A Machine-Checked Correctness Proof of Normalization by Evaluation for Simply Typed Lambda Calculus

Author: András Kovács Advisor: Ambrus Kaposi

Budapest, 2017

Contents

1 Introduction

1.1 Overview

Normalization animates the syntax of typed λ-calculi, making them useful as programming languages. Also, in contrast to general terms, normal forms possess a more restricted structure which simplifies formal reasoning. Moreover, normalization allows one to decide convertibility of terms, which is required during type checking polymorphic and dependent type theories.[1]

This article presents an efficient and verified implementation of $\beta\eta$-normalization for simply typed λ-calculus (STLC) in Agda, an implementation of constructive type theory. In this setting, the syntax can be seen as an embedded language, and manipulating embedded terms is a limited form of metaprogramming. Normalization can be reused in the implementation of proof tactics, decision procedures or domain-specific languages, therefore efficiency is desirable. It is also preferable that the syntax remains as simple and self-contained as possible, both for pedagogical purposes and for the ease of reuse in other Agda developments. For this reason we choose an intrinsic syntax with de Bruijn indices and implicit substitutions.

We are not concerned with mapping the algorithm to lower-level abstract machines or hardware. Thus we are free to make use of the most convenient evaluation mechanism at hand: that of the metalanguage. This way, we can gloss over implementation details of efficient higher-order evaluation. This is the core idea of normalization by evaluation (NbE) from an operational point of view. Also, totality of evaluation in the metatheory is always implicitly assumed from the inside. This lets us implement normalization in a structurally recursive way, making its totality trivial. NbE also naturally supports η-normalization.

If we are to trust a particular normalization algorithm, we need to prove the following properties (borrowing terminology from [3]):

[1] In particular, the types of System F_ω largely correspond to simply typed λ-terms.

- *Completeness*: normalization maps a term to a normal form which is convertible to the term.
- *Soundness*: normalization maps convertible terms to equal normal forms.

Additionally, we require stability, which establishes that normalization is aligned with the specification of normal forms:

- *Stability*: normalization acts as the identity function on normal terms.

Section 2 describes the metatheory we work in. In Section 3, we formalize the syntax of STLC along with $\beta\eta$-conversion. Section 4 presents the normalization function, and Section 5 proves its correctness. Section 6 describes two alternative formalizations, one with more efficient evaluation and one with more concise correctness proofs. We discuss the results and possible future research in the final section.

The content of this article is fully formalized in Agda. The formalization can be found at *https://github.com/AndrasKovacs/stlc-nbe*.

1.2 Related Work

Martin-Löf [23] first used NbE (though the term had not been coined yet) to show decidability of type checking for the 1975 version of his intuitionistic type theory. Berger and Schwichtenberg reintroduced NbE in 1991 [10] as a normalization method for lambda calculi. Abel's work [1] gives a comprehensive overview of NbE, and also develops it for a range of type theories. The approach differs markedly from ours, as it uses extrinsic syntax and separate typed and untyped phases of evaluation.

Catarina Coquand's work [17] is the most closely related to our development. She formally proves soundness and completeness of NbE for simple type theory. However, she considers a nameful syntax with explicit substitutions while we use intrinsic de Bruijn variables with implicit substitutions.

Altenkirch and Chapman [3] formalize a big-step normalization algorithm in Agda. This development uses an explicit substitution calculus and implements normalization using an environment machine with first-order closures. The algorithm works similarly to ours on the common syntactic fragment, but it is not structurally recursive and hence specifies evaluation as a inductive relation, using the Bove-Capretta method [11].

Altenkirch and Kaposi [5] use a glued presheaf model ("presheaf logical predicate") for NbE for a minimal dependent type theory. This development provided the initial inspiration for the formalization in this article; it seemed plausible that "scaling down" dependently typed NbE to simple type theory would yield a formalization that is more compact than existing ones. This turned to be the case; however the discussion of resulting development is relegated to Section 6.1, because using a Kripke model has the advantage of clean separation of the actual algorithm and its naturality proofs, which was deemed preferable to the more involved presheaf construction. Also, our work uses

a simple presheaf model rather than the glued model used in Ibid. or in the work of Altenkirch, Hoffman and Streicher [4].

2 Metatheory

In this section the metatheory used for formalization is presented, first broadly, then its specific implementation in Agda.

2.1 Type Theory

The basic system we use is intensional Martin-Löf type theory (MLTT). For an overview and tutorial see the first chapter of the Homotopy Type Theory book [31]. However, there is no canonical definition of intensional type theory. There are numerous variations concerning type universes and the range of allowed type definitions. Our development uses the following features:

- *Strictly positive inductive definitions.* These are required for an intrinsically well-typed syntax. We do not require induction-recursion, induction-induction, higher induction or large inductive types.
- *Two predicative universes, named* `Set₀` *and* `Set₁`*, with large elimination into* `Set₀`. `Set₀` is the base universe in Agda, i. e. the universe of small types. Large elimination is needed for recursive definitions of semantic types, since inductive definitions for them would not be positive and hence would be illegal. We only need `Set₁` to type the large eliminations. In Agda `Set` is a synonym for `Set₀`.
- *Function extensionality as an axiom.* This posits that two functions are equal if they take equal arguments to equal results. We could eschew it by using setoid reasoning with semantic equivalence, as C. Coquand does [17]. However, we only use function extensionality in correctness proofs, and the normalization function does not refer to it or to any other postulate. We are content with *logical* as opposed to computational content for correctness proofs. This way, the numerous congruence lemmas needed for setoid reasoning can be skipped. Also, function extensionality has computational interpretation in cubical [14] and observational [7] type theories, to which this development could be plausibly ported in the future (when practical implementations of the mentioned theories become available).

In Agda 2.5.2, there is robust support for dependent pattern matching without Streicher's K axiom [29]. We use the `--without-K` language option for all of our code in the main development. However, we do use axiom K in Section 6.1 for the direct presheaf model, but only for technical convenience, as it can be avoided by additional uniqueness proofs for equalities.

2.2 Agda

Agda is a dependently typed programming language and proof assistant. Its design and core features are described in Ulf Norell's thesis [26]. The latest documentation is available online [20]. For a book-length introduction geared towards beginners, see [30]. We summarize here the core constructions and our notational conventions and liberties.

2.2.1 Basic Constructions

Two essential artifacts in Agda code are *inductive type definitions* and *function definitions*.

Inductive definitions introduce new types. Agda has *open type universes*, which means that programmers may freely add new types as long as they preserve logical consistency[2]. An inductive definition involves a type constructor declaration followed by zero or more term constructor declarations. For example, the type of natural numbers is defined the following way:

```
data ℕ : Set where
  zero : ℕ
  suc  : ℕ → ℕ
```

Inductive definitions may have *parameters* and *indices*. The former are implicitly quantified over the term constructors, but must be uniform in the constructors' return types. The latter must be explicitly quantified in term constructors, but are allowed to vary. The definition for length-indexed vectors exhibits both:

```
data Vec (A : Set) : ℕ → Set where
  nil  : Vec A zero
  cons : {n : ℕ} → A → Vec A n → Vec A (suc n)
```

In a type constructor declaration, parameters are listed left to the colon, while indices are to the right. `A : Set` is a parameter, so it is implicitly quantified and is the same in the return types of `nil` and `cons`. In contrast, the length index is quantified in `cons`. We can use brackets to make parameters implicit; here `cons` has an implicit first parameter, and can be used like `cons zero nil` for a value of `Vec ℕ (suc zero)`. Implicit arguments are filled in by Agda's unification algorithm. Alternatively, the `∀` symbol can be used to leave the types of parameters implicit, and we could define `cons` as follows:

```
cons : ∀ {n} → A → Vec A n → Vec A (suc n)
```

Additionally, Agda has records, providing syntactic sugar and namespace management for iterated Σ types. For example semigroups can be defined as elements of the following record type:

```
record Semigroup : Set₁ where
  field
```

[2]Agda checks *strict positivity* as a sufficient condition for consistency.

```
S     : Set
_⊗_   : M → M → M
assoc : ∀ m₁ m₂ m₃ → (m₁ ⊗ (m₂ ⊗ m₃)) ≡ ((m₁ ⊗ m₂) ⊗ m₃)
```

Elements of `Semigroup` can be given as `record {S = ...; _⊗_ = ...; assoc = ...}`. Whenever we have some `(sg : Semigroup)` in context, we can bring all of its fields into the current namespace by `open Semigroup sg`. Also, given an `(sg : Semigroup)` in scope, we can use field names as projections, as in `(S sg)` or `(assoc sg)`. We will sometimes omit record instances and simply write `S` or `assoc` (or whatever record fields we are working with) when it is clear from where we are projecting from.

Function definitions are given by pattern matching. In Agda, pattern matching implements branching evaluation as it is usual in functional languages, but it also refines type indices based on particular selected cases. The main practical difference between parameters and indices is that we can can gain information about indices by pattern matching, while we can't infer anything about parameters just by having a parameterized term.

For example, in the following definition of concatenation for `Vec`, the result type is refined depending on whether the first argument is empty:

```
_+_ : ℕ → ℕ → ℕ
zero  + m = m
suc n + m = suc (n + m)

_++_ : ∀ {A n m} → Vec A n → Vec A m → Vec A (n + m)
nil       ++ ys = ys
cons x xs ++ ys = cons x (xs ++ ys)
```

In the `cons x xs ++ ys` case, the result type is refined to `Vec A (suc (n + m))`, allowing us to build a `suc`-long result on the right hand side with `cons`.

Function definitions must be total, hence pattern matching must be exhaustive and recursive calls must be structurally decreasing. Dependent pattern matching with structural recursion allows us to write the same proofs as inductive eliminators, but with a more convenient interface and less administrative burden. For instance, injectivity and disjointedness of term constructors is implicit in pattern matching while they have to be separately proven when using eliminators.

As a notational convention, we shall take cues from the Idris [12] programming language and sometimes leave implicit parameters implicit even in type declarations of functions or constructors. In this style, declaring `_++_` looks as follows:

```
_++_ : Vec A n → Vec A m → Vec A (n + m)
```

This is not valid Agda, but we shall do this whenever the types and binding status of parameters are obvious. Sometimes we also omit type annotations from bindings when they can be inferred or not relevant for exposition.

2.2.2 Included Code Examples

The project can be found at *https://github.com/AndrasKovacs/stlc-nbe*. Most Agda code listings in this article are also included in the formal development, although the versions here may include syntactic liberties and abbreviations. We indicate the names of relevant source files in the examples as Agda comments or in expository text.

2.2.3 Standard Library

The formal development does not have any external dependencies. We need less than a hundred lines of code from the Agda standard library [21], but we choose to include it alongside the project in `Lib.agda`, for portability, and also make some changes.

Most of `Lib.agda` concerns equality and equational reasoning. Propositional equality is the same as in standard Agda:

```
data _≡_ {i}{A : Set i} (x : A) : A → Set i where
  refl : x ≡ x
```

We also postulate function extensionality, for functions with both implicit and explicit arguments (the "i" in `fexti` stands for "implicit"):

```
postulate
  fext  : (∀ x → f x ≡ g x) → f ≡ g
  fexti : (∀ x → f {x} ≡ g {x}) → (λ {x} → f {x}) ≡ (λ {x} → g {x})
```

Note that the types of `f` and `g` are left implicit above, in accordance with out notational liberties. In full Agda we write:

```
fext :
  ∀ {i j}{A : Set i}{B : A → Set j}{f g : (x : A) → B x}
  → ((x : A) → f x ≡ g x) → f ≡ g
```

Here the quantification noise is increased by universe indices. Universe polymorphism is only used in `Lib.agda`, but we will henceforth omit universe indices anyway.

Our style of equational reasoning deviates from the standard library. First, we use a more compact notation for symmetry and transitivity, inspired by the Homotopy Type Theory book [31]:

```
_■_ : x ≡ y → y ≡ z → x ≡ z -- transitivity
refl ■ refl = refl


_⁻¹ : x ≡ y → y ≡ x -- symmetry
refl ⁻¹ = refl
```

`_⁻¹` is postfix, so if we have `(p : x ≡ y)`, then `(p ⁻¹)` has type `(y ≡ x)`.

Second, for reasoning about congruences we mainly use two operations. The first (named `ap` in the HoTT book and `cong` in the standard library) expresses that functions respect equality:

```
_&_ : (f : A → B) → x ≡ y → f x ≡ f y
f & refl = refl
infixl 9 _&_
```

The second one expresses that *non-dependent function application* respects equality as well:

```
_⊗_ : f ≡ g → x ≡ y → f x ≡ g y
refl ⊗ refl = refl
infixl 8 _⊗_
```

Using these two operators, we can lift non-dependent functions with arbitrary arities to congruences. For example, if we have `(p : a₀ ≡ a₁)`, `(q : b₀ ≡ b₁)` and `(f : A → B → C)`, then `(f & p ⊗ q)` has type `(f a₀ b₀ ≡ f a₁ b₁)`. Note that both operators associate left, and `_&_` binds more strongly. This is reminiscent of the lifting syntax with applicative functors [25] in Haskell programming.

Furthermore, we use coercion instead of transport (see HoTT book [31, pp. 72], also named `subst` in the Agda standard library):

```
coe : A ≡ B → A → B
coe refl a = a
```

`transport`/`subst` can be recovered by `_&_` and `coe`:

```
subst : (P : A → Set) → x ≡ y → P x → P y
subst P eq = coe (P & eq)
```

The choice for `coe` is mainly stylistic and rather subjective. `coe (P & eq)` does not look worse than `subst P eq`, but `coe` by itself is used often and is more compact than `subst id`.

In Agda, sometimes an explicit "equational reasoning" syntax is used. For example, commutativity for addition may look like the following:

```
+-comm : (m n : ℕ) → m + n ≡ n + m
+-comm zero    n = +-right-identity n ⁻¹
+-comm (suc m) n = begin
    suc m + n    ≡⟨⟩
    suc (m + n)  ≡⟨ suc & +-comm m n ⟩
    suc (n + m)  ≡⟨ +-suc n m ⁻¹ ⟩
    n + suc m    QED
```

In formal development we do not use this style and instead keep the intermediate expressions implicit:

```
+-comm (suc m) n = (suc & +-comm n m) ■ (+-suc n m ⁻¹)
```

This is because in an interactive environment one can step through the intermediate points fairly easily, and writing them out adds significant visual clutter. However, we will use informal equational reasoning when it aids explanation.

We also borrow `Σ` (dependent sum), `⊤` (unit type), `⊥` (empty type) and `_⊎_` (disjoint union) from the standard library. Elements of `Σ` are constructed as `(a , b)` pairs, and the projections are `proj₁` and `proj₂`. `A × B` is defined in terms of `Σ` as the non-dependent pair type. `⊤` has `tt` ("trivially true") as sole element. Injection into disjoint unions is done with `inj₁` and `inj₂`. The *ex falso* principle is named `⊥-elim`, with type `{A : Set} → A → A`.

