The Virtues of Eta-expansion

C. Barry Jay [1]

Computing Methods Unit
School of Computing Sciences
University of Technology Sydney
PO Box 123
Broadway
2007
Australia.
e-mail: cbj@socs.uts.edu.au

Neil Ghani

LFCS, Department of Computer Science
University of Edinburgh
The King's Buildings, Mayfield Road
Edinburgh, UK, EH9 3JZ
e-mail: ng@dcs.ed.ac.uk

January 8, 1994

[1] Research supported by The Royal Society of Edinburgh/BP.

Abstract

Interpreting η-conversion as an expansion rule in the simply-typed λ-calculus maintains the confluence of reduction in a richer type structure. This use of expansions is supported by categorical models of reduction, where β-contraction, as the local counit, and η-expansion, as the local unit, are linked by local triangle laws. The latter form reduction loops, but strong normalisation (to the long $\beta\eta$-normal forms) can be recovered by "cutting" the loops.

1 Introduction

Extensional equality for terms of the simply-typed λ-calculus requires η-conversion, whose interpretation as a reduction rule is usually a contraction

$$\lambda x.fx \Rightarrow f$$

If the type structure contains only arrow and product types, whose η-reduction is

$$\langle \pi c, \pi' c\rangle \Rightarrow c$$

then the resulting system, including the usual β-reductions, has the properties of being:

- a rewriting system (i.e. allowing subject reduction);
- confluent;
- strongly normalising.

so that reduction provides an effective procedure for deciding the equality of terms.

These properties typically fail, however, if the calculus is extended by the addition of more datatypes. Even the introduction of a unit type (necessary for defining types with given constants, such as booleans and lists) is problematic. Specifically, if $* : 1$ is the given constant of unit type, with η-rule

$$a \Rightarrow * \qquad \text{if } a : 1$$

then confluence is lost. This is because for any variable $f : A \to 1$ the term $\lambda x.fx$ has two normal forms

$$\lambda x.* \ \Leftarrow \ \lambda x.fx \ \Rightarrow \ f$$

(as noticed by Obtulowicz, and reported in [20]).

Lambek and Scott [20] handle this problem by suppressing all terms of unit type other than $*$, while Curien and Di Cosmo [5] recover confluence by applying the Knuth-Bendix algorithm to obtain additional reduction rules, while only preserving weak normalisation.

The solution considered here is simply to interpret η-conversions as expansions

$$\begin{array}{lcll} t & \Rightarrow & \lambda x.tx & \text{if } t : A \to B \\ c & \Rightarrow & \langle \pi c, \pi' c\rangle & \text{if } c : A \times B \\ a & \Rightarrow & * & \text{if } a : 1 \end{array}$$

Note that in each case the amount of type information which can be inferred from a term is increased.

The result is a confluent rewriting system which is not normalising, as can be seen from the following infinite reduction sequence

$$t \Rightarrow \lambda x.tx \Rightarrow \lambda y.(\lambda x.tx)y \Rightarrow \ \ldots \qquad (1)$$

The correct interpretation of these infinite reductions, and of expansions generally, can be found within category theory, and its models of reduction. Types are interpreted by objects, terms by morphisms, and reductions by 2-cells. For example, if $(\lambda z.t)a$ is a term of type Y containing one free variable of type X, then its β-reduction is represented

$$
\begin{tikzcd}
X \arrow[r, bend left, "(\lambda z.t)a"] \arrow[r, bend right, "t[a/z]"'] \arrow[r, phantom, "\Downarrow"] & Y
\end{tikzcd}
$$

Labelling the 2-cells yields a 2-category [24, 23] while leaving them unlabelled yields an ordered category [16].

Just as in the denotational semantics of λ-terms, models of reduction should be cartesian closed categories, where cartesian closure, or better, the underlying theory of adjunctions, is interpreted in a sense appropriate to the presence of 2-cells. Exactly how this should be done is an area of active research [8, 19, 13, 14, 3, 15, 11] but most developments share the following properties. There is a local counit, corresponding to β-contraction, and a local unit, corresponding to η-expansion, which are linked by local triangle laws. When function types are interpreted by exponentials these are

$$
\begin{array}{rcccl}
\lambda x.t & \Rightarrow & \lambda y.(\lambda x.t)y & \Rightarrow & \lambda y.t[y/x] \equiv \lambda x.t \\
ta & \Rightarrow & (\lambda x.tx)a & \Rightarrow & ta
\end{array}
\tag{2}
$$

Such reductions, from a term to itself, are called *loops*. The triangle laws for the product type are

$$
\begin{array}{rcccl}
\langle a,b\rangle & \Rightarrow & \langle \pi\langle a,b\rangle, \pi'\langle a,b\rangle\rangle & \Rightarrow & \langle a,b\rangle \\
\pi c & \Rightarrow & \pi\langle \pi c, \pi' c\rangle & \Rightarrow & \pi c \\
\pi' c & \Rightarrow & \pi'\langle \pi c, \pi' c\rangle & \Rightarrow & \pi' c
\end{array}
\tag{3}
$$

where the latter pair of loops are thought of as a single loop in the product category. Those for the unit type are

$$
* \Rightarrow * \Rightarrow * \tag{4}
$$

and the trivial loop in the terminal category.

These mathematically inspired loops yield infinite reduction sequences; normalisation thus requires these loops be "cut". As β-reduction is inviolate, we must ban those expansions which give rise to them. Thus, terms of function type may be expanded provided they are neither λ-abstractions nor applied: terms of product type may be expanded if they are neither pairs nor projected. In fact, these restrictions are enough to recover strong normalisation, i.e. it is the triangle laws which prevent normalisation. For example, in the reduction sequence (1) all steps after the first are prohibited. Reduction is then confluent and strongly normalising, but does not form a rewriting system, since the second triangle law depends on the context.

These restrictions on expansions were first exploited by Mints [21] for technical purposes, but omitted from his later writings; his approach is being revived by Čubrić [4]. Di Cosmo and Kesner [6] and Akama [1] have also been studying this system recently. The use of unrestricted η-expansion, and its primacy over the restricted system appear to be new.

