

# On the exp-log normal form of types

Danko Ilik

► **To cite this version:**

Danko Ilik. On the exp-log normal form of types: Decomposing extensional equality and representing terms compactly. Proceedings of the 44th ACM SIGPLAN Symposium on Principles of Programming Languages, Jan 2017, Paris, France. pp.387-399. <hal-01167162v2>

**HAL Id: hal-01167162**

**<https://hal.inria.fr/hal-01167162v2>**

Submitted on 17 Aug 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

# The exp-log normal form of types

## Decomposing extensional equality and representing terms compactly

Danko Ilik

Inria & LIX, Ecole Polytechnique  
91128 Palaiseau Cedex, France  
danko.ilik@inria.fr

### Abstract

Lambda calculi with algebraic data types lie at the core of functional programming languages and proof assistants, but conceal at least two fundamental theoretical problems already in the presence of the simplest non-trivial data type, the sum type. First, we do not know of an explicit and implemented algorithm for deciding the beta-eta-equality of terms—and this in spite of the first decidability results proven two decades ago. Second, it is not clear how to decide when two types are essentially the same, i.e. isomorphic, in spite of the meta-theoretic results on decidability of the isomorphism.

In this paper, we present the exp-log normal form of types—derived from the representation of exponential polynomials via the unary exponential and logarithmic functions—that any type built from arrows, products, and sums, can be isomorphically mapped to. The type normal form can be used as a simple heuristic for deciding type isomorphism, thanks to the fact that it is a systematic application of the high-school identities.

We then show that the type normal form allows to reduce the standard beta-eta equational theory of the lambda calculus to a specialized version of itself, while preserving the completeness of equality on terms.

We end by describing an alternative representation of normal terms of the lambda calculus with sums, together with a Coq-implemented converter into/from our new term calculus. The difference with the only other previously implemented heuristic for deciding interesting instances of eta-equality by Balat, Di Cosmo, and Fiore, is that we exploit the type information of terms substantially and this often allows us to obtain a canonical representation of terms without performing sophisticated term analyses.

**Categories and Subject Descriptors** Software and its engineering [Language features]: Abstract data types; Software and its engineering [Formal language definitions]: Syntax; Theory of computation [Program constructs]: Type structures

**Keywords** sum type, eta equality, normal type, normal term, type isomorphism, type-directed partial evaluation

### 1. Introduction

The lambda calculus is a notation for writing functions. Be it simply-typed or polymorphic, it is also often presented as the core of modern functional programming languages. Yet, besides functions as first-class objects, another essential ingredient of these languages are algebraic data types that typing systems supporting only the  $\rightarrow$ -type and polymorphism do not model directly. A natural model for the core of functional languages should at least include direct support for a simplest case of variant types, *sums*, and of records i.e. *product* types. But, unlike the theory of the  $\{\rightarrow\}$ -typed lambda calculus, the theory of the  $\{\rightarrow, +, \times\}$ -typed one is not all roses.

**Canonicity of normal terms and  $\eta$ -equality** A first problem is canonicity of normal forms of terms. Take, for instance, the term  $\lambda xy.yx$  of type  $\tau + \sigma \rightarrow (\tau + \sigma \rightarrow \rho) \rightarrow \rho$ , and three of its  $\eta$ -long representations,

$$\begin{aligned} &\lambda x.\lambda y.y\delta(x, z.\iota_1 z, z.\iota_2 z) \\ &\lambda x.\lambda y.\delta(x, z.y(\iota_1 z), z.y(\iota_2 z)) \\ &\lambda x.\delta(x, z.\lambda y.y(\iota_1 z), z.\lambda y.y(\iota_2 z)), \end{aligned}$$

where  $\delta$  is a pattern matching construct, i.e. a *case*-expression analysing the first argument, with branches of the pattern matching given via the variable  $z$  in the second and third argument.

These three terms are all equal with respect to the standard equational theory  $=_{\beta\eta}$  of the lambda calculus (Figure 1), but why should we prefer any one of them over the others to be a *canonical* representative of the class of equal terms?

Or, consider the following two terms of type  $(\tau_1 \rightarrow \tau_2) \rightarrow (\tau_3 \rightarrow \tau_1) \rightarrow \tau_3 \rightarrow \tau_4 + \tau_5 \rightarrow \tau_2$  (example taken from (Balat et al. 2004)):

$$\begin{aligned} &\lambda xyzu.x(yz) \\ &\lambda xyzu.\delta(\delta(u, x_1.\iota_1 z, x_2.\iota_2(yz)), y_1.x(yy_1), y_2.xy_2). \end{aligned}$$

These terms are  $\beta\eta$ -equal, but can one easily notice the equality? In order to do so, since both terms are  $\beta$ -normal, one would need to do non-trivial  $\beta$ - and  $\eta$ -expansions (see Example 2 in Section 4).

For the lambda calculus over the restricted language of types—when the sum type is absent—these problems do not exist, since  $\beta$ -normalization followed by an  $\eta$ -expansion is deterministic and produces a canonical representative for any class of  $\beta\eta$ -equal terms. Deciding  $=_{\beta\eta}$  for that restricted calculus amounts to comparing canonical forms up to syntactic identity.

In the presence of sums, we only have a notion of canonical *interpretation* of terms in the category of sheaves for the Grothendieck topology over the category of constrained environments (Altenkirch et al. 2001), as well as the sophisticated normal form of terms due to Balat, Di Cosmo, and Fiore which is not canonical (unique) syntactically (Balat et al. 2004). Balat et al. also

provide an implementation of a type-directed partial evaluator that normalizes terms to their normal form, and this represented up to now the only implemented *heuristic* for deciding  $\beta\eta$ -equality—it is not a full decision procedure, because the normal forms are not canonical. We shall discuss these and the other decidability results some more in Related work of Section 5.

Treating *full*  $\beta\eta$ -equality is hard, even if, in practice, we often only need to treat special cases of it, such as certain commuting conversions.

**Recognizing isomorphic types** If we leave aside the problems of canonicity of and equality between terms, there is a further problem at the level of *types* that makes it hard to determine whether two type signatures are essentially the same one. Namely, although for each of the type languages  $\{\rightarrow, \times\}$  and  $\{\rightarrow, +\}$  there is a very simple algorithm for deciding *type isomorphism*, for the whole of the language  $\{\rightarrow, +, \times\}$  it is only known that type isomorphism is decidable when types are to be interpreted as finite structures, and that without a practically implementable algorithm in sight (Ilik 2014).

The importance of deciding type isomorphism for functional programming has been recognized early on by Rittri (Rittri 1991), who proposed to use it as a criterium for searching over a library of functional subroutines. Two types being isomorphic means that one can switch programs and data back and forth between the types without loss of information. Recently, type isomorphisms have also become popular in the community around homotopy type theory.

It is embarrassing that there are no algorithms for deciding type isomorphism for such an ubiquitous type system. Finally, even if finding an implementable decision procedure for the *full* type language  $\{\rightarrow, +, \times\}$  were hard, might we simply be able to cover fragments that are important in practice?

**Organization of this paper** In this paper, we shall be treating the two kinds of problems explained above simultaneously, not as completely distinct ones: traditionally, studies of canonical forms and deciding equality on terms have used very little of the type information annotating the terms (with the exceptions mentioned in the concluding Section 5).

We shall start by introducing in Section 2 a normal form for *types*—called the *exp-log normal form* (ENF)—that preserves the isomorphism between the source and the target type; we shall also give an implementation, a purely functional one, that can be used as a heuristic procedure for deciding isomorphism of two types.

Even if reducing a type to its ENF does not present a complete decision procedure for isomorphism of *types*, we shall show in the subsequent Section 3 that it has dramatic effects on the theory of  $\beta\eta$ -equality of *terms*. Namely, one can reduce the problem of showing equality for the standard  $=_{\beta\eta}$  relation to the problem of showing it for a new equality theory  $=_{\beta\eta}^e$  (Figure 2)—this later being a *specialization* of  $=_{\beta\eta}$ . That is, a complete axiomatization of  $\beta\eta$ -equality that is a strict subset of the currently standard one is possible.

In Section 4, we shall go further and describe a minimalist calculus of terms—*compact terms* at ENF type—that can be used as an alternative to the usual lambda calculus with sums. With its properties of a syntactic simplification of the later (for instance, there is no lambda abstraction), the new calculus allows a more canonical representation of terms. We show that, for a number of interesting examples, converting lambda terms to compact terms and comparing the obtained terms for syntactic identity provides a simple heuristic for deciding  $=_{\beta\eta}$ .

The paper is accompanied by a prototype normalizing converter between lambda- and compact terms implemented in Coq.

## 2. The exp-log normal form of types

The trouble with sums starts already at the level of types. Namely, when we consider types built from function spaces, products, and disjoint unions (sums),

$$\tau, \sigma ::= \chi_i \mid \tau \rightarrow \sigma \mid \tau \times \sigma \mid \tau + \sigma,$$

where  $\chi_i$  are atomic types (or type variables), it is not always clear when two given types are essentially the same one. More precisely, it is not known *how* to decide whether two types are isomorphic (Ilik 2014). Although the notion of isomorphism can be treated abstractly in Category Theory, in bi-Cartesian closed categories, and without committing to a specific term calculus inhabiting the types, in the language of the standard syntax and equational theory of lambda calculus with sums (Figure 1), the types  $\tau$  and  $\sigma$  are isomorphic when there exist coercing lambda terms  $M : \sigma \rightarrow \tau$  and  $N : \tau \rightarrow \sigma$  such that

$$\lambda x.M(Nx) =_{\beta\eta} \lambda x.x \quad \text{and} \quad \lambda y.N(My) =_{\beta\eta} \lambda y.y.$$

In other words, data/programs can be converted back and forth between  $\tau$  and  $\sigma$  without loss of information.

The problem of isomorphism is in fact closely related to the famous Tarski High School Identities Problem (Burriss and Yeats 2004; Fiore et al. 2006). What is important for us here is that *types can be seen as just arithmetic expressions*: if the type  $\tau \rightarrow \sigma$  is denoted by the binary arithmetic exponentiation  $\sigma^\tau$ , then every type  $\rho$  denotes at the same time an *exponential* polynomial  $\rho$ . The difference with ordinary polynomials is that the exponent can now also contain a (type) variable, while exponentiation in ordinary polynomials is always of the form  $\sigma^n$  for a concrete  $n \in \mathbb{N}$  i.e.  $\sigma^n = \underbrace{\sigma \times \dots \times \sigma}_{n\text{-times}}$ . Moreover, we have that

$$\tau \cong \sigma \text{ implies } \mathbb{N}^+ \models \tau = \sigma,$$

that is, type isomorphism implies that arithmetic equality holds for any substitution of variables by positive natural numbers.

This hence provides an procedure for proving *non-isomorphism*: given two types, prove they are not equal as exponential polynomials, and that means they cannot possibly be isomorphic. But, we are interested in a positive decision procedure. Such a procedure exists for both the languages of types  $\{\rightarrow, \times\}$  and  $\{\times, +\}$ , since then we have an equivalence:

$$\tau \cong \sigma \text{ iff } \mathbb{N}^+ \models \tau = \sigma.$$

Indeed, in these cases type isomorphism can not only be decided, but also effectively built. In the case of  $\{\times, +\}$ , the procedure amounts to transforming the type to disjunctive normal form, or the (*non-exponential*) polynomial to canonical form, while in that of  $\{\rightarrow, \times\}$ , there is a canonical normal form obtained by type transformation that follows currying (Rittri 1991).