3 Syntax

In this section we present the syntax of simply typed lambda calculus - in our case intrinsic syntax with implicit substitutions and de Bruijn indices. We take term conversion to be part of the syntax and therefore include it in this section.[3] Since the definition of conversion refers to substitution and weakening, they are also presented here.

3.1 Core Syntax

The complete definition is the following:

```
-- Syntax.agda
data Ty : Set where
  ι   : Ty
  _⇒_ : Ty → Ty → Ty

data Con : Set where
  •   : Con
  _,_ : Con → Ty → Con

data _∈_ (A : Ty) : Con → Set where
  vz : A ∈ (Γ , A)
  vs : A ∈ Γ → A ∈ (Γ , B)

data Tm Γ : Ty → Set where
  var : A ∈ Γ → Tm Γ A
```

[3]We take the view that the "proper" definition of the syntax would be an initial object in some suitable category of models, e. g. the initial category with families. In this way, models include conversion as propositional equations, so the syntax is quotiented by conversion as well; see e. g. [13] and [6]. We do not use this definition because quotients are not yet natively supported in Agda, and we find value in developing a conventional presentation first.

```
  lam : Tm (Γ , A) B → Tm Γ (A ⇒ B)
  app : Tm Γ (A ⇒ B) → Tm Γ A → Tm Γ B
```

There is a base type `ι` (Greek iota) and function types. Contexts are just lists of types. `_∈_` is an inductively defined membership relation on contexts. It has the same structure as de Bruijn indices; for this reason they are named `vz` for "zero" and `vs` for "suc". Terms are parameterized by a `Γ` context and indexed by a type. `var` picks an entry from the contexts, `lam` is abstraction while `app` is function application.

As an example, the identity function on the base type is represented as:

```
id-ι : Tm • (ι ⇒ ι)
id-ι = lam (var vz) -- with nameful syntax: λ (x : ι) . x
```

Th presented syntax is *intrinsic*. In other words, only the well-typed terms are defined. An *extrinsic* definition would first define untyped *preterms* and separately a typing relation on preterms and contexts:

```
data Tm : Set where -- omitted constructors
data _⊢_∈_ (Γ : Con) → Tm → Ty → Set where -- omitted constructors
```

Here, `Γ ⊢ t ∈ A` expresses that the preterm `t` is well-formed and has type `A` in context `Γ`.

There are advantages and disadvantages to both intrinsic and extrinsic definitions. With intrinsic syntax, well-formedness and type preservation come for free. On the other hand, proofs and computations which do not depend on types are easier to formulate with preterms. In the case of STLC, the type system is simple enough so that carrying around type information never becomes burdensome, so in this setting intrinsic definitions are all-around more convenient. For polymorphic systems such as System F, intrinsic typing introduces significant bureaucratic overhead, by tagging terms with type coercions; see Benton et al. [9] for approaches to dealing with them. For extrinsic syntax, powerful automation for reasoning about substitution is available in Coq, allowing spectacularly compact proofs [28]. It remains to be seen if it can be practically adapted to intrinsic syntax. For dependent type theories, intrinsic typing with quotient inductive-inductive definitions seems to be most promising, as it requires a relatively compact set of rules [6], in contrast to extrinsic approaches which suffer from a veritable explosion of well-formedness conditions and congruence rules.

3.2 Context Embeddings and Substitutions

In order to specify β-conversion, we need a notion of substitution. Also, the η-rule must refer to a notion of *weakening* or *embedding*: it mentions the "same" term under and over a λ binder, but the two mentions cannot be definitionally the same, since they are in different contexts and thus have different types. We need to express the notion that whenever one has a term in a context, one can construct essentially the same term

in a larger context. We use *order-preserving-embeddings* for this. Embeddings enable us to state the η-rule, but we also make use of them for defining substitutions.

3.2.1 Embeddings

We define order-preserving embeddings the following way:

```
-- Embedding.agda
data OPE : Con → Con → Set where
  •    : OPE • •
  drop : OPE Γ Δ → OPE (Γ , A) Δ
  keep : OPE Γ Δ → OPE (Γ , A) (Δ , A)
```

Elements of `OPE Γ Δ` express that `Δ` can be obtained from `Γ` by dropping zero or more entries in order.

Some choices can be made here. C. Coquand [18] uses an explicit identity embedding constructor with type `(∀ {Γ} → OPE Γ Γ)` instead of our `•`. This slightly simplifies some proofs and slightly complicates others. It does not have propositionally unique identity embeddings[4], but overall the choice between this definition and ours is fairly arbitrary.

An alternative solution is using *renamings*. Renamings are essentially substitutions containing only variables, alternatively definable as functions mapping variables of smaller contexts to variables of larger contexts [24, pp. 24]:

```
Ren : Con → Con → Set
Ren Γ Δ = {A : Ty} → A ∈ Δ → A ∈ Γ
```

However, renamings are able to express *permutations* of contexts as well, which we have no need for in this development. Thus, renamings are larger than what is strictly necessary, which may or may not result in technical difficulties. We aim to choose the smallest workable representation, as a general principle.

Embeddings have action on variables and terms, reconstructing them in larger contexts. We denote embedding by lowercase "e" subscripts:

```
∈ₑ : OPE Γ Δ → A ∈ Δ → A ∈ Γ
∈ₑ •        v      = v
∈ₑ (drop σ) v      = vs (∈ₑ σ v)
∈ₑ (keep σ) vz     = vz
∈ₑ (keep σ) (vs v) = vs (∈ₑ σ v)

Tmₑ : OPE Γ Δ → Tm Δ A → Tm Γ A
Tmₑ σ (var v)   = var (∈ₑ σ v)
Tmₑ σ (lam t)   = lam (Tmₑ (keep σ) t)
Tmₑ σ (app f a) = app (Tmₑ σ f) (Tmₑ σ a)
```

[4]Since wrapping the identity embedding with any number of `keep`-s also yields identity embeddings.

Identity embeddings keep every entry:

```
idₑ : ∀ {Γ} → OPE Γ Γ
idₑ {•}     = •
idₑ {Γ , A} = keep (idₑ {Γ})
```

We define `wk` as shorthand for the embedding that only drops the topmost entry:

```
wk : ∀ {A Γ} → OPE (Γ , A) Γ
wk = drop idₑ
```

3.2.2 Substitutions

A salient feature of our syntax is the lack of explicit substitution, i. e. we define substitutions separately as lists of terms, and their action on terms is given as a recursive function. We choose implicit substitutions for two reasons:

1. It is the canonical presentation of STLC in textbooks [27, 22].
2. Lighter formalization. With implicit substitutions, the definition of conversion becomes significantly smaller. With explicit substitutions, we need congruence closure of β and η conversion on terms and substitutions, plus rules for the action of substitution on terms. With implicit substitutions, only terms need to be considered. For illustration, Altenkirch and Chapman's definition [3] of conversion for explicit substitution calculus has twenty-four rules, while this development has only seven. It also follows that with our syntax, equational reasoning about substitutions involves definitional or propositional equalities, which are automatically respected by all constructions. There is minor complication involving semantic interpretation of substitutions, discussed in Section 5.5.

Strictly speaking, stating β-conversion only requires single substitutions. However, simultaneous substitution is easier to define and reason about. Thus, we shall define the former using the latter and identity substitutions. We have substitutions as lists of terms:

```
-- Substitution.agda
data Sub (Γ : Con) : Con → Set where
  •   : Sub Γ •
  _,_ : Sub Γ Δ → Tm Γ A → Sub Γ (Δ , A)
```

Elements of `Sub Γ Δ` contain a `Tm Γ A` for each `A` type in `Δ`. In other words, an element of `Sub Γ Δ` assigns to each variable in Δ a term in Γ.

Again, there are some choices in defining `Sub`. Identity substitutions (which have the identity action on terms) could be represented as an explicit constructor, as well as `Sub` composition and weakenings. Our `Sub` can be seen as a normal form of substitutions: explicit composition and weakening forms tree-like structures, but we flatten them to just lists of terms. This eliminates redundancy, and allows us to define compositions and identities recursively, which provides us with more definitional equalities.

We could also give a functional definition (see [24, pp. 24]), similarly to how we did for renamings before:

```
Sub : Con → Con → Set
Sub Γ Δ = {A : Ty} → A ∈ Δ → Tm Γ A
```

However, this contains many definitionally distinct functions with the same action on terms, and so we dismiss it on grounds of unnecessary largeness.

Before we can define identity substitutions and the action of substitution on terms, we need an operation with type `(∀ {A Γ Δ} → Sub Γ Δ → Sub (Γ , A) (Δ , A))`. This enables pushing substitutions under λ-binders when recursing on terms. Constructing this substitution requires that we embed all terms in the input substitution into the extended `(Γ , A)` context. We define this as right composition with an embedding:

```
_∘ᵉ_ : Sub Δ Σ → OPE Γ Δ → Sub Γ Σ
•       ∘ᵉ δ = •
(σ , t) ∘ᵉ δ = (σ ∘ᵉ δ) , Tmᵉ δ t
```

The notation `_∘ᵉ_` expresses that we have a substitution on the left and an embedding on the right. Its type resembles that of ordinary function composition, if we consider both `Sub` and `OPE` as context morphisms. We will expand on the categorical interpretation of this in Section 5.2.

There is a canonical injection `⌜_⌝ᵒᵖᵉ` from `OPE` to `Sub`:

```
dropˢ : Sub Γ Δ → Sub (Γ , A) Δ
dropˢ σ = σ ∘ᵉ wk

keepˢ : Sub Γ Δ → Sub (Γ , A) (Δ , A)
keepˢ σ = dropˢ σ , var vz

⌜_⌝ᵒᵖᵉ : OPE Γ Δ → Sub Γ Δ
⌜ •       ⌝ᵒᵖᵉ = •
⌜ drop σ ⌝ᵒᵖᵉ = dropˢ ⌜ σ ⌝ᵒᵖᵉ
⌜ keep σ ⌝ᵒᵖᵉ = keepˢ ⌜ σ ⌝ᵒᵖᵉ
```

We note that `keepˢ` is the needed operation for pushing substitutions under binders. Now the action of substitution on terms and variables can be defined:

```
∈ˢ : Sub Γ Δ → A ∈ Δ → Tm Γ A
∈ˢ (σ , t) vz     = t
∈ˢ (σ , t) (vs v) = ∈ˢ σ v

Tmˢ : Sub Γ Δ → Tm Δ A → Tm Γ A
Tmˢ σ (var v)   = ∈ˢ σ v
Tmˢ σ (lam t)   = lam (Tmˢ (keepˢ σ) t)
Tmˢ σ (app f a) = app (Tmˢ σ f) (Tmˢ σ a)
```

`∈ₛ` simply looks up a term from a substitution[5], while `Tmₛ` recurses into terms to substitute all variables. Identity substitutions can be defined as well:

```
idₛ : {Γ : Con} → Sub Γ Γ
idₛ {•}     = •
idₛ {Γ , A} = keepₛ idₛ
```

Single substitution with a `(t : Tm Γ A)` term is given by `(idₛ , t)`. This assigns the `t` term to the zeroth de Bruijn variable and leaves all other variables unchanged.

3.3 Conversion

The conversion relation is given as:

```
-- Conversion.agda
data _~_ {Γ} : ∀ {A} → Tm Γ A → Tm Γ A → Set where
  η     : t ~ lam (app (Tmₑ wk t) (var vz))
  β     : app (lam t) t' ~ Tmₛ (idₛ , t') t

  lam   : t ~ t' → lam t ~ lam t'
  app   : f ~ f' → a ~ a' →  app f a ~ app f' a'

  ~refl : t ~ t
  _~⁻¹  : t ~ t' → t' ~ t
  _~■_  : t ~ t' → t' ~ t'' → t ~ t''
```

`η` and `β` are the actual conversion rules. We push `t` under a lambda with `(Tmₑ wk)` in `η`, and use single substitution for `β`. The rest are rules for congruence (`lam`, `app`) and equivalence closure (`~refl`, `_~⁻¹`, `_~■_`). The syntax for equivalence rules is the same as for propositional equality, except for the ~ prefixes.

4 Normalization

In this section we specify normal forms and implement normalization. Then, we discuss Kripke models and how normalization can be expressed in terms of them.

[5]The immediate reason for not naming `∈ₛ` simply "lookup" is that we will need several different lookup functions for various purposes. For any sizable formal development, naming schemes eventually emerge, for better or worse. The author has found that uniformly encoding salient information about types of operations in their names is worthwhile to do. It is also not easy to hit the right level of abstraction; too little and we repeat ourselves too much or neglect to include known structures, but too much abstraction may cause the project to be less accessible and often also less convenient to develop in the first place. Many of the operations and proofs here could be defined explicitly using categorical language, i. e. bundling together functors' actions on objects, morphisms and category laws in records, then projecting out required components. Our choice is to keep the general level of abstraction low for this development.