The normal forms of the restricted system have a longer history, as Prawitz' *expanded normal forms* [22], and as Huet's *long $\beta\eta$-normal forms* [12]. Unlike the present system, reduction proceeded there in two stages; first do β-reduction, then η-expansion, subject to restrictions that preserve β-normality. These forms appear, for example, in the development of the LF logical framework [9] and in the study of type classes [10]. The long $\beta\eta$-normal forms still satisfy a universal property in the unrestricted system as they are *essentially normal* i.e. any reduction from such a term is reversible. Thus every term is reducible to an essential normal form.

To summarise, the various interpretations of η-conversion give reduction systems with any two of the three properties listed above, namely, subject reduction, confluence and strong normalisation. The unrestricted system is required for a categorical semantics of reduction, in which type constructions are interpreted by adjoints, while its restricted form provides an effective procedure for deciding the equality of λ-terms.

2 Eta-expansion

The simply-typed λ-calculus over a set of base types has *types* freely generated by

$$T := T\times T \mid T\to T \mid N \mid 1 \mid \mathcal{C}$$

where C is any base type. The *atomic* types are $1, N$ and the base types. For each type T, there are disjoint sets of variables $Var(T)$, and constants $Con(T)$ such that $* \in Con(1)$ and $0 \in Con(N)$. The *well-formed terms* (with their associated types) are generated by the following inference system

$$\frac{}{x : A}\; x \in Var(A) \qquad \frac{}{* : 1} \qquad \frac{}{c : A}\; c \in Con(A)$$

$$\frac{b : B}{\lambda x.b : A\to B}\; x \in Var(A) \qquad \frac{f : A\to B \quad a : A}{\mathtt{app}(f,a) : B}$$

$$\frac{a : A \quad b : B}{\langle a,b\rangle : A\times B} \qquad \frac{c : A\times B}{\pi c : A} \qquad \frac{c : A\times B}{\pi' c : B}$$

$$\frac{}{0 : N} \qquad \frac{n : N}{Sn : N} \qquad \frac{a : A \quad h : A\to A \quad n : N}{\mathtt{It}(a,h,n)}$$

The term $\texttt{app}(f,a)$ may be written fa. A typical term may be denoted $\mathcal{T}(u_i)$ where $\mathcal{T}$ is the outer term constructor and the u_i are its arguments, e.g. a variable x may be expressed as $x()$. Basic familiarity with λ-calculus is assumed (e.g. [2, 20, 7]), e.g. terms equivalent under α-conversion will be identified.

The *expansionary rewrite relation* has the following set of basic reductions:

$$
\begin{array}{lrcl}
(\beta_{\rightarrow}) & (\lambda x.b)a & \Rightarrow & b[a/x] \\
(\eta_{\rightarrow}) & t & \Rightarrow & \lambda x.tx \\
(\beta_{\times,1}) & \pi\langle a,b\rangle & \Rightarrow & a \\
(\beta_{\times,2}) & \pi'\langle a,b\rangle & \Rightarrow & b \\
(\eta_{\times}) & c & \Rightarrow & \langle \pi c, \pi' c\rangle \\
(\eta_{1}) & a & \Rightarrow & * \\
(\beta_{N,1}) & \texttt{It}(a,h,0) & \Rightarrow & a \\
(\beta_{N,2}) & \texttt{It}(a,h,Sn) & \Rightarrow & h\texttt{It}(a,h,n)
\end{array}
$$

where: (i) the η-rules have implicit type restrictions, e.g. $\eta_{\rightarrow}$ is restricted to terms of function type; (ii) $\eta_{\rightarrow}$ requires that x not be free in t, and; (iii) $\beta_{\rightarrow}$ involves implicit α-conversion whenever substitution threatens to capture free variables of a. Closing basic reductions under the term constructors of the language yields the 1-step rewrite relation, whose reflexive, transitive closure is denoted $\Rightarrow$.

Rewrites built solely from the various η-rules are called *expansions*. Conversely, those built without expansions are *contractions*. The rewrite relation on terms obtained by restricting the basic reductions to expansions (respectively, contractions) is denoted η (respectively, β).

3 An Abstract Confluence Theorem

Confluence of the expansionary system and its restricted fragment will be proved using the following variant of a theorem by Kahrs [18].

Let R be a relation on some set. Denote its reflexive closure by $R^{=}$ and its reflexive transitive closure by R^*. Another relation S (on the same set) *satisfies the diamond lemma relative to* R if every divergence $t_1 \overset{S}{\leftarrow} t \overset{S}{\rightarrow} t_2$ can be completed to

$$
\begin{tikzcd}
t \arrow[r, "S"] \arrow[d, "S"'] & t_2 \arrow[r, "S"] & {} \arrow[d, "R^*"] \\
t_1 \arrow[r, "S"'] & {} \arrow[r, "R^*"'] & {}
\end{tikzcd}
$$

Of course, this is weaker than requiring S satisfy the diamond lemma relative to the identity relation, i.e. the usual diamond lemma.

S is *R-extendable* if every divergence $t_1 \overset{S}{\leftarrow} t \overset{R}{\rightarrow} t_2$ can be completed to

$$\begin{tikzcd}
t \arrow[r, "R"] \arrow[d, "S"'] & t_2 \arrow[r, "R^*"] & t_3 \arrow[d, "S"] \\
t_1 \arrow[rr, "R^*"'] & & t_4
\end{tikzcd}$$

Lemma 3.1 and Theorem 3.2 below remain valid if the definition of extendability is weakened to permit further R-rewrites after the S-rewrite from t_3 to t_4. However the examples do not require such generality and so the more limited definition is used. Assume that R is confluent and strongly normalising for the rest of this section.

Lemma 3.1 *If S is R-extendable then it is R^*-extendable.*

Proof It suffices to prove that any divergence $t_1 \overset{S}{\leftarrow} t \overset{R^*}{\rightarrow} t_2$ can be completed to

$$\begin{tikzcd}
t \arrow[r, "R^*"] \arrow[d, "S"'] & t_2 \arrow[r, "R^*"] & t_3 \arrow[d, "S"] \\
t_1 \arrow[rr, "R^*"'] & & t_4
\end{tikzcd}$$

which is done by induction on the R-rank of t.

