

# An Introduction to Logical Relations

## Proving Program Properties Using Logical Relations

Lau Skorstengaard  
lask@cs.au.dk

August 30, 2019

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Simply Typed Lambda Calculus . . . . .	3
1.2	Logical Relations . . . . .	5
1.3	Categories of Logical Relations . . . . .	7
<b>2</b>	<b>Normalization of the Simply Typed Lambda Calculus</b>	<b>7</b>
2.1	Strong Normalization of STLC . . . . .	7
2.2	Exercises . . . . .	13
<b>3</b>	<b>Type Safety for STLC</b>	<b>13</b>
3.1	Type safety - the classical treatment . . . . .	13
3.2	Type safety - using logical predicate . . . . .	14
3.3	Exercises . . . . .	18
<b>4</b>	<b>Universal Types and Relational Substitutions</b>	<b>18</b>
4.1	System F (STLC with universal types) . . . . .	19
4.2	Properties of System-F: Free Theorems . . . . .	20
4.3	Contextual Equivalence . . . . .	21
4.4	A Logical Relation for System F . . . . .	23
4.5	Compatibility Lemmas . . . . .	27
4.6	Exercises . . . . .	31

<b>5</b>	<b>Existential types</b>	<b>32</b>
5.1	STLC with Existential Types . . . . .	33
5.2	Example . . . . .	34
5.3	Logical Relation . . . . .	36
<b>6</b>	<b>Recursive Types and Step Indexing</b>	<b>37</b>
6.1	Simply Typed Lambda Calculus with Recursive Types . . . . .	40
6.2	A Step-Indexed Logical Predicate . . . . .	41
6.3	Exercises . . . . .	45
<b>7</b>	<b>References and Worlds</b>	<b>45</b>
7.1	STLC with References . . . . .	46
7.2	Properties of STLC with References . . . . .	47
7.3	Logical Predicate . . . . .	49
7.4	Safety Proof . . . . .	54
7.5	Exercises . . . . .	61
7.6	Further Reading . . . . .	63
<b>A</b>	<b>Landin’s Knot</b>	<b>64</b>

## 1 Introduction

The term logical relations stems from Gordon Plotkin’s memorandum *Lambda-definability and logical relations written* in 1973. However, the spirit of the proof method can be traced back to Wiliam W. Tait who used it to show strong normalization of *System T* in 1967.

Names are a curious thing. When I say “chair”, you immediately get a picture of a chair in your head. If I say “table”, then you picture a table. The reason you do this is because we denote a chair by “chair” and a table by “table”, but we might as well have said “giraffe” for chair and “Buddha” for table. If we encounter a new word composed of known words, it is natural to try to find its meaning by composing the meaning of the components of the name. Say we encounter the word “tablecloth” for the first time, then if we know what “table” and “cloth” denotes we can guess that it is a piece of cloth for a table. However, this approach does not always work. For instance, a “skyscraper” is not a scraper you use to scrape the sky. Likewise for logical relations, it may be a fool’s quest to try to find meaning in the name. Logical relations are relations, so that part of the name makes sense. They are also defined in a way that has a small resemblance to a logic, but trying to give meaning to logical relations only from the parts of the name will not help

you understand them. A more telling name might be Type Indexed Inductive Relations. However, Logical Relations is a well-established name and easier to say, so we will stick with it (no one would accept “giraffe” to be a chair).

The majority of this note is based on the lectures of Amal Ahmed at the Oregon Programming Languages Summer School, 2015. The videos of the lectures can be found at <https://www.cs.uoregon.edu/research/summerschool/summer15/curriculum.html>.

## 1.1 Simply Typed Lambda Calculus

The language we use to present logical predicates and relations is the simply typed lambda calculus (STLC). In the first section, it will be used in its basic form, and later it will be used as the base language when we study new constructs and features. We will later leave it implicit that it is STLC that we extend with new constructs. STLC is defined as follows:

**Types:**

$$\tau ::= \text{bool} \mid \tau \rightarrow \tau$$

**Terms:**

$$e ::= x \mid \text{true} \mid \text{false} \mid \text{if } e \text{ then } e \text{ else } e \mid \lambda x : \tau. e \mid e e$$

**Values:**

$$v ::= \text{true} \mid \text{false} \mid \lambda x : \tau. e$$

**Evaluation Context:**

$$E ::= \square \mid \text{if } E \text{ then } e \text{ else } e \mid E e \mid v E$$

**Evaluations:**

$$\frac{}{\text{if true then } e_1 \text{ else } e_2 \rightarrow e_1}$$

$$\frac{}{\text{if false then } e_1 \text{ else } e_2 \rightarrow e_2}$$

$$\frac{e \rightarrow e'}{E[e] \rightarrow E[e']}$$

$$\frac{}{(\lambda x : \tau. e) v \rightarrow e[v/x]}$$

**Typing contexts:**

$$\Gamma ::= \bullet \mid \Gamma, x : \tau$$

**Typing rules:**

$$\frac{}{\Gamma \vdash \text{false} : \text{bool}} \text{T-FALSE} \quad \frac{}{\Gamma \vdash \text{true} : \text{bool}} \text{T-TRUE} \quad \frac{\Gamma(x) = \tau}{\Gamma \vdash x : \tau} \text{T-VAR}$$

$$\frac{\Gamma, x : \tau_1 \vdash e : \tau_2}{\Gamma \vdash \lambda x : \tau_1. e : \tau_1 \rightarrow \tau_2} \text{T-ABS} \quad \frac{\Gamma \vdash e_1 : \tau_2 \rightarrow \tau \quad \Gamma \vdash e_2 : \tau_2}{\Gamma \vdash e_1 e_2 : \tau} \text{T-APP}$$

$$\frac{\Gamma \vdash e : \text{bool} \quad \Gamma \vdash e_1 : \tau \quad \Gamma \vdash e_2 : \tau}{\Gamma \vdash \text{if } e \text{ then } e_1 \text{ else } e_2 : \tau} \text{T-IF}$$

For the typing contexts, it is assumed that the binders ( $x$ ) are distinct. That is, if  $x \in \text{dom}(\Gamma)$ , then  $\Gamma, x : \tau$  is not a legal context.

For readers unfamiliar with inference rules: A rule

$$\frac{A \quad B}{A \wedge B}$$

is read as if  $A$  and  $B$  is the case, then we can conclude  $A \wedge B$ . This means that the typing rule for application

$$\frac{\Gamma \vdash e_1 : \tau_2 \rightarrow \tau \quad \Gamma \vdash e_2 : \tau_2}{\Gamma \vdash e_1 e_2 : \tau} \text{T-APP}$$

says that an application  $e_1 e_2$  has the type  $\tau$  under the typing context  $\Gamma$  when  $e_2$  has type  $\tau_2$  under  $\Gamma$  and  $e_1$  has type  $\tau_2 \rightarrow \tau$  also under  $\Gamma$ .

## 1.2 Logical Relations

A logical relation is a proof method that can be used to prove properties of programs written in a particular programming language. Proofs for properties of programming languages often go by induction on the typing or evaluation judgement. A logical relations add a layer of indirection by constructing a collection of programs that all have the property we are interested in. We will see this in more detail later. As a motivation, here are a number of examples of properties that can be proven with a logical relation:

- Termination (Strong normalization)
- Type safety
- Program equivalences
  - Correctness of programs
  - Representation independence
  - Parametricity and free theorems, e.g.

$$f : \forall \alpha. \alpha \rightarrow \alpha$$

The program cannot inspect  $\alpha$  as it has no idea which type it will be, therefore  $f$  must be the identity function.

$$\forall \alpha. \text{int} \rightarrow \alpha$$

A function with this type cannot exist (the function would need to return something of type  $\alpha$ , but it only has something of type  $\text{int}$  to work with, so it cannot possibly return a value of the proper type).

- Security-Typed Languages (for Information Flow Control (IFC))

Example: All types in the code snippet below are labelled with their security level. A type can be labelled with either  $L$  for *low* or  $H$  for *high*. We do not want any information flowing from variables with a *high* label to a variable with a *low* label. The following is an example of an insecure program because it has an *explicit flow* of information from *low* to *high*:

```
x : intL
y : intH
x = y    //This assignment is insecure.
```

Information may also leak through a *side channel*. There are many varieties of side channels, and they vary from language to language depending on their features. One of the perhaps simplest side channels is the following: Say we have two variable  $x$  and  $y$ . They are both of integer type, but the former is labelled with *low* and the latter with *high*. Now say the value of  $x$  depends on the value of  $y$ , e.g.  $x = 0$  when  $y > 0$  and  $x = 1$  otherwise. In this example, we may not learn the exact value of  $y$  from  $x$ , but we will have learned whether  $y$  is positive. The side channel we just sketched looks as follows:

```
x : intL
y : intH
if y > 0 then x = 0 else x = 1
```

Generally, speaking the property we want is non-interference:

$$P(v_L, v_{1H}) \approx_L P(v_L, v_{2H}) \quad \text{for } \vdash P : \text{int}^L \times \text{int}^H \rightarrow \text{int}^L$$

That is for programs that generate *low* results, we want “*low-equivalent*” results. *Low-equivalence* means: if we execute  $P$  twice with the same *low* value but two different *high* values, then the *low* results of the two executions should be equal. In other words, the execution cannot have depended on the *high* value which means that no information was leaked to the *low* results.

## 1.3 Categories of Logical Relations

We can split logical relations into two: logical predicates and logical relations. Logical predicates are unary and are usually used to show properties of programs. Logical relations are binary and are usually used to show equivalences:

Logical Predicates	Logical Relations
(Unary)	(Binary)
$P_\tau(e)$	$R_\tau(e_1, e_2)$
- One property	- Program Equivalence
- Strong normalization	
- Type safety	

There are some properties that we want logical predicates and relation to have in general. We describe these properties for logical predicates as they easily generalize to logical relations. In general, we want the following things to hold true for a logical predicate that contains expressions  $e^1$ :

1. The expressions is closed and well-types, i.e.  $\bullet \vdash e : \tau$ .
2. The expression has the property we are interested in.
3. The property of interest is preserved by eliminating forms.

## 2 Normalization of the Simply Typed Lambda Calculus

### 2.1 Strong Normalization of STLC

In this section, we prove strong normalization for the simply typed lambda calculus which means that every term is strongly normalizing. Normalization of a term is the process of reducing it to its normal form (where it can be reduced no further). If a term is strongly normalizing, then it always reduces to its normal form. In our case, the normal forms of the language are the values of the language.

#### A first attempt at proving strong normalization for STLC

We will first attempt a syntactic proof of the strong normalization property of STLC to demonstrate how it fails. However, first we need to state what we mean by strong normalization.

---

<sup>1</sup>Note: these are rules of thumb. For instance, one exception to the rule is the proof of type safety where the well-typedness condition is weakened to only require  $e$  to be closed.

**Definition 1.** For expression  $e$  and value  $v$ :

$$e \Downarrow v \stackrel{\text{def}}{=} e \rightarrow^* v$$

$$e \Downarrow \stackrel{\text{def}}{=} \exists v. e \Downarrow v \quad \blacksquare$$

**Theorem 1** (Strong Normalization). If  $\bullet \vdash e : \tau$ , then  $e \Downarrow$  ■

*Proof.* ; This proof gets stuck and is not complete. !

Induction on the structure of the typing derivation.

**Case**  $\bullet \vdash \text{true} : \text{bool}$ , this term has already terminated.

**Case**  $\bullet \vdash \text{false} : \text{bool}$ , same as for **true**.

**Case**  $\bullet \vdash \text{if } e \text{ then } e_1 \text{ else } e_2 : \tau$ , simple, but requires the use of canonical forms of  $\text{bool}^2$ .

**Case**  $\bullet \vdash \lambda x : \tau_1. e : \tau_1 \rightarrow \tau_2$ , it is a value already and it has terminated.

**Case** 
$$\frac{\Gamma \vdash e_1 : \tau_2 \rightarrow \tau \quad \Gamma \vdash e_2 : \tau_2}{\Gamma \vdash e_1 e_2 : \tau} \text{ T-APP}$$

by the induction hypothesis, we get  $e_1 \Downarrow v_1$  and  $e_2 \Downarrow v_2$ . By the type of  $e_1$ , we conclude  $e_1 \Downarrow \lambda x : \tau_2. e'$ . What we need to show is  $e_1 e_2 \Downarrow$ . By the evaluation rules, we know  $e_1 e_2$  takes the following steps:

$$\begin{aligned} e_1 e_2 &\rightarrow^* (\lambda x : \tau_2. e') e_2 \\ &\rightarrow^* (\lambda x : \tau_2. e') v_2 \\ &\rightarrow e'[v_2/x] \end{aligned}$$

Here we run into an issue a know nothing about  $e'$ . Our induction hypothesis is simply not strong enough<sup>3</sup>. ⊠

The direct style proof did not work in this case, and it is not clear what to do to make it work.

## Proof of strong normalization using a logical predicate

Now that the direct proof failed, we try using a logical predicate. First, we define the predicate  $\text{SN}_\tau(e)$ :

$$\begin{aligned} \text{SN}_{\text{bool}}(e) &\Leftrightarrow \bullet \vdash e : \text{bool} \wedge e \Downarrow \\ \text{SN}_{\tau_1 \rightarrow \tau_2}(e) &\Leftrightarrow \bullet \vdash e : \tau_1 \rightarrow \tau_2 \wedge e \Downarrow \wedge (\forall e'. \text{SN}_{\tau_1}(e') \implies \text{SN}_{\tau_2}(e e')) \end{aligned}$$

---

<sup>2</sup>See Pierce's Types and Programming Languages [Pierce, 2002] for more about canonical forms.