4.1 Normal terms

Our definition of normal forms is entirely standard: they are either lambdas or *neutral* terms of base type, and neutral terms are variables applied to zero of more normal arguments:

```
-- NormalForm.agda
mutual
  data Nf (Γ : Con) : Ty → Set where
    ne  : Ne Γ ι → Nf Γ ι
    lam : Nf (Γ , A) B → Nf Γ (A ⇒ B)

  data Ne (Γ : Con) : Ty → Set where
    var : A ∈ Γ → Ne Γ A
    app : Ne Γ (A ⇒ B) → Nf Γ A → Ne Γ B
```

Agda allows us to overload `var`, `lam` and `app` for normal forms. β-normality is obvious, since only variables can be applied. Normal forms are also η-long: since `ne` only injects neutrals of base type, all normal terms of function type must in fact be lambdas. Only requiring β-normality would be as simple as having `(ne : ∀ {A} → Ne Γ A → Nf Γ A)`. Alternatively, normal forms can be given as

```
data Nf (Γ : Con) : (A : Ty) → Tm Γ A → Set
```

with the same construction as before except that it is a predicate on general terms, expressing their normality. This definition would be roughly as convenient as the one we use. We also need actions of context embeddings:

```
Nfₑ : OPE Γ Δ → Nf Δ A → Nf Γ A -- definitions omitted
Neₑ : OPE Γ Δ → Ne Δ A → Ne Γ A -- definitions omitted
```

Normal and neutral terms can be injected back to terms. Also, injection commutes with embedding (i. e. injection is natural).

```
⌜_⌝Nf : Nf Γ A → Tm Γ A
⌜_⌝Ne : Ne Γ A → Tm Γ A

⌜⌝Ne-nat : ∀ σ n → ⌜ Neₑ σ n ⌝Ne ≡ Tmₑ σ ⌜ n ⌝Ne
⌜⌝Nf-nat : ∀ σ n → ⌜ Nfₑ σ n ⌝Nf ≡ Tmₑ σ ⌜ n ⌝Nf
```

4.2 Preliminaries: Models of STLC

Clearly, STLC is a small fragment of Agda, so we should be able to interpret the syntax back to Agda types and constructions in a straightforward way. From a semantic viewpoint, the most straightforward interpretation of the syntax is called the *standard model*. From an operational viewpoint, the standard model is just a well-typed interpreter [8] for STLC as an embedded language. It is implemented as follows:

```
-- Misc/StdModel.agda
Tyˢ : Ty → Set
Tyˢ ι       = ⊥
Tyˢ (A ⇒ B) = Tyˢ A → Tyˢ B

Conˢ : Con → Set
Conˢ •       = ⊤
Conˢ (Γ , A) = Conˢ Γ × Tyˢ A

∈ˢ : ∀ {Γ A} → A ∈ Γ → (Conˢ Γ → Tyˢ A)
∈ˢ vz     Γˢ = proj₂ Γˢ
∈ˢ (vs v) Γˢ = ∈ˢ v (proj₁ Γˢ)

Tmˢ : ∀ {Γ A} → Tm Γ A → (Conˢ Γ → Tyˢ A)
Tmˢ (var v)   Γˢ = ∈ˢ v Γˢ
Tmˢ (lam t)   Γˢ = λ aˢ → Tmˢ t (Γˢ , aˢ)
Tmˢ (app f a) Γˢ = Tmˢ f Γˢ (Tmˢ a Γˢ)
```

Traditionally, notation for semantics uses double brackets, e. g. the interpretation of types would look like this:

```
⟦_⟧ : Ty → Set
⟦ ι     ⟧ = ⊥
⟦ A ⇒ B ⟧ = ⟦ A ⟧ → ⟦ B ⟧
```

We instead opt for notating semantic interpretations with superscripts on type constructors (contrast *subscripts*, which we use for denoting operations for embeddings and substitutions). Syntactic overloading in Agda is not convenient for our purposes here, so we have to distinguish different constructions (terms or types, etc.) and different models in any case. Therefore it makes sense to just reuse the names of types in the syntax, and distinguish different models by different superscripts. Similarly, we annotate names of function arguments with superscripts, if they are elements of certain semantic sets. For example, for some model `M` we use `Γᴹ`, `Δᴹ` etc. for elements of semantic contexts, and `aᴹ, bᴹ` etc. for semantic terms.

As to the implementation of the model: we interpret the base type with the empty type, and functions as functions. Contexts are interpreted as lists of semantic values. Terms are interpreted as functions from semantic contexts to semantic types. Now, normalizing `(Tmˢ (lam (var vz)) tt)` yields the `(λ aˢ → aˢ)` Agda term, so indeed we interpret a syntactic identity function with a metatheoretic identity function.

The standard model already has a key feature of NbE: it uses metatheoretical functions for evaluation. However, the standard model does not allow one to go back to syntax from semantics. If we have an `(f : Aˢ → Bˢ)` semantic function, all we can do with it is to apply it to an argument and extract a `Bˢ`. *Weak evaluation* of programs assumes that all variables are mapped to semantic values. This corresponds to the usual notion of program evaluation in common programming languages. In contrast, strong

normalization requires reducing terms inside function bodies, under binders. Bound variables are not mapped to any semantic value, so they block evaluation. In STLC, variables block function application; in richer systems variables may block case analysis or elimination of inductively defined data as well. See [1, pp. 12] for an overview for various approaches for implementing computation under binders.

NbE adds additional structure to semantic types, internalizing computation with blocking variables, which allows one to get back to syntax, moreover, to a subset of syntax containing only normal terms.

Adding progressively more structure to models allows us to prove more properties about the syntax. The standard model yields a simple proof of consistency, namely that there is no term with base type in the empty context:

```
consistency : Tm • ι → ⊥
consistency t = Tmˢ t tt
```

Kripke models contain slightly more structure than the standard model, allowing us to implement normalization, which is a *completeness* theorem from a logical viewpoint [19], as per the Curry-Howard correspondence. Adding yet more structure yields *presheaf models*, which enable correctness proofs for normalization as well. We summarize the computational and logical interpretations below.

Table 1: Overview of models

Model	Enabled computation	Yields proof of
Standard	Type-safe interpreter	Consistency
Kripke	Normalizer	Completeness
Presheaf	Normalizer	Proof-relevant completeness

4.3 Implementation

We denote the model for normalization with ^{N} superscripts. The implementation follows a similar shape as the standard model. The key difference - from which others follow - is in the interpretation of functions:

```
-- Normalization.agda
Tyᴺ : Ty → Con → Set
Tyᴺ ι       Γ = Nf Γ ι
Tyᴺ (A ⇒ B) Γ = ∀ {Δ} → OPE Δ Γ → Tyᴺ A Δ → Tyᴺ B Δ
```

Each type is mapped to a `(Con → Set)` predicate. Hence, semantic types are not just sets anymore, but families of sets indexed by contexts. The base type is mapped to the family of its normal forms. Functions are mapped to semantic functions which take semantic inputs to semantic outputs in *any context larger than* Γ. The additional `OPE` parameter is the key to the algorithm. As it was noted in Section 4.2, semantic

functions must be able to take blocking variables as inputs. Blocking variables must be necessarily "fresh" with respect to the Γ context of a semantic function. Here, we can apply a semantic function to suitable embedding into a larger context, allowing to subsequently apply it to semantic terms containing variables pointing into that context. In the simplest case (and the only case we will actually use), if we have

```
fᴺ : ∀ {Δ} → OPE Δ Γ → Tyᴺ A Δ → Tyᴺ B Δ
```

then we also have

```
fᴺ (wk {A}) : Tyᴺ A (Γ , A) → Tyᴺ B (Γ , A)
```

The interpretations of contexts and variables are analogous to those in the standard model: contexts become lists of semantic values and variables look up semantic values from semantic contexts. However, we define semantic contexts inductively rather than recursively, for technical convenience.

```
data Conᴺ : Con → Con → Set where
  •   : Conᴺ • Δ
  _,_ : Conᴺ Γ Δ → Tyᴺ A Δ → Conᴺ (Γ , A) Δ

∈ᴺ : ∀ {Γ A} → A ∈ Γ → ∀ {Δ} → Conᴺ Γ Δ → Tyᴺ A Δ
∈ᴺ vz     (Γᴺ , tᴺ) = tᴺ
∈ᴺ (vs v) (Γᴺ , _ ) = ∈ᴺ v Γᴺ
```

When interpreting terms, variables and applications are easy. In the latter case we just apply the recursive result to `idₑ` - we can choose *not* to conjure up fresh variables if there is no need to. We run into a bit of a bump in the interpretation of lambdas:

```
Tmᴺ : ∀ {Γ A} → Tm Γ A → ∀ {Δ} → Conᴺ Γ Δ → Tyᴺ A Δ
Tmᴺ (var v)   Γᴺ = ∈ᴺ v Γᴺ
Tmᴺ (lam t)   Γᴺ = λ σ aᴺ → Tmᴺ t (? , aᴺ)  -- ? marks the bump
Tmᴺ (app f a) Γᴺ = Tmᴺ f Γᴺ idₑ (Tmᴺ a Γᴺ)
```

Clearly, the idea is to evaluate the inner `t` term in `(Γᴺ , aᴺ)` as we did in the standard model, but `Γᴺ` does not have the right type. It has type `Conᴺ Γ Δ`, but we need to return in some `Σ` extended context, with `(σ : OPE Σ Δ)` witnessing the extension. So, what we need is an action of context embedding on semantic contexts, and on semantic terms as well, since the former are just lists of the latter. To implement it, we first need *composition* for context embeddings:

```
_∘ₑ_ : OPE Δ Σ → OPE Γ Δ → OPE Γ Σ
σ      ∘ₑ •      = σ
σ      ∘ₑ drop δ = drop (σ ∘ₑ δ)
drop σ ∘ₑ keep δ = drop (σ ∘ₑ δ)
keep σ ∘ₑ keep δ = keep (σ ∘ₑ δ)
```

It witnesses transitivity for the embedding relation. Then, embeddings on semantic terms is easy enough:

```
Tyᴺₑ : OPE Δ Γ → Tyᴺ A Γ → Tyᴺ A Δ
Tyᴺₑ {ι}     σ tᴺ = Nfₑ σ tᴺ
Tyᴺₑ {A ⇒ B} σ tᴺ = λ δ aᴺ → tᴺ (σ ∘ₑ δ) aᴺ

Conᴺₑ : OPE Σ Δ → Conᴺ Γ Δ → Conᴺ Γ Σ
Conᴺₑ σ •         = •
Conᴺₑ σ (Γᴺ , tᴺ) = Conᴺₑ σ Γᴺ , Tyᴺₑ σ tᴺ
```

For `Tyᴺₑ {ι}`, we use embedding of normal forms. For `Tyᴺₑ {A ⇒ B}`, we compose the externally given `σ` embedding with the input embedding `δ`. `Conᴺₑ` is just pointwise application of `Tyᴺₑ`. `Tmᴺ` can be now given as:

```
Tmᴺ : ∀ {Γ A} → Tm Γ A → ∀ {Δ} → Conᴺ Γ Δ → Tyᴺ A Δ
Tmᴺ (var v)   Γᴺ = ∈ᴺ v Γᴺ
Tmᴺ (lam t)   Γᴺ = λ σ aᴺ → Tmᴺ t (Conᴺₑ σ Γᴺ , aᴺ)
Tmᴺ (app f a) Γᴺ = Tmᴺ f Γᴺ idₑ (Tmᴺ a Γᴺ)
```

We still need a `quote` function to transform semantic terms to normal terms. We also need to mutually define an `unquote` function which sends neutral terms to semantic terms, for reasons shortly illuminated. Then, normalization is given as evaluation followed by quotation, where evaluation uses a semantic context given by unquoting each variable in the context.

```
mutual
  qᴺ : ∀ {A Γ} → Tyᴺ A Γ → Nf Γ A
  qᴺ {ι}     tᴺ = tᴺ
  qᴺ {A ⇒ B} tᴺ = lam (qᴺ (tᴺ wk (uᴺ (var vz))))

  uᴺ : ∀ {A Γ} → Ne Γ A → Tyᴺ A Γ
  uᴺ {ι}     n = ne n
  uᴺ {A ⇒ B} n = λ σ aᴺ → uᴺ (app (Neₑ σ n) (qᴺ aᴺ))

uᶜᴺ : ∀ {Γ} → Conᴺ Γ Γ
uᶜᴺ {•}     = •
uᶜᴺ {Γ , A} = Conᴺₑ wk uᶜᴺ , uᴺ (var vz)

nf : ∀ {Γ A} → Tm Γ A → Nf Γ A
nf t = qᴺ (Tmᴺ t uᶜᴺ)
```

At first, it can be difficult to build a mental model of the operation of the algorithm. On the highest level, normalization alternates between evaluation and quoting: a term is first evaluated, then if it has a function type, its semantic value is applied to a "blocking" input and we quote the result, which can be a semantic function again. This evaluate-quote alternation proceeds until the result is not a function but a base value. Note that although `qᴺ` does not directly call `Tmᴺ`, it does apply semantic functions, and this application results in an *internal* call to `Tmᴺ`, accordingly to the definition of semantic application.

In fact, this encapsulation of recursive `Tmᴺ` calls is a crucial detail which makes this whole definition structurally recursive and thus total. In the `(Tmᴺ (lam t) Γᴺ = λ σ aᴺ → Tmᴺ t (Conᴺₑ σ Γᴺ , aᴺ))` clause, the recursive `Tmᴺ` call is evidently structural, and because it happens under a *metatheoretic* lambda, semantic functions can be applied to *any* value in any definition without compromising structurality.

To explain the previous scare quotes around "blocking": in the `(tᴺ wk (uᴺ (var vz)))` application, the `(uᴺ (var vz))` term plays the role of blocking input, but it is not quite entirely blocking. `uᴺ` performs an operation best described as *semantic η-expansion*: it acts as identity on neutral base terms, and from neutral function it produces semantic function which build up neutral applications from their inputs. Thus, `(uᴺ {ι} (var vz))` simply reduces to `(ne (var vz))`, while `(uᴺ {ι ⇒ i} (var vz))` reduces to `(λ σ aᴺ → ne (app (var (∈ₑ σ vz)) aᴺ))`. In short, `uᴺ` returns semantic values which yield properly η-expanded normal forms when quoted. As convoluted this may seem, it is actually a very elegant solution.

4.4 Kripke Models

We show now how Kirpke models relate to the above definition of normalization. Following the presentation of Altenkirch [2], Kripke models can be defined in Agda for the syntax of STLC as follows (omitting universe polymorphism):

```
-- Misc/Kripke.agda
record KripkeModel : Set₁ where
  field
    W       : Set
    _≤_     : W → W → Set
    ≤refl   : ∀ {w} → w ≤ w
    _≤■_    : ∀ {w₁ w₂ w₃} → w₁ ≤ w₂ → w₂ ≤ w₃ → w₁ ≤ w₃
    _||-ι   : W → Set
    ι-mono  : ∀ {w₁ w₂} → w₁ ≤ w₂ → w₁ ||-ι → w₂ ||-ι
```

A Kripke model consists of a preordered set of "worlds", denoted `W`, along with a forcing predicate for the base type, and a *monotonicity condition* `ι-mono`. On a concrete level, we can obtain the definition of `KripkeModel` by noticing the details in our previous `N` model which can be abstracted out, and packing them into a record. Indeed, the interpretation of function types, contexts and terms can be derived from this amount of data. Agda allow us to include them in the record as well, as definitions which depend on the record fields:

```
record KripkeModel : Set₁ where
  field
    ... -- the same as before

  _||-Ty_ : W → Ty → Set
```

```
w ||-Ty ι       = w ||-ι
w ||-Ty (A ⇒ B) = ∀ {w'} → w ≤ w' → w' ||-Ty A → w' ||-Ty B

_||-Con_ : W → Con → Set
w ||-Con •       = ⊤
w ||-Con (Γ , A) = (w ||-Con Γ) × (w ||-Ty A)

_||-_ : Con → Ty → Set
Γ ||- A = ∀ w → w ||-Con Γ → w ||-Ty A

Ty-mono : ∀ {A w w'} → w ≤ w' → w ||-Ty A → w' ||-Ty A
Ty-mono {ι}     σ p = ι-mono σ p
Ty-mono {A ⇒ B} σ p δ q = p (σ ≤■ δ) q

Con-mono : ∀ {Γ}{w w'} → w ≤ w' → w ||-Con Γ → w' ||-Con Γ
Con-mono {•}     σ p = p
Con-mono {Γ , A} σ p = Con-mono σ (proj₁ p) , Ty-mono {A} σ (proj₂ p)

||-Tm : ∀ {Γ A} → Tm Γ A → Γ ||- A
||-Tm (var vz)     w (_ , q) = q
||-Tm (var (vs v)) w (p , q) = ||-Tm (var v) w p
||-Tm (lam t)      w p σ a   = ||-Tm t _ (Con-mono σ p , a)
||-Tm (app f a)    w p       = ||-Tm f w p ≤refl (||-Tm a w p)
```

With this, everything up to $\mathtt{Tm}^{\mathtt{N}}$ in our previous implementation can be obtained by defining the corresponding model:

```
N : KripkeModel
N = record {
    W      = Con
  ; _≤_    = λ Γ Δ → OPE Δ Γ
  ; ≤refl  = idₑ
  ; _≤■_   = _∘ₑ_
  ; _||-ι  = λ Γ → Nf Γ ι
  ; ι-mono = Nfₑ }
```

This is an interesting exercise in abstraction, but what is the deeper significance? STLC can be viewed as a simple intuitionistic propositional logic, with a single atomic propopsition ι. What is a right notion of semantics, though? We could abstract out the interpetation of base types from the standard model as given in Section 4.2, and get another notion of models; would not that suffice? From a logician's point of view, the issue is that it does not give us completeness. Semantics in general is necessary because we use logic to talk about concepts about which we have underlying intuition, but we have no *a priori* reason to believe that syntactic constructions talk meaningfully about any intuitive concept.

