

and

$$\begin{aligned} E = \{X_1 &= \lambda yz.z(X_3 y\mathbf{K}), \\ X_2 &= X_1\mathbf{IKIKK}, \\ X_3 &= \lambda y_2 y_3 y_4.y_4(X_2 y_2 \mathbf{I}), \\ X_4 &= \lambda x.X_2\mathbf{IK}, \\ X_5 &= \lambda yz.X_6\mathbf{KK}, \\ X_6 &= \lambda yz.X_5\mathbf{KII}\}. \end{aligned}$$

https://doi.org/10.1017/S0960129500003091 Published online by Cambridge University Press

Then

$$\begin{aligned} hnf(S) &= \{X_1 = \lambda yz.z(X_3 y\mathbf{K}),\ X_3 = \lambda y_2 y_3 y_4.y_4(X_2 y_2 \mathbf{I})\} \\ u\text{-}hnf(S) &= \{X_2 = X_1\mathbf{IKIKK},\ X_4 = \lambda x.X_2\mathbf{IK},\ X_5 = \lambda yz.X_6\mathbf{KK},\ X_6 = \lambda yz.X_5\mathbf{KII}\} \end{aligned}$$

Definition 2.8. Let $S = (E, V)$ be a system in standard form. S is *in final form* if $hnf(S) = \emptyset$.

So, if S is not empty, S is in final form if and only if for all $Q \in R(E)$, Q is neither a constant nor in hnf, but Q is in u-hnf.

Convention 2.2. The empty system is in final form.

Example 2.3. Let $S = (E, V)$ be a system with $V = \{X_1, X_2, X_3\}$ and

$$\begin{aligned} E = \{X_1 &= \lambda y.X_3 y, \\ X_2 &= X_3\mathbf{IKK}, \\ X_3 &= \lambda y_2.X_2 y_2 \mathbf{I}\}. \end{aligned}$$

S is in final form.

Definition 2.9. Let $S = (E, V)$ be a system and $E = \{X_i = M_i \mid M_i \in \Gamma(V)\}_{0 \leqslant i \leqslant m}$. Let $E_1 = \{X_i = \widetilde{M_i} \mid \widetilde{M_i}$ is an hnf or a u-hnf of $M_i\}_{0 \leqslant i \leqslant m}$.

1 The system $S_1 = (E_1, V(S))$ is said to be a *head normal form* of S.
2 The system S is *in head normal form* if $\forall Q \in R(E)$, Q is either in hnf or in u-hnf.

Example 2.4. Let $S_1 = (E_1, V_1)$ be a system with $V_1 = \{X_1, X_2, X_3, X_4, X_5, X_6\}$ and

$$\begin{aligned} E_1 = \{X_1 &= (\lambda xyz.z(xy\mathbf{K}))X_3, \\ X_2 &= X_1\mathbf{IKIKK}, \\ X_3 &= (\lambda y_1 y_2 y_3 y_4.y_4(y_1 y_2 \mathbf{I}))X_2, \\ X_4 &= (\lambda xyz_1 z_2.z_1 z_2)X_1 X_2 \mathbf{K}, \\ X_5 &= \lambda xyz.X_4\mathbf{I}X_1\mathbf{KK}, \\ X_6 &= (\lambda z.X_4)X_5\mathbf{KK}\}. \end{aligned}$$

Then the system $S_2 = (E_2, V_2)$ with $V_2 = \{X_1, X_2, X_3, X_4, X_5, X_6\}$ and

$$\begin{aligned} E_2 = \{X_1 &= \lambda yz.z(X_3 y\mathbf{K}), \\ X_2 &= X_1\mathbf{IKIKK}, \\ X_3 &= \lambda y_2 y_3 y_4.y_4(X_2 y_2 \mathbf{I}), \\ X_4 &= \lambda z_2 y.z_2, \\ X_5 &= \lambda xyz.X_4\mathbf{I}X_1\mathbf{KK}, \\ X_6 &= X_4\mathbf{KK}\} \end{aligned}$$

is in head normal form and it is a head normal form of S_1.

Definition 2.10. (Barendregt 1984, Definitions 8.3.9 and 8.3.10).

https://doi.org/10.1017/S0960129500003091 Published online by Cambridge University Press

1 If $M \in \Gamma(V)$ is of the form

$$M \equiv \lambda x_1 \cdots x_n.(\lambda x.M_0)M_1 \cdots M_m$$

$n \geqslant 0$, $m \geqslant 1$, then $(\lambda x.M_0)M_1$ is called the *head redex* of M.

2 Suppose M has Δ as head redex. Write

$$M \underset{h}{\longmapsto} N$$

if, $M \xrightarrow{\Delta} N$, that is, N results from M by contracting Δ. Then $\underset{h}{\longmapsto}$ is called *one step head reduction.*

3 The *head reduction* (*path*) of M is the uniquely determined sequence $M_0, M_1 \ldots$ such that $M \equiv M_0 \underset{h}{\longmapsto} M_1 \underset{h}{\longmapsto} \cdots$.

If M_n is an hnf or a u-hnf, then the head reduction of M is said to *terminate* at M_n. Otherwise M has an *infinite* head reduction.

The proof of the following theorem does not depend on the presence of unknowns.

Theorem 2.1. (Barendregt 1984, Theorem 8.3.11). M has either an $h$nf or a u-hnf iff the head reduction path of M terminates.

Theorem 2.2. Let $S = (E, V)$ be a system that has a solution in normal form. Then there exists a system S' such that S' is a head normal form of S. Moreover, S' is such that S and S' have the same solutions. So, in particular:

1 S has either one or infinite solutions iff S' has either one or infinite solutions.

2 S' has a solution in normal form.

Proof. We consider a generic equation of S:

$$X_i = M_i$$

We may assume that the unknowns (which play no role) have been replaced by fresh free variables.

If the sequence of head-reductions in M_i does not terminate, then it does not terminate with any sequence of terms replaced for the fresh free variables. But this is not possible since there is a solution in nf.

So $E(S') = \{X_i = \widetilde{M_i}\}_{1 \leqslant i \leqslant m}$, where $\widetilde{M_i}$ is either an hnf or a u-hnf of M_i. The proof that S' has the additional properties is immediate since any β-reduction preserves solutions. □

3. Transformations of systems of equations

In this section we show how to transform our systems of equations with the aim of proving that some suitable form can always be obtained. In showing this, we use a procedural approach, which also proves that the final form can be effectively constructed under the hypothesis that a solution in normal form exists. So, in the following we always start from a given system S and define two kind of (one step) transformation, namely:

— elimination of constants,
— expansion of unknowns.