Given that it is not known whether one can find such a canonical normal form for the full language of types (Ilik 2014), what we can hope to do in practice is to find at least a *pseudo-canonical* normal form. We shall now define such a type normal form.

The idea is to use the decomposition of the binary exponential function  $\sigma^\tau$  through unary exponentiation and logarithm. This is a well known transformation in Analysis, where for the natural logarithm and Euler's number  $e$  we would use

$$\sigma^\tau = e^{\tau \times \log \sigma} \quad \text{also written} \quad \sigma^\tau = \exp(\tau \times \log \sigma).$$

The systematic study of such normal forms by Du Bois-Reymond described in the book (Hardy 1910) served us as inspiration.

But how exactly are we to go about using this equality for types when it uses logarithms i.e. transcendental numbers? Luckily, we do not have to think of real numbers at all, because what is described above can be seen through the eyes of abstract Algebra, in

$$\begin{aligned}
M, N ::= & x^\tau \mid (M^{\tau \rightarrow \sigma} N^\sigma)^\sigma \mid (\pi_1 M^{\tau \times \sigma})^\tau \mid (\pi_2 M^{\tau \times \sigma})^\sigma \mid \delta(M^{\tau + \sigma}, x_1^\tau.N_1^\rho, x_2^\sigma.N_2^\rho)^\rho \mid \\
& |(\lambda x^\tau.M^\sigma)^{\tau \rightarrow \sigma} \mid \langle M^\tau, N^\sigma \rangle^{\tau \times \sigma} \mid (\iota_1 M^\tau)^{\tau + \sigma} \mid (\iota_2 M^\sigma)^{\tau + \sigma} \\
& (\lambda x.N)M =_\beta N\{M/x\} \tag{\beta_\rightarrow} \\
& \pi_i \langle M_1, M_2 \rangle =_\beta M_i \tag{\beta_\times} \\
& \delta(\iota_i M, x_1.N_1, x_2.N_2) =_\beta N_i\{M/x_i\} \tag{\beta_+} \\
& N =_\eta \lambda x.Nx \quad x \notin \text{FV}(N) \tag{\eta_\rightarrow} \\
& N =_\eta \langle \pi_1 N, \pi_2 N \rangle \tag{\eta_\times} \\
& N\{M/x\} =_\eta \delta(M, x_1.N\{\iota_1 x_1/x\}, x_2.N\{\iota_2 x_2/x\}) \quad x_1, x_2 \notin \text{FV}(N) \tag{\eta_+}
\end{aligned}$$

**Figure 1.** Terms of the  $\{\rightarrow, +, \times\}$ -typed lambda calculus and axioms of the equational theory  $=_{\beta\eta}$  between typed terms.

exponential fields, as a pair of mutually inverse homomorphisms  $\exp$  and  $\log$  between the multiplicative and additive group, satisfying

$$\begin{aligned}
\exp(\tau_1 + \tau_2) &= \exp \tau_1 \times \exp \tau_2 & \exp(\log \tau) &= \tau \\
\log(\tau_1 \times \tau_2) &= \log \tau_1 + \log \tau_2 & \log(\exp \tau) &= \tau.
\end{aligned}$$

In other words,  $\exp$  and  $\log$  can be considered as macro expansions rather than unary type constructors. Let us take the type  $\tau + \sigma \rightarrow (\tau + \sigma \rightarrow \rho) \rightarrow \rho$  from Section 1, assuming for simplicity that  $\tau, \sigma, \rho$  are atomic types. It can be normalized in the following way:

$$\begin{aligned}
\tau + \sigma \rightarrow (\tau + \sigma \rightarrow \rho) \rightarrow \rho &= \\
&= (\rho^{\tau + \sigma})^{\tau + \sigma} = \\
&= \exp((\tau + \sigma) \log\{\exp\{\exp((\tau + \sigma) \log \rho) \log \rho\}\}) \rightsquigarrow \\
&\rightsquigarrow \exp((\tau + \sigma) \log\{\exp\{\exp(\tau \log \rho) \exp(\sigma \log \rho) \log \rho\}\}) \rightsquigarrow \\
&\rightsquigarrow \exp((\tau + \sigma) \exp(\tau \log \rho) \exp(\sigma \log \rho) \log \rho) \rightsquigarrow \\
&\rightsquigarrow \exp(\tau \exp(\tau \log \rho) \exp(\sigma \log \rho) \log \rho) \\
&\exp(\sigma \exp(\tau \log \rho) \exp(\sigma \log \rho) \log \rho) = \\
&= \rho^{\tau \rho^\tau} \rho^\sigma \rho^{\sigma \rho^\sigma} \\
&= (\tau \times (\tau \rightarrow \rho) \times (\sigma \rightarrow \rho) \rightarrow \rho) \times (\sigma \times (\tau \rightarrow \rho) \times (\sigma \rightarrow \rho) \rightarrow \rho).
\end{aligned}$$

As the  $\exp$ - $\log$  transformation of arrow types is at the source of this type normalization procedure, we call the obtained normal form *the exp-log normal form (ENF)*. Be believe the link to abstract algebra is well work keeping in mind, since it may give rise to further cross-fertilization between mathematics and the theory of programming languages. However, from the operational point of view, all this transformation does is that it *prioritized* and *orients* the high-school identities,

$$\begin{aligned}
(f + g) + h &\rightsquigarrow f + (g + h) & (1) \\
(fg)h &\rightsquigarrow f(gh) & (2) \\
f(g + h) &\rightsquigarrow fg + fh & (3) \\
(f + g)h &\rightsquigarrow fh + gh & (4) \\
f^{g+h} &\rightsquigarrow f^g f^h & (5) \\
(fg)^h &\rightsquigarrow f^h g^h & (6) \\
(f^g)^h &\rightsquigarrow f^{hg}, & (7)
\end{aligned}$$

all of which are valid as type isomorphisms. We can thus also compute the *isomorphic* normal form of the type directly, for instance

for the second example of Section 1:

$$\begin{aligned}
(\tau_1 \rightarrow \tau_2) \rightarrow (\tau_3 \rightarrow \tau_1) \rightarrow \tau_3 \rightarrow \tau_4 + \tau_5 \rightarrow \tau_2 &= \\
&= \left( \left( \left( \tau_2^{\tau_4 + \tau_5} \right)^{\tau_3} \right)^{\tau_1 \tau_3} \right)^{\tau_2 \tau_1} \rightsquigarrow \\
&\rightsquigarrow \tau_2^{\tau_2 \tau_1} \tau_1^{\tau_3} \tau_3 \tau_4 \tau_2^{\tau_2 \tau_1} \tau_1^{\tau_3} \tau_3 \tau_5 = \\
&= (\tau_4 \times \tau_3 \times (\tau_3 \rightarrow \tau_1) \times (\tau_1 \rightarrow \tau_2) \rightarrow \tau_2) \times \\
&\quad (\tau_5 \times \tau_3 \times (\tau_3 \rightarrow \tau_1) \times (\tau_1 \rightarrow \tau_2) \rightarrow \tau_2).
\end{aligned}$$

Of course, some care needs to be taken when applying the rewrite rules, in order for the procedure to be deterministic, like giving precedence to the type rewrite rules and normalizing sub-expressions. To be precise, we provide a purely functional Coq implementation below. This is just one possible implementation of the rewriting rules, but being purely functional and structurally recursive (i.e. terminating) it allows us to understand the restrictions imposed on types in normal form, as it proves the following theorem.

**Theorem 1.** *If  $\tau$  is a type in exp-log normal form, then  $\tau \in \text{ENF}$ , where*

$$\text{ENF} \ni e ::= c \mid d,$$

where

$$\begin{aligned}
\text{DNF} \ni d, d_i &::= c_1 + (c_2 + (\dots + n) \dots) & n \geq 2 \\
\text{CNF} \ni c, c_i &::= (c_1 \rightarrow b_1) \times (\dots \times (c_n \rightarrow b_n) \dots) & n \geq 0 \\
\text{Base} \ni b, b_i &::= p \mid d,
\end{aligned}$$

and  $p$  denotes atomic types (type variables).

Assuming a given set of atomic types,

Parameter Proposition : Set.

the goal is to map the unrestricted language of types, given by the inductive definition,<sup>1</sup>

Inductive Formula : Set :=  
| prop : Proposition → Formula  
| disj : Formula → Formula → Formula  
| conj : Formula → Formula → Formula

<sup>1</sup>May the reader to forgive us for the implicit use of the Curry-Howard correspondence in the Coq code snippets, where we refer to types and type constructors as formulas and formula constructors.

```
| impl : Formula → Formula → Formula.
```

into the exp-log normal form which fits in the following inductive signature.

```
Inductive CNF : Set :=
| top
| con : CNF → Base → CNF → CNF
with DNF : Set :=
| two : CNF → CNF → DNF
| dis : CNF → DNF → DNF
with Base : Set :=
| prp : Proposition → Base
| bd : DNF → Base.
Inductive ENF : Set :=
| cnf : CNF → ENF
| dnf : DNF → ENF.
```

The  $con\ c_1\ b\ c_2$  constructor corresponds to  $b^{c_1}c_2$  or the type  $(c_1 \rightarrow b) \times c_2$  from Theorem 1. The normalization function,  $enf(\cdot)$ ,

```
Fixpoint enf (f : Formula) {struct f} : ENF :=
match f with
| prop p ⇒ cnf (p2c p)
| disj f0 f1 ⇒ dnf (nplus (enf f0) (enf f1))
| conj f0 f1 ⇒ distrib (enf f0) (enf f1)
| impl f0 f1 ⇒ cnf (explogn (enf2cnf (enf f1)) (enf f0))
end.
```

is defined using the following fixpoints:

**nplus** which makes a flattened  $n$ -ary sum out of two given  $n$ -ary sums, i.e. implements the  $+$ -associativity rewriting (1),