Soundness and completeness together express that syntactic and semantic entailment

are logically equivalent; the lack of completeness indicates that there is a mismatch, that semantics is larger than necessary. In Agda, we define soundness and completeness for Kripke models as follows[6]:

```
sound    = ∀ {Γ A} → Tm Γ A → (∀ M → let open KripkeModel M in Γ ||- A)
complete = ∀ {Γ A} → (∀ M → let open KripkeModel M in Γ ||- A) → Tm Γ A
```

The `let open KripkeModel M` incantation makes locally available all fields and definitions of `M`. We give here a brief informal proof for soundness and completeness.

Theorem. *STLC is sound and complete with respect to Kripke models.*

Proof. Soundness follows immediately from `||-Tm`. For completeness, we instantiate the hypothesis with the previously defined `N` model, then define quoting and unquoting the same way as in Section 4.3. Using them, we can produce a `(Nf Γ A)`, which can be canonically injected back to `(Tm Γ A)`. □

With the standard set-theoretic semantics for first-order logic, the intuitive content is clear: formulas talk about truth and falsehood. What is an intuitive meaning of Kripke semantics, though? A possible answer is that it talks about *knowledge*. `W` worlds can be understood as states of knowledge, and `_≤_` denotes increasing knowledge. Monotonicity expresses that if something is proven in some state of knowledge, it will never be invalidated, not matter how "wiser" we get.

Note that in the formal development the direct `ᴺ` definitions are used instead of Kripke models. This is primarily to make the algorithm as transparent as possible to lay readers. In the `Normalization.agda` file, the core algorithm is laid out in about 50 lines, and together with the definition of syntax and embedding it is about a hundred lines.

5 Correctness

In this section, we prove completeness, soundness and stability for the algorithm.

5.1 Categorical Notions

First, we shall introduce a modest number of concepts from category theory. The prime motivation is that they allow us to compactly describe the contents of our proofs. They also make it easier to mentally keep track of proof obligations, since often a few words of categorical definitions can be unpacked into a dozen lemmas. When proofs align with category theory, that is usually also indicative that one has the right approach. That said, we shall keep category theory to a minimum and only introduce definitions that are directly applicable to our work.

[6]Note that these are wholly different from soundness and completeness as defined for a normalization algorithm in Section 1.1.

A *category* is defined as follows.

```
-- Misc/Category.agda
record Category : Set₁ where
  field
    Obj   : Set             -- "objects"
    morph : Obj → Obj → Set -- "morphisms"
    id    : morph I I
    _∘_   : morph J K → morph I J → morph I K
    idl   : f ∘ id ≡ f
    idr   : id ∘ f ≡ f
    ass   : f ∘ (g ∘ h) ≡ (f ∘ g) ∘ h
```

In short, categories have a set of objects and sets of morphisms between objects, with identity morphisms, composition, and associativity and identity laws for composition. Note that this is a naive definition which does not take into account size issues and subtleties arising from the ambient type theory; for a more rigorous treatment see [31, pp. 319].

In written exposition, $obj : C$ expresses that obj is an object of C, while $f : C(A, B)$ means that f is a morphism of C from A to B.

Most importantly, the syntax of STLC itself forms a category, which is sometimes called the *syntacic category*. We denote it as **STLC**. Its objects are contexts, its morphisms are substitutions with the identity substitution as identity morphism. We have not yet defined composition for substitution, which provides composition for **STLC**, nor the composition laws. Providing these will be a significant part of our proof obligations about substitutions.

Another important category is the category of sets, denoted **Set**. In our setting, **Set** has as objects Agda types and functions as morphisms (again, ignoring subtleties).

Functors are structure-preserving mappings between categories:

```
record Functor (C D : Category) : Set₁ where
  field
    Obj⇒   : Obj C → Obj D
    morph⇒ : morph I J → morph (Obj⇒ I) (Obj⇒ J)
    id⇒    : morph⇒ id ≡ id
    ∘⇒     : morph⇒ (f ∘ g) ≡ morph⇒ f ∘ morph⇒ g
```

We use $F : C \to D$ to express that F is a functor from C to D. Functors map objects to objects and morphisms to morphisms. Also, they must map identities to identities and compositions to compositions. We will use the terms *action on objects* and *action on morphisms* to denote the first two, and *functor laws* for the latter two.

A *contravariant functor* flips morphisms, i. e. it has

```
morph⇒ : morph I J → morph (Obj⇒ J) (Obj⇒ I))
```

The composition law changes accordingly. Contravariant functors are often denoted as $F : C^{op} \to D$ (meaning that F's domain is the *opposite* category of C, where all morphisms are flipped). Regular non-flipping functors are sometimes called *covariant functors*.

A *presheaf* is a contravariant functor from some C category to **Set**, i. e. it maps objects to sets and morphisms to functions, with morphism arrows being flipped in the process.

Natural transformations are mappings between $F, G : C \to D$ functors. They consist of a ϕ family of morphisms for each I object of C, and a *naturality* condition, which expresses that ϕ commutes with any lifted C-morphism.

```
record Nat {C D : Category}(F G : Functor C D) : Set₁ where
  field
    Φ   : {I : Obj C} → morph (Obj⇒ F I) (Obj⇒ G I)
    nat : {I J : Obj C}{f : morph I J} → ψ ∘ morph⇒ F f ≡ morph⇒ G f ∘ ψ
```

$F : C \to D$ functors themselves form a category, where functors are objects and natural transformations are morphisms.

This development mostly utilizes presheaves and natural transformations between presheaves. More concretely, our setting will be mostly $PSh(\mathbf{OPE})$, the category of presheaves on **OPE**, where **OPE** is the category of contexts and order-preserving context embeddings. We examine **OPE** in detail in the next section.

5.2 Laws for Embedding and Substitution

Before we attempt to prove correctness we should pin down the laws of embedding and substitution. People who set out to write normalization proofs soon find that a jumble of twenty-odd substitution lemmas is required to make headway. The naive proving process works by repeatedly hitting roadblocks and reacting by adding more lemmas. A more prudent way is to characterize all of them beforehand in categorical terms, which lends us confidence that a given set of lemmas is complete.

5.2.1 Embeddings

Contexts and order-preserving embeddings form a category, which we call **OPE**. Objects are contexts, morphisms are elements of the `OPE` type. Identity and composition are as previously defined in Section 3.2.1 and Section 4.3:

```
idₑ  : ∀ {Γ} → OPE Γ Γ
_∘ₑ_ : OPE Δ Σ → OPE Γ Δ → OPE Γ Σ
```

The category laws can be proven with simple induction.

```
idlₑ : idₑ ∘ₑ σ ≡ σ                  -- left identity
idrₑ : σ ∘ₑ idₑ ≡ σ                  -- right identity
assₑ : (σ ∘ₑ δ) ∘ₑ ν ≡ σ ∘ₑ (δ ∘ₑ ν) -- associativity
```

Additionally, variables and terms of a given type are presheaves on **OPE**. `(λ Γ → A ∈ Γ)` maps each object of **OPE** to an object of **Set**, that is, an Agda type. Likewise does `(λ Γ → Tm Γ A)`. They also have action on morphisms: these are precisely the embedding operations on variables and terms, defined in Section 3.2.1.

```
∈ₑ  : OPE Γ Δ → (A ∈ Δ → A ∈ Γ)
Tmₑ : OPE Γ Δ → (Tm Δ A → Tm Γ A)
```

Note that the input is a morphism in **OPE**, i. e. a context embedding, and the output is a morphism in **Set**, which is an Agda function. Also note the contravariance: the order of `Γ` and `Δ` is flipped in the output. Functor laws for variables and terms are as follows:

```
-- Embedding.agda
∈-idₑ  : ∀ v → ∈ₑ idₑ v ≡ v
∈-∘ₑ   : ∀ σ δ v → ∈ₑ (σ ∘ₑ δ) v ≡ ∈ₑ δ (∈ₑ σ v)
Tm-idₑ : ∀ t → Tmₑ idₑ t ≡ t
Tm-∘ₑ  : ∀ σ δ t → Tmₑ (σ ∘ₑ δ) t ≡ Tmₑ δ (Tmₑ σ t)
```

Here, the functor laws are given in a pointwise form. The categorically precise type of `∈-idₑ` would be the following:

```
∈-idₑ : ∈ₑ idₑ ≡ (λ v → v)
```

This expresses that `idₑ` is mapped to the identity function in **Set**. Instead of using this definition, we apply both sides of the equation to a variable, and similarly transform the types for the other laws. The reason for this is that proving the categorically precise forms requires function extensionality which we would like to avoid when possible.

Normal and neutral terms are also presheaves on **OPE** in a similar manner:

```
-- NormalForm.agda
Nfₑ    : OPE Γ Δ → Nf Δ A → Nf Γ A
Neₑ    : OPE Γ Δ → Ne Δ A → Ne Γ A
Nf-∘ₑ  : ∀ σ δ t → Nfₑ (σ ∘ₑ δ) t ≡ Nfₑ δ (Nfₑ σ t)
Ne-∘ₑ  : ∀ σ δ t → Neₑ (σ ∘ₑ δ) t ≡ Neₑ δ (Neₑ σ t)
Nf-idₑ : ∀ t → Nfₑ idₑ t ≡ t
Ne-idₑ : ∀ t → Neₑ idₑ t ≡ t
```

This is to be expected, since neutral and normal terms are just terms with restricted structure.

5.2.2 Substitutions

We previously introduced **STLC** as the syntactic category, containing contexts as objects and substitutions as morphisms. However, we then presented **OPE** as yet another category with contexts as objects. Why single out **STLC** as the category which characterizes syntax? Well, the conversion relation is an integral part of the syntax, and **OPE** does not contain the morphisms needed to define β-conversion. In contrast, **STLC** contains all substitutions as well as all embeddings, since embeddings can be viewed as restricted substitutions. **OPE** is a *wide subcategory* of **STLC**: it contains all the objects of **STLC** but only some of its morphisms.

We review identity substitution here (previously defined in Section 3.2.2). Composition is elementwise substitution of terms in a substitution.

```
idₛ : ∀ {Γ} → Sub Γ Γ
idₛ {•}     = •
idₛ {Γ , A} = keepₛ idₛ

_∘ₛ_ : Sub Δ Σ → Sub Γ Δ → Sub Γ Σ
•       ∘ₛ δ = •
(σ , t) ∘ₛ δ = (σ ∘ₛ δ) , Tmₛ δ t
```

We also have substitution operations for terms and variables:

```
∈ₛ  : Sub Γ Δ → A ∈ Δ → Tm Γ A
Tmₛ : Sub Γ Δ → Tm Δ A → Tm Γ A
```

We aim to establish that **STLC** is a category, and that `(Tm _ A)` and `(A ∈ _)` are presheaves on it. This is significantly more complicated than what we have seen for **OPE**. Identity and composition for substitutions are actually defined using embeddings, thus proving properties of substitution involves a proving a variety of lemmas about their interaction with embedding.

The categorical structure of proof obligations is less clear here; although we have clear goals (category and presheaf laws), this author is unaware of a compact and abstract explanation of the process of building **STLC** around **OPE**. In [9], substitution laws are constructed the same way as here, but the authors of Ibid. also do not provide a semantic explanation. For contrast, see [5], where substitutions are given first as part of the syntax, and renamings (an alternative of embeddings) are defined later as a subcategory.