<sup>3</sup>:(

Now recall the three conditions from Section 1.3 that a logical predicate should satisfy. It is easy to verify that  $\text{SN}_\tau(e)$  only accepts closed well-typed terms. Further, the predicate also requires terms to have the property we are interested in proving, namely  $e \Downarrow$ . Finally, it should satisfy that “*the property of interest is preserved by eliminating forms*”. In STLC lambdas are eliminated by application which means that application should preserve strong normalization when the argument is strongly normalizing.

The logical predicate is defined over the structure of  $\tau$  which has *bool* as a base type, so the definition is well-founded<sup>4</sup>. We are now ready to prove strong normalization using  $\text{SN}_\tau(e)$ . To this end, we have the following lemmas:

**Lemma 1.** *If  $\bullet \vdash e : \tau$ , then  $\text{SN}_\tau(e)$*  ■

**Lemma 2.** *If  $\text{SN}_\tau(e)$ , then  $e \Downarrow$*  ■

These two lemmas are common for proofs using a logical predicate (or relation). We first prove that all well-typed terms are in the predicate, and then we prove that all terms in the predicate have the property we want to show (in this case strong normalization).

The proof of Lemma 2 is by induction on  $\tau$ . This proof is straightforward because the strong normalization was baked into the predicate. It is generally a straightforward proof as our rules of thumb guide us to bake the property of interest into the predicate.

To prove Lemma 1, we could try induction over  $\bullet \vdash e : \tau$ , but the case we will fail to show the case for T-ABS. Instead we prove a generalization of Lemma 1:

**Theorem 2** (Lemma 1 generalized). *If  $\Gamma \vdash e : \tau$  and  $\gamma \models \Gamma$ , then  $\text{SN}_\tau(\gamma(e))$*  ■

This theorem uses a substitution  $\gamma$  to close off the expression  $e$ . In order for  $\gamma$  to close off  $e$ , it must map all the possible variables of  $e$  to strongly normalizing terms. When we prove this lemma, we get a stronger induction hypothesis than we have when we try to prove Lemma 1. In Lemma 1, the induction hypothesis can only be used with a closed term; but in this lemma, we can use an open term provided we have a substitution that closes it.

To be more specific, a substitution  $\gamma = \{x_1 \mapsto v_1, \dots, x_n \mapsto v_n\}$  works as follows:

---

<sup>4</sup>This may seem like a moot point as it is so obvious, but for some of the logical relations we see later it is not so, so we may as well start the habit of checking this now.

**Definition 2.**

$$\begin{aligned}\emptyset(e) &= e \\ \gamma[x \mapsto v](e) &= \gamma(e[v/x])\end{aligned}\quad \blacksquare$$

<sup>5</sup>and  $\gamma \models \Gamma$  is read “the substitution  $\gamma$  satisfies the type environment  $\Gamma$ ”, and it is defined as follows:

$$\begin{aligned}\gamma \models \Gamma \\ \text{iff} \\ \text{dom}(\gamma) = \text{dom}(\Gamma) \wedge \forall x \in \text{dom}(\Gamma). \text{SN}_{\Gamma(x)}(\gamma(x))\end{aligned}$$

To prove Theoremq 2 we need two further lemmas

**Lemma 3** (Substitution Lemma). *If  $\Gamma \vdash e : \tau$  and  $\gamma \models \Gamma$ , then  $\bullet \vdash \gamma(e) : \tau$*  ■

*Proof.* Left as an exercise. □

**Lemma 4** (SN preserved by forward/backward reduction). *Suppose  $\bullet \vdash e : \tau$  and  $e \rightarrow e'$*

1. *if  $\text{SN}_{\tau}(e')$ , then  $\text{SN}_{\tau}(e)$*
2. *if  $\text{SN}_{\tau}(e)$ , then  $\text{SN}_{\tau}(e')$*

■

*Proof.* Left as an exercise. □

*Proof. (1 Generalized).* Proof by induction on  $\Gamma \vdash e : \tau$ .

**Case T-TRUE,**

Assume:

- $\Gamma \vdash \text{true} : \text{bool}$
- $\gamma \models \Gamma$

We need to show:

$$\text{SN}_{\text{bool}}(\gamma(\text{true}))$$

There is no variable, so the substitution does nothing, and we just need to show  $\text{SN}_{\text{bool}}(\text{true})$  which is true as  $\text{true} \Downarrow \text{true}$ .

**Case T-FALSE,** similar to the **true** case.

**Case T-VAR,**

Assume:

---

<sup>5</sup>We do not formally define substitution ( $e[v/x]$ ). We refer to Pierce [2002] for a formal definition.

- $\Gamma \vdash x : \tau$
- $\gamma \models \Gamma$

We need to show:

$$\text{SN}_\tau(\gamma(x))$$

This case follows from the definition of  $\gamma \models \Gamma$ . We know that  $x$  is well-typed, so it is in the domain of  $\Gamma$ . From the definition of  $\gamma \models \Gamma$ , we get  $\text{SN}_{\Gamma(x)}(\gamma(x))$ . From well-typedness of  $x$ , we have  $\Gamma(x) = \tau$  which then gives us what we needed to show.

**Case T-IF**, left as an exercise.

**Case T-APP**,

Assume:

- $\Gamma \vdash e_1 e_2 : \tau$
- $\gamma \models \Gamma$

We need to show:

$$\text{SN}_\tau(\gamma(e_1 e_2))$$

which amounts to  $\text{SN}_\tau(\gamma(e_1) \gamma(e_2))$ . By the induction hypothesis we have

$$\text{SN}_{\tau_2 \rightarrow \tau}(\gamma(e_1)) \tag{1}$$

$$\text{SN}_{\tau_2}(\gamma(e_2)) \tag{2}$$

By the 3rd property of (1),  $\forall e'. \text{SN}_{\tau_2}(e') \implies \text{SN}_\tau(\gamma(e_1) e')$ , instantiated with (2), we get  $\text{SN}_\tau(\gamma(e_1) \gamma(e_2))$  which is the result we need.

Note this was the case we got stuck in when we tried to do the direct proof. With the logical predicate, it is easily proven because we made sure to bake information about  $e_1 e_2$  into  $\text{SN}_{\tau_2 \rightarrow \tau}$  when we follow the rule of thumb: “*The property of interest is preserved by eliminating forms*”.

**Case T-ABS**,

Assume:

- $\Gamma \vdash \lambda x : \tau_1. e : \tau_1 \rightarrow \tau_2$
- $\gamma \models \Gamma$

We need to show:

$$\text{SN}_{\tau_1 \rightarrow \tau_2}(\gamma(\lambda x : \tau_1. e))$$

which amounts to  $\text{SN}_{\tau_1 \rightarrow \tau_2}(\lambda x : \tau_1. \gamma(e))$ . Our induction hypothesis in this case reads:

$$\Gamma, x : \tau_1 \vdash e : \tau_2 \wedge \gamma' \models \Gamma, x : \tau_1 \implies \text{SN}_{\tau_2}(\gamma'(e))$$

It suffices to show the following three things:

1.  $\bullet \vdash \lambda x : \tau_1. \gamma(e) : \tau_1 \rightarrow \tau_2$
2.  $\lambda x : \tau_1. \gamma(e) \Downarrow$
3.  $\forall e'. \text{SN}_{\tau_1}(e') \implies \text{SN}_{\tau_2}((\lambda x : \tau_1. \gamma(e)) e')$

If we use the substitution lemma (Lemma 3) and push the  $\gamma$  in under the  $\lambda$ -abstraction, then we get 1. The lambda-abstraction is a value, so by definition 2 is true.

It only remains to show 3. To do this, we want to somehow apply the induction hypothesis for which we need a  $\gamma'$  such that  $\gamma' \models \Gamma, x : \tau_1$ . We already have  $\gamma$  and  $\gamma \models \Gamma$ , so our  $\gamma'$  should probably have the form  $\gamma' = \gamma[x \mapsto v_?]$  for some  $v_?$  of type  $\tau_1$ . Let us move on and see if any good candidates for  $v_?$  present themselves.

Let  $e'$  be given and assume  $\text{SN}_{\tau_1}(e')$ . We then need to show  $\text{SN}_{\tau_2}((\lambda x : \tau_1. \gamma(e)) e')$ . From  $\text{SN}_{\tau_1}(e')$ , it follows that  $e' \Downarrow v'$  for some  $v'$ .  $v'$  is a good candidate for  $v_?$  so let  $v_? = v'$ . From the forward part of the preservation lemma (Lemma 4), we can further conclude  $\text{SN}_{\tau_1}(v')$ . We use this to conclude  $\gamma[x \mapsto v'] \models \Gamma, x : \tau_1$  which we use with the assumption  $\Gamma, x : \tau_1 \vdash e : \tau_2$  to instantiate the induction hypothesis and get  $\text{SN}_{\tau_2}(\gamma[x \mapsto v'](e))$ .

Now consider the following evaluation:

$$\begin{aligned} (\lambda x : \tau_1. \gamma(e)) e' &\rightarrow^* (\lambda x : \tau_1. \gamma(e)) v' \\ &\rightarrow \gamma(e)[v'/x] \equiv \gamma[x \mapsto v'](e) \end{aligned}$$

We already concluded that  $e' \rightarrow^* v'$ , which corresponds to the first series of steps. We can then do a  $\beta$ -reduction to take the next step, and finally we get something that is equivalent to  $\gamma[x \mapsto v'](e)$ . That is we have the evaluation

$$(\lambda x : \tau_1. \gamma(e)) e' \rightarrow^* \gamma[x \mapsto v'](e)$$

From  $\text{SN}_{\tau_1}(e')$ , we have  $\bullet \vdash e' : \tau_1$  and we already argued that  $\bullet \vdash \lambda x : \tau_1. \gamma(e) : \tau_1 \rightarrow \tau_2$ , so from the application typing rule we get  $\bullet \vdash (\lambda x : \tau_1. \gamma(e)) e' : \tau_2$ . We can use this with the above evaluation and the forward part of the preservation lemma (Lemma 4) to argue that every intermediate expressions in the steps down to  $\gamma[x \mapsto v'](e)$  are closed and well typed.

If we use  $\text{SN}_{\tau_2}(\gamma[x \mapsto v'](e))$  with  $(\lambda x : \tau_1. \gamma(e)) e' \rightarrow^* \gamma[x \mapsto v'](e)$  and the fact that every intermediate step in the evaluation is closed and well typed, then we can use the backward reduction part of the SN preservation lemma to get  $\text{SN}_{\tau_2}((\lambda x : \tau_1. \gamma(e)) e')$  which is the result we wanted.  $\square$

## 2.2 Exercises

1. Prove SN preserved by forward/backward reduction (Lemma 4).
2. Prove the substitution lemma (Lemma 3).
3. Go through the cases of the proof for Theorem 2 by yourself.
4. Prove the T-IF case of Theorem 2.
5. Extend the language with pairs and adjust the proofs.

Specifically, how do you apply the rules of thumb for the case of pairs? Do you need to add anything for the third clause (eliminating forms preservation), or does it work without doing anything for it like it did for case of booleans?

recu

## 3 Type Safety for STLC

In this section, we prove type safety for simply typed lambda calculus using a logical predicate.

First we need to consider what type safety is. The classical mantra for type safety is: “Well-typed programs do not *go wrong*.” It depends on the language and type system what *go wrong* actually means, but in our case a program has *gone wrong* when it is stuck<sup>6</sup> (an expression is stuck if it is irreducible but not a value).

### 3.1 Type safety - the classical treatment

Type safety for simply typed lambda calculus is stated as follows:

**Theorem 3** (Type Safety for STLC). *If  $\bullet \vdash e : \tau$  and  $e \rightarrow^* e'$ , then  $\text{Val}(e')$  or  $\exists e''. e' \rightarrow e''$ .* ■

Traditionally, the type safety proof uses two lemmas: progress and preservation.

**Lemma 5** (Progress). *If  $\bullet \vdash e : \tau$ , then  $\text{Val}(e)$  or  $\exists e'. \rightarrow e'$ .* ■

Progress is normally proved by induction on the typing derivation.

---

<sup>6</sup>In the case of language-based security and information flow control, the notion of *go wrong* would be that there is an undesired flow of information.

**Lemma 6** (Preservation). *If  $\bullet \vdash e : \tau$  and  $e \rightarrow e'$ , then  $\bullet \vdash e' : \tau$ .* ■

Preservation is normally proved by induction on the evaluation. Preservation is also known as *subject reduction*. Progress and preservation talk about one step, so to prove type safety we have to do induction on the evaluation. Here we do not want to prove type safety the traditional way, but if you are unfamiliar with it and want to learn more, then we refer to Pierce's *Types and Programming Languages* [Pierce, 2002].

We will use a logical predicate (as it is a unary property) to prove type safety.

### 3.2 Type safety - using logical predicate

We define the predicate in a slightly different way compared to Section 2. We define it in two parts: a value interpretation and an expression interpretation. The value interpretation is a function from types to the power set of closed values:

$$\mathcal{V}[\![-]\!] : \text{Type} \rightarrow \mathcal{P}(\text{ClosedVal})$$

The value interpretation is defined as:

$$\begin{aligned} \mathcal{V}[\![\text{bool}]\!] &= \{\mathbf{true}, \mathbf{false}\} \\ \mathcal{V}[\![\tau_1 \rightarrow \tau_2]\!] &= \{\lambda x : \tau_1. e \mid \forall v \in \mathcal{V}[\![\tau_1]\!]. e[v/x] \in \mathcal{E}[\![\tau_2]\!]\} \end{aligned}$$

We define the expression interpretation as:

$$\mathcal{E}[\![\tau]\!] = \{e \mid \forall e'. e \rightarrow^* e' \wedge \text{irred}(e') \implies e' \in \mathcal{V}[\![\tau]\!]\}$$

Notice that neither  $\mathcal{V}[\![\tau]\!]$  nor  $\mathcal{E}[\![\tau]\!]$  requires well-typedness. Normally, the logical predicate would require this as our guidelines suggest it. However, as the goal is to prove type safety we do not want it as a part of the predicate. In fact, if we did include a well-typedness requirement, then we would end up proving preservation for some of the proofs to go through. We do, however, require the value interpretation to only contain closed values.