If the R^*-reduction of t is vacuous then the result is trivial. Otherwise we have the following completion

$$\begin{tikzcd}
t \arrow[r, "R"] \arrow[ddd, "S"'] & \cdot \arrow[r, "R^*"] \arrow[d, "R^*"'] & \cdot \arrow[d, "R^*"] \\
& \cdot \arrow[r, "R^*"'] \arrow[dd, "S"'] & \cdot \arrow[d, "R^*"] \\
& & \cdot \arrow[d, "S"] \\
t_1 \arrow[r, "R^*"'] & \cdot \arrow[r, "R^*"'] & \cdot
\end{tikzcd}$$

where the left cell exists by the R-extendability of S, the top right cell by R-confluence and the bottom right cell by the induction hypothesis. □

Theorem 3.2 *If S satisfies the diamond lemma relative to R and is R-extendable then $R \cup S$ is confluent.*

Proof It suffices to show that $R^*; S^=; R^*$ satisfies the diamond lemma, which is done by induction on the R-rank and repeated use of the premises and previous lemma. Details are given for those cases in which S always appears; the others are simpler.

If both of the first R^*-reductions are of length 0 we can construct

$$\begin{tikzcd}
\cdot \arrow[d, "R^*"'] & \cdot \arrow[l, "R^*"'] \arrow[dd, "S"'] & \cdot \arrow[l, "S"'] \arrow[r, "S"] & \cdot \arrow[r, "R^*"] \arrow[dd, "S"] & \cdot \arrow[d, "R^*"] \\
\cdot \arrow[d, "S"'] & & & & \cdot \arrow[d, "S"] \\
\cdot & \cdot \arrow[l, "R^*"'] \arrow[r, "R^*"] & \cdot & \cdot \arrow[l, "R^*"'] \arrow[r, "R^*"] & \cdot
\end{tikzcd}$$

and the confluence of R yields the desired completion.

Otherwise the divergence is completed to

$$\begin{tikzcd}
\cdot \arrow[r, "R^*"] \arrow[d, "R^*"'] & \cdot \arrow[rr, "S"] \arrow[d, "R^*"] & & \cdot \arrow[r, "R^*"] \arrow[d, "R^*"'] & \cdot \arrow[d, "R^*"] \\
\cdot \arrow[r, "R^*"'] \arrow[dd, "S"'] & \cdot \arrow[r, "R^*"'] \arrow[d, "R^*"] & \cdot \arrow[r, "S"'] & \cdot \arrow[r, "R^*"'] & \cdot \arrow[d, "R^*"] \\
& \cdot \arrow[d, "S"] & & & \cdot \arrow[d, "S"] \\
\cdot \arrow[r, "R^*"] \arrow[d, "R^*"'] & \cdot \arrow[d, "R^*"] & & & \cdot \arrow[d, "R^*"] \\
\cdot \arrow[r, "R^*"'] & \cdot \arrow[r, "R^*"'] & \cdot \arrow[r, "S"'] & \cdot \arrow[r, "R^*"'] & \cdot
\end{tikzcd}$$

□

This theorem will be applied with R given by β-reduction, which is confluent and strongly normalising [7]. Neither η-expansion nor its restricted form satisfy the premises for S, however, so broader notions of expansion must be employed.

4 Confluence of the Expansion System

Define $\eta(t)$ to be the basic expansion of a term t of product, exponent or unit type. A *neutral* term is one which is not a λ-abstraction, pair or $*$.

Lemma 4.1 *The following statements hold.*

(i) $\eta(t)[t'/x] = \eta(t[t'/x])$

(ii) If t *is not neutral then* $\eta(t) \rightarrow_{\beta*} t$.

(iii) $\eta(\eta(t)) \rightarrow_{\beta*} \eta(t)$

(iv) If $t \rightarrow_{\beta*} t'$ *then* $\eta(t) \rightarrow_{\beta*} \eta(t')$.

Proof Trivial. Note that the number of β-reductions may increase in (iv) as the $\eta_\times$-rule duplicates its argument. □

Parallel expansion (denoted $\twoheadrightarrow$) is the smallest relation on terms closed under

(i) Congruence: For each term constructor $\mathcal{T}$

$$\frac{u_i \twoheadrightarrow u_i'}{\mathcal{T}(u_i) \twoheadrightarrow \mathcal{T}(u_i')}$$

(ii) Expansion: For any term u of product, exponent or unit type

$$\frac{u \twoheadrightarrow u'}{u \twoheadrightarrow \eta(u')}$$

(iii) Substitution:

$$\frac{u \twoheadrightarrow u' \quad v \twoheadrightarrow v'}{u[v/z] \twoheadrightarrow u'[v'/z]}$$

A formal proof d of $u \twoheadrightarrow u'$ is called a *derivation* and is denoted $u \twoheadrightarrow u'$. The *height* of d is the length of its longest branch.

Lemma 4.2 *Parallel expansion is a reflexive relation. Hence if* $t \twoheadrightarrow t'$ *then* $\eta(t) \twoheadrightarrow \eta(t')$.

Proof The proof is by induction on term structure. If t is a variable $x()$, then the congruence rule applies, while the inductive step is straight forward. Congruence is also sufficient to establish the second result. □

Lemma 4.3 *If there is a derivation* $t \twoheadrightarrow t'$ *whose only use of substitution is its last step then there is one which does not use substitution at all.*

Proof Let the final substitution be

$$\frac{u \twoheadrightarrow u' \quad v \twoheadrightarrow v'}{u[v/z] \twoheadrightarrow u'[v'/z]}$$

The proof is by induction on the height of the derivation $u \twoheadrightarrow u'$. Its last rule is either a congruence or expansion. First, consider a congruence

$$\frac{\dfrac{u_i \twoheadrightarrow u_i'}{\mathcal{T}(u_i) \twoheadrightarrow \mathcal{T}(u_i')} \quad v \twoheadrightarrow v'}{\mathcal{T}(u_i)[v/z] \twoheadrightarrow \mathcal{T}(u_i')[v'/z]}$$

If z is not free in $\mathcal{T}(u_i)$ or $\mathcal{T}(u_i) = z$ then the result is trivial. Otherwise the derivation can be replaced by

$$\frac{\dfrac{u_i \twoheadrightarrow u_i' \quad v \twoheadrightarrow v}{u_i[v/z] \twoheadrightarrow u_i'[v'/z]}}{\mathcal{T}(u_i[v/z]) \twoheadrightarrow \mathcal{T}(u_i'[v'/z])}$$

(taking care that $\mathcal{T}$ does not bind any free variables of v and v').