There are four different operations of composing embeddings and substitutions, since there are two choices for both arguments. We define the missing `_∘ₑ∘ₛ_` operation here:

```
_∘ₑ_   : OPE Δ Σ → OPE Γ Δ → OPE Γ Σ -- defined in 4.3
_∘ₛ_   : Sub Δ Σ → Sub Γ Δ → Sub Γ Σ -- defined in this section
_ₛ∘ₑ_ : Sub Δ Σ → OPE Γ Δ → Sub Γ Σ -- defined in 3.2.2

_∘ₑ∘ₛ_ : OPE Δ Σ → Sub Γ Δ → Sub Γ Σ
```

```
•       ₑ∘ₛ δ       = δ
drop σ ₑ∘ₛ (δ , t) = σ ₑ∘ₛ δ
keep σ ₑ∘ₛ (δ , t) = (σ ₑ∘ₛ δ) , t
```

There are variations of the category laws where occurrences of composition can be any of the four versions, and it appears that some (not all) of these laws have to be proven. Similarly, functor laws for terms and variables have multiple versions differing in the mapped composition. All in all, we need to prove the following theorems in order:

```
-- Substitution.agda
assₛₑₑ : ∀ σ δ ν → (σ ₛ∘ₑ δ) ₛ∘ₑ ν ≡ σ ₛ∘ₑ (δ ∘ₑ ν)
assₑₛₑ : ∀ σ δ ν → (σ ₑ∘ₛ δ) ₛ∘ₑ ν ≡ σ ₑ∘ₛ (δ ₛ∘ₑ ν)

idlₑₛ : ∀ σ → idₑ ₑ∘ₛ σ ≡ σ
idlₛₑ : ∀ σ → idₛ ₛ∘ₑ σ ≡ ⌜ σ ⌝ope
idrₑₛ : ∀ σ → σ ₑ∘ₛ idₛ ≡ ⌜ σ ⌝ope

∈-ₑ∘ₛ  : ∀ σ δ v → ∈ₛ (σ ₑ∘ₛ δ) v ≡ ∈ₛ δ (∈ₑ σ v)
Tm-ₑ∘ₛ : ∀ σ δ t → Tmₛ (σ ₑ∘ₛ δ) t ≡ Tmₛ δ (Tmₑ σ t)

∈-ₛ∘ₑ  : ∀ σ δ v → ∈ₛ (σ ₛ∘ₑ δ) v ≡ Tmₑ δ (∈ₛ σ v)
Tm-ₛ∘ₑ : ∀ σ δ t → Tmₛ (σ ₛ∘ₑ δ) t ≡ Tmₑ δ (Tmₛ σ t)

assₛₑₛ : ∀ σ δ ν → (σ ₛ∘ₑ δ) ∘ₛ ν ≡ σ ∘ₛ (δ ₑ∘ₛ ν)
assₛₛₑ : ∀ σ δ ν → (σ ∘ₛ δ) ₛ∘ₑ ν ≡ σ ∘ₛ (δ ₛ∘ₑ ν)

-- Functor laws for (A ∈ _) and (Tm _ A)
∈-∘ₛ   : ∀ σ δ v → ∈ₛ (σ ∘ₛ δ) v ≡ Tmₛ δ (∈ₛ σ v)
Tm-∘ₛ  : ∀ σ δ t → Tmₛ (σ ∘ₛ δ) t ≡ Tmₛ δ (Tmₛ σ t)
∈-idₛ  : ∀ v → ∈ₛ idₛ v ≡ var v
Tm-idₛ : ∀ t → Tmₛ idₛ t ≡ t

-- Category laws for STLC
idrₛ   : ∀ σ → σ ∘ₛ idₛ ≡ σ
idlₛ   : ∀ σ → idₛ ∘ₛ σ ≡ σ
assₛ   : ∀ σ δ ν → (σ ∘ₛ δ) ∘ₛ ν ≡ σ ∘ₛ (δ ∘ₛ ν)
```

The proofs are moderately interesting and involve straightforward induction and equational reasoning. They comprise about 110 lines of Agda.

5.3 Completeness

In this section we prove completeness of normalization. It expresses that the output of normalization is convertible to its input.

```
complete : ∀ {Γ A}(t : Tm Γ A) → t ~ ⌜ nf t ⌝Nf
```

`nf` refers to the function defined in Section 4.3. In the return type, the injection `⌜_⌝Nf` converts `(Nf Γ A)` back to `(Tm Γ A)`.

Since `nf` is defined with a Kripke model, it is clear that proving properties about it have to involve similar model structures. The overall structure of the proof is very similar to normalization itself; we will have interpretation of types, contexts, terms, and quoting and unquoting. In general, proofs about properties of functions must have the same inductive "shape" as the functions in question. The main difference here is that types are interpreted as *relations*:

```
-- Completeness.agda
_≈_ : ∀ {A Γ} → Tm Γ A → Tyᴺ A Γ → Set
_≈_ {ι}        t tᴺ = t ~ ⌜ tᴺ ⌝Nf
_≈_ {A ⇒ B}{Γ} t tᴺ =
    ∀ {Δ}(σ : OPE Δ Γ){a aᴺ} → a ≈ aᴺ → app (Tmₑ σ t) a ≈ tᴺ σ aᴺ
```

`_≈_` is a *Kripke logical relation*. Here, "logical relation" simply means that it is a relation defined by induction on types. "Kripke" indicates that the interpretation of function types is generalized to work in extended contexts. The purpose of the generalization is the same as with evaluation: it provides us with a "fresh variable" supply, allowing us to quote semantic functions and go under binders.

`_≈_` essentially extends conversion to relate syntactic and semantic terms. At the base type, it is precisely conversion; at function types, it expresses that syntactic and semantic application respect `_≈_`.

We extend `_≈_` to a relation between substitutions and semantic contexts:

```
data _≈ᶜ_ {Γ} : ∀ {Δ} → Sub Γ Δ → Conᴺ Δ Γ → Set where
  •   : • ≈ᶜ •
  _,_ : σ ≈ᶜ δ → t ≈ t' → (σ , t) ≈ᶜ (δ , t')
```

Substitutions are list of syntactic terms while semantic context are lists of semantic terms, so they can be elementwise related by `_≈_`.

Mirroring `ᴺ`, we need monotonicity for semantic terms and contexts:

```
≈ₑ  : ∀ {A : Ty} σ → t ≈ tᴺ → Tmₑ σ t ≈ Tyᴺₑ σ tᴺ
≈ᶜₑ : ∀ σ → δ ≈ᶜ νᴺ → δ ₛ∘ₑ σ ≈ᶜ Conᴺₑ σ νᴺ
```

The only interesting case here is the base type:

```
≈ₑ {ι} σ t≈tᴺ = ?
```

Here, the goal has type `(Tmₑ σ t ~ ⌜ Nfₑ σ tᴺ ⌝Nf)`, and we have `(t≈tᴺ : t ~ ⌜ tᴺ ⌝Nf)`. First, we witness `(Tmₑ σ t ~ Tmₑ σ ⌜ tᴺ ⌝Nf)` by showing that embedding respects conversion:

```
-- Embedding.agda
~ₑ : ∀ σ → t ~ t' → Tmₑ σ t ~ Tmₑ σ t'
```

This can be proven by induction on `(t ~ t')`, with the help of substitution laws defined in Section 5.2. Now, `(~ₑ σ t≈tᴺ : Tmₑ σ t ~ Tmₑ σ ⌜ tᴺ ⌝Nf)`, and the type can be shown to be equal to the goal using `⌜⌝Nf-nat`.

The only needed lemma which has no direct counterpart in `ᴺ` is the following:

```
_~■≈_ : ∀ {A Γ}{t t'}{tᴺ : Tyᴺ A Γ} → t ~ t' → t' ≈ tᴺ → t ≈ tᴺ
_~■≈_ {ι}     p q = p ~■ q
_~■≈_ {A ⇒ B} p q = λ σ a≈aᴺ → app (~ₑ σ p) ~refl ~■≈ q σ a≈aᴺ
```

Now, the interpretation of terms can be defined. This is often called the "fundamental theorem" of a logical relation interpretation.

```
Tm≈ : ∀ (t : Tm Γ A){σ}{δᴺ : Conᴺ Γ Δ} → σ ≈ᶜ δᴺ → Tmₛ σ t ≈ Tmᴺ t δᴺ
```

The proof is by induction on the input term. Witnesses for variables are simply looked up from `(σ ≈ᶜ δᴺ)`, the same way as with other models. If the input term is an application, we use the inductive hypotheses the same way as in `Tmᴺ`, although first we need to rewrite the goal type with a functor law:

```
Tm≈ (app f a) {σ} σ≈δᴺ
  rewrite Tm-idₑ (Tmₛ σ f) ⁻¹ = Tm≈ f σ≈δᴺ idₑ (Tm≈ a σ≈δᴺ)
```

Only the case for `(lam t)` is non-trivial:

```
Tm≈ (lam t) {σ} {δ} σ≈δᴺ ν {a} {aᴺ} a≈aᴺ = ?
```

The goal is

```
app (lam (Tmₑ (keep ν) (Tmₛ (keepₛ σ) t))) a ≈ Tmᴺ t (Conᴺₑ ν δᴺ , aᴺ)
```

By using the induction hypotheses, we get the following:

```
Tm≈ t (≈ᶜₑ ν σ≈δᴺ , a≈aᴺ)
  : Tmₛ ((σ ₛ∘ₑ ν) , a) t ≈ Tmᴺ t (Conᴺₑ ν δᴺ , aᴺ)
```

The left hand side of the `_≈_` does not yet match the goal. We need to use `_~■≈_` to compose it with a conversion proof on the left. The β rule can be used, but the result type does not line up on the right hand side:

```
β (Tmₑ (keep ν) (Tmₛ (keepₛ σ) t)) a
  : app (lam (Tmₑ (keep ν) (Tmₛ (keepₛ σ) t))) a ~
    Tmₛ (idₛ , a) (Tmₑ (keep ν) (Tmₛ (keepₛ σ) t))
```

Hence we need to coerce this type to the desired one, using embedding and substitution laws. We omit the equality proof here.[7]

```
Tm≈ (lam t) {σ} {δᴺ} σ≈δᴺ ν {a} {aᴺ} a≈aᴺ =
  coe eq (β (Tmₑ (keep ν) (Tmₛ (keepₛ σ) t)) a) -- eq omitted
```

[7]Generally, explicit reasoning about substitutions is omitted from textbooks as well, because it tends to be technical and uninteresting.

```
  ~■≈
  Tm≈ t (≈ᶜₑ ν σ≈δᴺ , a≈aᴺ)
```

Now, quoting, unquoting and completeness can be defined as follows:

```
mutual
  q≈ : ∀ {A} → t ≈ tᴺ → t ~ ⌜ qᴺ tᴺ ⌝Nf
  q≈ {ι}     t≈tᴺ = t≈tᴺ
  q≈ {A ⇒ B} t≈tᴺ = η _ ~■ lam (q≈ (t≈tᴺ wk (u≈ (var vz))))

  u≈ : ∀ {A}(n : Ne Γ A) → ⌜ n ⌝Ne ≈ uᴺ n
  u≈ {ι}     n = ~refl
  u≈ {A ⇒ B} n σ {a} {aᴺ} a≈aᴺ
    rewrite ⌜⌝Ne-nat σ n ⁻¹ =
      app ~refl (q≈ a≈aᴺ) ~■≈ u≈ (app (Neₑ σ n) (qᴺ aᴺ))

uᶜ≈  : ∀ {Γ} → idₛ {Γ} ≈ᶜ uᶜᴺ
uᶜ≈ {•}     = •
uᶜ≈ {Γ , A} = ≈ᶜₑ wk uᶜ≈ , u≈ (var vz)

complete : ∀ {Γ A}(t : Tm Γ A) → t ~ ⌜ nf t ⌝Nf
complete t = coe eq (q≈ (Tm≈ t uᶜ≈)) -- eq omitted
```

Since, `nf` was defined as evaluation followed by quotation, we define `complete` this way as well. Expanding the type of `complete`, we get:

```
complete : ∀ {Γ A}(t : Tm Γ A) → t ~ ⌜ qᴺ (Tmᴺ t uᶜᴺ) ⌝Nf
```

However, `(q≈ (Tm≈ t uᶜ≈))` has type `(Tmₛ idₛ t ~ ⌜ qᴺ (Tmᴺ t uᶜᴺ) ⌝Nf)`, hence we need an extra coercion to rewrite the result type along `Tm-idₛ`. This completes the proof. □

Now, since a large part of the proof was just reiterating the structure of ᴺ, would not it be better to implement normalization and completeness in one go? For instance, the following could be implemented:

```
nf : ∀ {Γ A}(t : Tm Γ A) → Σ (Nf Γ A) λ n → t ~ ⌜ n ⌝Nf
```

However, this has a disadvantage: the proof terms threaded through add considerable noise to any subsequent proof about normalization. It is also contrary to our goal of implementing the algorithm as simply as possible.

5.4 Presheaf Refimenent

For soundness, one might be tempted to try directly defining another logical relation, and proceeding similarly as in the previous section. That approach would fall short. The reason is that at some point one has to prove naturality for evaluation, quoting and unquoting, i. e. that they commute with embeddings.

```
Tmᴺ-nat : ∀ t σ Γᴺ → Tmᴺ t (Conᴺₑ σ Γᴺ) ≡ Tyᴺₑ σ (Tmᴺ t Γᴺ)
qᴺ-nat  : ∀ σ tᴺ → Nfₑ σ (qᴺ tᴺ) ≡ qᴺ (Tyᴺₑ σ tᴺ)
uᴺ-nat  : ∀ σ n → Tyᴺₑ σ (uᴺ n) ≡ uᴺ (Neₑ σ n)
```

Unfortunately, these are not actually provable, because `Tyᴺ` is too large: it contains "unnatural" semantic terms. Recall its definition:

```
Tyᴺ : Ty → Con → Set
Tyᴺ ι       Γ = Nf Γ ι
Tyᴺ (A ⇒ B) Γ = ∀ {Δ} → OPE Δ Γ → Tyᴺ A Δ → Tyᴺ B Δ
```

`(Tyᴺ ι Γ)` is always fine, but `(Tyᴺ (A ⇒ B) Γ)` contains ill-behaved values, for instance a semantic function which returns different values depending on the length of the input context `Δ`. Such a function clearly does not commute with embedding. We would like to restrict semantic functions to a a well-behaved subset.

A *presheaf model* of STLC interprets types and contexts as presheaves. In particular, it interprets function types as *presheaf exponentials* [15] whose definition expands to a semantic function together with the naturality condition we are looking for. There is a presheaf model over **OPE** which can be used for normalization and which is a straightforward refinement of the Kripke model `ᴺ`. Recall that a presheaf is a contravariant set-valued functor, hence the interpretation of types has to include four components: action on objects, action on morphisms and the two functor laws. Below, for the interpretation of types the action on objects is defined and we give the types of the other components:

```
-- action on objects
Tyᴾ : Ty → Con → Set
Tyᴾ ι       Γ = Nf Γ ι
Tyᴾ (A ⇒ B) Γ =
  Σ (∀ {Δ} → OPE Δ Γ → Tyᴾ A Δ → Tyᴾ B Δ) λ fᴾ →
  ∀ {Δ Σ}(σ : OPE Δ Γ)(δ : OPE Σ Δ) aᴾ
    → fᴾ (σ ∘ₑ δ) (Tyᴾₑ δ aᴾ) ≡ Tyᴾₑ δ (fᴾ σ aᴾ)

-- action on morphisms
Tyᴾₑ    : OPE Δ Γ → Tyᴾ A Γ → Tyᴾ A Δ

-- functor laws
Tyᴾ-idₑ : ∀ tᴾ → Tyᴾₑ idₑ tᴾ ≡ tᴾ
Tyᴾ-∘ₑ  : ∀ tᴾ σ δ → Tyᴾₑ (σ ∘ₑ δ) tᴾ ≡ Tyᴾₑ δ (Tyᴾₑ σ tᴾ)
```

The key change is in the interpretation of function types, where we have a `Σ` of a function and a naturality condition.

Note that the action on morphisms was already present in the Kripke model under the name of "monotonicity condition". In fact, every presheaf model is a Kripke model (but not vice versa). In Kripke models, types are also interpreted as functors, but from a *preordered set* rather than a general category.

Preordered sets can be viewed as categories with at most a single morphism between any two objects [16]. If $I \leq J$, then there is a morphism between I and J, otherwise there is no such morphism. Identity and composition of morphisms comes from reflexivity and transitivity of preordering.