An expression is irreducible if it is unable to take any reduction steps according to the evaluation rules. The predicate *irred* captures whether an expression is irreducible:

$$\begin{aligned} &\text{irred}(e) \\ &\text{iff} \\ &\nexists e'. e \rightarrow e' \end{aligned}$$

The sets are defined on the structure of the types.  $\mathcal{V}[\tau_1 \rightarrow \tau_2]$  contains  $\mathcal{E}[\tau_2]$ , but  $\mathcal{E}[\tau_2]$  uses  $\tau_2$  directly in  $\mathcal{V}[\tau_2]$ , so the definition is structurally well-founded. To prove type safety, we first define a new predicate, *safe*:

$$\begin{aligned} & \text{safe}(e) \\ & \text{iff} \\ & \forall e'. e \rightarrow^* e' \implies \text{Val}(e') \vee \exists e''. e' \rightarrow e'' \end{aligned}$$

An expression  $e$  is *safe* if it can take a number of steps and end up either as a value or as an expression that can take another step.

We are now ready to prove type safety. Just like we did for strong normalization, we use two lemmas:

**Lemma 7.** *If  $\bullet \vdash e : \tau$ , then  $e \in \mathcal{E}[\tau]$*  ■

**Lemma 8.** *If  $e \in \mathcal{E}[\tau]$ , then  $\text{safe}(e)$*  ■

Rather than proving Lemma 7 directly, we prove a more general theorem and get Lemma 7 as a consequence. We are not yet in a position to state the generalization. First, we need to define the interpretation of environments:

$$\begin{aligned} \mathcal{G}[\bullet] &= \{\emptyset\} \\ \mathcal{G}[\Gamma, x : \tau] &= \{\gamma[x \mapsto v] \mid \gamma \in \mathcal{G}[\Gamma] \wedge v \in \mathcal{V}[\tau]\} \end{aligned}$$

Further, we need to define semantic type safety:

$$\begin{aligned} \Gamma \models e : \tau \\ & \text{iff} \\ & \forall \gamma \in \mathcal{G}[\Gamma]. \gamma(e) \in \mathcal{E}[\tau] \end{aligned}$$

This definition should look familiar because we use the same trick as we did for strong normalization: Instead of only considering closed terms, we consider all terms but require a substitution that closes it.

We can now define our generalized version of Lemma 7:

**Theorem 4** (Fundamental Property). *If  $\Gamma \vdash e : \tau$ , then  $\Gamma \models e : \tau$*  ■

A theorem like this would typically be the first you prove after defining a logical relation. In this case, the theorem says that syntactic type safety implies semantic type safety.

We also alter Lemma 8 to fit with Theorem 4:

**Lemma 9.** *If  $\emptyset \models e : \tau$ , then  $\text{safe}(e)$*  ■

*Proof.* Suppose  $e \rightarrow^* e'$  for some  $e'$ , then we need to show  $\text{Val}(e')$  or  $\exists e''. e' \rightarrow e''$ . We proceed by case on whether or not  $\text{irred}(e')$ :

**Case**  $\neg \text{irred}(e')$ , this case follows directly from the definition of  $\text{irred}$ :  $\text{irred}(e')$  is defined as  $\nexists e''. e' \rightarrow e''$  and as the assumption is  $\neg \text{irred}(e')$ , we get  $\exists e''. e' \rightarrow e''$ .

**Case**  $\text{irred}(e')$ , by assumption we have  $\bullet \models e : \tau$ . As the typing context is empty, we choose the empty substitution and get  $e \in \mathcal{E}[\tau]$ . We now use the definition of  $e \in \mathcal{E}[\tau]$  with the two assumptions  $e \rightarrow^* e'$  and  $\text{irred}(e')$  to conclude  $e' \in \mathcal{V}[\tau]$ . As  $e'$  is in the value interpretation of  $\tau$ , we can conclude  $\text{Val}(e')$ . □

To prove the Fundamental Property (Theorem 4), we need a substitution lemma:

**Lemma 10** (Substitution). *Let  $e$  be syntactically well-formed term, let  $v$  be a closed value and let  $\gamma$  be a substitution that maps term variables to closed values, and let  $x$  be a variable not in the domain of  $\gamma$ , then*

$$\gamma[x \mapsto v](e) = \gamma(e)[v/x] \quad \blacksquare$$

*Proof.* By induction on the size of  $\gamma$ .

**Case**  $\gamma = \emptyset$ , this case is immediate by definition of substitution. That is by definition we have  $[x \mapsto v]e = e[v/x]$ .

**Case**  $\gamma = \gamma'[y \mapsto v']$ ,  $x \neq y$ , in this case our induction hypothesis is:

$$\gamma'[x \mapsto v]e = \gamma'(e)[v/x]$$

We wish to show

$$\gamma'[y \mapsto v']([x \mapsto v]e) = \gamma'[y \mapsto v'](e)[v/x]$$

$$\gamma'[y \mapsto v']([x \mapsto v]e) = \gamma'[x \mapsto v][y \mapsto v']e \tag{3}$$

$$= \gamma'[x \mapsto v](e[v'/y]) \tag{4}$$

$$= \gamma'(e[v'/y])[v/x] \tag{5}$$

$$= \gamma'[y \mapsto v'](e)[v/x] \tag{6}$$

In the first step (3), we swap the two mappings. It is safe to do so as both  $v$  and  $v'$  are closed values, so we know that no variable capturing will occur. In the second step (4), we just use the definition of substitution (Definition 2). In the third step (5), we use the induction hypothesis<sup>7</sup>. Finally in the last step (6), we use the definition of substitution to get the  $y$  binding out as an extension of  $\gamma'$ . □

<sup>7</sup>The induction hypothesis actually has a number of premises, as an exercise convince yourself that they are satisfied.

*Proof. (Fundamental Property, Theorem 4).* Proof by induction on the typing judgement.

**Case T-ABS,**

assuming  $\Gamma \vdash \lambda x : \tau_1. e : \tau_1 \rightarrow \tau_2$ , we need to show  $\Gamma \models \lambda x : \tau_1. e : \tau_1 \rightarrow \tau_2$ . Suppose  $\gamma \in \mathcal{G}[\Gamma]$  and show

$$\gamma(\lambda x : \tau_1. e) \in \mathcal{E}[\tau_1 \rightarrow \tau_2] \equiv (\lambda x : \tau_1. \gamma(e)) \in \mathcal{E}[\tau_1 \rightarrow \tau_2]$$

Now suppose that  $\lambda x : \tau_1. \gamma(e) \rightarrow^* e'$  and  $\text{irred}(e')$ . We then need to show  $e' \in \mathcal{V}[\tau_1 \rightarrow \tau_2]$ .  $\lambda x : \tau_1. \gamma(e)$  is irreducible because it is a value, and we can conclude it takes no steps. In other words  $e' = \lambda x : \tau_1. \gamma(e)$ . This means we need to show  $\lambda x : \tau_1. \gamma(e) \in \mathcal{V}[\tau_1 \rightarrow \tau_2]$ . Now suppose  $v \in \mathcal{V}[\tau_1]$ , then we need to show  $\gamma(e)[v/x] \in \mathcal{E}[\tau_2]$ .

Keep the above proof goal in mind and consider the induction hypothesis:

$$\Gamma, x : \tau_1 \models e : \tau_2$$

Instantiate this with  $\gamma[x \mapsto v]$ . We have  $\gamma[x \mapsto v] \in \mathcal{G}[\Gamma, x : \tau_1]$  because of assumptions  $\gamma \in \mathcal{G}[\Gamma]$  and  $v \in \mathcal{V}[\tau_1]$ . The instantiation gives us  $\gamma[x \mapsto v](e) \in \mathcal{E}[\tau_2] \equiv \gamma(e)[v/x] \in \mathcal{E}[\tau_2]$ . The equivalence is justified by the substitution lemma we proved. This is exactly the proof goal we kept in mind.

**Case T-APP,** Show this case as an exercise.

The remaining cases are straightforward.  $\square$

Now consider what happens if we add pairs to the language (exercise 5 in exercise section 2.2). We need to add a clause to the value interpretation:

$$\mathcal{V}[\tau_1 \times \tau_2] = \{\langle v_1, v_2 \rangle \mid v_1 \in \mathcal{V}[\tau_1] \wedge v_2 \in \mathcal{V}[\tau_2]\}$$

There is nothing surprising in this addition to the value relation, and it should not be a challenge to show the pair case of the proofs.

If we extend our language with sum types.

$$e ::= \dots \mid \text{inl } v \mid \text{inr } v \mid \text{case } e \text{ of } \text{inl } x \Rightarrow e_1 \quad \text{inr } x \Rightarrow e_2$$

Then we need to add the following clause to the value interpretation:

$$\mathcal{V}[\tau_1 + \tau_2] = \{\text{inl } v \mid v \in \mathcal{V}[\tau_1]\} \cup \{\text{inr } v \mid v \in \mathcal{V}[\tau_2]\}$$

It turns out this clause is sufficient. One might think that it is necessary to require the body of the match to be in the expression interpretation, which looks something like  $\forall e \in \mathcal{E}[\tau]$ . This requirement will, however, give well-foundedness problems, as  $\tau$  is not a structurally smaller type than  $\tau_1 + \tau_2$ . It may come as a surprise that we do not need to relate the expressions as the slogan for logical relations is: “Related inputs to related outputs.”

### 3.3 Exercises

1. Prove the T-APP case of the Fundamental Property (Theorem 4).
2. Verify the remaining cases of Theorem 4 (T-TRUE, T-FALSE, T-VAR, and T-IF).

## 4 Universal Types and Relational Substitutions

In the previous sections, we considered the unary properties and safety and termination, but now we shift our focus to relational properties and specifically program equivalences. Generally speaking, we use logical relations rather than predicates for proving relational properties. A program equivalence is a relational property, so we are going to need a logical relation.

We will consider the language System F which is STLC with universal types. Universal types allow us to write generic functions. For instance, say we have a function that sorts integer lists:

$$\text{sortint} : \text{list int} \rightarrow \text{list int}$$

The function takes a list of integers and returns the sorted list. Say we now want a function `sortstring` that sorts lists of strings. Instead of implementing a completely new sorting function for this, we could factor out the generic code responsible for sorting from `sortint` and make a generic function. The type signature for a generic sort function would be:

$$\text{sort} : \forall \alpha. (\text{list } \alpha) \times (\alpha \times \alpha \rightarrow \text{bool}) \rightarrow \text{list } \alpha$$

The generic sort function takes a type, a list of elements of this type, and a comparison function that compares to elements of the given type. The result of the sorting function is a list sorted according to the comparison function. An example of an application of this function could be

$$\text{sort [int]} (3, 7, 5) <$$

Whereas `sort` instantiated with the string type, but given an integer list would not be a well typed instantiation.

$$\text{sort [string]} \overset{("a", "c", "b")}{\cancel{(3, 7, 5)}} \text{string} <$$

Here the application with the list  $(3, 7, 5)$  is not well typed, but if we instead use a list of strings, then it type checks.

We want to extend the simply typed lambda calculus with functions that abstract over types in the same way lambda abstractions,  $\lambda x : \tau. e$ , abstract over terms. We do that by introducing a type abstraction:

$$\Lambda\alpha. e$$

This function abstracts over the type  $\alpha$  which allows  $e$  to depend on  $\alpha$ .

## 4.1 System F (STLC with universal types)

We extend STLC as follows to get System F:

$$\begin{aligned} \tau &::= \dots \mid \forall\alpha. \tau \\ e &::= \dots \mid \Lambda\alpha. e \mid e[\tau] \\ v &::= \dots \mid \Lambda\alpha. e \\ E &::= \dots \mid E[\tau] \end{aligned}$$

New evaluation:

$$\overline{(\Lambda\alpha. e)[\tau] \rightarrow e[\tau/\alpha]}$$

Type environment:

$$\Delta ::= \bullet \mid \Delta, \alpha$$

The type environment consists of type variables, and they are assumed to be distinct type variables. That is, the environment  $\Delta, \alpha$  is only well-formed if  $\alpha \notin \text{dom}(\Delta)$ <sup>8</sup>. With the addition of type environments, we update the form of the typing judgement as follows

$$\Delta, \Gamma \vdash e : \tau$$

We now need a notion of well-formed types. If  $\tau$  is well formed with respect to  $\Delta$ , then we write:

$$\Delta \vdash \tau$$

We do not include the formal rules here, but they amount to  $\text{FTV}(\tau) \subseteq \Delta$ , where  $\text{FTV}(\tau)$  is the set of free type variables in  $\tau$ .

We further introduce a notion of well-formed typing contexts. A context is well formed if all the types that appear in the range of  $\Gamma$  are well formed.

$$\begin{aligned} \Delta \vdash \Gamma \\ \text{iff} \\ \forall x \in \text{dom}(\Gamma). \Delta \vdash \Gamma(x) \end{aligned}$$

---

<sup>8</sup>We do not annotate  $\alpha$  with a kind as we only have one kind in this language.