For each one, we show how the set of solutions is correspondingly transformed. Then we consider sequences of transformation steps and prove that they eventually stop on a final form.

3.1. *Elimination of constants*

In this subsection we show how to eliminate constants.

Definition 3.1. Let $S = (E, V)$ be a system. Let $\{X_1 = T_1, \ldots, X_k = T_k\} \subseteq E(S)$ be such that for $1 \leqslant i \leqslant k$, we have T_i is a constant. S_1 is said to be *obtained by elimination of constants* of S if

$$V_1 = V \setminus \{X_1, \ldots, X_k\},$$
$$E_1 = \{X_i = \widetilde{M_i}\}_{k+1 \leqslant i \leqslant m} \text{ where } \widetilde{M_i} = M_i[T_1/X_1, \ldots, T_k/X_k].$$

We recall that a term T is a constant if in T there are no occurrences of unknowns.

Now we show how the set of solutions of a system S is changed by an elimination of constants.

The proof of the following propositions is immediate.

Proposition 3.1. Let S be a system. Let $\{X_i = T_i\}_{1 \leqslant i \leqslant k} \subseteq E(S)$ with T_i constant for $1 \leqslant i \leqslant k$. Let S_1 be a system obtained by elimination of constants of S.

1 If S is closed, then S_1 is closed.
2 For every solution $(T_1, \ldots, T_k, H_{k+1}, \ldots, H_m)$ of S, we have $(H_{k+1}, \ldots, H_m)$ is a solution of S_1.
3 For every solution $(H_{k+1}, \ldots, H_m)$ of S_1, we have $(T_1, \ldots, T_k, H_{k+1}, \ldots, H_m)$ is a solution of S.

Proposition 3.2. Let $S = (E, V)$ be a system and S_1 be a system obtained by elimination of constants of S. Then:

1 S has either one or infinite solutions iff S_1 has either one or infinite solutions;
2 S has a solution in normal form iff S_1 has a solution in normal form.

Remark 3.1. The special case of *empty solution* arises by elimination of constant unknowns, when all unknowns turn out to be constant (and such constant terms are the unique solution of the original system). Also, observe that when eliminating a constant unknown, we are losing information. Therefore, to ensure the effectiveness of the process, we must remember all the substitutions performed in the transformation. See also Section 4.

3.2. *Expansions of unknowns.*

This is our main transformation rule. The idea is, starting from, for example,

$$X_1 = \lambda yz.z(X_3 y\mathbf{K})$$

https://doi.org/10.1017/S0960129500003091 Published online by Cambridge University Press

to observe that every solution is constrained to have the form $X_1 = \lambda yz.z(Yy)$ for some unknown Y. In turn, Y must satisfy the equation $Y = \lambda y.(X_3 y\mathbf{K})$. If we eliminate X_1 in favour of Y, we in some sense go down into the Böhm tree of any solution. When there is a solution in nf, this process must eventually stop. Now we turn to formal definitions.

Definition 3.2. Let $S = (E, V)$ be a system such that $hnf(S) \neq \emptyset$.

1 An equation of $hnf(S)$ is said to be *expansive*.
2 An unknown X_j such that $X_j = M_j \in hnf(S)$ is said to be *expansive*.

Let $S = (E, V)$ be a system such that $hnf(S) \neq \emptyset$.
Let

$$X_i = M_i \tag{1}$$

be an expansive equation of S. Thus

$$M_i = \lambda y_1 \cdots y_n.zM_{i_1} \cdots M_{i_{k_i}}$$

with $n \geqslant 0$, $k_i \geqslant 1$, $M_{i_j} \in \Gamma(V)$ and Equation (1) has the form

$$X_i = \lambda y_1 \cdots y_n.zM_{i_1} \cdots M_{i_{k_i}}.$$

Given M_{i_j}, $1 \leqslant j \leqslant k_i$, let $\vec{v}_{i_j} = y_{i_{j_1}} \cdots y_{i_{j_k}}$ where $y_{i_{j_1}}, \dots, y_{i_{j_k}} \in \{y_1, \dots, y_n\}$ are the variables occurring in M_{i_j}, put in the abstraction order.

So, by introducing the new unknowns $X_{i_1}, \dots, X_{i_{k_i}}$, we can write X_i in the new form:

$$X_i = \lambda y_1 \cdots y_n.z(X_{i_1}\vec{v}_{i_1}) \cdots (X_{i_{k_i}}\vec{v}_{i_{k_i}}) \tag{2}$$

Equation (2) is called the *associated equation* of Equation (1).

Of course, each new unknown X_{i_j} has to satisfy the equation $X_{i_j} = \lambda\vec{v}_{i_j}.\widetilde{M}_{i_j}$ where $\widetilde{M}_{i_j} \equiv M_{i_j}[\lambda y_1 \cdots y_n.z(X_{i_1}\vec{v}_{i_1}) \cdots (X_{i_{k_i}}\vec{v}_{i_{k_i}})/X_i]$.

Definition 3.3. The set of equations $\mathrm{Exp}(X_i = M_i) = \{X_{i_1} = \lambda\vec{v}_{i_1}.\widetilde{M}_{i_1}, \dots, X_{i_{k_i}} = \lambda\vec{v}_{i_{k_i}}.\widetilde{M}_{i_{k_i}}\}$ is called the *expansion* of equation $X_i = M_i$, where

$$\widetilde{M}_i \equiv M_i[\lambda y_1 \cdots y_n.z(X_{i_1}\vec{v}_{i_1}) \cdots (X_{i_{k_i}}\vec{v}_{i_{k_i}})/X_i].$$

$X_{i_1}, \dots, X_{i_{k_i}}$ are called the *unknowns of the expansion* of the equation $X_i = M_i$.