Second, Lemma 4.1 allows us to replace an expansion

$$\frac{\dfrac{u \twoheadrightarrow u''}{u \twoheadrightarrow \eta(u'')} \quad v \twoheadrightarrow v'}{u[v/z] \twoheadrightarrow \eta(u'')[v'/z]}$$

by

$$\frac{\dfrac{u \twoheadrightarrow u'' \quad v \twoheadrightarrow v'}{u[v/z] \twoheadrightarrow u''[v'/z]}}{u[v/z] \twoheadrightarrow \eta(u''[v'/z])}$$

In each case the height of the left-hand derivation of the new substitutions is reduced and so can be eliminated. □

Theorem 4.4 (Substitution Elimination) *If there is a derivation $t \twoheadrightarrow t'$, then there is one which doesn't involve substitution.*

Proof Use induction on the number of substitutions in the derivation and apply Lemma 4.3. □

Corollary 4.5 *Let $\mathcal{T}$ be a term constructor of arity n. If*

$$u = \mathcal{T}(u_1, \ldots, u_n) \twoheadrightarrow u'$$

then there are derivations $u_i \twoheadrightarrow u_i'$ and a number k such that

$$u' = \eta^k(\mathcal{T}(u_1', \ldots, u_n'))$$

In particular, if u is a variable or constant and $u \twoheadrightarrow u'$, then $u' = \eta^k(u)$.

Proof Induction on the height of the derivation. □

Corollary 4.6 *There are inclusions* $\rightarrow_\eta \;\subset\; \twoheadrightarrow \;\subset\; \rightarrow_{\eta^*}$ *which are strict.*
Proof The strictness of the first inclusion is trivial. The second inclusion follows from Corollary 4.5, which also implies strictness, since, for any variable f of type $N \times N \rightarrow N$, there is an expansion

$$f \rightarrow_{\eta^*} \lambda x. f\langle \pi x, \pi' x\rangle$$

whose reduct is not of the form $\eta^k(f)$. $\square$

Proposition 4.7 *Parallel expansion satisfies the diamond lemma and so is confluent. Thus* $\rightarrow_\eta$ *is also confluent.*
Proof The proof is by induction on term structure. Consider a pair of rewrites of $\mathcal{T}(u_i)$ to $\eta^j(\mathcal{T}(u_i'))$ and $\eta^k(\mathcal{T}(u_i''))$ where $u_i \twoheadrightarrow u_i'$ and $u_i \twoheadrightarrow u_i''$. By induction, each u_i' and u_i'' have a common expansion v_i and so $\eta^{j+k}(\mathcal{T}(v_i))$ is the desired completion. Now Corollary 4.6 shows $\rightarrow_\eta$ is confluent. $\square$

Lemma 4.8 *Parallel expansion is β-extendable.*
Proof Consider a divergence $t_1 \twoheadleftarrow t \rightarrow_\beta t_2$. The proof is by induction on the height of the derivation of its parallel expansion. There are two cases to consider.

First, if the parallel expansion is an application of the expansion rule to $t \twoheadrightarrow t_1'$, then by induction there is a completion

$$\begin{tikzcd} t \arrow[r, "\beta"] \arrow[d, two heads] & t_2 \arrow[r, "\beta^*"] & t_3 \arrow[d, two heads] \\ t_1' \arrow[rr, "\beta^*"'] & & t_4' \end{tikzcd}$$

Setting $t_4 = \eta(t_4')$ implies that $t_1 \rightarrow_{\beta^*} t_4$ by Lemma 4.1 and $t_3 \twoheadrightarrow t_4$ by expansion.

Second, if the last rule of the parallel expansion is a congruence then perform a case analysis on the β-reduction. If it is $(\lambda x.\phi)(\psi) \rightarrow_\beta \phi[\psi/x]$ then the parallel expansion must be of the form $(\lambda x.\phi)(\psi) \twoheadrightarrow \sigma_1\psi_1$ where $\lambda x.\phi \twoheadrightarrow \sigma_1$ and $\psi \twoheadrightarrow \psi_1$. Hence, by Corollary 4.5, there is a term ϕ_1 such that $\phi \twoheadrightarrow \phi_1$ and

$$\sigma_1 = \eta^k(\lambda x.\phi_1)$$

and so $\sigma_1 \rightarrow_{\beta^*} \lambda x.\phi_1$ by Lemma 4.1. Thus the divergence can be completed to

$$\begin{tikzcd} (\lambda x.\phi)(\psi) \arrow[r, "\beta"] \arrow[d, two heads] & \phi[\psi/x] \arrow[r, phantom, "="] & \phi[\psi/x] \arrow[d, two heads] \\ \sigma_1\psi_1 \arrow[r, "\beta^*"'] & (\lambda x.\phi_1)(\psi_1) \arrow[r, "\beta"'] & \phi_1[\psi_1/x] \end{tikzcd}$$

Other basic β-reductions are handled similarly.