In a Kripke model, functor laws hold by default, since the uniqueness of morphisms implies that identities necessarily map to identities and compositions to compositions. However, our **OPE** as it is defined in Agda is actually not a preorder. It can be informally considered as such if we only care about the mere existence of an embedding between two contexts. Formally, we could use *propositional truncation* [31, pp. 117] on the sets of morphisms of **OPE** to get an actual preorder, but that is beyond the capabilities of Agda and the scope of this article.

In the presheaf model, terms are interpreted as morphisms from semantic contexts to semantic types, in other words, natural transformations. Concretely, this means that the evaluation function[8] should be mutually defined with the proof that it commutes with embeddings, i. e. its naturality. Similarly, quoting and unquoting are defined mutually with their naturality proofs.

In this section, we do not use the $^{\mathsf{P}}$ model - it is described in Section 6.1 in more detail. Instead, the previous definition of normalization using $^{\mathsf{N}}$ is kept unchanged, but we shall need additional machinery to prove the necessary naturalities. The solution is the following: first, we define a logical predicate on semantic values which restricts them to natural inhabitants. Then, we give all naturality proofs, using restricted semantic values when required, and also prove that the $\mathtt{Tm}^{\mathsf{N}}$ evaluation function outputs natural values.

The naturality predicate is defined below. We use superscript $^{\mathsf{P}}$ notation for definitions related to this predicate. Here, $^{\mathsf{P}}$ means "presheaf refinement", not "presheaf model".

```
-- PresheafRefinement.agda
Ty^P : ∀ {Γ A} → Ty^N A Γ → Set
Ty^P {Γ} {ι}     t^N = ⊤
Ty^P {Γ} {A ⇒ B} t^N =
  ∀ {Δ}(σ : OPE Δ Γ) {a^N : Ty^N A Δ}
  → Ty^P a^N
  → (∀ {Σ} (δ : OPE Σ Δ) → t^N (σ ∘e δ) (Ty^Ne δ a^N) ≡ Ty^Ne δ (t^N σ a^N))
    × Ty^P (t^N σ a^N)
```

At base type every semantic value is natural. A semantic function is natural if when applied to a natural argument it commutes with embedding and yields a natural result.

We extend naturality to semantic contexts and prove that it is preserved by embedding:

```
Ty^Pe : ∀ σ → Ty^P t^N → Ty^P (Ty^Ne σ t^N)
```

[8]Recall that the evaluation function is the interpretation of terms.

```
data Conᴾ : ∀ {Γ Δ} → Conᴺ Γ Δ → Set where
  •   : Conᴾ •
  _,_ : Conᴾ Γᴺ → Tyᴾ tᴺ → Conᴾ (Γᴺ , tᴺ)

Conᴾₑ : ∀ σ → Conᴾ Γᴺ → Conᴾ (Conᴺₑ σ Γᴺ)
```

Next are functor laws for types and contexts. These are provable without reference to naturality. However, function extensionality is required.

```
Tyᴺ-idₑ : ∀ {A}(tᴺ : Tyᴺ A Γ) → Tyᴺₑ idₑ tᴺ ≡ tᴺ
Tyᴺ-idₑ {A = ι}     tᴺ = Nf-idₑ tᴺ
Tyᴺ-idₑ {A = A ⇒ B} tᴺ = fexti λ Δ → fext λ δ → tᴺ & idlₑ δ

Tyᴺ-∘ₑ : ∀ {A}(tᴺ : Tyᴺ A Σ) σ δ → Tyᴺₑ (σ ∘ₑ δ) tᴺ ≡ Tyᴺₑ δ (Tyᴺₑ σ tᴺ)
Tyᴺ-∘ₑ {A = ι}     tᴺ σ δ = Nf-∘ₑ σ δ tᴺ
Tyᴺ-∘ₑ {A = A ⇒ B} tᴺ σ δ = fexti λ Ξ → fext λ ν → tᴺ & assₑ σ δ ν

-- proofs for Con are just elementwise extensions as always
Conᴺ-idₑ : ∀ σᴺ → Conᴺₑ idₑ σᴺ ≡ σᴺ
Conᴺ-∘ₑ  : ∀ σ δ ν → Conᴺₑ (σ ∘ₑ δ) ν ≡ Conᴺₑ δ (Conᴺₑ σ ν)
```

We proceed to naturality for evaluation, quoting and unquoting, which are provable this time, thanks to the additional ᴾ assumptions:

```
Tmᴾ     : ∀ t → Conᴾ Γᴺ → Tyᴾ (Tmᴺ t Γᴺ)
Tmᴺ-nat : ∀ t σ → Conᴾ Γᴺ → Tmᴺ t (Conᴺₑ σ Γᴺ) ≡ Tyᴺₑ σ (Tmᴺ t Γᴺ)
uᴾ      : ∀ n → Tyᴾ (uᴺ n)
qᴺ-nat  : ∀ σ tᴺ → Tyᴾ tᴺ → Nfₑ σ (qᴺ tᴺ) ≡ qᴺ (Tyᴺₑ σ tᴺ)
uᴺ-nat  : ∀ σ n → Tyᴺₑ σ (uᴺ n) ≡ uᴺ (Neₑ σ n)
uᶜᴾ     : ∀ {Γ} → Conᴾ (uᶜᴺ {Γ})
```

The proofs are obtained using mutual induction and substitution laws.

5.5 Soundness

Soundness expresses that normalization maps convertible terms to the same normal form:

```
sound : t ~ t' → nf t ≡ nf t'
```

Unfolding `nf` in the return type yields `(qᴺ (Tmᴺ t) ≡ qᴺ (Tmᴺ t'))`. It is clear that `(Tmᴺ t)` should be in some relation to `(Tmᴺ t')` such that we can prove that `qᴺ` takes related inputs to equal outputs.

The simplest option is to define that relation as propositional equality. Since we assume function extensionality, propositional equality on semantic values is equivalent to the following logical relation:

```
_≈_ : ∀ {A Γ} → Tyᴺ A Γ → Tyᴺ A Γ → Set
_≈_ {ι}     {Γ} t t' = t ≡ t'
_≈_ {A ⇒ B}{Γ} t t' = ∀ {Δ}(σ : OPE Δ Γ){a a'} → a ≈ a' → t σ a ≈ t' σ a'
```

In other words, equality of semantic values is just propositional equality at base types and pointwise equality for semantic functions. However, we are again foiled by requirements of naturality: `_≈_` relates unnatural values as well, but some needed equalities can only be shown on natural domains. Hence, we amend the above definition:

```
-- Soundness.agda
_≈_ : ∀ {A Γ} → Tyᴺ A Γ → Tyᴺ A Γ → Set
_≈_ {ι}     {Γ} t t' = t ≡ t'
_≈_ {A ⇒ B}{Γ} t t' =
  ∀ {Δ}(σ : OPE Δ Γ){a a'} → Tyᴾ a → Tyᴾ a' → a ≈ a' → t σ a ≈ t' σ a'
```

Alternatively, `_≈_` and `Tyᴾ` could be merged into a single *partial equivalence relation*. In this definition, two functions are related if both are natural and take related inputs to related outputs. The naturality predicate can be recovered by setting `(Tyᴾ t)` to be `(t ≈ t)`. This alternative was explored in the formalization but appeared to be less convenient than separate naturality predicates.

To prove soundness, first we extend `_≈_` to semantic contexts and prove monotonicity, similarly as in the completeness proof:

```
data _≈ᶜ_ : ∀ {Γ Δ} → Conᴺ Γ Δ → Conᴺ Γ Δ → Set where
  •   : • ≈ᶜ •
  _,_ : σ ≈ᶜ δ → t ≈ t' → (σ , t) ≈ᶜ (δ , t')

≈ₑ  : ∀ σ → t ≈ t' → Tyᴺₑ σ t ≈ Tyᴺₑ σ t'
≈ᶜₑ : ∀ σ → δ ≈ᶜ ν → Conᴺₑ σ δ ≈ᶜ Conᴺₑ σ ν
```

Next, we prove that evaluating the same term in related semantic contexts yields related semantic values.

```
-- ∈≈ just looks up a witness, as before.
∈≈ : (v : A ∈ Γ) → σ ≈ᶜ δ → ∈ᴺ v σ ≈ ∈ᴺ v δ

Tm≈ : (t : Tm Γ A) → Conᴾ Γᴺ → Conᴾ Γᴺ' →  Γᴺ ≈ᶜ Γᴺ' → Tmᴺ t Γᴺ ≈ Tmᴺ t Γᴺ'
Tm≈ (var v)   Γᴾ Γᴾ' σ≈δ = ∈≈ v σ≈δ
Tm≈ (lam t)   Γᴾ Γᴾ' σ≈δ =
  λ ν aᴾ aᴾ' aᴺ≈aᴺ'
    → Tm≈ t (Conᴾₑ ν Γᴾ , aᴾ) (Conᴾₑ ν Γᴾ' , aᴾ') (≈ᶜₑ ν σ≈δ , a≈a')
Tm≈ (app f a) Γᴾ Γᴾ' σ≈δ =
  Tm≈ f Γᴾ Γᴾ' σ≈δ idₑ (Tmᴾ a Γᴾ) (Tmᴾ a Γᴾ') (Tm≈ a Γᴾ Γᴾ' σ≈δ)
```

The proof is entirely mechanical, although it is complicated by the need to carry along `Conᴾ` proofs, in order to produce `(Tyᴾ a Γ)` naturality proofs for inputs in the `(app f a)` case.

Next, we prove the more general lemma expressing that convertible terms are evaluated to related values:

```
~≈ : t ~ t' → Conᴾ Γᴺ → Conᴾ Γᴺ' → Γᴺ ≈ᶜ Γᴺ' → Tmᴺ t Γᴺ ≈ Tmᴺ t' Γᴺ'
```

This is the bulk of the work. We proceed with induction on `(t ~ t')`. First, let us consider the constructors for the equivalence closure, namely reflexivity, symmetry and transitivity. The latter two can be proven in general for `_≈_` and `_≈ᶜ_`:

```
_≈⁻¹  : t ≈ t' → t' ≈ t
_≈ᶜ⁻¹ : σ ≈ᶜ δ → δ ≈ᶜ σ
_≈■_  : t ≈ t' → t' ≈ t'' → t ≈ t''
_≈ᶜ■_ : σ ≈ᶜ δ → δ ≈ᶜ ν → σ ≈ᶜ ν
```

Reflexivity is not generally provable, but we only need it on the image of `Tmᴺ`, which is already covered by `Tm≈`. We can implement the cases for equivalence closure in `~≈`:

```
~≈ {t = t} ~refl Γᴾ Γᴾ' Γᴺ≈Γᴺ' =
  Tm≈ t Γᴾ Γᴾ' Γᴺ≈Γᴺ'
~≈ (t'~t ~⁻¹) Γᴾ Γᴾ' Γᴺ≈Γᴺ' =
  ~≈ t'~t Γᴾ' Γᴾ (Γᴺ≈Γᴺ' ≈ᶜ⁻¹) ≈⁻¹
~≈ (t~t' ~■ t'~t'') Γᴾ Γᴾ' Γᴺ≈Γᴺ' =
  ~≈ t~t' Γᴾ Γᴾ (Γᴺ≈Γᴺ' ≈ᶜ■ (Γᴺ≈Γᴺ' ≈ᶜ⁻¹)) ≈■ ~≈ t'~t'' Γᴾ Γᴾ' Γᴺ≈Γᴺ'
```

The `app` and `lam` congruences are also straightforward:

```
~≈ (lam t~t') Γᴾ Γᴾ' Γᴺ≈Γᴺ' =
  λ ν aᴾ aᴾ' a≈a' →
    ~≈ t~t' (Conᴾₑ ν Γᴾ , aᴾ) (Conᴾₑ ν Γᴾ' , aᴾ') (≈ᶜₑ ν Γᴺ≈Γᴺ' , a≈a')
~≈ (app {f = f}{f'}{a = a}{a'} t~t' a~a') Γᴾ Γᴾ' Γᴺ≈Γᴺ' =
  ~≈ t~t' Γᴾ Γᴾ' Γᴺ≈Γᴺ' idₑ (Tmᴾ a Γᴾ) (Tmᴾ a' Γᴾ') (~≈ a~a' Γᴾ Γᴾ' Γᴺ≈Γᴺ')
```

The real work happens with the β and η rules. Let us see η:

```
~≈ (η t) {Γᴺ} {Γᴺ'} Γᴾ Γᴾ' Γᴺ≈Γᴺ' ν {a} {a'} aᴾ aᴾ' a≈a' = ?
```

The goal has the following type:

```
Tmᴺ t Γᴺ ν a ≈ Tmᴺ (Tmₑ wk t) (Conᴺₑ ν Γᴺ' , a') idₑ a'
```

Making full use of the induction hypothesis gives us the following:

```
Tm≈ t Γᴾ Γᴾ' Γᴺ≈Γᴺ' ν aᴾ aᴾ' a≈a' : Tmᴺ t Γᴺ ν a ≈ Tmᴺ t Γᴺ' ν a'
```

Hence, we need to rewrite (or coerce) the right hand side of the result type with an equality.

```
eq : Tmᴺ (Tmₑ wk t) (Conᴺₑ ν Γᴺ' , a') idₑ a' ≡ Tmᴺ t Γᴺ' ν a'
```

This should give us a moment of pause. Obviously, we need to prove something about the action of `Tmᴺ` on *embedded terms* such as `(Tmₑ wk t)`. However, we haven't even defined an interpretation for embeddings in any model. Embedding and substitution is implicit in our syntax, therefore we are not obliged to interpret them. Note that in

the presheaf model in Section 5.4, we interpreted *types* as presheaves, and for each type we defined its action on embeddings, with embeddings considered as **OPE** morphisms, but that is not the same as interpreting embeddings themselves, considered as **STLC** morphisms.

We interpret embeddings as natural transformations between semantic contexts:

```
OPEᴺ : OPE Γ Δ → ∀ {Σ} → Conᴺ Γ Σ → Conᴺ Δ Σ
OPEᴺ •        Γᴺ        = δᴺ
OPEᴺ (drop σ) (Γᴺ , _)  = OPEᴺ σ Γᴺ
OPEᴺ (keep σ) (Γᴺ , tᴺ) = OPEᴺ σ Γᴺ , tᴺ

OPEᴺ-nat : ∀ σ δ Γᴺ → OPEᴺ σ (Conᴺₑ δ Γᴺ) ≡ Conᴺₑ δ (OPEᴺ σ Γᴺ)
-- proof omitted
```

While an embedding encodes that one context can be obtained from another by dropping zero or more elements, a semantic embedding actually performs the dropping of elements on semantic contexts. For instance, `(OPEᴺ wk)` acts as a "tail" function, dropping the topmost element.