**ntimes** which is analogous to ‘nplus’, but for products, implementing (2),

**distrib** which performs the distributivity rewriting, (3) and (4), and

**explogn** which performs the rewriting involving exponentiations, (5), (6), and (7).

```
Fixpoint nplus1 (d : DNF)(e2 : ENF) {struct d} : DNF :=
match d with
| two c c0 ⇒ match e2 with
| cnf c1 ⇒ dis c (two c0 c1)
| dnf d0 ⇒ dis c (dis c0 d0)
end
| dis c d0 ⇒ dis c (nplus1 d0 e2)
end.
Definition nplus (e1 e2 : ENF) : DNF :=
match e1 with
| cnf a ⇒ match e2 with
| cnf c ⇒ two a c
| dnf d ⇒ dis a d
end
| dnf b ⇒ nplus1 b e2
end.
Fixpoint ntimes (c1 c2 : CNF) {struct c1} : CNF :=
match c1 with
| top ⇒ c2
```

```
| con c10 d c13 ⇒ con c10 d (ntimes c13 c2)
end.
```

```
Fixpoint distrib0 (c : CNF)(d : DNF) : ENF :=
match d with
| two c0 c1 ⇒ dnf (two (ntimes c c0) (ntimes c c1))
| dis c0 d0 ⇒ dnf match distrib0 c d0 with
| cnf c1 ⇒ two (ntimes c c0) c1
| dnf d1 ⇒ dis (ntimes c c0) d1
end
end.
```

```
Definition distrib1 (c : CNF)(e : ENF) : ENF :=
match e with
| cnf a ⇒ cnf (ntimes c a)
| dnf b ⇒ distrib0 c b
end.
```

```
Fixpoint explog0 (d : Base)(d2 : DNF) {struct d2} : CNF :=
match d2 with
| two c1 c2 ⇒ ntimes (con c1 d top) (con c2 d top)
| dis c d3 ⇒ ntimes (con c d top) (explog0 d d3)
end.
```

```
Definition explog1 (d : Base)(e : ENF) : CNF :=
match e with
| cnf c ⇒ con c d top
| dnf d1 ⇒ explog0 d d1
end.
```

```
Fixpoint distribn (d : DNF)(e2 : ENF) {struct d} : ENF :=
match d with
| two c c0 ⇒ dnf (nplus (distrib1 c e2) (distrib1 c0 e2))
| dis c d0 ⇒ dnf (nplus (distrib1 c e2) (distribn d0 e2))
end.
```

```
Definition distrib (e1 e2 : ENF) : ENF :=
match e1 with
| cnf a ⇒ distrib1 a e2
| dnf b ⇒ distribn b e2
end.
```

```
Fixpoint explogn (c : CNF)(e2 : ENF) {struct c} : CNF :=
match c with
| top ⇒ top
| con c1 d c2 ⇒
ntimes (explog1 d (distrib1 c1 e2)) (explogn c2 e2)
end.
```

```
Definition p2c : Proposition → CNF :=
fun p ⇒ con top (prp p) top.
```

```
Definition b2c : Base → CNF :=
fun b ⇒
match b with
| prp p ⇒ p2c p
| bd d ⇒ con top (bd d) top
end.
```

```
Fixpoint enf2cnf (e : ENF) {struct e} : CNF :=
match e with
| cnf c ⇒ c
| dnf d ⇒ b2c (bd d)
end.
```

From the inductive characterization of the previous theorem, it is immediate to notice that the exp-log normal form (ENF) is in fact a combination of disjunctive- (DNF) and conjunctive normal forms

(CNF), and their extension to also cover the function type. We shall now apply this simple and loss-less transformation of types to the equational theory of terms of the lambda calculus with sums.

### 3. $\beta\eta$ -Congruence classes at ENF type

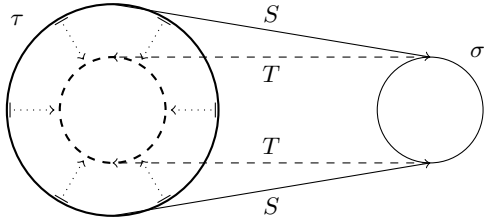
The virtue of type isomorphisms is that they preserve the equational theory of the term calculus: an isomorphism between  $\tau$  and  $\sigma$  is witnessed by a pair of lambda terms

$$T : \sigma \rightarrow \tau \quad \text{and} \quad S : \tau \rightarrow \sigma$$

such that

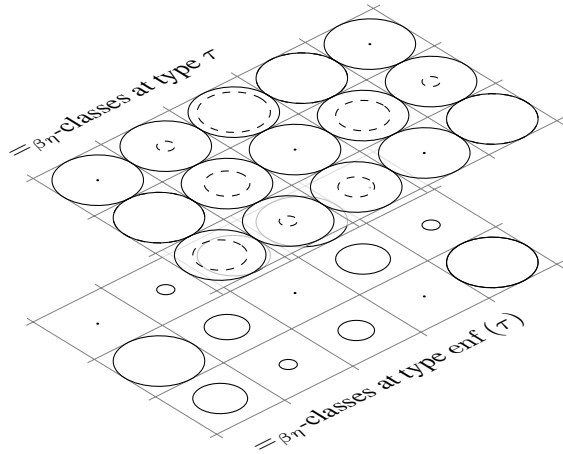
$$\lambda x.T(Sx) =_{\beta\eta} \lambda x.x \quad \text{and} \quad \lambda y.S(Ty) =_{\beta\eta} \lambda y.y.$$

Therefore, when  $\tau \cong \sigma$ , and  $\sigma$  happens to be more canonical than  $\tau$ —in the sense that to any  $\beta\eta$ -equivalence class of type  $\tau$  corresponds a smaller one of type  $\sigma$ —one can reduce the problem of deciding  $\beta\eta$ -equality at  $\tau$  to deciding it for a smaller subclass of terms.



In the case when  $\sigma = \text{enf}(\tau)$ , the equivalence classes at type  $\sigma$  will not be larger than their original classes at  $\tau$ , since the main effect of the reduction to exp-log normal form is to get rid of as many sum types on the left of an arrow as possible, and it is known that for the  $\{\times, \rightarrow\}$ -typed lambda calculus one can choose a single canonical  $\eta$ -long  $\beta$ -normal representative out of a class of  $\beta\eta$ -equal terms.

Thus, from the perspective of type isomorphisms, we can observe the partition of the set of terms of type  $\tau$  into  $=_{\beta\eta}$ -congruence classes as projected upon different parallel planes in three dimensional space, one plane for each type isomorphic to  $\tau$ . If we choose to observe the planes for  $\tau$  and  $\text{enf}(\tau)$ , we may describe the situation by the following figure.



The dashed circle depicts the compaction, if any, of a congruence class achieved by coercing to ENF type. The single point depicts the compaction to a singleton set, the case where a unique canonical representative of a class of  $\beta\eta$ -terms exists.

We do not claim that the plane of  $\text{enf}(\tau)$  is *always* the best possible plane to choose for deciding  $=_{\beta\eta}$ . Indeed, for concrete base types there may well be further type isomorphisms to apply

(think of the role of the unit type 1 in  $(1 \rightarrow \tau + \sigma) \rightarrow \rho$ ) and hence a better plane than the one for  $\text{enf}(\tau)$ . However, it is a reasonably good default choice.

For the cases of types where the sum can be completely eliminated, such as the two examples of Section 1, the projection amounts to compacting the  $\beta\eta$ -congruence class to a single point, a canonical normal term of type  $\text{enf}(\tau)$ .

Assuming  $\tau, \sigma, \tau_i$  are base types, the canonical representatives for the two  $\beta\eta$ -congruence classes of Section 1 are

$$\langle \lambda x.(\pi_1(\pi_2 x))(\pi_1 x), \lambda x.(\pi_2(\pi_2 x))(\pi_1 x) \rangle$$

and

$$\langle \lambda x.(\pi_1 x)((\pi_1 \pi_2 x)(\pi_1 \pi_2 \pi_2 x)), \lambda x.(\pi_1 x)((\pi_1 \pi_2 x)(\pi_1 \pi_2 \pi_2 x)) \rangle.$$

Note that, unlike (Balat et al. 2004), we do not need *any* sophisticated term analysis to derive a canonical form in this kind of cases. One may either apply the standard terms witnessing the isomorphisms by hand, or use our normalizer described in Section 4.

The natural place to pick a canonical representative is thus the  $\beta\eta$ -congruence class of terms at the normal type, not the class at the original type! Moreover, beware that even if it may be tempting to map a canonical representative along isomorphic coercions back to the original type, the obtained representative may not be truly canonical since there is generally more than one way to specify the terms  $S$  and  $T$  that witness a type isomorphism.

Of course, not always can all sum types be eliminated by type isomorphism, and hence not always can a class be compacted to a single point in that way. Nevertheless, even in the case where there are still sums remaining in the type of a term, the ENF simplifies the set of applicable  $=_{\beta\eta}$ -axioms.

We can use it to get a restricted set of equations,  $=_{\beta\eta}^e$ , shown in Figure 2, which is still complete for proving full  $\beta\eta$ -equality, as made precise in the following theorem.

**Theorem 2.** *Let  $P, Q$  be terms of type  $\tau$  and let  $S : \tau \rightarrow \text{enf}(\tau), T : \text{enf}(\tau) \rightarrow \tau$  be a witnessing pair of terms for the isomorphism  $\tau \cong \text{enf}(\tau)$ . Then,  $P =_{\beta\eta} Q$  if and only if  $SP =_{\beta\eta}^e SQ$  and if and only if  $T(SP) =_{\beta\eta} T(SQ)$ .*

*Proof.* Since the set of terms of ENF type is a subset of all typable terms, it suffices to show that all  $=_{\beta\eta}$ -equations that apply to terms of ENF type can be derived already by the  $=_{\beta\eta}^e$ -equations.

Notice first that  $\eta_x^e$  and  $\eta_\pi^e$  are special cases of  $\eta_+$ , so, in fact, the only axiom missing from  $=_{\beta\eta}^e$  is  $\eta_+$  itself,

$$N\{M/x\} =_{\eta}^e \delta(M, x_1.N\{\iota_1 x_1/x\}, x_2.N\{\iota_2 x_2/x\}) \\ (x_1, x_2 \notin \text{FV}(N)),$$

when  $N$  is of type  $c$ ; the case of  $N$  of type  $d$  is covered directly by the  $\eta_+^e$ -axiom. We thus show that the  $\eta_+$ -axiom is derivable from the  $=_{\beta\eta}^e$ -ones by induction on  $c$ .