If the β-reduction of $t = \mathcal{T}(u_i)$ is of a proper subterm then the divergence is of the form

$$\eta^k(\mathcal{T}(u_i')) \twoheadleftarrow \mathcal{T}(u_i) \rightarrow_\beta \mathcal{T}(u_i'')$$

where $u_i \twoheadrightarrow u_i'$ and $u_i \rightarrow_{\beta=} u_i''$ by assumption. The induction hypothesis is applied to each of these divergences to obtain terms v_i and w_i such that $u_i'' \rightarrow_{\beta*} v_i \twoheadrightarrow w_i$ and $u_i' \rightarrow_{\beta*} w_i$ and so $\eta^k(\mathcal{T}(w_i))$ provides the desired completion. □

Theorem 4.9 *The expansionary system is confluent.*
Proof It has the same reflexive transitive closure as $\rightarrow_\beta \cup \twoheadrightarrow$ which is confluent by Theorem 3.2. □

5 Confluence of the Restricted System

A rewrite $t \Rightarrow t'$ is *reversible* if $t' \Rightarrow t$. Among these are the *triangular expansions*, i.e. the expansions appearing in the triangle laws (2,3,4). The 1-step *restricted reduction system* consists of those 1-step rewrites of the expansionary relation which are *not* triangular expansions.

The proof of confluence for the restricted system is obtained by imposing restrictions on parallel expansion which reflect those on expansions generally. Expansions are triangular if they are either of non-neutral terms, or of terms which are projected or applied etc. The first restriction is directly incorporated into the definition below. The second, context-sensitive, restriction appears indirectly, via the notion of principle subterm, which we now define.

The *principle subterm* of an application tu is t; that of the projections πt and $\pi' t$ is t. Other terms do not have a principle subterm.

Restricted parallel expansion (denoted $\twoheadrightarrow_r$) is the smallest relation on terms closed under

(i) Expansion: If u' is neutral and of product, exponent or unit type then

$$\frac{u \twoheadrightarrow_r u'}{u \twoheadrightarrow_r \eta(u')}$$

(ii) Congruence: For each term constructor $\mathcal{T}$

$$\frac{u_i \twoheadrightarrow_r u_i'}{\mathcal{T}(u_i) \twoheadrightarrow_r \mathcal{T}(u_i')}$$

provided that if u_i is the principle subterm of $\mathcal{T}(u_i)$ then there is a derivation of $u_i \twoheadrightarrow_r u_i'$ whose last step is not an expansion.

Lemma 5.1 *Restricted parallel expansions lie in the reflexive transitive closure of restricted expansion.*

Proof An inductive proof will establish the stronger claim that if $u \twoheadrightarrow_r v$ then there is a reduction $u \rightarrow_{\eta^*} v$ such that if the former derivation does not end in an expansion then the latter does not contain a basic expansion. If the last step of the derivation of $u \twoheadrightarrow_r v$ is an expansion then the result is immediate. If the last step yields $\mathcal{T}(u_i) \twoheadrightarrow_r \mathcal{T}(v_i)$ by congruence then the existence of the reduction follows by induction (and the restrictions on principle sub-terms) and the resulting reduction contains no basic expansions since all reductions act on the u_i's. $\square$

Lemma 5.2 *If $t_1 \twoheadrightarrow t_2$ then there is a β-reduction $t_2 \rightarrow_{\beta^*} t_3$ such that $t_1 \twoheadrightarrow_r t_3$.*

$$\begin{tikzcd} t_1 \arrow[r, two heads] \arrow[d, two heads, "r"'] & t_2 \arrow[dl, "\beta^*"] \\ t_3 & \end{tikzcd}$$

Proof The proof is by induction on the height of the derivation. If the last step is of the form

$$\frac{t_1 \twoheadrightarrow t_2'}{t_1 \twoheadrightarrow \eta(t_2')}$$

then the induction hypothesis gives a term t_3' such that $t_2' \rightarrow_{\beta^*} t_3'$ and $t_1 \twoheadrightarrow_r t_3'$. If t_3' is neutral, then $t_3 = \eta(t_3')$ is as required. Otherwise $\eta(t_2') \rightarrow_{\beta^*} \eta(t_3') \rightarrow_{\beta^*} t_3'$ by Lemma 4.1(iv,ii). Now set $t_3 = t_3'$.

Alternatively, if the last step of $t_1 \twoheadrightarrow t_2$ is a congruence, say

$$\frac{u_i \twoheadrightarrow u_i'}{\mathcal{T}(u_i) \twoheadrightarrow \mathcal{T}(u_i')}$$

then the induction hypothesis yields reductions $u_i \twoheadrightarrow_r v_i$ and $u_i' \rightarrow_{\beta^*} v_i$ for each i. Hence $\mathcal{T}(u_i') \rightarrow_{\beta^*} \mathcal{T}(v_i)$ and $\mathcal{T}(u_i) \twoheadrightarrow_r \mathcal{T}(v_i)$ unless $\mathcal{T}(u_i)$ has a principle sub-term u_j for which $u_j \twoheadrightarrow_r v_j$ ends with v_j' being expanded. Then $\mathcal{T}(v_i')$ where $v_i' = v_i$ if $i \neq j$ is the required term. $\square$

Lemma 5.3 *Restricted parallel expansion satisfies the diamond lemma relative to β-reduction.*

Proof Given two $\twoheadrightarrow_r$-reducts of t, there is a completion of the form

$$\begin{tikzcd}
{} \arrow[r, two heads, "r"] \arrow[d, two heads, "r"'] & {} \arrow[r, two heads, "r"] \arrow[d, two heads] & {} \arrow[dd, "\beta^*"] \\
{} \arrow[r, two heads] \arrow[d, two heads, "r"'] & {} \arrow[ur, "\beta^*"'] \arrow[dl, "\beta^*"] & \\
{} \arrow[rr, "\beta^*"'] & & {}
\end{tikzcd}$$

where the top square arises from the diamond lemma for parallel expansion, the triangular cells are instances of Lemma 5.2, and the bottom cell is given by confluence of β-reduction. $\square$

Lemma 5.4 *Restricted parallel expansion is β-extendable.*
Proof Any divergence of the form $t_1 \leftarrow_\beta t \twoheadrightarrow_r t_2$ can be completed to

$$\begin{tikzcd} t \arrow[r, "\beta"] \arrow[d, two heads, "r"'] & t_1 \arrow[r, "\beta^*"] & t_3 \arrow[r, two heads, "r"] \arrow[d, two heads] & t_5 \\ t_2 \arrow[rr, "\beta^*"'] & & t_4 \arrow[ur, "\beta^*"'] & \end{tikzcd}$$

where the first cell is the β-extendability of $\twoheadrightarrow$ and the second is by Lemma 5.2. $\square$

Theorem 5.5 *The restricted expansionary system is confluent.*
Proof The confluence of $\rightarrow_\beta \cup \twoheadrightarrow_r$ is an application of Theorem 3.2 so that it suffices to prove that the restricted system has the same reflexive transitive closure. That restricted parallel expansion is reflexive and contains restricted expansion follows as in the unrestricted case, and Lemma 5.1 establishes the rest. $\square$

6 Strong Normalisation

Theorem 6.1 (Mints) *The restricted reduction system is strongly normalising to Huet's long $\beta\eta$-normal forms [12].* $\square$.