For any type judgment  $\Delta, \Gamma \vdash e : \tau$ , we have as an invariant that  $\tau$  is well formed in  $\Delta$  and  $\Gamma$  is well formed in  $\Delta$ . The old typing system modified to use the new form of the typing judgment looks like this:

$$\begin{array}{c}
\frac{}{\Delta; \Gamma \vdash \mathbf{false} : \mathit{bool}} \text{T-FALSE} \qquad \frac{}{\Delta; \Gamma \vdash \mathbf{true} : \mathit{bool}} \text{T-TRUE} \\
\\
\frac{\Gamma(x) = \tau}{\Delta; \Gamma \vdash x : \tau} \text{T-VAR} \qquad \frac{\Delta; \Gamma \vdash e : \mathit{bool} \quad \Delta; \Gamma \vdash e_1 : \tau \quad \Delta; \Gamma \vdash e_2 : \tau}{\Delta; \Gamma \vdash \mathbf{if } e \mathbf{ then } e_1 \mathbf{ else } e_2 : \tau} \text{T-IF} \\
\\
\frac{\Delta; \Gamma, x : \tau_1 \vdash e : \tau_2}{\Delta; \Gamma \vdash \lambda x : \tau_1. e : \tau_1 \rightarrow \tau_2} \text{T-ABS} \qquad \frac{\Delta; \Gamma \vdash e_1 : \tau_2 \rightarrow \tau \quad \Delta; \Gamma \vdash e_2 : \tau_2}{\Delta; \Gamma \vdash e_1 e_2 : \tau} \text{T-APP}
\end{array}$$

Notice that the only thing that has changed is that  $\Delta$  has been added to the environment in the judgments. We further extend the typing rules with the following two rules to account for our new language constructs:

$$\frac{\Delta; \Gamma \vdash e : \forall \alpha. \tau \quad \Delta \vdash \tau'}{\Delta; \Gamma \vdash e[\tau'] : \tau[\tau'/\alpha]} \text{T-TAPP} \qquad \frac{\Delta, \alpha; \Gamma \vdash e : \tau}{\Delta; \Gamma \vdash \Lambda \alpha. e : \forall \alpha. \tau} \text{T-TABS}$$

## 4.2 Properties of System-F: Free Theorems

In System-F, certain types reveal the behavior of the functions with that type. Let us consider terms with the type  $\forall \alpha. \alpha \rightarrow \alpha$ . Recall from Section 1.2 that this has to be the identity function. We can now phrase this as a theorem:

**Theorem 5.** *If all of the following hold*

- $\bullet; \bullet \vdash e : \forall \alpha. \alpha \rightarrow \alpha$
- $\bullet \vdash \tau$
- $\bullet; \bullet \vdash v : \tau$ ,

then

$$e[\tau] v \rightarrow^* v \quad \blacksquare$$

This is a free theorem in this language. Another free theorem from Section 1.2 is for expressions of type  $\forall \alpha. \alpha \rightarrow \mathit{bool}$ . All expressions with this type has to be constant functions. We can also phrase this as a theorem:

**Theorem 6.** *If all of the following hold*

- $\bullet; \bullet \vdash e : \forall \alpha. \alpha \rightarrow \text{bool}$
- $\bullet \vdash \tau$
- $\bullet; \bullet \vdash v_1 : \tau$
- $\bullet; \bullet \vdash v_2 : \tau$

then

$$e[\tau] v_1 \approx^{ctx} e[\tau] v_2 \quad \blacksquare$$

<sup>9</sup> We can even allow the type abstraction to be instantiated with different types:

**Theorem 7.** *If all of the following hold*

- $\bullet; \bullet \vdash e : \forall \alpha. \alpha \rightarrow \text{bool}$
- $\bullet \vdash \tau$
- $\bullet \vdash \tau'$
- $\bullet; \bullet \vdash v_1 : \tau$
- $\bullet; \bullet \vdash v_2 : \tau'$

then

$$e[\tau] v_1 \approx^{ctx} e[\tau'] v_2 \quad \blacksquare$$

We get these free theorems because the functions do not know the type of the argument which means that they have no way to inspect it. The function can only treat its argument as an unknown “blob”, so it has no choice but to return the same value every time.

The question now is: how do we prove these free theorems? The two last theorems both talk about program equivalence, so the proof technique of choice is a logical relation. The first theorem did not mention a program equivalence, but it can also be proven with a logical relation.

### 4.3 Contextual Equivalence

To define a contextual equivalence, we first define the notion of a program context. A program context is a complete program with exactly one hole in it:

$$C ::= [\cdot] \mid \text{if } C \text{ then } e \text{ else } e \mid \text{if } e \text{ then } C \text{ else } e \mid \text{if } e \text{ then } e \text{ else } C \mid \\ \lambda x : \tau. C \mid C e \mid e C \mid \Lambda \alpha. C \mid C[\tau]$$

---

<sup>9</sup>We have not defined  $\approx^{ctx}$  yet. For now, it suffices to know that it equates programs that are behaviorally the same.

We need a notion of context typing. For simplicity, we just introduce it for simply typed lambda calculus. The context typing is written as:

$$\frac{\Gamma \vdash e : \tau \quad \Gamma' \vdash C[e] : \tau'}{C : (\Gamma \vdash \tau) \Longrightarrow (\Gamma' \vdash \tau')}$$

This means that for any expression  $e$  of type  $\tau$  under  $\Gamma$  if we embed it into  $C$ , then the type of the embedding is  $\tau'$  under  $\Gamma'$ .

Informally we want contextual equivalence to express that two expressions give the same result no matter what program context they are plugged into. In other words, two expressions are contextually equivalent when no program context is able to observe any difference in behavior between the two expressions. For this reason, contextual equivalence is also known as *observational equivalence*. A hole has to be plugged with a term of the correct type, so we annotate the equivalence with the type of the hole which means that the two contextually equivalent expressions must have that type.

$$\begin{aligned} & \Delta; \Gamma \vdash e_1 \approx^{ctx} e_2 : \tau \\ & \text{iff} \\ & \forall C : (\Delta; \Gamma \vdash \tau) \Longrightarrow (\bullet; \bullet \vdash bool). (C[e_1] \Downarrow v \iff C[e_2] \Downarrow v) \end{aligned}$$

This definition assumes that  $e_1$  and  $e_2$  have type  $\tau$  under the specified contexts. Notice  $C$  plugged with a term is required to be closed, so it can be evaluated. When plugged with an expression,  $C$  is required to be of type *bool*. If we allowed  $C$  to be of function type when plugged with an expression, then the two functions produced would have to be equivalent. This especially becomes interesting when the program context is just a hole because we need to decide whether the two functions are equivalent. This is, however, what we are trying to define with contextual equivalence<sup>10</sup>.

Contextual equivalence can be used to argue about correctness of a program. For instance, say we have two implementations of a stack, one is implemented using an array and the other using a list. If we can show that the two implementations are contextual equivalent, then we can use either stack implementation in any context with the same result. In particular, if one of the stack implementations is highly optimized, then we would like to use that implementation. However,

---

<sup>10</sup>In non-terminating languages, the definition of contextual equivalence only mentions termination and not whether the values produced are equal. For non-terminating languages this makes sense because if a context is able to differentiate two expressions, then it is easy to construct another context that terminates when plugged with one expression but diverges when plugged with the other.

it is not always clear whether such highly optimized programs do what they are supposed to. In order to argue this, one could implement a stack reference or specification implementation which is slower but where it is easy to see that it behaves like a stack simply by inspecting the implementation. A proof of contextual equivalence between the highly optimized stack implementation and the specification implementation could be seen as a correctness proof *with respect to* the specification.

In the next part, we will introduce a logical relation such that

$$\Delta; \Gamma \vdash e_1 \approx^{LR} e_2 : \tau \implies \Delta; \Gamma \vdash e_1 \approx^{ctx} e_2 : \tau$$

That is we want to show that the logical relation is sound with respect to contextual equivalence.

If we can prove the above soundness, then we can state our free theorems with respect to  $\approx^{LR}$  and get the result with respect to  $\approx^{ctx}$  as a corollary. In fact, it is very difficult to prove contextual equivalences directly as the proof has to consider every possible program context. Such a proof would of course be by induction, but the lambda-abstraction case would be very difficult.

## 4.4 A Logical Relation for System F

In this section, we construct the logical relation for System F, so we can prove the free theorems of Section 4.2. We are defining a relation, so the value interpretation consists of pairs. All pairs in the value relation must be closed and well-typed:

$$\mathcal{V}[\tau] = \{(v_1, v_2) \mid \bullet; \bullet \vdash v_1 : \tau \wedge \bullet; \bullet \vdash v_2 : \tau \wedge \dots\}$$

For succinctness, we leave this implicit when we define the rest of the value relation. As a first try, we naively construct the logical relation in the same way we constructed the logical predicates in the previous sections:

$$\begin{aligned} \mathcal{V}[\mathit{bool}] &= \{(\mathit{true}, \mathit{true}), (\mathit{false}, \mathit{false})\} \\ \mathcal{V}[\tau \rightarrow \tau'] &= \left\{ (\lambda x : \tau. e_1, \lambda x : \tau. e_2) \mid \right. \\ &\quad \left. \forall (v_1, v_2) \in \mathcal{V}[\tau]. (e_1[v_1/x], e_2[v_2/x]) \in \mathcal{E}[\tau'] \right\} \end{aligned}$$

The value interpretation of the function type is defined based on the slogan for logical relations: “Related inputs to related outputs.” If we had chosen to use equal inputs rather than related, then our definition would be more restrictive than necessary.

We did not define a value interpretation for type variables in the previous sections, so let us try to push on without defining that part.

The next type is  $\forall\alpha. \tau$ . When we define the value interpretation, we consider the elimination forms which in this case is type application (cf. the guidelines in Section 1.3). Before we proceed, let us revisit Theorem 7, the last free theorem of Section 4.2. There are some important points to note in this free theorem. First of all, we want to be able to apply  $\Lambda$ -terms to different types, so we will have to pick two different types in our value interpretation. Further in the lambda-abstraction case, we pick related values, so perhaps we should pick related types here. However, we do not have a notion of related types, and the theorem does not relate the two types, so relating them might not be a good idea after all. With these points in mind, we make a first attempt at constructing the value interpretation of  $\forall\alpha. \tau$ :

$$\mathcal{V}[\forall\alpha. \tau] = \{(\Lambda\alpha. e_1, \Lambda\alpha. e_2) \mid \forall\tau_1, \tau_2. (e_1[\tau_1/\alpha], e_2[\tau_2/\alpha]) \in \mathcal{E}[\tau[?/\alpha]]\}$$

Now the question is what type to relate the two expressions under. We need to substitute  $?$  for some type, but if we use either  $\tau_1$  or  $\tau_2$ , then the well-typedness requirement will be broken for one of the two expressions. The solution is to scrap the substitution keep  $\tau$  unchanged. We do, however, need to keep track of the two types we picked for  $\alpha$ . To this end, we use a relational substitution:

$$\rho = \{\alpha_1 \mapsto (\tau_{11}, \tau_{12}), \dots\}$$

Which we parameterize the interpretations with:

$$\mathcal{V}[\forall\alpha. \tau]_\rho = \{(\Lambda\alpha. e_1, \Lambda\alpha. e_2) \mid \forall\tau_1, \tau_2. (e_1[\tau_1/\alpha], e_2[\tau_2/\alpha]) \in \mathcal{E}[\tau]_{\rho[\alpha \mapsto (\tau_1, \tau_2)]}\}$$

We need to parameterize the entire logical relation with the relational substitution, so we can close off values and use the information when polymorphic type variables are interpreted. By getting rid of the substitution, we are now interpreting types with free type variables, so we need to have a value interpretation of type variable  $\alpha$ :

$$\mathcal{V}[\alpha]_\rho = \{(v_1, v_2) \mid \rho(\alpha) = (\tau_1, \tau_2) \dots\}$$

However, the big question is: How should these values be related? To find the answer, we look to Theorem 7 again. In the theorem, the expression  $e$  is contextually related to itself when applied to different types and values. That is  $e[\tau] v_1$  is contextually equivalent to  $e[\tau'] v_2$  which means that  $e$  must treat  $v_1$  and  $v_2$  the same. This would suggest that,  $v_1$  and  $v_2$  should somehow be related under type  $\alpha$ . To achieve this in the value interpretation, we pick a relation  $R$  in the interpretation of type abstractions. Relation  $R$  relates values of type  $\tau$  and  $\tau'$  which in the example of the free theorem would allow us to relate  $v_1$  and  $v_2$ . We need this relation when we interpret type variables, so  $R$  is passed around as part of the

relational substitution. In the interpretation of type variables, we simply look-up  $R$  in the relational substitution and require the values to be related in the relation. With this in mind, we update the interpretation of type abstractions:

$$\mathcal{V}[\forall\alpha. \tau]_\rho = \left\{ (\Lambda\alpha. e_1, \Lambda\alpha. e_2) \left| \begin{array}{l} \forall\tau_1, \tau_2, R \in \text{Rel}[\tau_1, \tau_2]. \\ (e_1[\tau_1/\alpha], e_2[\tau_2/\alpha]) \in \mathcal{E}[\tau]_{\rho[\alpha \mapsto (\tau_1, \tau_2, R)]} \end{array} \right. \right\}$$

We do not require much of the relation,  $R$ . It has to be a set of pairs of values, and the values in every pair of the relation have to be closed and well typed under the corresponding type. So we define  $\text{Rel}[\tau_1, \tau_2]$  as:

$$\text{Rel}[\tau_1, \tau_2] = \left\{ R \in \mathcal{P}(\text{Val} \times \text{Val}) \left| \begin{array}{l} \bullet \vdash \tau_1 \wedge \bullet \vdash \tau_2 \wedge \\ \forall (v_1, v_2) \in R. \bullet \vdash v_1 : \tau_1 \wedge \bullet \vdash v_2 : \tau_2 \end{array} \right. \right\}$$