**Case for  $N$  of type  $(c \rightarrow b) \times c_0$ .**

$$\begin{aligned} & N\{M/x\} \\ &=_{\eta}^e \langle \pi_1(N\{M/x\}), \pi_2(N\{M/x\}) \rangle \quad \text{by } \eta_x^e \\ &= \langle (\pi_1 N)\{M/x\}, (\pi_2 N)\{M/x\} \rangle \\ &=_{\eta}^e \langle \delta(M, x_1.(\pi_1 N)\{\iota_1 x_1/x\}, x_2.(\pi_1 N)\{\iota_2 x_2/x\}), \\ &\quad \delta(M, x_1.(\pi_2 N)\{\iota_1 x_1/x\}, x_2.(\pi_2 N)\{\iota_2 x_2/x\}) \rangle \quad \text{by IH} \\ &=_{\beta\eta}^e \langle \pi_1(\delta(M, x_1.N\{\iota_1 x_1/x\}, x_2.N\{\iota_2 x_2/x\})), \\ &\quad \pi_2(\delta(M, x_1.N\{\iota_1 x_1/x\}, x_2.N\{\iota_2 x_2/x\})) \rangle \quad \text{by } \eta_\pi^e \\ &=_{\eta}^e \delta(M, x_1.N\{\iota_1 x_1/x\}, x_2.N\{\iota_2 x_2/x\}) \quad \text{by } \eta_x^e \end{aligned}$$

$$\begin{aligned}
M, N ::= & x^e \mid (M^{c \rightarrow b} N^c)^b \mid (\pi_1 M^{(c \rightarrow b) \times c_0})^{c \rightarrow b} \mid (\pi_2 M^{(c \rightarrow b) \times c_0})^{c_0} \mid \delta(M^{c+d}, x_1^c.N_1^e, x_2^d.N_2^e)^e \\
& \mid (\lambda x^c.M^b)^{c \rightarrow b} \mid \langle M^{b \rightarrow c}, N^{c_0} \rangle^{(b \rightarrow c) \times c_0} \mid (\iota_1 M^c)^{c+d} \mid (\iota_2 M^d)^{c+d} \\
& (\lambda x^c.N^b)M =_{\beta}^e N\{M/x\} & (\beta_{\rightarrow}^e) \\
& \pi_i \langle M_1^{b \rightarrow c}, M_2^{c_0} \rangle =_{\beta}^e M_i & (\beta_{\times}^e) \\
& \delta(\iota_i M, x_1.N_1, x_2.N_2)^e =_{\beta}^e N_i\{M/x_i\} & (\beta_{+}^e) \\
& N^{c \rightarrow b} =_{\eta}^e \lambda x.Nx & x \notin \text{FV}(N) \quad (\eta_{\rightarrow}^e) \\
& N^{(c \rightarrow b) \times c_0} =_{\eta}^e \langle \pi_1 N, \pi_2 N \rangle & (\eta_{\times}^e) \\
& N^b\{M^d/x\} =_{\eta}^e \delta(M, x_1.N\{\iota_1 x_1/x\}, x_2.N\{\iota_2 x_2/x\})^b & x_1, x_2 \notin \text{FV}(N) \quad (\eta_{+}^e) \\
& \pi_i \delta(M, x_1.N_1, x_2.N_2) =_{\eta}^e \delta(M, x_1.\pi_i N_1, x_2.\pi_i N_2)^c & (\eta_{\pi}^e) \\
& \lambda y.\delta(M, x_1.N_1, x_2.N_2) =_{\eta}^e \delta(M, x_1.\lambda y.N_1, x_2.\lambda y.N_2)^{c \rightarrow b} & y \notin \text{FV}(M) \quad (\eta_{\lambda}^e)
\end{aligned}$$

**Figure 2.** Lambda terms of ENF type and the equational theory  $=_{\beta\eta}^e$ .

**Case for  $N$  of type  $c \rightarrow b$ .**

$$\begin{aligned}
& N\{M/x\} \\
& =_{\eta}^e \lambda y.(N\{M/x\})y \quad \text{by } \eta_{\rightarrow}^e \\
& = \lambda y.(Ny)\{M/x\} \quad \text{for } y \notin \text{FV}(N\{M/x\}) \\
& =_{\eta}^e \lambda y.\delta(M, x_1.(Ny)\{\iota_1 x_1/x\}, x_2.(Ny)\{\iota_2 x_2/x\}) \quad \text{by } \eta_{+}^e \\
& =_{\eta}^e \delta(M, x_1.(\lambda y.Ny)\{\iota_1 x_1/x\}, x_2.(\lambda y.Ny)\{\iota_2 x_2/x\}) \quad \text{by } \eta_{\lambda}^e \\
& =_{\eta}^e \delta(M, x_1.N\{\iota_1 x_1/x\}, x_2.N\{\iota_2 x_2/x\}) \quad \text{by } \eta_{\rightarrow}^e
\end{aligned}$$

□

The transformation of terms to ENF type thus allows to simplify the (up to now) standard axioms of  $=_{\beta\eta}$ . The new axioms are complete for  $=_{\beta\eta}$  in spite of them being only *special cases* of the old ones. A notable feature is that we get to disentangle the left-hand side and right-hand side of the equality axioms: for instance, the right-hand side of  $\beta_{\rightarrow}$ -axiom can no longer overlap with the left-hand side of the  $\eta_{+}$ -axiom, due to typing restrictions on the term  $M$ .

One could get rid of  $\eta_{\pi}^e$  and  $\eta_{\lambda}^e$  if one had a version of  $\lambda$ -calculus resistant to these permuting conversions. The syntax of such a lambda calculus would further be simplified if, instead of binary, one had  $n$ -ary sums and products. In that case, there would be no need for variables of sum type at all (currently they can only be introduced by the second branch of  $\delta$ ). We would in fact get a calculus with only variables of type  $c \rightarrow b$ , and that would still be suitable as a small theoretical core of functional programming languages.

#### 4. A compact representation of terms at ENF type

It is the subject of this section to show that the desiderata for a more canonical calculus from the previous paragraph can in fact be achieved. We shall define a new representation of lambda terms, that we have isolated as the most compact syntax possible during the formal Coq development of a normalizer of terms at ENF type. The description of the normalizer itself will be left for the second part of this section, Subsection 4.1. In the first part of the section, we shall demonstrate the value of representing terms in our calculus on a number of examples. Comparing our normal form for *syntactical identity* provides a first such heuristic for deciding  $=_{\beta\eta}$  in the presence of sums.

Before we continue with the presentation of the new calculus, for the sake of precision, we give the formal representation of

terms of the two term calculi. First, we represent the usual lambda calculus with sums.

**Inductive ND** : list Formula  $\rightarrow$  Formula  $\rightarrow$  Set :=

- | hyp :  $\forall \{Gamma A\}$ ,  
ND (A :: Gamma) A
- | wkn :  $\forall \{Gamma A B\}$ ,  
ND Gamma A  $\rightarrow$  ND (B :: Gamma) A
- | lam :  $\forall \{Gamma A B\}$ ,  
ND (A :: Gamma) B  $\rightarrow$  ND Gamma (impl A B)
- | app :  $\forall \{Gamma A B\}$ ,  
ND Gamma (impl A B)  $\rightarrow$  ND Gamma A  $\rightarrow$  ND Gamma B
- | pair :  $\forall \{Gamma A B\}$ ,  
ND Gamma A  $\rightarrow$  ND Gamma B  $\rightarrow$  ND Gamma (conj A B)
- | fst :  $\forall \{Gamma A B\}$ ,  
ND Gamma (conj A B)  $\rightarrow$  ND Gamma A
- | snd :  $\forall \{Gamma A B\}$ ,  
ND Gamma (conj A B)  $\rightarrow$  ND Gamma B
- | inl :  $\forall \{Gamma A B\}$ ,  
ND Gamma A  $\rightarrow$  ND Gamma (disj A B)
- | inr :  $\forall \{Gamma A B\}$ ,  
ND Gamma B  $\rightarrow$  ND Gamma (disj A B)
- | cas :  $\forall \{Gamma A B C\}$ ,  
ND Gamma (disj A B)  $\rightarrow$   
ND (A :: Gamma) C  $\rightarrow$  ND (B :: Gamma) C  $\rightarrow$   
ND Gamma C.

The constructors are self-explanatory, except for *hyp* and *wkn*, which are in fact used to denote de Bruijn indices: *hyp* denotes 0, while *wkn* is the successor. For instance, the term  $\lambda xyz.y$  is represented as *lam (lam (lam (wkn hyp)))* i.e. *lam (lam (lam 1))*.

DeBruijn indices creep in as the simplest way to work with binders in Coq, and although they may reduce readability, they solve the problem with  $\alpha$ -conversion of terms.

Next is our compact representation of terms, defined by the following simultaneous inductive definition of terms at base type (*HSb*), together with terms at product type (*HSc*). These later are simply finite lists of *HSb*-terms.

**Inductive HSc** : CNF  $\rightarrow$  Set :=

- | tt : HSc top
- | pair :  $\forall \{c1 b c2\}$ , HSc c1 b  $\rightarrow$  HSc c2  $\rightarrow$  HSc (con c1 b c2)

```

with HSb : CNF → Base → Set :=
| app : ∀ {p c0 c1 c2},
  HSc (explogn c1 (cnf (ntimes c2 (con c1 (prp p) c0)))) →
  HSb (ntimes c2 (con c1 (prp p) c0)) (prp p)
| cas : ∀ {d b c0 c1 c2 c3},
  HSc (explogn c1 (cnf (ntimes c2 (con c1 (bd d) c0)))) →
  HSc (explogn (explog0 b d)
    (cnf (ntimes c3 (ntimes c2 (con c1 (bd d) c0)))))) →
  HSb (ntimes c3 (ntimes c2 (con c1 (bd d) c0))) b
| wkn : ∀ {c0 c1 b1 b},
  HSb c0 b → HSb (con c1 b1 c0) b
| inl_two : ∀ {c0 c1 c2},
  HSc (explogn c1 (cnf c0)) → HSb c0 (bd (two c1 c2))
| inr_two : ∀ {c0 c1 c2},
  HSc (explogn c2 (cnf c0)) → HSb c0 (bd (two c1 c2))
| inl_dis : ∀ {c0 c d},
  HSc (explogn c (cnf c0)) → HSb c0 (bd (dis c d))
| inr_dis : ∀ {c0 c d},
  HSb c0 (bd d) → HSb c0 (bd (dis c d)).

```

For a more human-readable notation of our calculus, we are going to use the following one,

$$P, Q ::= \langle M_1, \dots, M_n \rangle \quad (n \geq 0)$$

$$M, M_i ::= x_n P \mid \delta(x_n P, Q) \mid wM \mid \iota_1 P \mid \iota_2 P \mid \iota'_1 P \mid \iota'_2 M,$$

with typing rules as follows:

$$\frac{M_1 : (c_1 \vdash b_1) \quad \dots \quad M_n : (c_n \vdash b_n)}{\langle M_1, \dots, M_n \rangle : (c_1 \rightarrow b_1) \times \dots \times (c_n \rightarrow b_n)}$$

$$\frac{P : (c_2 \times (c_1 \rightarrow p) \times c_0 \Rightarrow c_1)}{x_n P : (c_2 \times (c_1 \rightarrow p) \times c_0 \vdash p)}$$

$$\frac{P : (c_2 \times (c_1 \rightarrow d) \times c_0 \Rightarrow c_1) \quad Q : (c_3 \times c_2 \times (c_1 \rightarrow d) \times c_0 \Rightarrow (d \rightrightarrows b))}{\delta(x_n P, Q) : (c_3 \times c_2 \times (c_1 \rightarrow d) \times c_0 \vdash b)}$$

$$\frac{M : (c_0 \vdash b)}{wM : ((c_1 \rightarrow b_1) \times c_0 \vdash b)}$$

$$\frac{P : (c_0 \Rightarrow c_1)}{\iota_1 P : (c_0 \vdash c_1 + c_2)} \quad \frac{P : (c_0 \Rightarrow c_2)}{\iota_2 P : (c_0 \vdash c_1 + c_2)}$$

$$\frac{P : (c_0 \Rightarrow c)}{\iota'_1 P : (c_0 \vdash c + d)} \quad \frac{M : (c_0 \vdash d)}{\iota'_2 M : (c_0 \vdash c + d)}.$$

The typing rules above involve two kinds of typing judgments.

**Judgments at base type:** Denoted  $M : (c \vdash b)$ , this is the main judgment kind, the conclusion of all but the first typing rule. It says that  $M$  is a term of type  $b$  (i.e. either an atomic  $p$  or a disjunction type  $d$ ) in the typing context  $c$ . This context  $c$  takes over the place of the usual context  $\Gamma$  and allows only hypotheses (variables) of type  $c_i \rightarrow b_i$  to be used inside the term  $M$ .