Example 3.1. Let $S = (E, V)$ be a system with $V = \{X_1, X_2, X_3\}$ and

$$\begin{aligned} E = \{X_1 &= \lambda yz.z(X_3 y\mathbf{K})(zX_2), \\ X_2 &= X_1\mathbf{IKIKK}, \\ X_3 &= \lambda y_1 y_2 y_3.y_3(X_2 y_1 \mathbf{I})\} \end{aligned}$$

— $X_1 = \lambda yz.z(X_3 y\mathbf{K})(zX_2)$ is an expansive equation,
— $X_1 = \lambda yz.z(X_{11}y)(X_{12}z)$ is its associated equation,

The expansion of the previous equation is

— $X_{11} = \lambda y.X_3 y\mathbf{K}$
— $X_{12} = \lambda z.zX_2$.

https://doi.org/10.1017/S0960129500003091 Published online by Cambridge University Press

Definition 3.4. Let S be a system and X_i be an expansive unknown of S. Then S' is said to be a system obtained by *expansion* of the unknown X_i if:

1 $V(S') = V(S) \setminus \{X_i\} \cup \{X_{i_1}, \dots, X_{i_{k_i}}\}$ where $\{X_{i_1}, \dots, X_{i_{k_i}}\}$ are the unknowns of the expansions of the equation $X_i = M_i$.
2 $E(S') = \mathrm{Exp}(X_i = M_i) \cup \{X_j = \tilde{M}_j\}_{1 \leqslant j \leqslant m \text{ and } j \neq i}$ where

$$\widetilde{M}_j \equiv M_j[\lambda y_1 \cdots y_n.z(X_{i_1}\vec{v}_{i_1}) \cdots (X_{i_{k_i}}\vec{v}_{i_{k_i}})/X_i].$$

Theorem 3.1. Let S be a system, $X_i = \lambda y_1 \cdots y_n.zM_{i_1} \cdots M_{i_{k_i}}$ be an expansive equation of S and $X_i = \lambda y_1 \cdots y_n.z(X_{i_1}\vec{v}_{i_1}) \cdots (X_{i_{k_i}}\vec{v}_{i_{k_i}})$ be its associated equation.
Moreover, let S_1 be a system obtained by expansion of the variable X_i of $V(S)$.

1 If S is closed, then S_1 is closed.
2 For every solution $(H_1, \dots, H_{i-1}, H_i, H_{i+1}, \dots, H_m)$ of S, H_i must have the form

$$H_i \equiv \lambda y_1 \cdots y_n.zH_{i_1} \cdots H_{i_{k_i}}$$

and $(H_1, \dots, H_{i-1}, \lambda\vec{v}_{i_1}.H_{i_1}, \dots, \lambda\vec{v}_{i_{k_i}}.H_{i_{k_i}}, H_{i+1}, \dots, H_m)$ is a solution of S_1.
3 For every solution $(H_1, \dots, H_{i-1}, \lambda\vec{v}_{i_1}.H_{i_1}, \dots, \lambda\vec{v}_{i_{k_i}}.H_{i_{k_i}}, H_{i+1}, \dots, H_m)$ of S_1, $(H_1, \dots, H_{i-1}, H_i, H_{i+1}, \dots, H_m)$ with $H_i \equiv \lambda y_1 \cdots y_n.zH_{i_1} \cdots H_{i_{k_i}}$ is a solution of S.

Proof.

1 Statement (1) is obvious.
2 Let a solution $(H_1, \dots, H_{i-1}, H_i, H_{i+1}, \dots, H_m)$ of S be fixed. Since

$$X_i = \lambda y_1 \cdots y_n.zM_{i_1} \cdots M_{i_{k_i}}$$

is an equation of S, the component H_i must satisfy the equation

$$X_i = \lambda y_1 \cdots y_n.z(M_{i_1}[H_1/X_1 \cdots H_m/X_m]) \dots (M_{i_{k_i}}[H_1/X_1 \cdots H_m/X_m]),$$

and therefore, up to convertibility, H_i must have the form

$$H_i \equiv \lambda y_1 \cdots y_n.zH_{i_1} \cdots H_{i_{k_i}}.$$

Let $\vec{v}_{i_j}$ be, as usual, the sequence of variables occurring free in M_{i_j}, with $1 \leqslant j \leqslant k_i$, put in the abstraction order. First we show that we may assume that each $\lambda\vec{v}_{i_j}.H_{i_j}$ is closed. Indeed, we may assume that $M_{i_j}[H_1/X_1 \cdots H_m/X_m] \longrightarrow^* H_{i_j}$ (otherwise we can choose a common reduct), and therefore all free variables in H_{i_j} must occur in $M_{i_j}[H_1/X_1 \cdots H_m/X_m]$. It is, moreover, obvious that

$$H_i \equiv \lambda y_1 \cdots y_n.z((\lambda\vec{v}_{i_1}.H_{i_1})\vec{v}_{i_1}) \cdots ((\lambda\vec{v}_{i_{k_i}}.H_{i_{k_i}})\vec{v}_{i_{k_i}}) \tag{3}$$

(where, by an abuse of notation, we have not renamed variables). Now let

$$\bar{M}_{i_j} \equiv M_{i_j}[\lambda y_1 \cdots y_n.z(X_{i_1}\vec{v}_{i_1}) \cdots (X_{i_{k_i}}\vec{v}_{i_{k_i}})/X_i],$$

and

$$\widetilde{\bar{M}}_{i_j} \equiv \bar{M}_{i_j}[H_1/X_1, \dots, H_{i-1}/X_{i-1}, (\lambda\vec{v}_{i_1}.H_{i_1})/X_{i_1}, \dots, (\lambda\vec{v}_{i_{k_i}}.H_{i_{k_i}})/X_{i_{k_i}}, H_{i+1}/X_{i+1}, \dots, H_m/X_m].$$

https://doi.org/10.1017/S0960129500003091 Published online by Cambridge University Press

We have to show

$$\lambda\vec{v}_{i_j}.H_{i_j} = \lambda\vec{v}_{i_j}.\widetilde{\bar{M}}_{i_j} \text{ for all } 1 \leqslant j \leqslant k_i. \tag{4}$$

By (3)

$$\widetilde{\bar{M}}_{i_j} \equiv \bar{M}_{i_j}[H_1/X_1, \ldots, H_{i-1}/X_{i-1}, H_i/X_i, H_{i+1}/X_{i+1}, \ldots, H_m/X_m],$$

and therefore $H_{i_j} = \widetilde{\bar{M}}_{i_j}$ holds since $(H_1, \ldots, H_{i-1}, H_i, H_{i+1}, \ldots, H_m)$ is a solution.

3 The proof is analogous to the proof of Statement (2). □

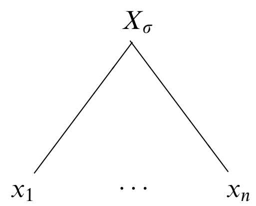
Corollary 3.1. Let S be a system and S_1 be a system obtained by expansion of some unknown of S. Then:

1 S has either one or infinite solutions iff S_1 has either one or infinite solutions.
2 S has a solution in normal form iff S_1 has a solution in normal form.