In the interpretation of  $\alpha$ , we require the values to be related under the relation we choose in the value interpretation of  $\forall\alpha. \tau$ :

$$\mathcal{V}[\alpha]_\rho = \{(v_1, v_2) \mid \rho(\alpha) = (\tau_1, \tau_2, R) \wedge (v_1, v_2) \in R\}$$

For convenience, we introduce the following notation: Given

$$\rho = \{\alpha_1 \mapsto (\tau_{11}, \tau_{12}, R_1), \alpha_2 \mapsto (\tau_{21}, \tau_{22}, R_2), \dots\}$$

define the following projections:

$$\begin{aligned} \rho_1 &= \{\alpha_1 \mapsto \tau_{11}, \alpha_2 \mapsto \tau_{21}, \dots\} \\ \rho_2 &= \{\alpha_1 \mapsto \tau_{12}, \alpha_2 \mapsto \tau_{22}, \dots\} \\ \rho_R &= \{\alpha_1 \mapsto R_1, \alpha_2 \mapsto R_2, \dots\} \end{aligned}$$

Notice that  $\rho_1$  and  $\rho_2$  now are type substitutions, so we write  $\rho_1(\tau)$  to mean  $\tau$  where all the type variables mentioned in the substitution have been substituted with the appropriate types. We can now write the value interpretation for type variables in a more succinct way:

$$\mathcal{V}[\alpha]_\rho = \rho_R(\alpha)$$

We need to add  $\rho$  to the other parts of the value interpretation as well. Moreover, as we now interpret open types, we require the pairs of values in the relation to be well typed under the type closed off using the relational substitution. So all value interpretations have the form

$$\mathcal{V}[\tau]_\rho = \{(v_1, v_2) \mid \bullet; \bullet \vdash v_1 : \rho_1(\tau) \wedge \bullet; \bullet \vdash v_2 : \rho_2(\tau) \wedge \dots\}$$

We further need to close of the type annotation of the variable in functions, so our value interpretations end up as:

$$\begin{aligned} \mathcal{V}[\mathit{bool}]_\rho &= \{(\mathbf{true}, \mathbf{true}), (\mathbf{false}, \mathbf{false})\} \\ \mathcal{V}[\tau \rightarrow \tau']_\rho &= \left\{ (\lambda x : \rho_1(\tau). e_1, \lambda x : \rho_2(\tau). e_2) \mid \right. \\ &\quad \left. \forall (v_1, v_2) \in \mathcal{V}[\tau]_\rho. (e_1[v_1/x], e_2[v_2/x]) \in \mathcal{E}[\tau']_\rho \right\} \end{aligned}$$

With the above challenge resolved, the remainder of the definitions are fairly straightforward. The expression interpretation is defined as follows:

$$\mathcal{E}[\tau]_\rho = \left\{ (e_1, e_2) \mid \begin{array}{l} \bullet; \bullet \vdash e_1 : \rho_1(\tau) \wedge \bullet; \bullet \vdash e_2 : \rho_2(\tau) \wedge \\ \exists v_1, v_2. e_1 \rightarrow^* v_1 \wedge e_2 \rightarrow^* v_2 \wedge (v_1, v_2) \in \mathcal{V}[\tau]_\rho \end{array} \right\}$$

In System F, the environment is extended with a type context  $\Gamma$ , so we need an interpretation for it.

$$\begin{aligned} \mathcal{D}[\bullet] &= \{\emptyset\} \\ \mathcal{D}[\Delta, \alpha] &= \{\rho[\alpha \mapsto (\tau_1, \tau_2, R)] \mid \rho \in \mathcal{D}[\Delta] \wedge R \in \text{Rel}[\tau_1, \tau_2]\} \end{aligned}$$

The interpretation of the type context is simply a relational substitution for the variables in the context. We also need an interpretation for the typing context  $\Gamma$  just like we had one in the previous sections.

$$\begin{aligned} \mathcal{G}[\bullet]_\rho &= \{\emptyset\} \\ \mathcal{G}[\Gamma, x : \tau]_\rho &= \{\gamma[x \mapsto (v_1, v_2)] \mid \gamma \in \mathcal{G}[\Gamma]_\rho \wedge (v_1, v_2) \in \mathcal{V}[\tau]_\rho\} \end{aligned}$$

We need the relational substitution in the interpretation of  $\Gamma$ , because  $\tau$  might contain free type variables now. We introduce a convenient notation for the projections of  $\gamma$  similar to the one we did for  $\rho$ :

$$\gamma = \{x_1 \mapsto (v_{11}, v_{12}), x_2 \mapsto (v_{21}, v_{22}), \dots\}$$

Define the projections as follows:

$$\begin{aligned} \gamma_1 &= \{x_1 \mapsto v_{11}, x_2 \mapsto v_{21}, \dots\} \\ \gamma_2 &= \{x_1 \mapsto v_{12}, x_2 \mapsto v_{22}, \dots\} \end{aligned}$$

We are now ready to define when two terms are logically related. We define it in a similar way to the logical predicate we already have defined. First we pick  $\rho$  and  $\gamma$

to close off the expressions, then we require the closed off expressions to be related under the expression interpretation of the type in question<sup>11</sup>:

$$\Delta; \Gamma \vdash e_1 \approx e_2 : \tau \stackrel{\text{def}}{=} \left\{ \begin{array}{l} \Delta; \Gamma \vdash e_1 : \tau \wedge \\ \Delta; \Gamma \vdash e_2 : \tau \wedge \\ \forall \rho \in \mathcal{D}[\Delta], \gamma \in \mathcal{G}[\Gamma]_\rho. \\ (\rho_1(\gamma_1(e_1)), \rho_2(\gamma_2(e_2))) \in \mathcal{E}[\tau]_\rho \end{array} \right.$$

With our logical relation defined, the first thing to prove is the fundamental property:

**Theorem 8** (Fundamental Property). *If  $\Delta; \Gamma \vdash e : \tau$  then  $\Delta; \Gamma \vdash e \approx e : \tau$*  ■

This theorem may seem a bit mundane, but it is actually quite strong. In the definition of the logical relation,  $\Delta$  and  $\Gamma$  can be seen as maps from place holders that need to be replaced in the expression. When we choose  $\rho$  and  $\gamma$ , we can pick different types and terms to put in the expression which can give us two very different programs.

In some presentations, this is also known as the parametricity lemma. It may even be stated without the short-hand notation for equivalence we use here.

We use a number of *compatibility lemmas* to prove the theorem. Each of the compatibility lemmas correspond to a typing rule and essentially prove one case of the induction proof for the theorem.

## 4.5 Compatibility Lemmas

We state a compatibility lemma for each of the typing rules. Each of the lemmas correspond to a case in the induction proof of the Fundamental Property (Theorem 8), so the theorem will follow directly from the compatibility lemmas. We state the compatibility lemmas as rules to illustrate the connection to the typing rules:

1.  $\overline{\Gamma; \Delta \vdash \text{true} \approx \text{true} : \text{bool}}$
2.  $\overline{\Gamma; \Delta \vdash \text{false} \approx \text{false} : \text{bool}}$
3.  $\overline{\Gamma; \Delta \vdash x \approx x : \Gamma(x)}$
4.  $\frac{\overline{\Delta; \Gamma \vdash e_1 \approx e_2 : \tau' \rightarrow \tau} \quad \overline{\Delta; \Gamma \vdash e'_1 \approx e'_2 : \tau'}}{\overline{\Delta; \Gamma \vdash e_1 e'_1 \approx e_2 e'_2 : \tau}}$

<sup>11</sup>From now on we drop the superscript for logically related terms and just write  $\approx$ .

- $$5. \frac{\Delta; \Gamma, x : \tau \vdash e_1 \approx e_2 : \tau'}{\Delta; \Gamma \vdash \lambda x : \tau. e_1 \approx \lambda x : \tau. e_2 : \tau \rightarrow \tau'}$$
- $$6. \frac{\Delta; \Gamma \vdash e_1 \approx e_2 : \forall \alpha. \tau \quad \Delta \vdash \tau'}{\Delta; \Gamma \vdash e_1[\tau'] \approx e_2[\tau'] : \tau[\tau'/\alpha]}$$

The rule for **if** has been omitted here. Further notice, some of the lemmas are more general than what we actually need. Take for instance the compatibility lemma for expression application. To prove the fundamental property, we really just needed to have the same expressions on both sides of the equivalence. However, the generalized version can also be used to prove that the logical relation is sound with respect to contextual equivalence.

We prove the compatibility lemma for type application here and leave the remaining proofs to the reader. To prove the compatibility lemma for type application, we are going to need the following lemma

**Lemma 11** (Compositionality). *If*

- $\Delta \vdash \tau'$
- $\Delta, \alpha \vdash \tau$
- $\rho \in \mathcal{D}[\Delta]$
- $R = \mathcal{V}[\tau']_\rho$

then

$$\mathcal{V}[\tau[\tau'/\alpha]]_\rho = \mathcal{V}[\tau]_{\rho[\alpha \mapsto (\rho_1(\tau'), \rho_2(\tau'), R)]} \quad \blacksquare$$

The lemma says: syntactically substituting some type for  $\alpha$  in  $\tau$  and then interpreting it is the same as semantically substituting the type for  $\alpha$ . To prove this lemma, we would need to show  $\mathcal{V}[\tau]_\rho \in \text{Rel}[\rho_1(\tau), \rho_2(\tau)]$  which is fairly easy given how we have defined our value interpretation.

*Proof.* (Compatibility, Lemma 6). What we want to show is

$$\frac{\Delta; \Gamma \vdash e_1 \approx e_2 : \forall \alpha. \tau \quad \Delta \vdash \tau'}{\Delta; \Gamma \vdash e_1[\tau'] \approx e_2[\tau'] : \tau[\tau'/\alpha]}$$

So we assume

- (1)  $\Delta; \Gamma \vdash e_1 \approx e_2 : \forall \alpha. \tau$
- (2)  $\Delta \vdash \tau'$

According to our definition of the logical relation, we need to show three things:

- i.  $\Delta; \Gamma \vdash e_1[\tau'] : \tau[\tau'/\alpha]$
- ii.  $\Delta; \Gamma \vdash e_2[\tau'] : \tau[\tau'/\alpha]$
- iii.  $\forall \rho \in \mathcal{D}[\Delta]. \forall \gamma \in \mathcal{G}[\Gamma]_\rho. (\rho_1(\gamma_1(e_1[\tau'])), \rho_2(\gamma_2(e_2[\tau']))) \in \mathcal{E}[\tau[\tau'/\alpha]]_\rho$

The two first follows from the well-typedness part of (1) together with (2) and the appropriate typing rule. So it only remains to show iii.

Suppose we have  $\rho \in \mathcal{D}[\Delta]$  and  $\gamma \in \mathcal{G}[\Gamma]_\rho$ . We then need to show:

$$(\rho_1(\gamma_1(e_1[\tau'])), \rho_2(\gamma_2(e_2[\tau']))) \in \mathcal{E}[\tau[\tau'/\alpha]]_\rho$$

By the definition of the  $\mathcal{E}$ -relation, we need to show that the two terms step to two related values. We keep this goal in mind and turn our attention to (1). By definition it gives us:

$$\forall \rho \in \mathcal{D}[\Delta]. \forall \gamma \in \mathcal{G}[\Gamma]_\rho. (\rho_1(\gamma_1(e_1)), \rho_2(\gamma_2(e_2))) \in \mathcal{E}[\forall \alpha. \tau]_\rho$$

If we instantiate this with our  $\rho$  and  $\gamma$ , then we get

$$(\rho_1(\gamma_1(e_1)), \rho_2(\gamma_2(e_2))) \in \mathcal{E}[\forall \alpha. \tau]_\rho$$

which means that  $e_1$  and  $e_2$  evaluate to some value  $v_1$  and  $v_2$  where  $(v_1, v_2) \in \mathcal{V}[\forall \alpha. \tau]_\rho$ . This means that  $v_1$  and  $v_2$  must be of type of type  $\forall \alpha. \tau$ . From this, we know that  $v_1$  and  $v_2$  are type abstractions, so there must exist  $e'_1$  and  $e'_2$  such that  $v_1 = \Lambda \alpha. e'_1$  and  $v_2 = \Lambda \alpha. e'_2$ . We can now instantiate  $(v_1, v_2) \in \mathcal{V}[\forall \alpha. \tau]_\rho$  with two types and a relation. We choose  $\rho_1(\tau')$  and  $\rho_2(\tau')$  as the two types for the instantiation and  $\mathcal{V}[\tau']_\rho$  as the relation<sup>12</sup>. This gives us

$$(e'_1[\rho_1(\tau')/\alpha], e'_2[\rho_2(\tau')/\alpha]) \in \mathcal{E}[\tau]_{\rho[\alpha \mapsto (\rho_1(\tau'), \rho_2(\tau'), \mathcal{V}[\tau']_\rho)]}$$

For convenience, we write  $\rho' = \rho[\alpha \mapsto (\rho_1(\tau'), \rho_2(\tau'), \mathcal{V}[\tau']_\rho)]$ . By definition of the  $\mathcal{E}$ -relation, we know that  $e'_1[\rho_1(\tau')/\alpha]$  and  $e'_2[\rho_2(\tau')/\alpha]$  evaluate to some values  $v_{1_f}$  and  $v_{2_f}$ , respectively, where  $(v_{1_f}, v_{2_f}) \in \mathcal{V}[\tau]_{\rho'}$ .