Additionally, since the identity embedding is defined recursively (as opposed to as a constructor), we have to prove that it is interpreted as the identity natural transformation:

```
OPEᴺ-idₑ : ∀ Γᴺ → OPEᴺ idₑ Γᴺ ≡ Γᴺ
```

Now, we can give the action of evaluation on embedded terms:

```
∈ₑᴺ  : ∀ σ v Γᴺ → ∈ᴺ (∈ₑ σ v)   Γᴺ ≡ ∈ᴺ  v (OPEᴺ σ Γᴺ)
Tmₑᴺ : ∀ σ t Γᴺ → Tmᴺ (Tmₑ σ t) Γᴺ ≡ Tmᴺ t (OPEᴺ σ Γᴺ)
```

Note that these are essentially *computation* rules, but given as propositional equalities. If the syntax had explicit substitutions (and, by extension, explicit embeddings), these rules would be actual definitional equalities in the evaluation function. It is a trade-off: we forgo these definitional equalities in favor of others stemming from the recursive definition of embedding and substitution, and in favor of the greatly reduced number of conversion rules.

We can turn our attention back to the equality proof required for the η case. By informal equational reasoning:

```
Tmᴺ (Tmₑ wk t) (Conᴺₑ ν Γᴺ' , a') idₑ a' ≡
  -- by (Tmₑᴺ wk t (Conᴺₑ ν Γᴺ' , a'))
Tmᴺ t (OPEᴺ idₑ (Conᴺₑ ν Γᴺ')) idₑ a' ≡
  -- by (OPEᴺ-idₑ (Conᴺₑ ν Γᴺ'))
Tmᴺ t (Conᴺₑ ν Γᴺ') idₑ a' ≡
  -- by (Tmᴺ-nat t ν {Γᴺ'} Γᴾ')
Tmᴺ t Γᴺ' (ν ∘ₑ idₑ) a' ≡
  -- by (idrₑ ν)
Tmᴺ t Γᴺ' ν a'
```

In Agda, we can rewrite the goal type with the above equality, then return the induction hypothesis:

```
~≈ (η t) {Γᴺ} {Γᴺ'} Γᴾ Γᴾ' Γᴺ≈Γᴺ' ν {a} {a'} aᴾ aᴾ' a≈a'
  rewrite
    Tmₑᴺ wk t (Conᴺₑ ν Γᴺ' , a')
  | OPEᴺ-idₑ (Conᴺₑ ν Γᴺ')
  | Tmᴺ-nat t ν {Γᴺ'} Γᴾ'
  | idrₑ ν
  = Tm≈ t Γᴾ Γᴾ' Γᴺ≈Γᴺ' ν aᴾ aᴾ' a≈a'
```

Now, consider the β case:

```
~≈ (β t t') {Γᴺ} {Γᴺ'} Γᴾ Γᴾ' σ≈δ = ?
```

The goal is `(Tmᴺ t (Conᴺₑ idₑ Γᴺ , Tmᴺ t' Γᴺ) ≈ Tmᴺ (Tmₛ (idₛ , t') t) Γᴺ')`. Note that the right hand side is the evaluation of a substituted expression. Similarly as with embeddings, an interpretation must be given for substitutions as well as the action of evaluation on substituted expressions.

Substitutions are also interpreted as natural transformations, however, naturality only holds on natural inputs:

```
Subᴺ : Sub Γ Δ → ∀ {Σ} → Conᴺ Γ Σ → Conᴺ Δ Σ
Subᴺ •       Γᴺ = •
Subᴺ (σ , t) Γᴺ = Subᴺ σ Γᴺ , Tmᴺ t Γᴺ

Subᴺ-nat : ∀ σ δ Γᴺ → Conᴾ Γᴺ → Subᴺ σ (Conᴺₑ δ Γᴺ) ≡ Conᴺₑ δ (Subᴺ σ Γᴺ)
```

`(Subᴺ σ Γᴺ)` evaluates each term in `σ` in `Γᴺ`, thereby converting a list of terms (substitution) into a list of values (semantic context). We also need two related lemmas:

```
Subᴺ-sₛₑ : ∀ σ δ Γᴺ → Subᴺ (σ sₛₑ δ) Γᴺ ≡ Subᴺ σ (OPEᴺ δ Γᴺ)
Subᴺ-idₛ : ∀ Γᴺ → Subᴺ idₛ Γᴺ ≡ Γᴺ
```

There is no deep reason why we need `_sₛₑ_` but not the other variants; only this lemma is needed in the rest of the soundness proof.

The action of evaluation on substituted terms is a bit more involved. Naturality must be assumed for the input semantic context, or else the proof does not go through - in fact, this is the reason why naturality predicates were introduced. This implies that the action of evaluation cannot be stated as propositional equality, only as `_≈_`, the previously defined equality of semantic values.

```
Tmₛᴺ : ∀ σ t Γᴺ → Conᴾ Γᴺ → Γᴺ ≈ᶜ Γᴺ → Tmᴺ (Tmₛ σ t) Γᴺ ≈ Tmᴺ t (Subᴺ σ Γᴺ)
```

Now, `Tmₛᴺ` can used to prove a `_≈_` where the right hand side matches the goal type for the β case.

```
(Tmₛᴺ (idₛ , t') t Γᴺ' Γᴾ' ((σ≈δ ≈ᶜ⁻¹) ≈ᶜ■ σ≈δ) ≈⁻¹)
  : Tmᴺ t (Subᴺ idₛ Γᴺ' , Tmᴺ t' Γᴺ') ≈ Tmᴺ (Tmₛ (idₛ , t') t) Γᴺ'
```

We can use transitivity of _≈_ and erase semantic $\mathtt{id}_e$ actions, after which the goal becomes

```
Tmᴺ t (Γᴺ , Tmᴺ t' Γᴺ) ≈ Tmᴺ t (Γᴺ' , Tmᴺ t' Γᴺ')
```

Note that the evaluation context differs on the sides, but the the evaluated term is the same. Thus, Tm≈ can be used to prove the goal, which completes the proof of ~≈.

```
(Tm≈ t (Γᴾ , Tmᴾ t' Γᴾ) (Γᴾ' , Tmᴾ t' Γᴾ') (σ≈δ , (Tm≈ t' Γᴾ Γᴾ' σ≈δ)))
  : Tmᴺ t (Γᴺ , Tmᴺ t' Γᴺ) ≈ Tmᴺ t (Γᴺ' , Tmᴺ t' Γᴺ')
```

After this, quoting, unquoting and soundness is easily given.

```
mutual
  q≈ : ∀ {A Γ}{t t' : Tyᴺ A Γ} → t ≈ t' → qᴺ t ≡ qᴺ t'
  q≈ {ι}     t≈t' = t≈t'
  q≈ {A ⇒ B} t≈t' = lam & q≈ (t≈t' (wk {A}) (uᴾ _) (uᴾ _) (u≈ refl))

  u≈ : ∀ {A Γ}{n n' : Ne Γ A} → n ≡ n' → uᴺ n ≈ uᴺ n'
  u≈ {ι}     p = ne & p
  u≈ {A ⇒ B} p = λ σ aᴾ aᴾ' a≈a' → u≈ (app & (Neₑ σ & p) ⊗ q≈ a≈a')

uᶜ≈ : ∀ {Γ} → uᶜᴺ {Γ} ≈ᶜ uᶜᴺ
uᶜ≈ {•}     = •
uᶜ≈ {Γ , A} = ≈ᶜₑ wk uᶜ≈ , u≈ refl

sound : ∀ {Γ A}{t t' : Tm Γ A} → t ~ t' → nf t ≡ nf t'
sound t~t' = q≈ (~≈ t~t' uᶜᴾ uᶜᴾ uᶜ≈)
```

□

5.6 Stability

Stability expresses that normalization has no action on normal terms:

```
stable : (n : Ne Γ A) → nf ⌜ n ⌝Nf ≡ n
```

It differs in several ways from the other two correctness properties. First, it is not quite as crucial; failing either soundness or completeness results in something else than a normalization algorithm, but instability is tolerable. A primary goal of normalization is to have a decision procedure for conversion, however, it is achievable without stability:

```
-- decidable Nf equality, given by simple induction
Nf≡? : (n n' : Nf Γ A) → (n ≡ n') ⊎ (n ≡ n' → ⊥)

decidableConversion : (t t' : Tm Γ A) → (t ~ t') ⊎ (t ~ t' → ⊥)
decidableConversion t t' with Nf≡? (nf t) (nf t')
... | inj₁ p =
```

```
  inj₁ (complete t ~■ coe ((λ x → ⌜ x ⌝Nf ~ t') & p ⁻¹) (complete t' ~⁻¹))
... | inj₂ p =
  inj₂ (λ q → p (sound q))
```

Rather, stability tells us whether a normalization algorithm matches up exactly with a definition of normal forms. In our case, if we leave normalization essentially unchanged, but do not mandate η-long normal forms as we currently do, there would be η-short normal forms which are η-expanded by normalization:

```
t : Tm (ι ⇒ ι) (ι ⇒ ι)
t = var vz

nf-η-expands : nf t ≡ (lam (app (var (vs vz)) (var vz)))
nf-η-expands = refl
```

Here, normalization "thinks" that normal forms are more restricted than they actually are. To regain stability, normalization would have to be modified so that it does not do any η-conversion and leaves neutral terms unchanged. That said, an unstable algorithm is still fine - it just does "more" than what is required. One could argue that instability should be avoided because unstable algorithms perform unnecessary computation.

Stability is much simpler to prove than previous correctness properties. Simple mutual induction on normal and neutral terms suffices, making use of previously proved naturalities and straightforward equational reasoning.

```
stableϵ : (v : A ϵ Γ) → ϵᴺ v uᶜᴺ ≡ uᴺ (var v)
stableϵ vz     = refl
stableϵ (vs v) =
    ϵᴺ-nat v wk uᶜᴺ
  ■ Tyᴺₑ wk & stableϵ v
  ■ uᴺ-nat wk (var v)
  ■ (λ x → uᴺ (var (vs x))) & ϵ-idₑ v

mutual
  stable : (n : Nf Γ A) → nf ⌜ n ⌝Nf ≡ n
  stable (ne n)  = stableNe n
  stable (lam n) = lam & stable n

  stableNe : (n : Ne Γ A) → Tmᴺ ⌜ n ⌝Ne uᶜᴺ ≡ uᴺ n
  stableNe (var v)   = stableϵ v
  stableNe (app f a) =
      (λ x → x idₑ (Tmᴺ ⌜ a ⌝Nf uᶜᴺ)) & stableNe f
    ■ (λ x → uᴺ (app (Neₑ idₑ f) x)) & stable a
    ■ (λ x → uᴺ (app x a)) & Ne-idₑ f
```

□

6 Variations

In this section we look at two alternate variants of the formalization with different design choices. The first one is more compact but includes a less transparent implementation of normalization. The second one contains a more efficient evaluation function at the cost of slightly larger proofs.

6.1 Direct Presheaf Model

In Section 5.4 the presheaf model of STLC was described, but not pursued to its full extent. Going for a full presheaf model lets us dispense with about 150 lines of Agda, which is a significant portion of the roughly 750 lines of the whole development. The formalization can be found at *https://github.com/AndrasKovacs/stlc-nbe/tree/master*.

The `ᴺ` model is changed to a presheaf model, therefore it includes all mutually defined naturality proofs. The interpretation of types is now the same as `Tyᴾ` in Section 5.4, although here we also use `ᴺ` superscripts for the presheaf model[9].

A benefit of this approach is that the `Tyᴾ` logical predicate for naturality can be entirely dropped, since naturality is already baked into `Tyᴺ`. The drawback is that all functor laws and naturalities must be proven even when defining the evaluation function. We list here the types (only the types, for brevity) of the definitions required to implement normalization, along with brief descriptions of categorical interpretations. Below, `(f : PSh(OPE)(A, B))` means that `f` is a natural transformation in `PSh(OPE)` between `A` and `B` presheaves.

```
-- Presheaf.agda

-- Tyᴺ A : PSh(OPE)
------------------------------------------------------------
Tyᴺ  : Ty → Con → Set              -- action on objects
Tyᴺₑ : OPE Δ Γ → Tyᴺ A Γ → Tyᴺ A Δ -- action on morphisms

Tyᴺ-idₑ : ∀ tᴺ → Tyᴺₑ idₑ tᴺ ≡ tᴺ
Tyᴺ-∘ₑ  : ∀ tᴺ σ δ → Tyᴺₑ (σ ∘ₑ δ) tᴺ ≡ Tyᴺₑ δ (Tyᴺₑ σ tᴺ)

-- Conᴺ Γ : PSh(OPE)
------------------------------------------------------------
Conᴺ  : Con → Con → Set               -- action on objects
Conᴺₑ : OPE Δ Σ → Conᴺ Γ Σ → Conᴺ Γ Δ -- action on morphisms

Conᴺ-idₑ : ∀ Γᴺ → Conᴺₑ idₑ Γᴺ ≡ Γᴺ
Conᴺ-∘ₑ  : ∀ σ δ Γᴺ → Conᴺₑ (σ ∘ₑ δ) Γᴺ ≡ Conᴺₑ δ (Conᴺₑ σ Γᴺ)
```

[9]It stands for "normalization", which it achieves, but by a different model than previously.

```
-- ∈ᴺ {Γ}{A} v : PSh(OPE)(Conᴺ Γ, Tyᴺ A)
--------------------------------------------------------------
∈ᴺ     : ∀ {Γ A} → A ∈ Γ → ∀ {Δ} → Conᴺ Γ Δ → Tyᴺ A Δ
∈ᴺ-nat : ∀ (v : A ∈ Γ) σ Γᴺ → ∈ᴺ v (Conᴺₑ σ Γᴺ) ≡ Tyᴺₑ σ (∈ᴺ v Γᴺ)

-- Tmᴺ {Γ}{A} t : PSh(OPE)(Conᴺ Γ, Tyᴺ A)
--------------------------------------------------------------
Tmᴺ     : ∀ {Γ A} → Tm Γ A → ∀ {Δ} → Conᴺ Γ Δ → Tyᴺ A Δ
Tmᴺ-nat : ∀ t σ Γᴺ → Tmᴺ t (Conᴺₑ σ Γᴺ) ≡ Tyᴺₑ σ (Tmᴺ t Γᴺ)

-- qᴺ {A} : PSh(OPE)(Tyᴺ A, Nf _ A)
--------------------------------------------------------------
qᴺ     : ∀ {A Γ} → Tyᴺ A Γ → Nf Γ A
qᴺ-nat : ∀ σ tᴺ → Nfₑ σ (qᴺ tᴺ) ≡ qᴺ (Tyᴺₑ σ tᴺ)

-- uᴺ {A} : PSh(OPE)(Ne _ A, Tyᴺ A)
--------------------------------------------------------------
uᴺ     : ∀ {A Γ} → Ne Γ A → Tyᴺ A Γ
uᴺ-nat : ∀ σ n → Tyᴺₑ σ (uᴺ n) ≡ uᴺ (Neₑ σ n)


--------------------------------------------------------------
uᶜᴺ : ∀ {Γ} → Conᴺ Γ Γ
nf  : ∀ {Γ A} → Tm Γ A → Nf Γ A
```