3.3. *Sequences of expansions.*

The notion of a *Böhm tree* in the presence of unknowns is defined as usual treating unknowns as variables. (Also related notions, such as *path* in the Böhm tree, *etc.* are as usual.)

In the following we abbreviate



by $X_\sigma\vec{x}$, where $\vec{x} = x_1 \cdots x_n$.

Definition 3.5. Let π : $S_0, S_1, \ldots, S_i, \ldots, S_k, \ldots$ be a sequence of systems such that S_i is obtained by expansion of S_{i-1}, or by elimination of constants of S_{i-1}.

Let X be an unknown and S_i the first system such that $X \in V(S_i)$. We consider, moreover, the subsequence τ: $S_i, \ldots, S_k, \ldots$ of π.

We define *the sequence of Böhm trees of X* $BT_{S_i}(X), \ldots, BT_{S_k}(X), \ldots$ corresponding to τ as follows:

1 $BT_{S_i}(X) = X$.
2 $\forall S_j$, with $j > i$, $BT_{S_j}(X)$ is obtained from $BT_{S_{j-1}}(X)$ as follows:

(a) **Expansion of unknowns.** Let σ be a path such that $(BT_{S_{j-1}}(X))_\sigma = X_\sigma\vec{v}_\sigma$ and $X_\sigma \in V(S_{j-1})$ and assume, moreover, that in S_{j-1}, X_σ is the unknown of an expansive equation

$$X_\sigma = \lambda y_1 \cdots y_m.zM_{\sigma 1} \cdots M_{\sigma k_\sigma} \in A(S_{j-1}),$$

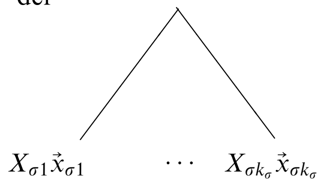
which has been expanded in S_j with associated equation

$$X_\sigma = \lambda y_1 \cdots y_m.y(X_{\sigma 1}\vec{y}_{\sigma 1}) \cdots (X_{\sigma k_\sigma}\vec{y}_{\sigma k_\sigma}).$$

https://doi.org/10.1017/S0960129500003091 Published online by Cambridge University Press

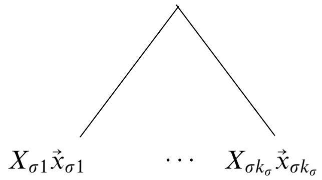
Then we set

$$(BT_{S_j}(X))_\sigma =_{\text{def}} \lambda z_1 \cdots z_k.x$$



where

$$\lambda z_1 \cdots z_k.x$$



is the Böhm tree of $(\lambda y_1 \cdots y_m.z(X_{\sigma 1}\vec{y}_{\sigma 1}) \cdots (X_{\sigma k_\sigma}\vec{y}_{\sigma k_\sigma}))\vec{v}_\sigma$ (so that every $\vec{x}_{\sigma i}$, $1 \leqslant i \leqslant k_\sigma$ is included in $\vec{v}_\sigma \cup \{y_1, \ldots y_n\}$). (Observe that $X_{\sigma n} \in V(S_j)$ and denoting by σn the path obtained by concatenating the direction n to the path σ, we have $(BT_{S_j}(X))_{\sigma n} = X_{\sigma n}\vec{x}_{\sigma n}$.)

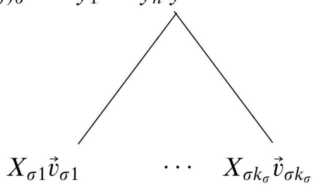
(b) **Elimination of constant.** Let again σ be a path such that $(BT_{S_{j-1}}(X))_\sigma = X_\sigma \vec{v}_\sigma$ and $X_\sigma \in V(S_{j-1})$, and assume, moreover, that in S_{j-1}, X_σ is the constant unknown of an equation $X_\sigma = T$, with T constant, which has been eliminated in S_j, by elimination of constants. Then we set $(BT_{S_j}(X))_\sigma =_{\text{def}} BT(T\vec{v}_\sigma)$, where $BT(T)$ denotes the Böhm tree of the term T.

(c) **Otherwise.** If σ corresponds to an unknown X_σ that is neither expansive nor constant, we put $(BT_{S_j}(X))_\sigma =_{\text{def}} X_\sigma \vec{v}_\sigma$.

Remark 3.2. $\vec{v}$ can be empty.

Lemma 3.1. Let $\pi : S_0, S_1, \ldots, S_i, \ldots, S_k, \ldots$ be a sequence of systems such that S_i is obtained by expansion of S_{i-1}. Let X be an unknown such that $X \in V(S_i)$ is in S_0, and let $(H_1, \ldots, H_{i-1}, H_i, H_{i+1}, \ldots, H_m)$ be a solution of S_0, where $X = H$. Let $BT_{S_0}(X), \ldots, BT_{S_k}(X), \ldots$ be the corresponding sequence of Böhm trees of X. Let σ be a path. Let $X_\sigma \vec{v}_\sigma = (BT_{S_{j-1}}(X))_\sigma$, with $X_\sigma \in V(S_{j-1})$, for some $j-1 > i$. Moreover, let S_j be obtained from S_{j-1} by expansion of X_σ, with

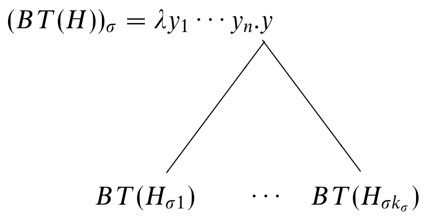
$$(BT_{S_j}(X))_\sigma = \lambda y_1 \cdots y_n.y$$



https://doi.org/10.1017/S0960129500003091 Published online by Cambridge University Press

where $X_{\sigma 1}, \ldots, X_{\sigma k_\sigma} \in V(S_j)$. Then

1



for some $H_{\sigma 1}, \ldots, H_{\sigma k_\sigma} \in \Lambda$.

2 $X_{\sigma 1} = \lambda \vec{v}_{\sigma 1}.H_{\sigma 1}, \ldots, X_{\sigma k_\sigma} = \lambda \vec{v}_{\sigma k_\sigma}.H_{\sigma k_\sigma}$ are components of a solution of S_j.