**Judgments at product type:** Denoted  $P : c$  or  $Q : c$ , this kind of judgment is only the conclusion of the first typing rule, whose sole purpose is to make a tuple of base type judgments.

However, the judgments at product type are used as *hypotheses* in the other typing rules, where their role is to allow  $n$  premises to the typing rule. For this usage, they are disguised as the macro-expansions  $c_1 \Rightarrow c_2$  or  $c \Rightarrow (d \rightrightarrows b)$ . Implemented by the Coq fixpoints *explogn* and *explogl*, these macro expansions

work as follows:

$$c_0 \Rightarrow (c_1 \rightarrow b_1) \times \dots \times (c_n \rightarrow b_n) \equiv (c_1 \times c_0 \rightarrow b_1) \times \dots \times (c_n \times c_0 \rightarrow b_n)$$

$$(c_1 + \dots + c_n) \rightrightarrows b \equiv (c_1 \rightarrow b) \times \dots \times (c_n \rightarrow b)$$

Note that the usage of  $d \rightrightarrows b$  in the typing rule for  $\delta$  allows the number of premises contained in  $Q$  to be determined by the size of the sum  $d$ .

The typing rules also rely on a implicit variable convention, where a variable  $x_n$  actually denotes the variable whose deBruijn index is  $n$  (we start counting from 0).

**Variables as deBruijn indices:** The variable  $x_n$  in the rules for  $x_n P$  and  $\delta(x_n P, Q)$  represent the hypothesis  $c_1 \rightarrow p$  and  $c_1 \rightarrow d$ . For concrete  $c_2, c_3$ , the subscript  $n$  means that the variable represents the  $n$ -th hypothesis of the form  $c \rightarrow b$ , counting from left to right and starting from 0, in the context of the term  $P$ , or the  $n + 1$ -st, in the context of the term  $Q$ .

We shall motivate our syntax in comparison to the syntax of the lambda calculus from Figure 2, by considering in order all term constructors of the later.

$x^e$  Since  $e \in \text{ENF}$ , either  $e = c \in \text{CNF}$  or  $e = d \in \text{DNF}$ . Variables of type  $d$  only appear as binders in the second branch of  $\delta$ , so if we have  $n$ -ary instead of binary  $\delta$ 's, the only type a variable  $x$  could have will be a  $c$ . But, since  $c$  is always of the form  $(c_1 \rightarrow b_1) \times \dots \times (c_n \rightarrow b_n)$ , a variable  $x^c$  could be written as a tuple of  $n$  variables  $x_i$  of types  $c_i \rightarrow b_i$ . Moreover, as we want our terms to always be  $\eta$ -expanded, and  $c_i \rightarrow b_i$  is an arrow type, we will not have a separate syntactic category of terms for variables  $x_i$  in the new calculus, but they will rather be encoded/merged with either the category of applications  $x_i P$  (when  $b$  is an atomic type  $p$ ), or the category of case analysis  $\delta(x_i P, Q)$  (when  $b \in \text{DNF}$ ), the two new constructors explained below.

$M^{c \rightarrow b} N^c$  We shall only need this term constructor at type  $b = p$ , since if  $b \in \text{DNF}$  the term  $MN$  would not be  $\eta$ -long (we want it to be represented by a  $\delta(MN, \dots)$ ). As we realized during our Coq development, we shall only need the case  $M = x$ , as there will be no other syntactic element of type  $c \rightarrow p$  (there will be no projections  $\pi_i$  left, while the  $\delta$  will only be necessary at type  $b$ ). In particular, the application  $x \langle \rangle$  can be used to represent the old category of variables, where  $\langle \rangle$  is the empty tuple of unit type 1 (the nullary product).

$(\pi_1 M^{(c \rightarrow b) \times c_0})^{c \rightarrow b}$  If  $M$  is  $\eta$ -expanded, as we want all terms to be, this term would only create a  $\beta$ -redex, and so will not be a part of the new syntax, as we are building a syntax for  $\beta$ -normal and  $\eta$ -long terms.

$(\pi_2 M^{(c \rightarrow b) \times c_0})^{c_0}$  When product types are represented as  $n$ -ary, the same reasoning as for  $\pi_1$  applies, so  $\pi_2$  will not be part of the new syntax.

$\delta(M^{c+d}, x_1^c.N_1^e, x_2^d.N_2^e)^e$  This constructor is only needed at the type  $e = b$ , a consequence of the fact that the  $\eta_+^e$ -axiom is specialized to type  $b$ : the axioms  $\eta_\pi^e$  and  $\eta_\lambda^e$  will not be expressible in the new syntax, since it will not contain  $\pi_i$ , as we saw, and it will not contain  $\lambda$ , as we shall see. We will also only need the scrutinee  $M$  to be of the form  $xN$ , like it the case of application; this additional restriction was not possible to see upfront, but only once we used Coq to analyze the terms needed for the normalizer.

The new constructor  $\delta(x_n P, Q)$  is thus like the old  $\delta(x_n P, \dots)$ , except that  $Q$  regroups in the form of an  $n$ -ary tuple all the



possible branches of the pattern matching (sum types will also be  $n$ -ary, not binary like before).

$(\lambda x^c.M^b)^{c \rightarrow b}$  This terms constructor is already severely restricted (for instance only one variable  $x$  can be abstracted), thanks to the restrictions on the left- and right-hand sides of the function type. But, as we found out during the Coq development, somewhat to our surprise, there is no need for  $\lambda$ -abstraction in our syntax. When reverse-normalizing from our calculus to the standard lambda calculus (see the six examples below),  $\lambda$ 's can be reconstructed thanks to the typing information.

$(M^{b \rightarrow c}, N^{c_0})^{(b \rightarrow c) \times c_0}$  This constructor will be maintained, corresponding to the only typing rule with conclusion a judgment of product type, but it will become  $n$ -ary,  $\langle M_1, \dots, M_n \rangle$ . In particular, we may have the *nullary* tuple  $\langle \rangle$  of the null product type (i.e. unit type 1).

$(\iota_1 M^c)^{c+d}, (\iota_2 M^d)^{c+d}$  These constructors will be maintained, but will be *duplicated*: the new  $\iota_1, \iota_2$  will only be used to construct a binary sum  $c_1 + c_2$  (this is the base case of sum constructors which must be at least binary by construction), while the  $\iota'_1, \iota'_2$  will be used to construct sums of the form  $c + d$ .

We shall now show a number of examples that our compact term representation manages to represent canonically. We will also show cases when  $\beta\eta$ -equality can *not* be decided using bringing terms to the compact normal form. For simplicity, all type variables ( $a, b, c, d, e, f, g, p, q, r, s, i, j, k, l$ ) are assumed to be of atomic type, none of them denoting members of *Base*, *CNF*, and *DNF*, anymore, and for the rest of this subsection.

**Convention 1.** We shall adopt the convention of writing the type  $1 \rightarrow p$  as  $p$  (1 is the unit type i.e. the nullary product type), writing the application to a nullary pair  $x_n \langle \rangle$  as  $x_n$ , and writing a *singleton* pair  $\langle M \rangle$  as just  $M$ . Hence, for instance, an application of some term  $M$  to a singleton pair, containing an application of a term  $N$  to a nullary pair,  $M \langle N \langle \rangle \rangle$ , will be written as the more readable  $MN$  corresponding to the usual  $\lambda$ -calculus intuitions.

**Example 1.** This is the first example from the introduction, concerning the relative positions of  $\lambda$ 's,  $\delta$ 's, and applications. The  $\beta\eta$ -equal terms

$$\lambda x.\lambda y.y\delta(x, z.\iota_1 z, z.\iota_2 z) \quad (8)$$

$$\lambda x.\lambda y.\delta(x, z.y(\iota_1 z), z.y(\iota_2 z)) \quad (9)$$

$$\lambda x.\delta(x, z.\lambda y.y(\iota_1 z), z.\lambda y.y(\iota_2 z)) \quad (10)$$

$$\lambda x.\lambda y.yx \quad (11)$$

at type

$$(p + q) \rightarrow ((p + q) \rightarrow r) \rightarrow r,$$

are all normalized to the same canonical representation

$$\langle x_0 x_2, x_1 x_2 \rangle \quad (12)$$

at the ENF type

$$\begin{aligned} & ((p \rightarrow r) \times (q \rightarrow r) \times p \rightarrow r) \times \\ & ((p \rightarrow r) \times (q \rightarrow r) \times q \rightarrow r) \end{aligned}$$

which can be reverse-normalized back to (9). However, the point is not that (9) is somehow better than the other 3 terms, but that a canonical representation should be sought at the ENF type, not the original type! This remark is valid in general, and in particular for the other examples below.

**Example 2.** This is the second example from the introduction (Example 6.2.4.2 from (Balat et al. 2004)). The  $\beta\eta$ -equal terms

$$\lambda x y z u. x(yz) \quad (13)$$

$$\lambda x y z u. \delta(u, x_1.x(yz), x_2.x(yz)) \quad (14)$$

$$\begin{aligned} & \lambda x y z u. \delta(u, x_1.\delta(\iota_1 z, y_1.x(y y_1), y_2.x y_2), \\ & x_2.\delta(\iota_2 y z, y_1.x(y y_1), y_2.x y_2)) \end{aligned} \quad (15)$$

$$\lambda x y z u. \delta(\delta(u, x_1.\iota_1 z, x_2.\iota_2(yz)), y_1.x(y y_1), y_2.x y_2) \quad (16)$$

at type

$$(a \rightarrow b) \rightarrow (c \rightarrow a) \rightarrow c \rightarrow (d + e) \rightarrow b,$$

are all normalized to the compact term

$$\langle x_3(x_2 x_1), x_3(x_2 x_1) \rangle \quad (17)$$

at the ENF type

$$\begin{aligned} & (d \times c \times (c \rightarrow a) \times (a \rightarrow b) \rightarrow b) \times \\ & (e \times c \times (c \rightarrow a) \times (a \rightarrow b) \rightarrow b), \end{aligned}$$

which can then be reverse-normalized to (14), if desired.

A reviewer once remarked that the two previous examples can be handled just by a CPS transformation. While our implementation *will* be based on continuations, the reason why these examples are handled by our method are not continuations, but rather the fact that all sum types can be eliminated, allowing us to choose a canonical term in the compact representation of the  $\{\rightarrow, \times\}$ -typed lambda calculus.