The proofs for completeness and stability are largely unchanged. For soundness, we again need to interpret embeddings and substitutions and define the action of evaluation on them.

```
-- Presheaf.agda

-- OPEᴺ {Γ}{Δ} σ : PSh(OPE)(Conᴺ Γ, Conᴺ Δ)
--------------------------------------------------------------
OPEᴺ     : ∀ {Γ Δ} → OPE Γ Δ → ∀ {Σ} → Conᴺ Γ Σ → Conᴺ Δ Σ
OPEᴺ-nat : ∀ σ δ Γᴺ → OPEᴺ σ (Conᴺₑ δ Γᴺ) ≡ Conᴺₑ δ (OPEᴺ σ Γᴺ)

OPEᴺ-idₑ : ∀ Γᴺ → OPEᴺ idₑ Γᴺ ≡ Γᴺ
Tmₑᴺ     : ∀ σ t Γᴺ → Tmᴺ (Tmₑ σ t) Γᴺ ≡ Tmᴺ t (OPEᴺ σ Γᴺ)

-- Subᴺ {Γ}{Δ} σ : PSh(OPE)(Conᴺ Γ, Conᴺ Δ)
--------------------------------------------------------------------------------
Subᴺ     : ∀ {Γ Δ} → Sub Γ Δ → ∀ {Σ} → Conᴺ Γ Σ → Conᴺ Δ Σ
Subᴺ-nat : ∀ σ δ Γᴺ → Subᴺ σ (Conᴺₑ δ Γᴺ) ≡ Conᴺₑ δ (Subᴺ σ Γᴺ)

Subᴺ-idₛ : ∀ Γᴺ → Subᴺ idₛ Γᴺ ≡ Γᴺ
Subᴺ-ₛ∘ₑ : ∀ σ δ Γᴺ → Subᴺ (σ ₛ∘ₑ δ) Γᴺ ≡ Subᴺ σ (OPEᴺ δ Γᴺ)
```

```
Tmₛᴺ     : ∀ σ t Γᴺ → Tmᴺ (Tmₛ σ t) Γᴺ ≡ Tmᴺ t (Subᴺ σ Γᴺ)
```

The soundness proof is dramatically simpler this time. The reason is that propositional equality is already right for semantic values, so there is no need for any additional _≈_ logical relation, and propositional equality is already an equivalence relation which respected by all constructions in the metatheory. Hence, soundness is simply given as follows:

```
-- Soundness.agda
~≈ : ∀ {t t' : Tm Γ A} → t ~ t' → (σ : Conᴺ Γ Δ) → Tmᴺ t σ ≡ Tmᴺ t' σ

sound : ∀ {Γ A}{t t' : Tm Γ A} → t ~ t' → nf t ≡ nf t'
sound t~t' = qᴺ & ~≈ t~t' uᶜᴺ
```

Here, the proof for ~≈ still makes use of `OPEᴺ` and `Subᴺ`, but it is much simpler than before and altogether takes 11 lines of Agda.

6.2 More Efficient Evaluation

We describe a more efficient evaluation function in this section. The correctness proofs are also formalized for this version, requiring only minor modifications. The formalization can be found at *https://github.com/AndrasKovacs/stlc-nbe/blob/efficient-appN*.

Let us review the definition of `Tmᴺ`:

```
Tmᴺ : ∀ {Γ A} → Tm Γ A → ∀ {Δ} → Conᴺ Γ Δ → Tyᴺ A Δ
Tmᴺ (var v)   Γᴺ = ∈ᴺ v Γᴺ
Tmᴺ (lam t)   Γᴺ = λ σ aᴺ → Tmᴺ t (Conᴺₑ σ Γᴺ , aᴺ)
Tmᴺ (app f a) Γᴺ = Tmᴺ f Γᴺ idₑ (Tmᴺ a Γᴺ)
```

Notice the `idₑ` in the `(app f a)` case. After each application to `idₑ`, the `Γᴺ` context is changed to `(Conᴺₑ idₑ Γᴺ)` during the evaluation of the internal `t` body of `(Tmᴺ f Γᴺ)`. This means that every application has an overhead proportional to the size of the current context. Also, recall `Tyᴺₑ` for functions:

```
Tyᴺₑ {A ⇒ B} σ tᴺ = λ δ aᴺ → tᴺ (σ ∘ₑ δ) aᴺ
```

`(Tyᴺₑ idₑ tᴺ)` creates a new closure around `tᴺ` which has essentially no effect, and these closures accumulate on the top of semantic functions in `Γᴺ` after each `(app f a)` evaluation. This is certainly bad for performance. We would like evaluation to be essentially as efficient for weak reduction as the standard model (as defined in Section 4.2). This can be achieved by defining `idₑ` and `Conᴺₑ` in a way such that `(Conᴺₑ idₑ Γᴺ)` immediately reduces to `Γᴺ`.

First, `idₑ` needs to be a primitive constructor:

```
-- Embedding.agda
data OPE : Con → Con → Set where
```

```
idₑ  : ∀ {Γ} → OPE Γ Γ
drop : ∀ {A Γ Δ} → OPE Γ Δ → OPE (Γ , A) Δ
keep : ∀ {A Γ Δ} → OPE Γ Δ → OPE (Γ , A) (Δ , A)
```

This slightly complicates the identity functor law for `(Tm _ A)` and `(A ∈ _)`, since identity embeddings are not unique anymore. We define a predicate on embeddings which picks out the identities:

```
Id-ish : ∀ {Γ} → OPE Γ Γ → Set
Id-ish idₑ      = ⊤
Id-ish (drop σ) = ⊥
Id-ish (keep σ) = Id-ish σ
```

`Id-ish` embeddings are `idₑ`-s wrapped in zero or more `keep`-s. We prove identity functor laws using `Id-ish`:

```
∈-idₑ  : ∀ (v : A ∈ Γ ){σ : OPE Γ Γ}{p : Idish σ} → ∈ₑ σ v ≡ v
Tm-idₑ : ∀ (t : Tm Γ A){σ : OPE Γ Γ}{p : Idish σ} → Tmₑ σ t ≡ t
```

Secondly, `Conᴺₑ` is redefined as follows:

```
Conᴺₑ : OPE Σ Δ → Conᴺ Γ Δ → Conᴺ Γ Σ
Conᴺₑ idₑ      Γᴺ        = Γᴺ
Conᴺₑ (drop σ) •         = •
Conᴺₑ (drop σ) (Γᴺ , tᴺ) = (Conᴺₑ (drop σ) Γᴺ) , Tyᴺₑ (drop σ) tᴺ
Conᴺₑ (keep σ) •         = •
Conᴺₑ (keep σ) (Γᴺ , tᴺ) = (Conᴺₑ (keep σ) Γᴺ) , (Tyᴺₑ (keep σ) tᴺ)
```

However, with this we lose two definitional equalities we had before, which are quite useful to have. We can resurrect them as propositional equalities:

```
-- PresheafRefinement.agda
Conᴺₑ-• : ∀ σ → Conᴺₑ σ • ≡ •
Conᴺₑ-, : ∀ σ Γᴺ tᴺ → Conᴺₑ σ (Γᴺ , tᴺ) ≡ (Conᴺₑ σ Γᴺ , Tyᴺₑ σ tᴺ)
```

We must manually rewrite along these equalities when necessary. It is not a significant burden though. All together, the changes described here result in about 20 additional lines in the formalization.

We eliminated the overhead on the evaluation of application, but what about quoting? It also contains semantic application:

```
qᴺ : ∀ {A Γ} → Tyᴺ A Γ → Nf Γ A
qᴺ {ι}     tᴺ = tᴺ
qᴺ {A ⇒ B} tᴺ = lam (qᴺ (tᴺ wk (uᴺ (var vz))))
```

Here, the `(tᴺ wk)` is essential, but it is far less of an overhead than the `idₑ`-s before. Notice that `qᴺ` is *type-directed*, and for each semantic value it is called as many times as there are function arguments in the value's type. Of course, `qᴺ` is applied to various sub-terms of a term during normalization (via `uᴺ`), so the overall number of `qᴺ` calls is likely greater than the number of arguments in a term's type. Still, it is only to be

expected that *normalization* is more costly than standard interpretation, since it does more work. *Evaluation* on its own is no less efficient than standard interpretation, with the current optimized implementation.

7 Discussion and Future Work

- Technical overhead: Vs big-step, hereditary subst, Coquand, Abel (?)
- Linecounts vs big-step and hsubst (email C. Coquand to get sources?)
- Usability as EDSL normalizer
- Scaling up the technique
- With implicit substitution and conversion relation: to System F, probably

References

[1] Andreas Abel. "Normalization by Evaluation: Dependent Types and Impredicativity". Habilitation thesis. PhD thesis. Ludwig-Maximilians-Universität München, 2013.

[2] Thorsten Altenkirch. "Normalisation by completeness". In: *Talk given at the Workshop on Normalization by Evaluation in Los Angeles*. 2009.

[3] Thorsten Altenkirch and James Chapman. "Big-step normalisation". In: *Journal of Functional Programming* 19.3-4 (2009), pp. 311–333.

[4] Thorsten Altenkirch, Martin Hofmann, and Thomas Streicher. "Categorical reconstruction of a reduction free normalization proof". In: *Category Theory and Computer Science*. Springer. 1995, pp. 182–199.

[5] Thorsten Altenkirch and Ambrus Kaposi. "Normalisation by evaluation for dependent types". In: *LIPIcs-Leibniz International Proceedings in Informatics*. Vol. 52. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. 2016.

[6] Thorsten Altenkirch and Ambrus Kaposi. "Type theory in type theory using quotient inductive types". In: *ACM SIGPLAN Notices*. Vol. 51. 1. ACM. 2016, pp. 18–29.

[7] Thorsten Altenkirch, Conor McBride, and Wouter Swierstra. "Observational equality, now!" In: *Proceedings of the 2007 workshop on Programming languages meets program verification*. ACM. 2007, pp. 57–68.

[8] Lennart Augustsson and Magnus Carlsson. *An exercise in dependent types: A well-typed interpreter*. 1999.

[9] Nick Benton et al. "Strongly typed term representations in Coq". In: *Journal of automated reasoning* 49.2 (2012), pp. 141–159.

[10] Ulrich Berger and Helmut Schwichtenberg. "An inverse of the evaluation functional for typed lambda-calculus". In: *Logic in Computer Science, 1991. LICS'91., Proceedings of Sixth Annual IEEE Symposium on*. IEEE. 1991, pp. 203–211.

[11] Ana Bove and Venanzio Capretta. "Modelling general recursion in type theory". In: *Mathematical Structures in Computer Science* 15.04 (2005), pp. 671–708.

[12] Edwin Brady. "Idris, a general-purpose dependently typed programming language: Design and implementation". In: *Journal of Functional Programming* 23.05 (2013), pp. 552–593.

[13] Simon Castellan. *Dependent type theory as the initial category with families*. 2014. URL: http://iso.mor.phis.me/archives/2011-2012/stage-2012-goteburg/report.pdf (visited on 03/10/2017).

[14] Cyril Cohen et al. "Cubical type theory: a constructive interpretation of the univalence axiom". In: *arXiv preprint arXiv:1611.02108* (2016).

[15] nLab contributors. *closed monoidal structure on presheaves*. 2017. URL: https://ncatlab.org/nlab/show/closed+monoidal+structure+on+presheaves (visited on 04/07/2017).

[16] nLab contributors. *preorder*. 2017. URL: https://ncatlab.org/nlab/show/preorder (visited on 04/07/2017).

[17] Catarina Coquand. "A formalised proof of the soundness and completeness of a simply typed lambda-calculus with explicit substitutions". In: *Higher-Order and Symbolic Computation* 15.1 (2002), pp. 57–90.

[18] Catarina Coquand. "A proof of normalization for simply typed lambda calculus written in ALF". In: *Proceedings of the Workshop on Types for Proofs and Programs*. Citeseer. 1992, pp. 85–92.

[19] Thierry Coquand and Peter Dybjer. "Intuitionistic model constructions and normalization proofs". In: *Mathematical Structures in Computer Science* 7.1 (1997), pp. 75–94.

[20] Agda developers. *Agda documentation*. 2017. URL: https://agda.readthedocs.io/en/v2.5.2/ (visited on 03/10/2017).

[21] Agda developers. *Agda standard library*. 2017. URL: https://github.com/agda/agda-stdlib (visited on 03/10/2017).

[22] Robert Harper. *Practical foundations for programming languages*. Cambridge University Press, 2016.

[23] Per Martin-Löf. "An intuitionistic theory of types: Predicative part". In: *Studies in Logic and the Foundations of Mathematics* 80 (1975), pp. 73–118.

[24] Conor McBride. *Datatypes of Datatypes*. 2015. URL: http://staff.mmcs.sfedu.ru/~ulysses/Edu/SSGEP/conor/conor.pdf.

[25] Conor McBride and Ross Paterson. "Applicative programming with effects". In: *Journal of functional programming* 18.01 (2008), pp. 1–13.

[26] Ulf Norell. *Towards a practical programming language based on dependent type theory*. Vol. 32. Citeseer, 2007.

[27] Benjamin C Pierce. *Types and programming languages*. MIT press, 2002.

[28] Steven Schäfer, Tobias Tebbi, and Gert Smolka. "Autosubst: Reasoning with de Bruijn Terms and Parallel Substitutions". In: *Interactive Theorem Proving - 6th International Conference, ITP 2015, Nanjing, China, August 24-27, 2015*. Ed. by Xingyuan Zhang and Christian Urban. LNAI. Springer-Verlag, Aug. 2015.

[29] Thomas Streicher. *Investigations into intensional type theory*. http://www.mathematik.tu-darmstadt.de/~streicher/HabilStreicher.pdf, 1993.

[30] Aaron Stump. *Verified Functional Programming in Agda*. New York, NY, USA: Association for Computing Machinery and Morgan & Claypool, 2016. ISBN: 978-1-97000-127-3.

[31] The Univalent Foundations Program. *Homotopy Type Theory: Univalent Foundations of Mathematics*. Institute for Advanced Study: https://homotopytypetheory.org/book, 2013.