Proof. We argue by induction on the length of the path σ. If σ is empty, then $X_\sigma = X$, and the lemma follows directly from Theorem 3.1. Otherwise, X_σ has been generated by some expansion step. Let H_σ be the subterm of H, determined by path σ. By the induction hypothesis, $\lambda \vec{v}_\sigma.H_\sigma$ is the X_σ-component of a solution of the system S_{j-1}. Since X_σ is an expansive unknown, with associated equation

$$X_\sigma = \lambda y_1 \cdots y_m.z(X_{\sigma 1}\vec{y}_{\sigma 1}) \cdots (X_{\sigma k_\sigma}\vec{y}_{\sigma k_\sigma}),$$

by Theorem 3.1, $\lambda \vec{v}_\sigma.H_\sigma$ must have the form $\lambda \vec{v}_\sigma.H_\sigma \equiv \lambda y_1 \cdots y_m.zH_{\sigma 1} \cdots H_{\sigma k_\sigma}$, and for every $1 \leqslant j \leqslant k_\sigma$, we have $\lambda \vec{y}_{\sigma j}.H_{\sigma j}$ is the $X_{\sigma j}$-component of a solution of S_j. So the Böhm tree of H at path σ, has the required shape.

Now we have to determine whether the variable array $\vec{v}_{\sigma j}$ is the one required to make $H_{\sigma j}$ a closed component of a solution of S_j. To see this, observe that in the definition of Böhm tree (Definition 3.5, item 2a) the array $\vec{v}_{\sigma j}$ is obtained from the array $\vec{y}_{\sigma j}$ of the associated equation by a replacement of variables by variables, in the right order. So abstracting on $\vec{v}_{\sigma j}$ is identical to abstracting on $\vec{y}_{\sigma j}$, up to a renaming. □

Example 3.2.

$$\begin{aligned} E(S_1) = \{X_1 &= \lambda xy.x(X_2x\mathbf{I})(\lambda z_1z_2.z_1(X_2y\mathbf{I})), \\ X_2 &= X_2\mathbf{I}\} \end{aligned}$$

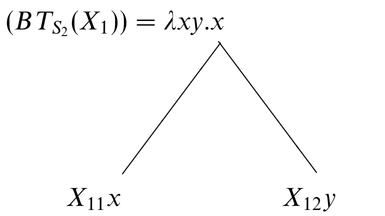
$BT_{S_1}(X_1) = X_1$
$BT_{S_1}(X_2) = X_2$

Expansion:
$X_1 = \lambda xy.x(X_2x\mathbf{I})(\lambda z_1z_2.z_1(X_2y\mathbf{I})) \Rightarrow$
$X_1 = \lambda xy.x(X_{11}x)(X_{12}y) \Rightarrow$

$$\begin{aligned} \mathrm{Exp}(X_1 = \lambda xy.x(X_2x\mathbf{I})(\lambda z_1z_2.z_1(X_2y\mathbf{I}))) = \{X_{11} &= \lambda x.X_2x\mathbf{I}, \\ X_{12} &= \lambda yz_1z_2.z_1(X_2y\mathbf{I})\} \end{aligned}$$

$$\begin{aligned} E(S_2) = \{X_{11} &= \lambda x.X_2x\mathbf{I}, \\ X_{12} &= \lambda yz_1z_2.z_1(X_2y\mathbf{I}), \\ X_2 &= X_2\mathbf{I}\} \end{aligned}$$

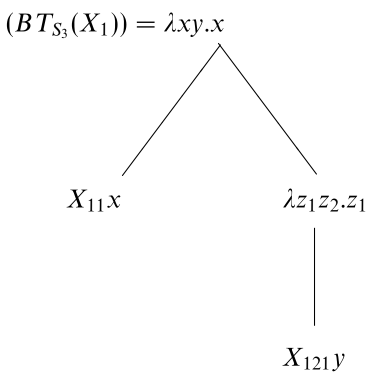
https://doi.org/10.1017/S0960129500003091 Published online by Cambridge University Press



$BT_{S_2}(X_2) = X_2$.

Expansion:
$X_{12} = \lambda y z_1 z_2.z_1(X_2 y\mathbf{I}) \Rightarrow$
$X_{12} = \lambda y z_1 z_2.z_1(X_{121} y) \Rightarrow$
$\mathrm{Exp}(X_{12} = \lambda y z_1 z_2.z_1(X_2 y\mathbf{I})) = \{X_{121} = \lambda y.X_2 y\mathbf{I}\}$

$$\begin{aligned} E(S_3) = \{X_{11} &= \lambda x.X_2 x\mathbf{I}, \\ X_{121} &= \lambda y.X_2 y\mathbf{I}, \\ X_2 &= X_2\mathbf{I}\} \end{aligned}$$



$BT_{S_3}(X_2) = X_2$

3.4. *The number of solutions*

Now we are in position to prove our main result. To this end, we have to consider *sequences* of systems that are obtained by repeatedly performing our transformation rules. A delicate point is that after each transformation step, the system that we obtain may be *not in standard form*. The following proposition shows that if we start with a system with a solution in normal form, after each transformation the resulting system can be put in standard form again.

Proposition 3.3. Let $S = (E, V)$ be a system with a solution in normal form. Then we can obtain from S by a finite number of elimination of constants a system $S' = (E', V')$ in standard form. Moreover, S' is such that:

1 S has either one or infinite solutions iff S' has either one or infinite solutions.
2 S' has a solution in normal form.

Proof. By Theorem 2.2, we may assume that S is in head normal form. If now we eliminate all the constant unknowns, then by Proposition 3.2 we obtain a system S_1 such that:

— S_1 has still one solution in normal form.
— S_1 has either one or infinite solutions iff S has either one or infinite solutions.
— S_1 has a smaller number of unknowns than S.

So, we can put S_1 in head normal form and repeat constant elimination. Since at each step the number of unknowns becomes smaller, we must eventually end with a system in standard form (possibly the empty system). □

Theorem 3.2. Let $S = (E, V)$ be a system that has a solution in normal form. Then we can obtain in a finite number of successive expansions and elimination of constants of a system $S' = (E', V')$ in final form. Moreover, S' is such that:

1 S has either one or infinite solutions iff S' has either one or infinite solutions.
2 S' has a solution in normal form.

Proof. Observe first that after any expansion of unknowns we can always obtain a system in standard form by Proposition 3.3, since by Corollary 3.1 we still have a system with a solution in normal form.