The long $\beta\eta$-normal forms are simply the terms which are irreducible in restricted system. Reduction to normal form can be achieved by first reducing to β-normal form, and then performing restricted expansions (from the inside out). (Čubrić's counter-example to this strategy arises from taking $a \Rightarrow *$ to be a β-rule.) The following lemmas are of use

Lemma 6.2 *(i) Let $t : A \times B$ be a term. If πt and $\pi' t$ are strongly normalisable then so is t.*

(ii) Let $t : A \rightarrow B$ be a term and $x : A$ be a variable not free in t. If $tx : B$ is strongly normalisable then so is t.

Proof For (i) it suffices to show that all the 1-step reducts of t are strongly normalisable, by induction on the rank of πt. The result $\langle \pi t, \pi' t \rangle$ of a basic expansion is strongly normalisable since its only reductions arise from those of its components. Any other 1-step reduction $t \Rightarrow t'$ induces another such of πt and $\pi' t$. The projections of t' are strongly normalisable, and of lower rank than those of t. Thus t' is strongly normalisable by induction. (ii) is proved similarly to (i) since reductions of λ-terms are all obtained by reductions of their body. $\square$

Strong normalisation of the restricted system is proved by Girard's reducibility techniques [7]. The presence of expansionary, context sensitive rewrite rules means that, although the reducibility predicates remain the same, much greater subtlety is required in establishing the lemmas of traditional reducibility proofs [7].

The set RED_T of *reducible terms* for each type T is defined by induction on the structure of T. Let $t : T$ be a term.

(i) If T is an atomic type then t is reducible if it is strongly normalisable.

(ii) If $T = U \times V$ then t is reducible if πt and $\pi' t$ are.

(iii) If $T = U \rightarrow V$ then t is reducible if $tu : V$ is reducible for all reducible $u : U$.

The proof hinges on simultaneously establishing the following three hypotheses

CR1 If $t \in \mathrm{RED}_T$ then t is strongly normalisable.

CR2 If $t \in \mathrm{RED}_T$ and $t \Rightarrow t'$ then $t' \in \mathrm{RED}_T$.

CR3′ If t is neutral and every 1-step reduct of t *other than those obtained by a basic expansion of* t is in RED_T then $t \in \mathrm{RED}_T$.

CR3′ differs from the original CR3 by the insertion of the italicised restriction, introduced to cope with the expansions. In particular, CR3′ implies that variables are reducible. Although the remaining modifications to the proof are minor, it is presented in full because of the basic delicacy of the arguments.

Lemma 6.3 *Let* CR1, 2, 3′ *hold for types A and B and let $u : A$ and $v : B$ be reducible. Then $\langle u, v\rangle$ is reducible.*
Proof Both u and v are strongly normalisable by CR1. Thus we can use induction on the sum of their ranks to prove that $t = \pi\langle u, v\rangle$ is reducible by CR3′. Its 1-step reducts (other than basic expansions) are

(i) u

(ii) $\pi\langle u', v\rangle$, where $u \Rightarrow u'$

(iii) $\pi\langle u, v'\rangle$ where $v \Rightarrow v'$

(i) u is reducible by assumption. (ii) u' is reducible by CR2 and has lower rank than u. Hence $\pi\langle u', v\rangle$ is reducible by induction. (iii) This is similar to (ii). □

Lemma 6.4 *Let* CR1, 2, 3′ *hold for all sub-types of A and B (inclusive) and let $t : B$ be a term and $x : A$ be a variable. If the term $t[u/x] : B$ is reducible whenever $u : A$ is then $\lambda x.t : A \rightarrow B$ is reducible.*
Proof Assume the result is true whenever the type of $\lambda x.t$ is a proper sub-type of $A \rightarrow B$. Since $t = t[x/x]$ is reducible it follows that it is strongly normalisable, as is any reducible $u : A$. Thus we can use induction on the sum of their ranks to prove that $(\lambda x.t)u : B$ is reducible by CR3′. Its 1-step reducts (other than basic expansions) are

(i) $t[u/x]$

(ii) $(\lambda x.t)u'$, where $u \Rightarrow u'$

(iii) $(\lambda x.t')u$ where $t \Rightarrow t'$

(i) $t[u/x]$ is reducible by assumption. (ii) u' has lower rank than u. (iii) Since t' has lower rank than t it suffices to show that $t'[u/x]$ is reducible, and apply induction.

If the reduction $t \Rightarrow t'$ induces a reduction $t[u/x] \Rightarrow t'[u/x]$ then we are done by CR2. A simple induction on the structure of t shows the only alternative to be that t' is obtained by a basic expansion of (one of the possibly many occurrences of) the free variable x in t and u is either a λ-abstraction or a pair. If $t[\eta(u)/x]$ is reducible, then we are done, since $t[\eta(u)/x] \Rightarrow t'[u/x]$ and CR2 holds for terms of type B. This in turn follows if $\eta(u)$ is reducible and there are two possibilities. Firstly if u is of product type, apply Lemma 6.3, while secondly if u is of functional type, we must prove that $\lambda y.uy$ is reducible (where y is not free in u). This follows from the induction hypothesis since $uv : D$ is reducible whenever $v : C$ is. $\square$

Theorem 6.5 CR1,CR2 *and* CR3$'$ *hold for every type* T.

Proof The proof is by induction on the structure of the type T. In each case two forms of 1-step reduction of a term $t : T$ are considered, namely the basic expansions and the others.