**Example 3** (Commuting conversions). The left and right hand sides of the common commuting conversions,

$$\begin{aligned} & \lambda x y z u. \delta(u, v_1.y v_1, v_2.z v_2) x =_{\beta\eta} \\ & =_{\beta\eta} \lambda x y z u. \delta(u, v_1.(y v_1) x, v_2.(z v_2) x), \end{aligned} \quad (18)$$

$$\begin{aligned} & \lambda x y z u v. \delta(\delta(x, x_1.y x_1, x_2.z x_2), w_1.u w_1, w_2.v w_2) =_{\beta\eta} \\ & =_{\beta\eta} \lambda x y z u v. \delta(x, x_1.\delta(y x_1, w_1.u w_1, w_2.v w_2), \\ & x_2.\delta(z x_2, w_1.u w_1, w_2.v w_2)) \end{aligned} \quad (19)$$

of types

$$s \rightarrow (p \rightarrow s \rightarrow r) \rightarrow (q \rightarrow s \rightarrow r) \rightarrow (p + q) \rightarrow r$$

and

$$\begin{aligned} & (p + q) \rightarrow (p \rightarrow r + s) \rightarrow (q \rightarrow r + s) \rightarrow \\ & (r \rightarrow a) \rightarrow (s \rightarrow a) \rightarrow a, \end{aligned}$$

are normalized to the compact terms

$$\langle x_2 \langle x_3, x_0 \rangle, x_1 \langle x_3, x_0 \rangle \rangle, \quad (18')$$

of ENF type

$$\begin{aligned} & (p \times (s \times q \rightarrow r) \times (s \times p \rightarrow r) \times s \rightarrow r) \times \\ & (q \times (s \times q \rightarrow r) \times (s \times p \rightarrow r) \times s \rightarrow r) \end{aligned}$$

and

$$\langle \delta(x_3 x_4, \langle x_2 x_0, x_1 x_0 \rangle), \delta(x_2 x_4, \langle x_2 x_0, x_1 x_0 \rangle) \rangle, \quad (19')$$

of ENF type

$$\begin{aligned} & ((s \rightarrow a) \times (r \rightarrow a) \times (q \rightarrow r + s) \times (p \rightarrow r + s) \times p \rightarrow a) \\ & \times ((s \rightarrow a) \times (r \rightarrow a) \times (q \rightarrow r + s) \times (p \rightarrow r + s) \times q \rightarrow a), \end{aligned}$$

which can be reverse-normalized to the right-hand sides of (18), and (19), respectively, if desired.

**Example 4** (Eta equations). Both the left- and the right-hand sides of the eta rules (represented as closed terms),

$$\lambda x.x =_{\beta\eta} \lambda xy.xy \quad (20)$$

$$\lambda x.x =_{\beta\eta} \lambda x.\langle \pi_1 x, \pi_2 x \rangle \quad (21)$$

$$\lambda xy.xy =_{\beta\eta} \lambda xy.\delta(y, x_1.x(\iota_1 x_1), x_2.x(\iota_2 x_2)) \quad (22)$$

$$\lambda xyz.\delta(z, z_1.\lambda u.xz_1, z_2.\lambda u.yz_2) =_{\beta\eta} \lambda xyzu.\delta(z, z_1.xz_1, z_2.yz_2) \quad (23)$$

$$\lambda xyz.\pi_1 \delta(z, z_1.xz_1, z_2.yz_2) =_{\beta\eta} \lambda xyz.\delta(z, z_1.\pi_1 xz_1, z_2.\pi_1 yz_2) \quad (24)$$

$$\lambda xyz.\pi_2 \delta(z, z_1.xz_1, z_2.yz_2) =_{\beta\eta} \lambda xyz.\delta(z, z_1.\pi_2 xz_1, z_2.\pi_2 yz_2) \quad (25)$$

of types

$$(p \rightarrow q) \rightarrow (p \rightarrow q) \quad (20)$$

$$(p \times q) \rightarrow (p \times q) \quad (21)$$

$$((p + q) \rightarrow r) \rightarrow ((p + q) \rightarrow r) \quad (22)$$

$$(p \rightarrow s) \rightarrow (q \rightarrow s) \rightarrow (p + q) \rightarrow r \rightarrow s \quad (23)$$

$$(p \rightarrow s \times r) \rightarrow (q \rightarrow s \times r) \rightarrow (p + q) \rightarrow s \quad (24)$$

$$(p \rightarrow s \times r) \rightarrow (q \rightarrow s \times r) \rightarrow (p + q) \rightarrow r \quad (25)$$

are mapped to the same compact term

$$x_1 x_0 \quad (20')$$

$$\langle x_0, x_1 \rangle \quad (21')$$

$$\langle x_1 x_0, x_2 x_0 \rangle \quad (22')$$

$$\langle x_3 x_1, x_2 x_1 \rangle \quad (23')$$

$$\langle x_3 x_0, x_1 x_0 \rangle \quad (24')$$

$$\langle x_4 x_0, x_2 x_0 \rangle, \quad (25')$$

of ENF types

$$p \times (p \rightarrow p) \rightarrow p \quad (20')$$

$$(p \times q \rightarrow p) \times (p \times q \rightarrow q) \quad (21')$$

$$(p \times (p \rightarrow r) \times (q \rightarrow r) \rightarrow r) \times \quad (22')$$

$$(q \times (p \rightarrow r) \times (q \rightarrow r) \rightarrow r) \quad (22')$$

$$(r \times p \times (q \rightarrow s) \times (p \rightarrow s) \rightarrow s) \times \quad (23')$$

$$(r \times q \times (q \rightarrow s) \times (p \rightarrow s) \rightarrow s) \quad (23')$$

$$(p \times (q \rightarrow s) \times (q \rightarrow r) \times (p \rightarrow s) \times (p \rightarrow r) \rightarrow s) \times \quad (24')$$

$$(q \times (q \rightarrow s) \times (q \rightarrow r) \times (p \rightarrow s) \times (p \rightarrow r) \rightarrow s) \quad (24')$$

$$(p \times (q \rightarrow s) \times (q \rightarrow r) \times (p \rightarrow s) \times (p \rightarrow r) \rightarrow r) \times \quad (25')$$

$$(q \times (q \rightarrow s) \times (q \rightarrow r) \times (p \rightarrow s) \times (p \rightarrow r) \rightarrow r), \quad (25')$$

and reverse-normalizing these compact terms produces always the right-hand side of the corresponding equation involving lambda terms.

Finally, as we shall see in the following two examples, our conversion to compact form does not guarantee a canonical representation for terms that are equal with respect to the strong forms of  $\beta\eta$ -equality used to duplicate subterms (Example 5) or change the order of case analysis of subterms (Example 6). Although such term transformations might not be desirable in the setting of real programming languages, for they change the order of evaluation, in a pure effect-free setting like a proof assistant, such transformation would be handy to have.

**Example 5.** The following  $\beta\eta$ -equal terms,

$$\lambda xyzu.\delta(uz, w.xw, w.yw) \quad (26)$$

$$\lambda xyzu.\delta(uz, w.\delta(uz, w'.xw', w'.yw'), w.yw), \quad (27)$$

of type

$$(f \rightarrow g) \rightarrow (h \rightarrow g) \rightarrow i \rightarrow (i \rightarrow f + h) \rightarrow g$$

are normalized to two *different* compact representations:

$$\delta(x_0 x_1, \langle x_4 x_0, x_3 x_0 \rangle) \quad (26')$$

$$\delta(x_0 x_1, \langle \delta(x_1 x_2, \langle x_5 x_0, x_4 x_0 \rangle), x_3 x_0 \rangle), \quad (27')$$

of ENF type

$$(i \rightarrow f + h) \times i \times (h \rightarrow g) \times (f \rightarrow g) \rightarrow g$$

which can then be reverse-normalized to the starting lambda terms themselves.

**Example 6.** The following  $\beta\eta$ -equal terms,

$$\lambda xyzuv.\delta(zv, x_1.\iota_1 x, x_2.\delta(uv, y_1.\iota_2 y, y_2.\iota_1 x)) \quad (28)$$

$$\lambda xyzuv.\delta(uv, y_1.\delta(zv, x_1.\iota_1 x, x_2.\iota_2 y), y_2.\iota_1 x), \quad (29)$$

of type

$$k \rightarrow l \rightarrow (f \rightarrow g + h) \rightarrow (f \rightarrow i + j) \rightarrow f \rightarrow k + l$$

are normalized to two *different* compact representations:

$$\langle \delta(x_2 x_0, \langle \iota_1 x_5, \delta(x_3 x_1, \langle \iota_2 x_5, \iota_1 x_6 \rangle) \rangle) \rangle \quad (28')$$

$$\langle \delta(x_1 x_0, \langle \delta(x_2 x_1, \langle \iota_1 x_6, \iota_2 x_5 \rangle), \iota_1 x_5 \rangle) \rangle, \quad (29')$$

of ENF type

$$f \times (f \rightarrow i + j) \times (f \rightarrow g + h) \times l \times k \rightarrow k + l$$

which can then be reverse-normalized to the starting lambda terms themselves.

**Comparison to the examples covered by the heuristic of (Balat et al. 2004)** In addition to Example 2 that was borrowed from (Balat et al. 2004), other examples that can be covered from that paper are examples 4.2.1–4.2.4 and Example 4.3.1. In these examples, not only are the input and the output of their TDPE represented uniquely, but also, in the cases when there are two distinct output normal forms according to Balat et al., our normalizer unifies the two normal forms into one, shown below:

$$x_0 \quad (4.2.1)$$

$$\langle \iota_1 x_0, \iota_2 x_0 \rangle \quad (4.2.2)$$

$$\langle \iota'_1 \langle x_0, x_1 \rangle, \iota'_2 \iota'_1 \langle x_0, x_1 \rangle, \iota'_2 \iota'_2 \iota_1 \langle x_0, x_1 \rangle, \iota'_2 \iota'_2 \iota_2 \langle x_0, x_1 \rangle \rangle \quad (4.2.3)$$

$$\langle \iota'_1 \langle x_1, x_0 \rangle, \iota'_2 \iota'_2 \iota_1 \langle x_1, x_0 \rangle, \iota'_2 \iota'_1 \langle x_1, x_0 \rangle, \iota'_2 \iota'_2 \iota_2 \langle x_1, x_0 \rangle \rangle \quad (4.2.4)$$

$$\langle \iota'_2 \iota'_2 \iota_1 x_1, \iota'_2 \iota'_1 x_1 \rangle. \quad (4.3.1)$$

Canonical representations could be obtained in these examples, because it was possible to represent the input and output terms in the fragment of the compact calculus which does not include  $\delta$ 's (although it still involves sum types).

On the other hand, there are also the examples where the input and output are *not* unified by our procedure.<sup>2</sup> Examples 4.3.2 and 4.3.3 are not handled because we do not permute the order of case analyses (as shown by our Example 6); Example 6.2.4.2 is not handled because we do not analyze if a subterm has been used twice in a term or not (as shown also by our Example 5); examples 4.3.4 and 4.4 are not even executable in our implementation, because we do not have a special treatment of the atomic empty type.

Of course, there is nothing stopping us from applying the program transformations that would allow to handle this kind of

<sup>2</sup>We shall not reproduce the compact representation for these examples in this paper, but they are available for inspection in the Coq formalization accompanying it.

cases—or nothing stopping Balat et al. from first applying our type-directed normalization procedure before performing their heuristic to unify the different normal outputs that they sometimes get. The point is that these two methodologies are orthogonal and they would ideally be used in combination inside a real-world application; for more comments about the two approaches, see Section 5.

#### 4.1 A converter for the compact term representation

In the remaining part of this section, we explain the high-level structure of our prototype normalizer of lambda terms into compact terms and vice versa. The full Coq implementation of the normalizer, together with the examples considered above, is given as a companion to this paper. This is only one possible implementation, using continuations, but all the previous material of this paper was written as generically as possible, so that it is useful if other implementation techniques are attempted in the future, such as rewriting based on evaluation contexts (i.e. the first-order reification of continuations), or abstract machines.