Now consider a fixed system S and all the systems S' that can be generated from S by iterating application of the transformation rules. For a contradiction, assume that no such system is in final form. This implies that we always reach a system S' such that $hnf(S') \neq \emptyset$. Moreover, since there cannot exist a infinite sequence $S'_0, S'_1, \ldots, S'_i, \ldots$ such that S'_{i+1} is obtained from S'_i by elimination of constants (without any intermediate expansion of variables), we have that in $hnf(S')$ there is always a variable that can be expanded.

So there is a variable X of the original system that has an infinite number of descendents X_σ. It follows by the Lemma 3.1 that the Böhm tree of any solution H for X has an infinite Böhm tree. This is impossible since there must be a solution in normal form for X.

To prove that S' has the additional properties (1) and (2), we observe that such properties are preserved in each transformation step by Proposition 3.2 and Corollary 3.1. □

Remark 3.3. We illustrate the previous result by considering the special case of the *empty solution*. As already noticed in Remark 3.1, the empty solution arises by elimination of constant unknowns, when all unknowns turn out to be constant. It follows that the system we started with has a unique solution. By our previous results, it follows that any transformation sequence must end in the empty solution. See also Example 3.4.

Definition 3.6. Let $H \equiv \lambda xy.\mathbf{K}(xxy)$ and $\mathbf{\Phi}_\mathbf{K} \equiv HH$.

Lemma 3.2.

$$\mathbf{\Phi_K}\lceil i \rceil \neq \mathbf{\Phi_K}\lceil j \rceil \text{ for } \lceil i \rceil \neq \lceil j \rceil$$

Proof. It easy to prove that $\mathbf{\Phi_K}\lceil i \rceil$ and $\mathbf{\Phi_K}\lceil j \rceil$ cannot have a common reduct. Indeed, starting with $\mathbf{\Phi_K}\lceil i \rceil$, for example, the only possible reductions are

$$\mathbf{\Phi_K}\lceil i \rceil \longrightarrow (\lambda y.\mathbf{K}(\mathbf{\Phi_K} y))\lceil i \rceil \longrightarrow \mathbf{K}(\mathbf{\Phi_K}\lceil i \rceil).$$

So, a proof by contradiction can be obtained immediately by induction on the number of reductions. □

Theorem 3.3. A system $S = (E, V)$ in final form, has either one or infinite solutions.

Proof. If $u\text{-}hnf(S) = \emptyset$, then S has only the empty solution. If $u\text{-}hnf(S) \neq \emptyset$, then

$$E(S) = \{X_1 = \lambda y_{1_1} \cdots y_{1_{k_1}}.X_{i_1} \cdots, \ldots, X_m = \lambda y_{m_1} \cdots y_{m_{k_m}}.X_{i_m} \cdots\}$$

where $X_{i_j} \in \{X_1, \ldots, X_m\}$. Then $X_1 = \mathbf{\Phi_K}\lceil i \rceil, \ldots, X_m = \mathbf{\Phi_K}\lceil i \rceil$ is a solution for S. This is easily seen using the equalities

$$\begin{aligned} \mathbf{\Phi_K}\lceil i \rceil x &= \mathbf{\Phi_K}\lceil i \rceil \\ \lambda x.\mathbf{\Phi_K}\lceil i \rceil &= \mathbf{\Phi_K}\lceil i \rceil. \end{aligned}$$

It follows that S has infinite solutions. □

Theorem 3.4.
The system $S = (E, V)$ with

$$E = \{X_1 = M_1 X_1 \cdots X_m, \ldots, X_m = M_m X_1 \cdots X_m\}$$

$m \geqslant 1$, which has a solution in normal form, has either one or infinite solutions.

Proof. By Theorem 3.2, there is a system S_f in final form obtained by successive expansions of S and such that S_f has either one or infinite solutions if and only if S has one or, respectively, infinite solutions. So the theorem follows by Theorem 3.3. □

Example 3.3. In Example 3.2, the system

$$\begin{aligned} E(S_1) = \{X_1 &= \lambda xy.x(X_2 x\mathbf{I})(\lambda z_1 z_2.z_1(X_2 y\mathbf{I})), \\ X_2 &= X_2\mathbf{I}\} \end{aligned}$$

has been brought to the final form

$$\begin{aligned} E(S_3) = \{X_{11} &= \lambda x.X_2 x\mathbf{I}, \\ X_{121} &= \lambda y.X_2 y\mathbf{I}, \\ X_2 &= X_2\mathbf{I}\}, \end{aligned}$$

which has an infinite number of solutions of the form

$$X_{11} = X_{121} = X_2 = \mathbf{\Phi_K}\lceil i \rceil.$$

https://doi.org/10.1017/S0960129500003091 Published online by Cambridge University Press

Since, moreover, $X_{12} = \lambda y z_1 z_2.z_1(X_{121}y)$ and $X_1 = \lambda xy.x(X_{11}x)(X_{12}y)$, we eventually get solutions of the form

$$X_1 = \lambda xy.x(\mathbf{\Phi_K}\lceil i \rceil)(\lambda z_1 z_2.z_1(\mathbf{\Phi_K}\lceil i \rceil)) \text{ and } X_2 = \mathbf{\Phi_K}\lceil i \rceil.$$

We again remark that such solutions are actually algorithmically computed, without any guess.

Example 3.4. Recall from the puzzle in the Introduction that we claimed that $F \equiv \lambda xy.y(x\mathbf{O})\mathbf{I}$ has an infinite range and a unique fixed point. Now we apply the previous results to compute such a fixed point.

Starting from the equation $X_1 = FX_1$, we get $X_1 = \lambda y.y(X_1\mathbf{O})\mathbf{I}$.

So we have an expansive equation and by expansion we obtain the system

$$\begin{aligned} X_1 &= \lambda y.yX_{11}X_{12} \\ X_{11} &= (\lambda y.yX_{11}X_{12})\mathbf{O} \\ X_{12} &= \mathbf{I}. \end{aligned}$$

By elimination of constants, we get $X_{11} = (\lambda y.yX_{11}\mathbf{I})\mathbf{O}$, that is $X_{11} = \mathbf{I}$, and we end with the empty system.

Going back through the transformation process, to the original system, we find its unique solution

$$X_1 = \lambda y.y\mathbf{II}.$$

Corollary 3.2. Let M be a term such that $N = MN$, for some N in nf. Then M has either one or infinite fixed points.

4. Conclusions and further research

There seems to be some possible improvements to the previous result. We are still working out the details, but the situation appears to be promising.