Let us take a step back and see what we have done. We have argued that the following evaluation takes place

$$\begin{aligned} \rho_i(\gamma_i(e_i))[\rho_i(\tau')] &\rightarrow^* (\Lambda \alpha. e'_i)[\rho_i(\tau')] \\ &\rightarrow e'_i[\rho_i(\tau')/\alpha] \\ &\rightarrow^* v_{i_f} \end{aligned}$$

<sup>12</sup>Here we use  $\mathcal{V}[\tau]_\rho \in \text{Rel}[\rho_1(\tau), \rho_2(\tau)]$  to justify using the value interpretation as our relation.

for  $i = 1, 2$ . The single step in the middle is justified by the type application reduction. The remaining steps are justified in our proof above. If we further note that  $\rho_i(\gamma_i(e_i[\tau'])) \equiv \rho_i(\gamma_i(e_i))[\rho_i(\tau')]$ , then we have shown that the two expressions from our goal in fact do evaluate to two values, and they are related. More precisely we have:

$$(v_{1_f}, v_{2_f}) \in \mathcal{V}[\tau]_{\rho'}$$

but that is not exactly what we wanted them to be related under. We are, however, in luck and can apply the compositionality lemma (Lemma 11) to obtain

$$(v_{1_f}, v_{2_f}) \in \mathcal{V}[\tau[\tau'/\alpha]]_{\rho}$$

which means that they are related under the relation we needed.  $\square$

We call theorems that follow as a consequence of parametricity for free theorems. The theorems from Section 4.2 are examples of free theorems. Next we will prove Theorem 5: All expressions of type  $\forall\alpha. \alpha \rightarrow \alpha$  must be the identity function. The theorem even says that any function of this type will terminate. This is, however, trivial as System-F is a terminating language<sup>13</sup>. In a non-terminating language such as System F with recursive types, we would state a weaker theorem where the expression would only have to be the identity function when it terminates. That is, divergence would also be acceptable behaviour.

*Proof of Theorem 5.* From the fundamental property and the well-typedness of  $e$ , we know  $\bullet \vdash e \approx e : \forall\alpha. \alpha \rightarrow \alpha$ . By definition this gives us

$$\forall \rho \in \mathcal{D}[\Delta]. \forall \gamma \in \mathcal{G}[\Gamma]_{\rho}. (\rho_1(\gamma_1(e)), \rho_2(\gamma_2(e))) \in \mathcal{E}[\forall\alpha. \alpha \rightarrow \alpha]_{\rho}$$

We instantiate this with an empty  $\rho$  and an empty  $\gamma$  to get  $(e, e) \in \mathcal{E}[\forall\alpha. \alpha \rightarrow \alpha]_{\emptyset}$ . By the definition of  $\mathcal{E}$ , we know that  $e$  evaluates to some value  $F$  and  $(F, F) \in \mathcal{V}[\forall\alpha. \alpha \rightarrow \alpha]_{\emptyset}$ . As  $F$  is a value of type  $\forall\alpha. \alpha \rightarrow \alpha$ , we know  $F = \Lambda\alpha. e_1$  for some  $e_1$ . Now instantiate  $(F, F) \in \mathcal{V}[\forall\alpha. \alpha \rightarrow \alpha]_{\emptyset}$  with the type  $\tau$  twice and the relation  $R = \{(v, v)\}$  (Note:  $R \in \text{Rel}[\tau, \tau]$ ). This gives us  $(e_1[\tau/\alpha], e_1[\tau/\alpha]) \in \mathcal{E}[\alpha \rightarrow \alpha]_{[\alpha \rightarrow (\tau, \tau, R)]}$ .

Let us take a quick intermission from the proof. To a large extent, the proof of a free theorem is just unfolding of definitions. However in order for the definitions to line up, it is important that the relation  $R$  is picked correctly which makes this an important non-trivial part of the proof. Before we chose the relation in this proof, we picked two types based on the theorem we want to show. In the theorem, we instantiate  $e$  with  $\tau$ , so we picked  $\tau$  for the proof. Likewise with

<sup>13</sup>For more on this see *Types and Programming Languages* by Benjamin Pierce[Pierce, 2002].

$R$ : in the theorem  $v$  is the argument for the function with the domain  $\alpha$ , so we picked the singleton relation  $\{(v, v)\}$ . Picking the correct relation requires some work, but the statement can guide the decision. Finally, if you are not sure what to pick for  $R$  try something and see if it works out. If it does not work, then you can simply pick a new relation and you may even have learned something to guide your next pick. Intermission over.

From  $(e_1[\mathcal{T}/\alpha], e_1[\mathcal{T}/\alpha]) \in \mathcal{E}[\alpha \rightarrow \alpha]_{[\alpha \mapsto (\tau, \tau, R)]}$ , we know that  $e_1[\mathcal{T}/\alpha]$  evaluates to some value  $g$  and  $(g, g) \in \mathcal{V}[\alpha \rightarrow \alpha]_{[\alpha \mapsto (\tau, \tau, R)]}$ . From the type of  $g$ , we know that it must be a  $\lambda$ -abstraction, so  $g = \lambda x : \tau. e_2$  for some expression  $e_2$ . Now instantiate  $(g, g) \in \mathcal{V}[\alpha \rightarrow \alpha]_{[\alpha \mapsto (\tau, \tau, R)]}$  with  $(v, v) \in \mathcal{V}[\alpha]_{[\alpha \mapsto (\tau, \tau, R)]}$  to get  $(e_2[v/x], e_2[v/x]) \in \mathcal{E}[\alpha]_{[\alpha \mapsto (\tau, \tau, R)]}$ . From this we know that  $e_2[v/x]$  steps to some value  $v_f$  and  $(v_f, v_f) \in \mathcal{V}[\alpha]_{[\alpha \mapsto (\tau, \tau, R)]}$ . We have that  $\mathcal{V}[\alpha]_{[\alpha \mapsto (\tau, \tau, R)]} \equiv R$  so  $(v_f, v_f) \in R$  which mean that  $v_f = v$  as  $(v, v)$  is the only pair in  $R$ .

Now let us take a step back and consider what we have shown above.

$$\begin{aligned}
e[\tau] v &\rightarrow^* F[\tau] v \\
&\equiv (\Lambda \alpha. e_1)[\tau] v \\
&\rightarrow (e_1[\mathcal{T}/\alpha]) v \\
&\rightarrow^* g v \\
&\equiv (\lambda x : \tau. e_2) v \\
&\rightarrow e_2[v/x] \\
&\rightarrow^* v_f \\
&\equiv v
\end{aligned}$$

First we argued that  $e[\tau]$  steps to some  $F$  and that  $F$  was a type abstraction,  $\Lambda \alpha. e_1$ . Then we performed the type application to get  $e_1[\mathcal{T}/\alpha]$ . We then argued that this steps to some  $g$  of the form  $\lambda x : \tau. e_2$  which further allowed us to do a  $\beta$ -reduction to obtain  $e_2[v/x]$ . We then argued that this reduced to  $v_f$  which was the same as  $v$ . In summation we argued  $e[\tau] v \rightarrow^* v$  which is the result we wanted.  $\square$

## 4.6 Exercises

1. Prove the following free theorem:

**Theorem 9.** *If  $\bullet; \bullet \vdash e : \forall \alpha. ((\tau \rightarrow \alpha) \rightarrow \alpha)$  and  $\bullet; \bullet \vdash k : \tau \rightarrow \tau_k$ , then*

$$\bullet; \bullet \vdash e[\tau_k] k \approx k(e[\tau] \lambda x : \tau. x) : \tau_k \quad \blacksquare$$

This theorem is a simplified version of one found in *Theorems For Free* by Philip Wadler[Wadler, 1989].

## 5 Existential types

In this section, we extend STLC with existential types and show how to make a logical relation for that language.

An existential type is reminiscent of a Java interface. It describes some functionality but leaves out the implementation. An implementation of an existential type must provide the functions and constants specified in the type. This means that there can be many different implementations of the same existential type. When an existential type is used, the user interacts with the exposed constants and functions without knowing the actual implementation.

Take for example a stack. We would expect a stack to have the following functions:

`mk` creates a new stack.

`push` puts an element on the top of the stack. It takes a stack and an element and returns the resulting stack.

`pop` removes the top element of the stack. It takes a stack and returns the new stack along with the element that was popped from it.

An interface would define the above signature which you then would go off and implement<sup>14</sup>. If we wanted to write an interface for a stack, it would look like this (this is meant to be suggestive, so it is in a non-formal notation):

$$\begin{aligned} \text{stack} = \exists \alpha. \langle & \text{mk} : 1 \rightarrow \alpha, \\ & \text{push} : \alpha \times \text{int} \rightarrow \alpha, \\ & \text{pop} : \alpha \rightarrow \alpha \times \text{int} \rangle \end{aligned}$$

where  $\alpha$  is the type of the stack used in the actual implementation. This means that `mk` is a function that takes unit and returns a stack. The `push` function takes a stack and an element and returns a stack. Finally, the `pop` function takes a stack and returns a stack and an element. The *stack* type all the  $\alpha$ 's which means that a client cannot see the actual type of the stack, so they cannot know how it is actually implemented.

We formally write existential types in a similar fashion to how we wrote universal types:

$$\exists \alpha. \tau$$

---

<sup>14</sup>There is a famous paper called *Abstract Data Types Have Existential Type* from '88 by Mitchell and Plotkin [Mitchell and Plotkin, 1988]. The title says it all.

Here  $\tau$  is the same as the record in the stack example. The interface is just a type, so now we need to define how one implements something of an existential type. If we were to implement the stack interface, then we would implement a package of functions that are supposed to be used together. This could look something like (again this is meant to be suggestive):

$$\begin{aligned} &\text{pack } \text{array}[\text{int}], \\ &\quad \langle \lambda x : \_ . \dots , \\ &\quad \lambda x : \_ . \dots , \\ &\quad \lambda x : \_ . \dots \rangle \end{aligned}$$

Here  $\text{array}[\text{int}]$  is the type we want to use for the concrete implementation and the record of functions is the concrete implementation that uses  $\text{array}[\text{int}]$  to implement a stack.

Existential types hide the implementation from the client, but that does not mean that they are the same. For instance, a *stack* implementation could use  $\text{int}$  for  $\alpha$ . In this case `mk` could return 0, `push` would throw away the  $\text{int}$  argument, and `pop` would just return a pair of its  $\alpha$  argument, which is an  $\text{int}$  argument here. This implementation clearly doesn't correspond to what we consider a stack, but it has the correct type. We would expect that all implementations of the interface that correspond to a stack would observationally behave the same. That is, we would expect them to be contextually equivalent. We want to use a logical relation to be able to establish contextual equivalence.

## 5.1 STLC with Existential Types

The syntactic additions to the language are `pack` and `unpack`: We briefly introduce the additions to the syntax and semantics:

$$\begin{aligned} \tau &::= \dots \mid \exists \alpha. \tau \\ e &::= \dots \mid \text{pack } \langle \tau, e \rangle \text{ as } \exists \alpha. \tau \mid \text{unpack } \langle \alpha, x \rangle = e \text{ in } e \\ v &::= \dots \mid \text{pack } \langle \tau, v \rangle \text{ as } \exists \alpha. \tau \\ E &::= \dots \mid \text{pack } \langle \tau, E \rangle \text{ as } \exists \alpha. \tau \mid \text{unpack } \langle \alpha, x \rangle = E \text{ in } e \end{aligned}$$

New packages are constructed with `pack` which makes it the introductory form for the existential type. A package consists of a *witness type* and an implementation of the existential type. Packages are opened by `unpack` which allows a client to use the components in the package. This makes `unpack` the elimination form. The

interaction between the two new constructs is expressed in the following evaluation rule:

$$\overline{\text{unpack } \langle \alpha, x \rangle = \text{pack } \langle \tau', v \rangle \text{ as } \exists \alpha. \tau \text{ in } e \rightarrow e[\tau' / \alpha][v / x]}$$

The form of the typing judgement remains the same but we add the following two rules:

$$\frac{\Delta; \Gamma \vdash e : \tau[\tau' / \alpha] \quad \Delta \vdash \tau'}{\Delta; \Gamma \vdash \text{pack } \langle \tau', e \rangle \text{ as } \exists \alpha. \tau : \exists \alpha. \tau}$$

$$\frac{\Delta; \Gamma \vdash e_1 : \exists \alpha. \tau \quad \Delta, \alpha; \Gamma, x : \tau \vdash e_2 : \tau_2 \quad \Delta \vdash \tau_2}{\Delta; \Gamma \vdash \text{unpack } \langle \alpha, x \rangle = e_1 \text{ in } e_2 : \tau_2}$$

Intuitively, the typing rule of `pack` says that provided an implementation of the existential type that implementation has to be well-typed when the witness type is plugged in for  $\alpha$ . It tells what type we substitute into. In the typing rule for `unpack`, it is important that  $\alpha$  is not free in  $\tau_2$  which is ensured by  $\Delta \vdash \tau_2$ . This makes sure that the package actually hides the witness type. The witness type can be pulled out of the package within a certain scope using `unpack`. If  $\alpha$  could be returned from that scope, then it would be exposed to the outer world which would defeat the purpose of hiding it. The `unpack`-expression takes out the components of  $e_1$  and calls them  $\alpha$  and  $x$ . The two components can then be used in the body,  $e_2$ , of the `unpack`-expression, but they are used with the aliases given to them by `unpack`.

## 5.2 Example

For the remainder of this section, we will use the following example: Take this existential type

$$\tau = \exists \alpha. \alpha \times (\alpha \rightarrow \text{bool})$$

and the following two expressions that we for now claim have type  $\tau$

$$e_1 = \text{pack } \langle \text{int}, \langle 1, \lambda x : \text{int}. x = 0 \rangle \rangle \text{ as } \tau$$

$$e_2 = \text{pack } \langle \text{bool}, \langle \text{true}, \lambda x : \text{bool}. \text{not } x \rangle \rangle \text{ as } \tau$$

where *int* and *bool* the *witness types*, respectively. We further claim that these two implementations of  $\tau$  are equivalent, and our goal in this note is to show this.