If T is atomic then CR1 is a tautology and CR2 holds trivially. For CR3$'$ it suffices to show that any basic expansion of t is also strongly normalisable but the only case is $t \Rightarrow *$.

Consider $T = A \times B$. If t is reducible then so are πt and $\pi' t$ which are then strongly normalisable by induction. Thus t is strongly normalisable by Lemma 6.2 and so CR1 holds.

CR2 for a basic expansion of t follows from Lemma 6.3 since the components of the pair $\langle \pi t, \pi' t \rangle$ are reducible by definition. Otherwise a 1-step reduction $t \Rightarrow t'$ yields $\pi t \Rightarrow \pi t'$ whence $\pi t'$ is reducible by CR2 for A. The analogous argument for $\pi' t'$ holds and so t' is reducible, as required.

For CR3$'$, let t be a neutral term whose 1-step reductions other than basic expansions produce reducible terms. Since t is neutral, a 1-step reduction of πt which is not a basic expansion must be of the form $\pi t \Rightarrow \pi t'$ where $t \Rightarrow t'$. The latter reduction is *not* a basic expansion since otherwise $\pi t \Rightarrow \pi t'$ would be triangular. Thus t' is reducible by hypothesis, whence $\pi t'$ is by definition. Thus πt is reducible by CR3$'$ for A. The analogous argument for $\pi' t$ holds and so t is reducible, as required.

Now consider $T = A \rightarrow B$. If t is a reducible term and $x : A$ is a variable not free in t then tx is reducible by definition, and so strongly normalisable by hypothesis. Thus t is strongly normalisable by Lemma 6.2.

CR2 for a basic expansion of t to $\lambda x.tx$ follows from Lemma 6.4 since its conditions are satisfied by assumption. The other 1-step reductions $t \Rightarrow t'$ induce $tu \Rightarrow t'u$

for any reducible term $u : A$ and so $t'u$ is reducible by CR2 for B. Thus t' is reducible by definition.

For CR3$'$, let t be a neutral term whose 1-step reductions other than basic expansion produce reducible terms. We will show that $tu : B$ is reducible for every reducible term $u : A$ by CR3$'$ for B and induction on the rank of u and . Consider a 1-step reduction of tu which is not a basic expansion. Since t is neutral, it is given by either $t \Rightarrow t'$ or $u \Rightarrow u'$. Now such a reduction $t \Rightarrow t'$ cannot be a basic expansion (since otherwise $tu \Rightarrow t'u$ would be triangular) and so t' is reducible, whence $t'u$ is. On the other hand, u' is reducible and of lower rank than u by CR1 and CR2. □

Proposition 6.6 *Let $t : T$ be any term, with free variables among $x_i : X_i$ for $i = 1, \ldots, n$ and let $u_i : X_i$ be reducible terms. Then $t[u_i/x_i]$ is reducible.*
Proof The proof is by induction over the structure of the term. The cases involving variables, $*$, projection and application are all trivial, while pairing is handled by Lemma 6.3. Finally, if $t = \lambda y.b : A \to B$ then $t[u_i/x_i]$ is reducible iff $b[u_i/x_i][v/y]$ is reducible for all reducible $v : A$ which follows by induction. □

Corollary 6.7 *In the restricted system every term is reducible, and so is strongly normalisable.*
Proof Apply the theorem with $u_i = x_i$. □

This normalisation result yields separate proofs of the earlier confluence theorems. That of the restricted system follows upon establishing weak confluence; the full system is covered by

Corollary 6.8 *If $t \Rightarrow t'$ in the full system then t and t' have the same normal form in the restricted system. Hence the full system is confluent.*
Proof Let t_0 be the normal form of t in the restricted system. It suffices, by induction on the length of the reduction sequence, to consider a 1-step reduction $t \Rightarrow t'$. If this step is in the restricted system then t' also has normal form t_0. Otherwise, $t_0 \Rightarrow t_1$ is a triangular expansion, and so is reversible, whence $t_1 \Rightarrow t_0 \Rightarrow t$ in the restricted system. Confluence follows directly. □

7 Essential Normalisation

As noted above, the presence of reduction loops means that no terms of higher type are normal in the full system, at least in the usual sense. There is, however, a weaker notion which "ignores" reversible reductions.

A term t is an *essential normal form* if any reduction of it is reversible, i.e. if $t \Rightarrow t'$ (not necessarily in one step) then $t' \Rightarrow t$. It is *essentially normalisable* if there is a number k, called an *essential bound* for t, such that each reduction sequence from t has at most k irreversible steps. The reduction system is *(strongly) essentially normalising* if every term is so, and *weakly essentially normalising* if every term reduces to an essentially normal term.

Theorem 7.1 *The full system is weakly essentially normalising to the long $\beta\eta$-normal forms.*

Proof If t is a long $\beta\eta$-normal form then by Corollary 6.8 its reducts in the full system reduce to its normal form in the restricted system, i.e. to t itself. Thus t is essentially normal. As every term reduces to a long $\beta\eta$-normal form in the restricted system, we are done. $\square$

Several plausible conjectures about this system do not hold, as will be seen from the following examples.

Not every reversible reduction is triangular. Let $f : N \rightarrow N \rightarrow N$ and $m, n : N$ all be variables. Then

$$fmn \Rightarrow (\lambda x.fx)mn \Rightarrow (\lambda x.\lambda y.fxy)mn \tag{5}$$

Now fmn is a long $\beta\eta$-normal form and so of course the first reduction must be a triangular expansion. The second expansion is reversible but not triangular. Similar possibilities arise with pairing.