The first improvement is to obtain that if a combinator F has a fixed point normal form, then either F has one fixed point or F has infinite fixed points with *head normal form.*

The second is to obtain the previous result with the weakened hypothesis that F has *a fixed point in hnf.*

At present, we do not know how to deal with the problem if we start from an *unsolvable* fixed point.

Appendix A. Terms with fixed points that are not universal generators

In this Appendix, we sketchily show how to make precise one possible approach to the NFP Question, already mentioned in the Introduction.

Given a combinator F, the idea is to consider a fixed point combinator Hx for F such that for some N, M the equality $HN = HM$ cannot hold unless F erases its argument.

As already pointed out, it is not clear whether such an H exists for a general term

F. Here, we shall prove that if a suitable H is *not a universal generator*, then H has the required property.

Definition A.1. Let F be given, we define the following combinators:

$$R \equiv \lambda xy.F(xxy)$$
$$H \equiv RR$$

We call H the *parametric fixed point combinator* for F.

We recall from Barendregt (1984) that a term T is a *universal generator* if for every N there is a reduct T' of T such that N is a subterm of T'. We need the following easy facts (where x is always a free fresh variable).

Lemma A.1. If T is not a universal generator, then $\lambda x.T(\mathbf{\Omega}x)$ and $\lambda x.T(\mathbf{\Omega}xx)$ are not universal generators, where x is a fresh free variable not occurring in T.

Proof. Let N be such that it is not a subterm of any reduct of T. We may freely assume that N is closed and has the shape $\mathbf{\Omega}M$ for some closed term M. So, since N does not begin with a λ, to be a subterm of some reduct of $\lambda x.T(\mathbf{\Omega}x)$, it must be a subterm of some reduct of $T(\mathbf{\Omega}x)$. Since N is closed it must be a subterm of some reduct of T, which is impossible. □

Lemma A.2. For every T, $\lambda x.T(\mathbf{\Omega}x) \neq \lambda x.T(\mathbf{\Omega}xx)$, unless Tx reduces to a term containing no occurrences of x, where x is a fresh free variable not occurring in T.

Proof. If T does not erase the subterms $\mathbf{\Omega}x$, $\mathbf{\Omega}xx$, there cannot be a common reduct since T cannot duplicate the variable x occurring only in $\mathbf{\Omega}x$. □

Theorem A.1. Let F be given, and let H be the parametric fixed point combinator for F. If H is not a universal generator, then F has either one or infinite fixed points.

Proof. By the previous lemma, $\lambda x.H(\mathbf{\Omega}x) \neq \lambda x.H(\mathbf{\Omega}xx)$ unless Hx reduces to a term P containing no occurrences of x. In this case, we claim that there exists an n such that $F^n x$ reduces to some Q with $x \notin Q$. To prove this, consider the reduction from Hx to P. To eliminate the free variable x, it is necessary to eliminate each occurrence of the term Hx, since the reduction of this term reproduces the term itself, via the reductions

$$Hx \longrightarrow (\lambda y.F(Hy))x \longrightarrow F(Hx). \tag{5}$$

Therefore, there is no loss of generality in treating the two reductions in (5) as a single reduction step

$$Hx \longrightarrow F(Hx), \tag{6}$$

since the intermediate redex is always either reduced or eliminated. We illustrate this by considering the case in which reducing $F(Hy)$ the variable y disappears. By the previous argument, this implies that $F(Hy)$ reduces to a term Q, such that in Q there is no occurrence of Hy. So we have that Hx reduces to $(\lambda y.Q)x$ and then to Q. By performing the reduction (6) directly, we still obtain Q, since $F(Hx)$ must reduce to Q.

Now we claim that we can simulate the reduction $D : Hx \longrightarrow^* P$ starting with $F^m x$, for

https://doi.org/10.1017/S0960129500003091 Published online by Cambridge University Press

a suitably large m. To see this, assume that the occurrences of F in $F^m x$ are *frozen*, that is, they cannot be reduced. Now we construct a derivation D' as follows. We start with $F^m x$ instead of Hx and then follow the reduction D. When Hx is duplicated or erased *etc.* we make all the same with $F^m x$, taking into account the correspondence between the copies of Hx and that of $F^m x$. When the reduction (6) is performed in D, we obtain $F(Hx)$; then in D' we *unfreeze* the F occurring at the top of the frozen Fs obtaining $F(F^n x)$, for some n. The unfrozen $F(\ldots)$ is then reduced in D' in the same way the external $F(\ldots)$ is reduced in the original reduction D. When Hx (or $(\lambda y.F(Hy))x$) is erased, duplicated *etc.*, the corresponding term $F^k x$, which contains all the still frozen F, is respectively erased, duplicated, *etc.*. It is clear that to perform the previous simulation, it suffices to start with m large enough. It follows that $F^m x \longrightarrow^* P$. Now let M be a fixed point of F. It follows that $M = F^m M = P$. So, F has a unique fixed point.

Having proved this, we consider the other case, that is, $\lambda x.H(\mathbf{\Omega} x) \neq \lambda x.H(\mathbf{\Omega} xx)$.

Now, by Lemma A.1, we have $\lambda x.H(\mathbf{\Omega} x)$ and $\lambda x.H(\mathbf{\Omega} xx)$ are not universal generators, since H is not a universal generator. By Barendregt (1984, Proposition 17.3.19), there exists a closed term Ξ such that $(\lambda x.H(\mathbf{\Omega} x))\Xi \neq (\lambda x.H(\mathbf{\Omega} xx))\Xi$. (In Barendregt (1984), the proof is given for terms that are not $\beta\eta$-universal generators in the $\lambda\eta$-calculus, but it works also in the $\lambda\beta$-calculus for terms that are not universal generators.)

It follows that the range of H is not a singleton. By the Range Theorem, F has infinite fixed points. □

Appendix B. The recursion theoretic approach

The proof of the range theorem

Let a combinator F be given, together with a finite number $P_0, \ldots, P_k$ of fixed points of F (with $k > 0$). We want to construct a fixed point of F different from all P_i, $1 \leqslant i \leqslant k$. In using the recursion theoretic approach, we have to construct a suitable recursive function χ such that

$$\begin{aligned} \chi(\ulcorner n \urcorner) &= \ulcorner P_{i+1} \urcorner \text{ if } \mathbf{E}\ulcorner n \urcorner = P_i \ (i = 0, \ldots, k \text{ and } i+1 \text{ is taken modulo } k+1) \\ \chi(\ulcorner n \urcorner) &= \mathbf{?}, \text{ otherwise,} \end{aligned}$$

where we follow the notation and terminology of Barendregt (1993) and, in particular, $\ulcorner n \urcorner$ is the n-th Church numeral, $\#P$ is the code of the term P and $\ulcorner P \urcorner = \ulcorner \#P \urcorner$ is the corresponding numeral. $\mathbf{E}$ is the Barendregt universal generator (see Barendregt (1984, Definition 8.2.7)).