We start by verifying that the two expressions are indeed well-typed and has type  $\tau$  by constructing their type derivation trees:

$$\begin{array}{c}
\frac{\frac{\frac{}{\bullet; \bullet \vdash 1 : int}}{\bullet; \bullet \vdash \langle 1, \lambda x : int. x = 0 \rangle : int \times (int \rightarrow bool)} \quad \frac{\frac{\frac{}{\bullet; x : int \vdash x : int} \quad \frac{}{\bullet; x : int \vdash 0 : int}}{\bullet; x : int \vdash x = 0 : bool}}{\bullet; \bullet \vdash \lambda x : int. x = 0 : int \rightarrow bool}}{\bullet; \bullet \vdash \langle 1, \lambda x : int. x = 0 \rangle : int \times (int \rightarrow bool)} \quad \frac{}{\bullet \vdash int}}{\bullet; \bullet \vdash \text{pack } \langle int, \langle 1, \lambda x : int. x = 0 \rangle \rangle \text{ as } \tau : \tau} \\
\\
\frac{\frac{\frac{}{\bullet; \bullet \vdash \text{true} : bool}}{\bullet; \bullet \vdash \langle \text{true}, \lambda x : bool. \text{not } x \rangle : bool \times (bool \rightarrow bool)} \quad \frac{\frac{\frac{}{\bullet; x : bool \vdash x : bool}}{\bullet; x : bool \vdash \text{not } x : bool}}{\bullet; \bullet \vdash \lambda x : bool. \text{not } x : bool \rightarrow bool}}{\bullet; \bullet \vdash \langle \text{true}, \lambda x : bool. \text{not } x \rangle : bool \times (bool \rightarrow bool)} \quad \frac{}{\bullet \vdash bool}}{\bullet; \bullet \vdash \text{pack } \langle bool, \langle \text{true}, \lambda x : bool. \text{not } x \rangle \rangle \text{ as } \tau : \tau}
\end{array}$$

To use a package, a client must open it with `unpack` first. For instance, one could attempt the following `unpack`:

`unpack <α, p> = e1 in (snd p) 5`

This use of `unpack` does, however, expose the inner workings of  $e_1$ . In particular, it exposes that  $e_1$  uses integers as an internal representation. The existential type was supposed to hide this information making the implementation opaque, but that is clearly not the case here. The `unpack`-expression is even syntactically well-formed, but it is not well-typed. The witness type  $int$  in  $e_1$  is not exposed by `unpack` in the typing rule. This means that the body of an `unpack` should use the witness type in an abstract way, i.e. it should use  $\alpha$ . When we type the `unpack`-expression, `(snd p)` takes an argument of type  $\alpha$ ; but in the above expression, we give it something of type  $int$  (if you don't see the issue, try to construct the type derivation tree). This means that we only can give provide `(fst p)` as an argument, i.e.

`unpack <α, p> = e1 in`  
~~`(snd p) 5`~~  
`(snd p) (fst p)`

It should come as no surprise that  $e_2$  only can be used in the exact same way:

`unpack <α, p> = e2 in`  
`(snd p) (fst p)`

It is also not type check if we return (`fst p`) (or (`snd p`) for that matter) as the body of an `unpack` cannot have the  $\alpha$  free in it.

We can now informally argue why  $e_1$  and  $e_2$  are observationally equivalent. In both expressions, there is only one value of type  $\alpha$ . In  $e_1$  it is `1`, and in  $e_2$  it is `true`. This means that they are going to be the only related values, i.e. we will define  $R = \{(1, \text{true})\}$ . As we have informally argued, they are the only values the two functions can take as arguments. This means that we can easily find the values that they can possibly return: In  $e_1$  it is  $(\lambda x : x = 0. ) 1 \rightarrow \text{false}$ , and in  $e_2$  it is  $(\lambda x : \text{not } x. ) \text{true} \rightarrow \text{false}$ . The only value that is ever exposed from the package is `false`. If this claim is true, then it is impossible for a client to distinguish the two packages from each other.

### 5.3 Logical Relation

To formally argue that  $e_1$  and  $e_2$  from the previous section are contextually equivalent, we need a logical relation. To this end, we extend our previous logical relation, so it also interprets  $\exists\alpha.\tau$ . The values we relate are of the form (`pack`  $\langle\tau_1, v_1\rangle$  `as`  $\exists\alpha.\tau$ , `pack`  $\langle\tau_2, v_2\rangle$  `as`  $\exists\alpha.\tau$ ) and as always our first instinct should be to look at the elimination form, so we want to consider `unpack`  $\langle\alpha, x\rangle = \text{pack } \langle\tau_i, v_i\rangle \text{ as } \exists\alpha.\tau$  in  $e_i$  for  $i = 1, 2$ . Now it would be tempting to relate the two bodies, but we get a similar issue to what we had for universal types. If we relate the two bodies, then what type should we relate them under? The type we get might be larger than the one we are interpreting which gives us a well-foundedness problem. So by analogy we do not require that the two `unpack` expressions have related bodies. Instead we relate  $v_1$  and  $v_2$  under some relation:

$$\mathcal{V}[\exists\alpha.\tau]_\rho = \left\{ \begin{array}{l} (\text{pack } \langle\rho_1(\tau_1), v_1\rangle \text{ as } \rho_1(\exists\alpha.\tau), \text{pack } \langle\rho_2(\tau_2), v_2\rangle \text{ as } \rho_2(\exists\alpha.\tau)) \mid \\ \exists R \in \text{Rel}[\rho_1(\tau_1), \rho_2(\tau_2)]. (v_1, v_2) \in \mathcal{V}[\tau]_{\rho[\alpha \mapsto (\rho_1(\tau_1), \rho_2(\tau_2), R)]} \end{array} \right\}$$

The relation turns out to be somewhat dual to the one for universal types. Instead of saying  $\forall\tau_1, \tau_2, R$ , we say  $\exists\tau_1, \tau_2, R$ , but as we get  $\tau_1$  and  $\tau_2$  directly from the values, we omit them in the definition. We also relate the two values at  $\tau$  and extend the relational substitution with the types we have for  $\alpha$ . Notice that we use  $\rho$  to close of the type variables in the two values we related.

With this extension to the value interpretation, we are ready to show that  $e_1$  and  $e_2$  are logically related. We reuse the definition of logical equivalence we defined previously. What we wish to show is:

**Theorem 10.**

$$\bullet; \bullet \vdash e_1 \approx e_2 : \exists\alpha. \alpha \times (\alpha \rightarrow \text{bool}) \quad \blacksquare$$

*Proof.* With an empty environment, this amounts to show  $(e_1, e_2) \in \mathcal{E}[\exists\alpha. \alpha \times (\alpha \rightarrow \text{bool})]_{\emptyset}$ . To show this, we need to establish that  $e_1$  and  $e_2$  evaluate to some value and that these two values are related under the same type and relational substitution. The expressions  $e_1$  and  $e_2$  are **pack**-expressions, so they are already values. In other words, it suffices to show  $(e_1, e_2) \in \mathcal{V}[\exists\alpha. \alpha \times (\alpha \rightarrow \text{bool})]_{\emptyset}$ . We now need to pick a relation and show that the implementations are related under  $\alpha \times (\alpha \rightarrow \text{bool})$  that is

$$(\langle 1, \lambda x : \text{int}. x = 0 \rangle, \langle \text{true}, \lambda x : \text{bool}. \text{not } x \rangle) \in \mathcal{V}[\alpha \times (\alpha \rightarrow \text{bool})]_{[\alpha \rightarrow (\text{int}, \text{bool}, R)]}$$

We pick  $R = \{(1, \text{true})\}$  as the relation. Two tuples are related if their components are related<sup>15</sup>. In other words, we need to show the following:

1.  $(1, \text{true}) \in \mathcal{V}[\alpha]_{\emptyset[\mapsto \alpha](\text{int}, \text{bool}, R)}$
2.  $(\lambda x : \text{int}. x = 0, \lambda x : \text{bool}. \text{not } x) \in \mathcal{V}[\alpha \rightarrow \text{bool}]_{[\alpha \rightarrow (\text{int}, \text{bool}, R)]}$

Item 1 amounts to showing  $(1, \text{true}) \in R$  which is true. In order to show 2, assume  $(v_1, v_2) \in \mathcal{V}[\alpha]_{[\alpha \rightarrow (\text{int}, \text{bool}, R)]}$  which is the same as  $(v_1, v_2) \in R$ . Due to our choice of  $R$ , we have  $v_1 = 1$  and  $v_2 = \text{true}$ . Now we need to show  $(v_1 = 0, \text{not } v_2) \in \mathcal{E}[\text{bool}]_{[\alpha \rightarrow (\text{int}, \text{bool}, R)]}$ . Which means that we need to show that the two expressions evaluate to two values related under  $\text{bool}$ .  $v_1 = 0$  evaluates to **false** as  $v_1$  is 1 and  $\text{not } v_2$  evaluates to **false** as well as  $v_2 = \text{true}$ . It remains to show  $(\text{false}, \text{false}) \in \mathcal{V}[\text{bool}]_{[\alpha \rightarrow (\text{int}, \text{bool}, R)]}$ . For base types, the value interpretation relates equal values, so this is indeed true.  $\square$

## 6 Recursive Types and Step Indexing

Consider the following program from the *untyped* lambda calculus:

$$\Omega = (\lambda x. x x) (\lambda x. x x)$$

The first bit of the evaluation of  $\Omega$  is a  $\beta$ -reduction followed by a substitution at which point we end up with  $\Omega$  again. In other words, the evaluation of  $\Omega$  diverges. Moreover, it is not possible to assign a type to  $\Omega$  (again the interested reader may try to verify this by attempting to assign a type). It can hardly come as a surprise that it cannot be assigned a type, as we previously proved that the simply typed lambda calculus is strongly normalizing, so if we could assign  $\Omega$  a type, then it would not diverge.

<sup>15</sup>We defined this for logical predicates, but not for logical relations.

To type  $\Omega$ , we need recursive types. If we are able to type  $\Omega$ , then we do not have strong normalization (as  $\Omega$  is not strongly normalizing). With recursive types, we can type structures that are inherently inductive such as lists, trees, and streams. In an ML-like language, a declaration of a tree type would look like this:

```
type tree = Leaf
         | Node of int * tree * tree
```

In Java, we could define a tree class with an int field and two tree fields for the sub trees:

```
class Tree {
    int value;
    Tree left, right;
}
```

In other words, we can define trees in common programming languages, but we cannot define them in the lambda calculus. Let us try to find a reasonable definition for recursive types by considering what properties are necessary to define trees. We want a type that can either be a node or a leaf. A leaf can be represented by unit (in our trees leaves carry no information), and a node is the product of an int and trees. We put the two constructs together with the sum type:

$$tree = 1 + (int * tree * tree)$$

This is what we want, but the definition is not well-founded as it is. We are defining the type *tree*, but *tree* appears on the right hand side of the definition which makes it circular. Instead of writing *tree*, we use a type variable  $\alpha$ :

$$\begin{aligned} \alpha &= 1 + (int \times \alpha \times \alpha) \\ &= 1 + (int \times (int \times \alpha \times \alpha) \times (int \times \alpha \times \alpha)) \\ &\vdots \end{aligned}$$

All sides of the above equations are equal, and they are all trees. We could keep going and get an infinite system of equations. If we keep substituting the definition of  $\alpha$  for  $\alpha$ , we keep getting bigger and bigger types. All of the types are trees, and all of them are finite. If we take the limit of this process, then we end up with an infinite tree, and that tree is the tree we conceptually have in our minds. So what we need is the fixed point of the above equation.

Let us define a recursive function for which we want to find a fixed point:

$$F = \lambda\alpha : type. 1 + (int \times \alpha \times \alpha)$$

We want the fixed point which by definition is  $t$  such that

$$t = F(t)$$

So we want

$$tree = F(tree)$$

The fixed point of this function is written:

$$\mu\alpha. F(\alpha)$$

Here  $\mu$  is a fixed-point type constructor. As the above is the fixed point by definition, it should be equal to  $F$  applied to it:

$$\mu\alpha. F(\alpha) = F(\mu\alpha. F(\alpha))$$

Now let us make this look a bit more like types by substituting  $F(\alpha)$  for  $\tau$ .

$$\mu\alpha. \tau = F(\mu\alpha. \tau)$$

The right hand side is really just  $\tau$  with  $\mu\alpha. \tau$  substituted with  $\tau$ :

$$\mu\alpha. \tau = \tau[\mu\alpha. \tau/\alpha]$$

We introduce the recursive type  $\mu\alpha. \tau$  to our language. With the recursive type, we can shift our view to an expanded version  $\tau[\mu\alpha. \tau/\alpha]$  and contract back to the original type. Expanding the type is called **unfold** and contracting is called **fold**.

$$\begin{array}{ccc} & \text{unfold} & \\ \mu\alpha. \tau & \xrightarrow{\quad} & \tau[\mu\alpha. \tau/\alpha] \\ & \xleftarrow{\quad} & \\ & \text{fold} & \end{array}$$

With recursive types in hand, we can now define our tree type:

$$tree \stackrel{\text{def}}{=} \mu\alpha. 1 + (int \times \alpha \times \alpha)$$

When we work with a recursive type, we need to be able to open the type and use whatever is under the  $\mu$ . Say we have an expression  $e$  of type  $tree$ . We need to be able to tell whether it is a leaf or a node. To do so, we need to peek under the  $\mu$  which is done with an **unfold**. The **unfold** gives us an expression where the type has been unfolded by one level. That is, the same expression but where  $\alpha$  has been substituted with the definition of  $tree$  and the outer  $\mu\alpha.$  has been removed. With the outer  $\mu\alpha.$  gone, we can match on the sum type to find out whether it is a leaf or a node. When we are done working with the type, we can fold it to get the original tree type.

$$\begin{aligned}
& tree = \mu\alpha. 1 + (int \times \alpha \times \alpha) \\
& \quad \text{fold} \begin{array}{c} \uparrow \\ \downarrow \end{array} \text{unfold} \\
& 1 + (int \times (\mu\alpha. 1 + (int \times \alpha \times \alpha)) \times (\mu\alpha. 1 + (int \times \alpha \times \alpha)))
\end{aligned}$$

This kind of recursive types is called iso-recursive types because there is an isomorphism between a  $\mu\alpha. \tau$  and its unfolding  $\tau[\mu\alpha. \tau/\alpha]$ .