In a nutshell, our implementation is a type-directed partial evaluator, written in continuation-passing style, with an intermediate phase between the evaluation and reification phases, that allows to map a ‘semantic’ representation of a term from a type to its ENF type, and vice versa. Such partial evaluators can be implemented very elegantly, and with getting certain correctness properties for free, using the GADTs from Ocaml’s type system, as shown by the recent work of Danvy, Keller, and Puech (Danvy et al. 2015). Nevertheless, we had chosen to carry out our implementation in Coq, because that allowed us to perform a careful interactive analysis of the necessary normal forms—hence the compact calculus introduced in the first part of this section.

Usual type-directed partial evaluation (TDPE), aka normalization-by-evaluation (NBE), proceeds in two phases. First an evaluator is defined which takes the input term and obtains its semantic representation, and then a reifier is used to map the semantic representation into an output syntactic term. Our TDPE uses an *intermediate* phase between the two phases, a phase where type isomorphisms are applied to the semantic domain so that the narrowing down of a class of equal terms, described in Section 2, is performed on the semantic annotation of a term.

The semantics that we use is defined by a continuation monad over a *forcing structure*, together with *forcing fixpoints* that map the type of the input term into a type of the ambient type theory. The forcing structure is an abstract signature (Coq module type), requiring a set  $K$  of possible worlds, a preorder relation on worlds,  $le$ , an interpretation of atomic types,  $pforces$ , and  $X$ , the return type of the continuation monad.

```
Module Type ForcingStructure.
  Parameter K : Set.
  Parameter le : K → K → Set.
  Parameter pforces : K → Proposition → Set.
  Parameter Answer : Set.
  Parameter X : K → Answer → Set.
End ForcingStructure.
```

The continuation monad is polymorphic and instantiable by a forcing fixpoint  $f$  and a world  $w$ . It ensures that the preorder relation is respected; intuitively, this has to do with preserving the monotonicity of context free variables: we cannot ‘forget’ a free variable i.e. contexts cannot decrease.

```
Definition Cont {class:Set} (f:K→class→Set)(w:K)(x:class)
  := ∀ (x0:Answer), ∀ {w'}, le w w' →
```

$$(\forall \{w''\}, le w' w'' \rightarrow f w'' x \rightarrow X w'' x0) \rightarrow X w' x0.$$

Next, the necessary forcing fixpoints are defined:  $bforces$ ,  $cforces$ , and  $dforces$ , which are used to construct the type of the continuation monad corresponding to *Base*, *CNF*, and *DNF*, respectively;  $sforces$  is used for constructing the type of the continuation monad corresponding to non-normalized types.

```
Fixpoint bforces (w:K)(b:Base) {struct b} : Set :=
  match b with
  | prp p ⇒ pforces w p
  | bd d ⇒ dforces w d
  end
with cforces (w:K)(c:CNF) {struct c} : Set :=
  match c with
  | top ⇒ unit
  | con c1 b c2 ⇒
    (∀ w', le w w' → Cont cforces w' c1 → Cont bforces w'
    b)
    × (Cont cforces w c2)
  end
with dforces (w:K)(d:DNF) {struct d} : Set :=
  match d with
  | two c1 c2 ⇒ (Cont cforces w c1) + (Cont cforces w c2)
  | dis c1 d2 ⇒ (Cont cforces w c1) + (Cont dforces w d2)
  end.
Fixpoint eforces (w:K)(e:ENF) {struct e} : Set :=
  match e with
  | cnf c ⇒ cforces w c
  | dnf d ⇒ dforces w d
  end.
Fixpoint sforces (w:K)(F:Formula) {struct F} : Set :=
  match F with
  | prop p ⇒ pforces w p
  | disj F G ⇒ (Cont sforces w F) + (Cont sforces w G)
  | conj F G ⇒ (Cont sforces w F) × (Cont sforces w G)
  | impl F G ⇒ ∀ w',
    le w w' → (Cont sforces w' F) → (Cont sforces w' G)
  end.
```

Given these definitions, we can write an evaluator for compact terms, actually two simultaneously defined evaluators  $evalc$  and  $evalb$ , proceeding by induction on the input term.

```
Theorem evalc {c} : (HSc c → ∀ {w}, Cont cforces w c)
with evalb {b c0} : (HSb c0 b → ∀ {w},
  Cont cforces w c0 → Cont bforces w b).
```

An evaluator for usual lambda terms can also be defined, by induction on the input term. A helper function  $lforces$  analogous to the list map function for  $sforces$  is necessary.

```
Fixpoint
  lforces (w:K)(Gamma:list Formula) {struct Gamma} :
  Set :=
  match Gamma with
  | nil ⇒ unit
  | cons A Gamma0 ⇒
    Cont sforces w A × lforces w Gamma0
  end.
```

**Theorem eval**  $\{A \text{ Gamma}\} : ND \text{ Gamma } A \rightarrow \forall \{w\},$   
 $\text{!forces } w \text{ Gamma} \rightarrow \text{Cont sforces } w A.$

The novelty of our implementation (besides isolating the compact term calculus itself), in comparison to previous type-directed partial evaluators for the lambda calculus with sums, consists in showing that one can go back and forth between the semantic annotation at a type  $F$  and the semantic annotation of the normal form  $\text{enf}(F)$ . The proof of this statement needs a number of auxiliary lemmas that we do not mention in the paper. We actually prove two statements simultaneously,  $f2f$  and  $f2f'$ , declared as follows.

**Theorem f2f** :  
 $(\forall F, \forall w, \text{Cont sforces } w F \rightarrow \text{Cont eforces } w (\text{enf } F))$   
**with f2f'** :  
 $(\forall F, \forall w, \text{Cont eforces } w (\text{enf } F) \rightarrow \text{Cont sforces } w F).$

As one can see from their type signatures,  $f2f$  and  $f2f'$  provide a link between the semantics of the standard lambda calculus for sums ( $ND$ ) and the semantics of our compact calculus ( $HSc/Hsb$ ).

We move forward to describing the reification phase. In this phase, two instantiations of a forcing structure are needed. Unlike the evaluators, which can work over an abstract forcing structure, the reifiers need concrete instantiations built from the syntax of the term calculus in order to produce syntactic normal forms.

The first instantiation is a forcing structure for the standard lambda calculus with sums. The set of worlds is the set of contexts (lists of types), the preorder on worlds is defined as the prefix relation on contexts, the forcing of an atomic type  $p$  is the set of terms of type  $p$  in the context  $w$ , and the answer type of the continuation monad is the set of terms of type  $F$  in the context  $w$ . One could be more precise, and instantiate the answer type by the set of *normal/neutral* terms, like it has been done in most other implementations of TDPE, and in our own prior works, but for the sake of simplicity, we do not make that distinction in this paper.

**Module structureND** <: *ForcingStructure*.  
**Definition**  $K := \text{list Formula}.$   
**Inductive**  $le\_ : \text{list Formula} \rightarrow \text{list Formula} \rightarrow \text{Set} :=$   
 $| le\_refl : \forall \{w\}, le\_ w w$   
 $| le\_cons : \forall \{w1 w2 F\},$   
 $le\_ w1 w2 \rightarrow le\_ w1 (\text{cons } F w2).$   
**Definition**  $le := le\_.$   
**Definition**  $le\_refl : \forall \{w\}, le w w.$   
**Definition**  $pforces := \text{fun } w p \Rightarrow ND w (\text{prop } p).$   
**Definition**  $\text{Answer} := \text{Formula}.$   
**Definition**  $X := \text{fun } w F \Rightarrow ND w F.$   
**End structureND.**

The second instantiation is a forcing structure for our calculus of compact terms. The set of worlds is the same as the set of CNFs, because our context are simply CNFs, the preorder is the prefix relation on CNFs, the forcing of atomic types are terms of atomic types, and the answer type of the continuation monad is the set of terms at base type.

**Module structureHS** <: *ForcingStructure*.  
**Definition**  $K := \text{CNF}.$

**Inductive**  $le\_ : \text{CNF} \rightarrow \text{CNF} \rightarrow \text{Set} :=$   
 $| le\_refl : \forall \{w\}, le\_ w w$   
 $| le\_cons : \forall \{w1 w2 c b\},$   
 $le\_ w1 w2 \rightarrow le\_ w1 (\text{con } c b w2).$

**Definition**  $le := le\_.$

**Definition**  $le\_refl : \forall \{w\}, le w w.$

**Definition**  $pforces := \text{fun } w p \Rightarrow Hsb w (\text{prp } p).$

**Definition**  $\text{Answer} := \text{Base}.$

**Definition**  $X := \text{fun } w b \Rightarrow Hsb w b.$

**End structureHS.**

Using the instantiated forcing structures, we can provide reification functions for terms of the lambda calculus,

**Theorem sreify** :  $(\forall F w, \text{Cont sforces } w F \rightarrow ND w F)$   
**with sreflect** :  $(\forall F w, ND w F \rightarrow \text{Cont sforces } w F).$

and for our compact terms:

**Theorem creify** :  
 $(\forall c w, \text{Cont cforces } w c \rightarrow HSc (\text{explogn } c (\text{cnf } w)))$   
**with creflect** :  $(\forall c w, \text{Cont cforces } (\text{ntimes } c w) c)$   
**with dreify** :  $(\forall d w, \text{Cont dforces } w d \rightarrow Hsb w (\text{bd } d))$   
**with dreflect** :  $(\forall d c1 c2 c3,$   
 $HSc (\text{explogn } c1 (\text{cnf } (\text{ntimes } c3 (\text{con } c1 (\text{bd } d) c2)))) \rightarrow$   
 $\text{Cont dforces } (\text{ntimes } c3 (\text{con } c1 (\text{bd } d) c2)) d).$

The reifier for atomic types,  $\text{preify}$ , is not listed above, because it is simply the ‘run’ operation on the continuation monad. As usually in TDPE, every reification function required its own simultaneously defined reflection function.

Finally, one can combine the reifiers, the evaluators, and the functions  $f2f$  and  $f2f'$ , in order to obtain both a normalizing converter of lambda terms into compact terms (called  $\text{nbe}$  in the Coq implementation), and a converter of compact terms into lambda terms (called  $\text{ebn}$  in the Coq implementation). One can, if one desires, also define only a partial evaluator of lambda terms and only a partial evaluator of compact terms.

## 5. Conclusion

**Summary of our results** We have brought into relation two distinct fundamental problems of the lambda calculi underlying modern functional programming languages, one concerning identity of types, and the other concerning identity of terms, and we have shown how improved understanding of the first problem can lead to improved understanding of the second problem.

We started by presenting a normal form of types, the exp-log normal form, that is a systematic ordering of the high-school identities allowing for a type to be mapped to normal form. This can be used as a simple heuristic for deciding type isomorphism, a first such result for the type language  $\{\rightarrow, +, \times\}$ . We believe that the link established to analysis and abstract algebra (the exp-log decomposition produces a pair of homomorphisms between the additive and the multiplicative group in an exponential field) may also be beneficial to programming languages theory in the future.