Proposition 7.2 *The full system is* not *strongly essentially normalising.*

Proof Let $s = (\lambda x.\langle y, y'\rangle)x : N \times N$ where $x, y, y' : N$ are variables. Then $\pi s \Rightarrow y$ and $\pi' s \Rightarrow y'$ and so

$$\begin{aligned}
\pi s &\Rightarrow \pi\langle \pi s, \pi' s\rangle \\
&\Rightarrow \pi\langle y, \pi' s\rangle \\
&\Rightarrow \pi\langle y, \pi'\langle \pi s, \pi' s\rangle\rangle \\
&\Rightarrow \pi\langle y, \pi'\langle \pi s, y'\rangle\rangle
\end{aligned}$$

Now set $r = \pi\langle y, \pi'\langle z, y'\rangle\rangle$ where $z : N$ is a variable and define

$$\begin{aligned}
s_0 &= \pi s \\
s_{n+1} &= r[s_n/z]
\end{aligned}$$

That $s_n \Rightarrow s_{n+1}$ was proved above for $n = 0$, while in general

$$s_{n+1} = r[s_n/u] \Rightarrow r[s_{n+1}/u] = s_{n+2}$$

The result follows upon showing the irreversibility of these reductions. More generally, we show, by induction on m, that if $s_m \Rightarrow s_n$ then $m \geq n$. The base case is vacuously true. For the induction step, consider a reduction $s_{m+1} = r[s_m/z] \Rightarrow s_n$.

Observe that if $r \Rightarrow r'$ then all occurrences of z in r' are as a component of a pair, since $r' \neq z$ by normalisation considerations and z may not be the subject of a projection.

Thus the reduction above decomposes into a one of $r \Rightarrow r'$ and, for each free occurrence of z in r', a reduction of s_m to a sub-term t of s_n. Now t has the same essential normal form y as s_m and so must be either y or s_p for some $p \leq n$ (by induction on n). Thus if t is s_p then the original induction hypothesis implies $n \geq p \geq m$ as required. Otherwise, each t is y and so $r'[y/z] = s_n$ which is impossible since the left-hand side is λ-free. $\square$

8 Acknowledgements

We are grateful to D. Čubrić who introduced us to the links with Mints' work when these results were first presented in 1989, and also to Sergei Soloviev who provided a translation of some of the key passages in [21]. We would also like to thank Pierre-Louis Curien, Stefan Kahrs, Edmund Kazmierczak, Michael Mendler, Gordon Plotkin, Alex Simpson and Terry Stroup for many stimulating conversations, and the members of the Applications of Categories Club and the Categorical Rewriting Club for their constructive criticism.

References

[1] Y.Akama, *On Mints' Reduction for ccc-Calculus* To appear in Typed Lambda Calculi and Applications 1993.

[2] H.P. Barendregt, *The Lambda Calculus Its Syntax and Semantics (Revised Edition)* Studies in Logic and the Foundations of Mathematics 103 (North Holland, 1984).

[3] A. Carboni, G.M. Kelly and R. Wood, *A 2-categorical approach to change of base and geometric morphisms I*, Sydney Category Seminar Reports 90-1 (1990).

[4] D. Čubrić, *On Free CCC*, manuscript.

[5] P-L. Curien and R. Di Cosmo, A confluent reduction for the λ-calculus with surjective pairing and terminal object, in: *Proceedings ICALP '91, Madrid* LNCS 510 (Springer, 1991) 291–302 (also appearing as Tech. Rep. LIENS–91–11).

[6] R.Di Cosmo and D.Kesner, *A confluent reduction for the extensional typed λ-calculus.*

[7] J-Y. Girard, P. Taylor and Y. Lafont, *Proofs and Types*, Cambridge Tracts in Theoretical Computer Science (Cambridge University Press, 1989).

[8] J.W. Gray, *Formal category theory: adjointness for 2-categories*, Lecture Notes in Mathematics 391 (Springer-Verlag, 1974).

[9] R. Harper, F. Honsell and G. Plotkin, *A framework for defining logics*, Technical Report LFCS-91-162 (LFCS, 1991).

[10] B.P. Hilken and D.E. Rydeheard, *Towards a categorical semantics of type classes*, Fundamenta Informatica, to appear.

[11] C.A.R. Hoare, He Jifeng & C.E. Martin, *Pre-adjunctions in Order Enriched Categories*, Oxford University Computing Laboratory (1989).

[12] G. Huet, *Résolution d'équations dans des langages d'ordre* $1, 2, \ldots, \omega$. Thèse d'Etat, Université de Paris VII, 1976.

[13] C.B. Jay, *Local adjunctions*, J. Pure and Appl. Alg. 53 (1988) 227–238.

[14] C.B. Jay, *Extendinging properties to categories of partial maps*, Tech. Rep. ECS- LFCS 90-107.

[15] C.B. Jay, Partial functions, ordered categories, limits and cartesian closure, in: G. Birtwistle (ed), *IV Higher Order Workshop, Banff, 1990*, (Springer, 1991) 151–161.

[16] C.B. Jay, Modelling reduction in confluent categories, in: *Proceedings of the Durham Symposium on Applications of Categories in Computer Science*, Lecture Note Series 177 (London Mathematics Society, 1992) 143–162.

[17] C.B. Jay, *Long $\beta\eta$ Normal Forms and Confluence*, Tech. Rep. ECS-LFCS-91-183 (and its revised version).

[18] S. Kahrs, *λ-rewriting*, PhD thesis, Universität Bremen, 1991. (in German)

[19] S. Kasangian, G.M. Kelly and F. Rossi, *Cofibrations and the realization of non deterministic automata*, Cahiers de Topologie et Géometrie Différentielle 24 (1983) 23–46.

[20] J. Lambek and P. Scott, *Introduction to higher order categorical logic*, Cambridge Studies in Advanced Mathematics 7 (Cambridge Univ. Press, 1986).

[21] G.E. Mints, Teorija categorii i teoria dokazatelstv.I., in: *Aktualnye problemy logiki i metodologii nauky , Kiev, 1979* 252-278.

[22] D. Prawitz, Ideas and results in proof theory, in: J.E. Fenstad (ed) *Proc. 2nd Scandinavian Logic Symp.* (North-Holland, 1971) 235–307.

[23] D.E. Rydeheard & J.G. Stell, *Foundations of equational deduction: A categorical treatment of equational proofs and unification algorithms*, in: Pitt et al, (eds), *Category Theory and Computer Science*, Lecture Notes in Computer Science 283 (Springer, 1987) 114 – 139.

[24] R.A.G. Seely, *Modelling computations: a 2-categorical framework*, in: *Proceedings of the Second Annual Symposium on Logic in Computer Science* (1987).