The problem is how to fill the item marked by **?**. Indeed, observe that by the equations

$$\chi(\ulcorner n \urcorner) = \ulcorner P_{i+1} \urcorner \text{ if } \mathbf{E}\ulcorner n \urcorner = P_i$$

we are not directly specifying a (partial) recursive function, since convertibility is not a recursive relation. So if n is such that $\mathbf{E}\ulcorner n \urcorner \neq P_i$ for $i = 0, \ldots, k$, then χ cannot terminate on input $\ulcorner n \urcorner$ since it has to go through all possible derivations to test $\mathbf{E}\ulcorner n \urcorner = P_i$, for some i. So, in Barendregt's proof, **? = undefined**.

To be more formal, we can define χ via an auxiliary function ψ such that

$$\psi(\lceil n \rceil, m) = \lceil P_{i+1} \rceil$$

if m is the number of a proof of $\mathbf{E}\lceil n \rceil = P_i$ ($i = 0, \ldots, k$ and $i+1$ is taken modulo $k+1$);

$$\psi(\lceil n \rceil, m) = \psi(\lceil n \rceil, m+1), \text{ otherwise,}$$

and then putting:

$$\chi(\lceil n \rceil) = \psi(\lceil n \rceil, 0)$$

In the proof of the range theorem, we then take the fixed point

$$Q = F(\mathbf{E}(G\lceil Q \rceil)) \tag{7}$$

where G represents χ, and conclude by contradiction that Q is different from all the P_i.

If we try to imitate this proof, we are not guaranteed by (7) that Q is a fixed point of F since $G\lceil Q \rceil$ may correspond to the **?** branch of χ.

So the **?** item is *overloaded*, since by the previous discussion it must be undefined to make the condition $\mathbf{E}\lceil n \rceil = P_i$ effective, and it must also be such that (7) is a fixed point of F for some Q. It seems hard to simultaneously fulfil all such requirements.

The 'Number of Fixed Points' question in the recursion theory

In this part of the Appendix, we very briefly discuss the NFP Question in the Recursion Theory, with respect to the celebrated Recursion Theorem, due to Kleene.

This seems to be appropriate since we have already mentioned that the recursion theoretic approach is a plausible attack to the NFP Question in Lambda Calculus. Moreover, there are strong similarities between the Recursion Theorem and the Fixed Point Theorem in Lambda Calculus, not only in the statements of the two theorems but also in their proofs. So, we hope that the following remarks will add perspective to our problem. However, we limit ourselves to state a number of facts and refer to Rogers (1967) and Odifreddi (1989) for details and proofs. In the following φ_n is the n-th function in a fixed effective enumeration of partial recursive functions.

Theorem B.1 (The Recursion Theorem). Let f be any recursive function; then there is an n such that $\varphi_{f(n)} = \varphi_n$. ($n$ is called a *fixed point* of f.) (See Rogers (1967, 11.2 Theorem I)).

Remark B.1. The equality $=$ means *equality of partial functions*, so that functions $\varphi_{f(n)}$ and φ_n are actually the same (partial) function.

Remark B.2. Quoting from Rogers (1967, 11.5 p. 192), 'Strictly speaking, the recursion theorem ... is not a fixed point theorem'. The basic reason is that if n is a fixed point of f and m is an index such that $\varphi_m = \varphi_n$ then the equality $\varphi_{f(m)} = \varphi_m$ may very well not hold (to find counterexamples is a simple exercise that we leave to the reader).

Remark B.3 (Number of fixed points). The previous remark is relevant also in the determination of the number of fixed points. That is, since the Recursion Theorem uses

https://doi.org/10.1017/S0960129500003091 Published online by Cambridge University Press

equality of functions, at least two different notions of equality are possible on indexes i, j:

— identity;

— equality of the corresponding functions: $i \cong j$ iff $\varphi_i = \varphi_j$.

If the first equality is chosen, then the number of fixed points of a recursive function f is always infinite (Odifreddi 1989, Exercise II.2.11). The proper meaning is that there are an infinite number of different *codes* such that their indexes are transformed by f into indexes of *equivalent codes* (that is, that compute the same function). However, if the second equality is considered, then there are recursive functions f such that they have a finite number of fixed points only (simply choose indexes $i_0, \ldots, i_k$ corresponding to pairwise different functions φ_{i_j}, $1 \leqslant j \leqslant k$, then set $f(i_j) = i_j$ and $f(m) = i_0$ for $m \neq i_j$).

Remark B.4 (Recursion theory vs lambda calculus). What may we learn from the previous remark with respect to our original NFP Question in Lambda Calculus?

The fact that there are infinitely many indexes n such that $\varphi_{f(n)} = \varphi_n$ is not so useful since we are doing everything up to convertibility.

Indeed, the latter fact may suggest the contrary: that a uniform recursion theoretic argument cannot exist.

Acknowledgements

We would like to thank the anonymous referees for their help in substantially improving previous versions of the paper.

References

Barendregt, H. P. (1984) *The Lambda Calculus, its Syntax and Semantics*, revised edition, Studies in Logic and the Foundations of Mathematics, North-Holland.

Barendregt, H. P. (1993) Constructive proofs of the range property in lambda calculus. *Theoretical Computer Science* **121** 59–69.

Böhm, C. and Tronci, E. (1991) About systems of equations, x-separability, and left-invertibility in the λ-calculus. *Information and Computation* **90** 1–32.

Odifreddi, P. G. (1989) *Classical Recursion Theory*, Studies in Logic and the Foundation of Mathematics, North-Holland.

Rogers, H. (1967) *Theory of Recursive Functions and Effective Computability*, McGraw-Hill.

Tronci, E. (1996) Equational programming in λ-calculus via SL-systems. Part 1. *Theoretical Computer Science* **160** 145–184.

Tronci, E. (1996) Equational programming in λ-calculus via SL-systems. Part 2. *Theoretical Computer Science* **160** 185–216.