The typing restrictions imposed to lambda terms in exp-log normal form allowed us to decompose the standard axioms for  $=_{\beta\eta}$  into a proper and simpler subset of themselves,  $=_{\beta\eta}^e$ . As far as we are aware, this simpler axiomatization has not been isolated before.

Even more pleasingly, the new axiomatization disentangles the old one, in the sense that left-hand sides and right-hand sides of the equality axioms can no longer overlap.

Finally, we ended by giving a compact calculus of terms that can be used as a more canonical alternative to the lambda calculus when modeling the core of functional programming languages: the new syntax does not allow for the  $\eta$ -axioms of Figure 2 even to be stated, with the exception of  $\eta_+^+$  that is still present, albeit with a restricted type. As our method exploits type information, it is orthogonal to the existing approaches that rely on term analysis (discussed below), and hence could be used in addition to them; we hope that it may one day help with addressing the part of  $\eta$ -equality that is still beyond decision procedures. We also implemented and described a prototype converter from/to standard lambda terms.

In the future, we would like to derive declarative rules to describe more explicitly the extent of the fragment of  $=_{\beta\eta}$  decided by our heuristic, although implicitly that fragment is determined by the reduction to ENF congruence classes explained in Section 3. It should be noted that in this respect our heuristic is no less explicitly described, than the only other published one (Balat et al. 2004) (reviewed below).

**Related work** Dougherty and Subrahmanyam (Dougherty and Subrahmanyam 1995) show that the equational theory of terms (morphisms) for almost bi-Cartesian closed categories is complete with respect to the set theoretic semantics. This presents a generalization of Friedman’s completeness theorem for simply typed lambda calculus without sums (Cartesian closed categories) (Friedman 1975).

Ghani (Ghani 1995) proves  $\beta\eta$ -equality of terms of the lambda calculus with sum types to be decidable, first proceeding by rewriting and eta-expansion, and then checking equality up to commuting conversions by interpreting terms as finite sets of quasi-normal forms; no canonical normal forms are obtained.

When sums are absent, the existence of a confluent and strongly normalizing rewrite system proves the existence of canonical normal forms, and then decidability is a simple check of syntactic identity of canonical forms. Nevertheless, even in the context when sums are absent, one may be interested in getting term representations that are canonical *modulo* type isomorphism, as in the recent works of Díaz-Caro, Dowek, and Martínez López (Díaz-Caro and Dowek 2015; Díaz-Caro and Martínez López 2016).

Altenkirch, Dybjer, Hofmann, and Scott (Altenkirch et al. 2001) give another proof of decidability of  $\beta\eta$ -equality for the lambda calculus with sums by carrying out a normalization-by-evaluation argument in category theory. They provide a canonical interpretation of the syntax in the category of sheaves for the Grothendieck topology over the category of constrained environments, and they claim that one can obtain an algorithm for a decision procedure by virtue of the whole development being formalizable in extensional Martin-Löf type theory.

In the absence of  $\eta_+$  (Dougherty 1993), or for the restriction of  $\eta_+$  to  $N$  being a variable (Di Cosmo and Kesner 1993), a confluent and strongly normalizing rewrite system exists, hence canonicity of normal forms for such systems follows.

In (Balat et al. 2004), Balat, Di Cosmo, and Fiore, present a notion of normal form which is a syntactic counterpart to the notion of normal forms in sheaves of Altenkirch, Dybjer, Hofmann, and Scott. However, the forms are not canonical, as there may be two different syntactic normal forms corresponding to a single semantic one. They also say they believe (without further analysis or proof) that one can get canonical normal forms if one considers an ordering of nested  $\delta$ -expressions.

The normal forms of Balat et al. are sophisticated and determining if something is a normal form relies on comparing sub-terms up to a congruence relation  $\approx$  on terms; essentially, this congruence allows to identify terms such as the ones of our Example 5

and Example 6. For determining if something is a normal form, in addition to the standard separation of neutral vs normal terms, one uses three additional criteria: (A)  $\delta$ -expressions that appear under a lambda abstraction must only case-analyze terms involving the abstracted variable; (B) no two terms which are equal modulo  $\approx$  can be case analyzed twice; in particular, no term can be case analyzed twice; (C) no case analysis can have the two branches which are equal modulo  $\approx$ . To enforce condition A, particular powerful control operators, `set/cupto`, are needed in the implementation, requiring a patch of the `ocaml toplevel`. Using our compact terms instead of lambda terms should help get rid of condition A (hence `set/cupto`), since as we showed keeping a constructor for  $\lambda$ ’s in the representation of normal forms is not necessary. On the other hand, we could profit from implementing checks such as B and C in our implementation; however, our goal was to see how far we can get in a purely type-directed way without doing any program analysis.

A final small remark about this line of works: in (Balat 2009), Balat used the word “canonical” to name his normal forms, but this does not preserve the usual meaning of that word, as showed in the previous article (Balat et al. 2004).

Lindley (Lindley 2007) presents another proof of decidability of  $\beta\eta$ -equality for the lambda calculus with sums, based on an original decomposition of the  $\eta_+$ -axiom into four axioms involving evaluation contexts (the proof of this decomposition, Proposition 1, is unfortunately only sketched); the proof of decidability uses rewriting modulo the congruence relation  $\approx$  of Balat et al.

Scherer (Scherer 2015) reinterprets Lindley’s rewriting approach to decidability in the setting of the structural proof theory of maximal multi-focusing, where he brings it in relation to the technique of preemptive rewriting (Chaudhuri et al. 2008). Scherer seems to derive canonicity of his normal forms for natural deduction from Lindley’s results, although the later does not seem to show canonical forms are a result of his rewriting decidability result.

The idea to apply type isomorphism in order to capture equality of terms has been used before (Ahmad et al. 2010), but only implicitly. Namely, in the *focusing* approach to sequent calculi (Liang and Miller 2007), one gets a more canonical representation of terms (proofs) by grouping all so called asynchronous proof rules into blocks called asynchronous phases. However, while all asynchronous proof rules are special kinds of type isomorphisms, not all possible type isomorphisms are accounted for by the asynchronous blocks: sequent calculi apply asynchronous proof rules superficially, by looking at the top-most connectives, but normalizing sequents (formulas) to their exp-log normal form applies proof rules deeply inside the proof tree. Our approach can thus also be seen as moving focusing proof systems into the direction of so called deep inference systems.

## References

- A. Ahmad, D. Licata, and R. Harper. Deciding coproduct equality with focusing. Manuscript, 2010.
- T. Altenkirch, P. Dybjer, M. Hofmann, and P. Scott. Normalization by evaluation for typed lambda calculus with coproducts. In *Logic in Computer Science, 2001. Proceedings. 16th Annual IEEE Symposium on*, pages 303–310, 2001.
- V. Balat. Keeping sums under control. In *Workshop on Normalization by Evaluation*, pages 11–20, Los Angeles, United States, Aug. 2009.
- V. Balat, R. Di Cosmo, and M. Fiore. Extensional normalisation and type-directed partial evaluation for typed lambda calculus with sums. In *Proceedings of the 31st ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL ’04*, pages 64–76, New York, NY, USA, 2004. ACM.
- S. N. Burris and K. A. Yeats. The saga of the high school identities. *Algebra Universalis*, 52:325–342, 2004.

- K. Chaudhuri, D. Miller, and A. Saurin. *Canonical Sequent Proofs via Multi-Focusing*, pages 383–396. Springer US, Boston, MA, 2008. ISBN 978-0-387-09680-3.
- O. Danvy, C. Keller, and M. Puech. Typeful Normalization by Evaluation. In P. L. Hugo Herbelin and M. Sozeau, editors, *20th International Conference on Types for Proofs and Programs (TYPES 2014)*, volume 39 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 72–88, Dagstuhl, Germany, 2015. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. ISBN 978-3-939897-88-0.
- R. Di Cosmo and D. Kesner. A confluent reduction for the extensional typed  $\lambda$ -calculus with pairs, sums, recursion and terminal object. In A. Lingas, R. Karlsson, and S. Carlsson, editors, *Automata, Languages and Programming*, volume 700 of *Lecture Notes in Computer Science*, pages 645–656. Springer Berlin Heidelberg, 1993.
- A. Díaz-Caro and G. Dowek. Simply typed lambda-calculus modulo type isomorphisms. Draft at <https://hal.inria.fr/hal-01109104>, 2015.
- A. Díaz-Caro and P. E. Martínez López. Isomorphisms considered as equalities: Projecting functions and enhancing partial application through an implementation of  $\lambda^+$ . In *IFL 2015: Symposium on the implementation and application of functional programming languages*. ACM, 2016. To appear. Preprint at [arXiv:1511.09324](https://arxiv.org/abs/1511.09324).
- D. Dougherty. Some lambda calculi with categorical sums and products. In *Rewriting Techniques and Applications*, pages 137–151. Springer, 1993.
- D. J. Dougherty and R. Subrahmanyam. Equality between functionals in the presence of coproducts. In *Proceedings of the 10th Annual IEEE Symposium on Logic in Computer Science, LICS '95*, pages 282–, Washington, DC, USA, 1995. IEEE Computer Society.
- M. Fiore, R. D. Cosmo, and V. Balat. Remarks on isomorphisms in typed lambda calculi with empty and sum types. *Annals of Pure and Applied Logic*, 141:35–50, 2006.
- H. Friedman. Equality between functionals. In *Logic Colloquium '73*, volume 453 of *Lecture Notes in Mathematics*, pages 22–37. Springer, 1975.
- N. Ghani.  $\beta\eta$ -equality for coproducts. In *Typed Lambda Calculi and Applications*, pages 171–185. Springer, 1995.
- G. H. Hardy. *Orders of Infinity. The 'Infinitärcalcul' of Paul Du Bois-Reymond*. Cambridge Tracts in Mathematic and Mathematical Physics. Cambridge University Press, 1910.
- D. Ilik. Axioms and decidability for type isomorphism in the presence of sums. In *Proceedings of the Joint Meeting of the Twenty-Third EACSL Annual Conference on Computer Science Logic (CSL) and the Twenty-Ninth Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, CSL-LICS '14, pages 53:1–53:7, New York, NY, USA, 2014. ACM.
- C. Liang and D. Miller. Focusing and polarization in intuitionistic logic. In J. Duparc and T. A. Henzinger, editors, *Computer Science Logic*, volume 4646 of *Lecture Notes in Computer Science*, pages 451–465. Springer Berlin Heidelberg, 2007. ISBN 978-3-540-74914-1.
- S. Lindley. Extensional rewriting with sums. In S. R. Della Rocca, editor, *Typed Lambda Calculi and Applications*, volume 4583 of *Lecture Notes in Computer Science*, pages 255–271. Springer Berlin Heidelberg, 2007.
- M. Rittri. Using types as search keys in function libraries. *J. Funct. Program.*, 1(1):71–89, 1991.
- G. Scherer. Multi-Focusing on Extensional Rewriting with Sums. In T. Altenkirch, editor, *13th International Conference on Typed Lambda Calculi and Applications (TLCA 2015)*, volume 38 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 317–331, Dagstuhl, Germany, 2015. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. ISBN 978-3-939897-87-3.