## 6.1 Simply Typed Lambda Calculus with Recursive Types

The syntax of STLC with recursive types is defined as follows:

$$\begin{aligned}
\tau & ::= \dots \mid \mu\alpha. \tau \\
e & ::= \dots \mid \text{fold } e \mid \text{unfold } e \\
v & ::= \dots \mid \text{fold } v \\
E & ::= \dots \mid \text{fold } E \mid \text{unfold } E
\end{aligned}$$

Further, the following evaluation rule is added:

$$\overline{\text{unfold } (\text{fold } v)} \rightarrow v$$

Finally, the following typing judgements are added:

$$\frac{\Gamma \vdash e : \tau[\mu\alpha. \tau/\alpha]}{\Gamma \vdash \text{fold } e : \mu\alpha. \tau} \text{T-FOLD} \qquad \frac{\Gamma \vdash e : \mu\alpha. \tau}{\Gamma \vdash \text{unfold } e : \tau[\mu\alpha. \tau/\alpha]} \text{T-UNFOLD}$$

This extension allows us to define a type for lists of integers:

$$int \text{ list} \stackrel{def}{=} \mu\alpha. 1 + (int \times \alpha)$$

In the lambda calculus with recursive types, the  $\Omega$  function has type  $(\mu\alpha. \alpha \rightarrow \alpha) \rightarrow (\mu\alpha. \alpha \rightarrow \alpha)$ . We do, however, need to rewrite it slightly as applies a function found under a  $\mu$ , so it must first unfold it:

$$\Omega = (\lambda x : \mu\alpha. \alpha \rightarrow \alpha. (\text{unfold } x) x)$$

It is indeed possible to type this version of  $\Omega$ :

$$\frac{\frac{\frac{x : \mu\alpha. \alpha \rightarrow \alpha \vdash x : \mu\alpha. \alpha \rightarrow \alpha}{x : \mu\alpha. \alpha \rightarrow \alpha \vdash \text{unfold } x : \tau_2 \rightarrow (\mu\alpha. \alpha \rightarrow \alpha)}}{\text{for } \tau_2 = \mu\alpha. \alpha \rightarrow \alpha} \quad x : \mu\alpha. \alpha \rightarrow \alpha \vdash x : \tau_2}{x : \mu\alpha. \alpha \rightarrow \alpha \vdash (\text{unfold } x) x : \mu\alpha. \alpha \rightarrow \alpha}}{\bullet \vdash \Omega : (\mu\alpha. \alpha \rightarrow \alpha) \rightarrow (\mu\alpha. \alpha \rightarrow \alpha)}$$

## 6.2 A Step-Indexed Logical Predicate

In a naive first attempt to construct the value interpretation, we could try:

$$\mathcal{V}[\mu\alpha. \tau] = \{\text{fold } v \mid \text{unfold } (\text{fold } v) \in \mathcal{E}[\tau[\mu\alpha. \tau/\alpha]]\}$$

We can simplify this slightly; first we use the fact that `unfold (fold v)` reduces to `v`. Next we use the fact that `v` must be a value and the fact that we want `v` to be in the expression interpretation of  $\tau[\mu\alpha. \tau/\alpha]$ . By unfolding the definition of the expression interpretation, we conclude that it suffices to require `v` to be in the value interpretation of the same type. With these simplifications we get the following definition:

$$\mathcal{V}[\mu\alpha. \tau] = \{\text{fold } v \mid v \in \mathcal{V}[\tau[\mu\alpha. \tau/\alpha]]\}$$

This is, however, not a well-founded definition. The value interpretation is defined inductively on the type, but  $\tau[\mu\alpha. \tau/\alpha]$  is not a structurally smaller type than  $\mu\alpha. \tau$ .

To solve this issue, we index the interpretation by a natural number,  $k$ , which we write as follows:

$$\mathcal{V}_k[\tau] = \{v \mid \dots\}$$

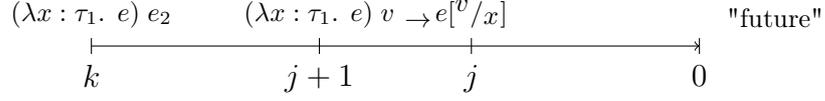
Hence  $v \in \mathcal{V}_k[\tau]$  is read as “`v` belongs to the interpretation of  $\tau$  for  $k$  steps.” We interpret this in the following way: given a value that we run for  $k$  or fewer steps (as in the value is used in some program context for fewer than  $k$  steps), then we will never notice that it does not have type  $\tau$ . If we use the same value in a program context that wants to run for more than  $k$  steps, then we might notice that it does not have type  $\tau$  which means that we might get stuck. This gives us an approximate guarantee.

We use the step as an inductive metric to make our definition well-founded. That is we define the interpretation inductively on the step-index followed by an inner induction on the type structure. Let us start by adding the step-index to our existing value interpretation:

$$\begin{aligned} \mathcal{V}_k[\text{bool}] &= \{\text{true}, \text{false}\} \\ \mathcal{V}_k[\tau_1 \rightarrow \tau_2] &= \{\lambda x : \tau_1. e \mid \forall j \leq k. \forall v \in \mathcal{V}_j[\tau_1]. e[v/x] \in \mathcal{E}_j[\tau_2]\} \end{aligned}$$

`true` and `false` are in the value interpretation of `bool` for any  $k$ , so `true` and `false` will for any  $k$  look like it has type `bool`. To illustrate how to understand the value interpretation of  $\tau_1 \rightarrow \tau_2$ , please consider the following time line:

$\lambda$  time-line



Here we start at index  $k$  and as we run the program, we use up steps until we at some point reach 0 and run out of steps. Say we at step  $k$ , we have an application of a lambda abstraction. The lambda abstraction is already ready to be applied, but the application may not happen right away. The  $\beta$ -reduction happens when the argument is a value, but the application may contain a non-value expression as its argument, i.e.  $(\lambda x : \tau_1. e) e_2$ . It takes a number of steps to reduce  $e_2$  to a value, and we have no way to tell how many it will take. All we can do is to assume that it has taken some number of steps to evaluate  $e_2$  to  $v$ , so we have  $j + 1$  steps left. At this time, we can perform the  $\beta$ -reduction which means that we have  $e[v/x]$  with  $j$  steps left.

We can now define the value interpretation of  $\mu\alpha.\tau$ :

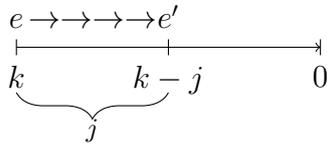
$$\mathcal{V}_k[\mu\alpha.\tau] = \{\text{fold } v \mid \forall j < k. v \in \mathcal{V}_j[\tau[\mu\alpha.\tau/\alpha]]\}$$

This definition is almost the same as the one we proposed above but step-indexed. In order to make the definition well-founded, we require  $j$  be *strictly* less than  $k$ . We do not define a value interpretation for type variables  $\alpha$ , as we have no polymorphism yet. The only place we have a type variable at the moment is in  $\mu\alpha.\tau$ , but in the interpretation we immediately close off the  $\tau$  under the  $\mu$ , so we will never encounter a free type variable.

Finally, we define the expression interpretation:

$$\mathcal{E}_k[\tau] = \{e \mid \forall j < k. \forall e'. e \rightarrow^j e' \wedge \text{irred}(e') \implies e' \in \mathcal{V}_{k-j}[\tau]\}$$

To illustrate what is going on here, consider the following time line:



We start with an expression  $e$ , then we take  $j$  steps and get to expression  $e'$ . At this point, if  $e'$  is irreducible, then it must be in the value interpretation of  $\tau$  for  $k - j$  steps. As explained above, the step-index approximates where we only can be sure that a value has a given type if we still have steps left. In other words, if

we allow the expression interpretation to spend all the steps, then we cannot say anything meaningful about the value.

We also need to lift the interpretation of type environments to step-indexing:

$$\begin{aligned}\mathcal{G}_k[\bullet] &= \{\emptyset\} \\ \mathcal{G}_k[\Gamma, x : \tau] &= \{\gamma[x \mapsto v] \mid \gamma \in \mathcal{G}_k[\Gamma] \wedge v \in \mathcal{V}_k[\tau]\}\end{aligned}$$

We are now in a position to lift the definition of semantic type safety to one with step-indexing.

$$\Gamma \models e : \tau \stackrel{\text{def}}{=} \forall k \geq 0. \forall \gamma \in \mathcal{G}_k[\Gamma]. \gamma(e) \in \mathcal{E}_k[\tau]$$

To actually prove type safety, we do it in two steps. First we state and prove the fundamental theorem:

**Theorem 11** (Fundamental property).

If  $\Gamma \vdash e : \tau$ , then  $\Gamma \models e : \tau$ . ■

When we have proven the fundamental property, we prove that it entails type safety, i.e.

$$\bullet \models e : \tau \implies \text{safe}(e)$$

Thanks to the way we defined the logical predicate, this second step should be trivial to prove.

The difficult part is to prove the fundamental property. This proof requires a lemma that says the value interpretation is monotone, i.e. if a value is in the value interpretation for some step, then it is also in there for any smaller step.

**Lemma 12** (Monotonicity).

If  $v \in \mathcal{V}_k[\tau]$  and  $j \leq k$ , then  $v \in \mathcal{V}_j[\tau]$ . ■

*Proof.* The proof is by case on  $\tau$ .

**Case**  $\tau = \text{bool}$ , assume  $v \in \mathcal{V}_k[\text{bool}]$  and  $j \leq k$ , we then need to show  $v \in \mathcal{V}_j[\text{bool}]$ . As  $v \in \mathcal{V}_k[\text{bool}]$ , we know that either  $v = \text{true}$  or  $v = \text{false}$ . If we assume  $v = \text{true}$ , then we immediately get what we want to show, as  $\text{true}$  is in  $\mathcal{V}_j[\text{bool}]$  for any  $j$ . Likewise for the case  $v = \text{false}$ .

**Case**  $\tau = \tau_1 \rightarrow \tau_2$ , assume  $v \in \mathcal{V}_k[\tau_1 \rightarrow \tau_2]$  and  $j \leq k$ , we then need to show  $v \in \mathcal{V}_j[\tau_1 \rightarrow \tau_2]$ . As  $v$  is a member of  $\mathcal{V}_k[\tau_1 \rightarrow \tau_2]$ , we can conclude that  $v = \lambda x : \tau_1. e$  for some  $e$ . By definition of  $v \in \mathcal{V}_j[\tau_1 \rightarrow \tau_2]$  we need to show

$$\forall i \leq j. \forall v' \in \mathcal{V}_i[\tau_1]. e[v'/x] \in \mathcal{E}_i[\tau_2]$$

Suppose  $i \leq j$  and  $v' \in \mathcal{V}_i[\tau_1]$ , we then need to show  $e[v'/x] \in \mathcal{E}_i[\tau_2]$ .

By assumption, we have  $v \in \mathcal{V}_k[\tau_1 \rightarrow \tau_2]$  which gives us

$$\forall n \leq k. \forall v' \in \mathcal{V}_n[\tau_1]. e[v'/x] \in \mathcal{E}_n[\tau_2]$$

By transitivity,  $j \leq k$ , and  $i \leq j$ , we get  $i \leq k$ . We use this with  $v' \in \mathcal{V}_i[\tau_1]$  to get  $e[v'/x] \in \mathcal{E}_i[\tau_2]$  which is what we needed to show.

**Case**  $\tau = \mu\alpha.x$ , assume  $v \in \mathcal{V}_k[\mu\alpha.\tau]$  and  $j \leq k$ , we then need to show  $v \in \mathcal{V}_j[\mu\alpha.\tau]$ . From  $v \in \mathcal{V}_k[\tau]$ , we conclude that there must exist a  $v'$  such that  $v = \mathbf{fold} v'$ . Now assume  $i < j$  and show  $v' \in \mathcal{V}_i[\tau[\mu\alpha.\tau/\alpha]]$ . From  $i < j$  and  $j \leq k$ , we can conclude  $i < k$ . We use with

$$\forall n < k. v' \in \mathcal{V}_n[\tau[\mu\alpha.\tau/\alpha]],$$

which we get from  $v \in \mathcal{V}_k[\mu\alpha.\tau]$ , to get  $v' \in \mathcal{V}_i[\tau[\mu\alpha.\tau/\alpha]]$ .  $\square$

*Proof (Fundamental Property, Theorem 11).* Proof by induction on the typing derivation.

**Case T-Fold,**

Assume

$$\Gamma \vdash \mathbf{fold} e : \mu\alpha.\tau$$

We need to show

$$\Gamma \models \mathbf{fold} e : \mu\alpha.\tau$$

So suppose we have  $k \geq 0$  and  $\gamma \in \mathcal{G}_k[\mu\alpha.\tau]$ , then we need to show  $\gamma(\mathbf{fold} e) \in \mathcal{E}_k[\mu\alpha.\tau]$  which amounts to showing  $\mathbf{fold} \gamma(e) \in \mathcal{E}_k[\mu\alpha.\tau]$ .

So suppose that  $j < k$  and that  $\mathbf{fold} \gamma(e) \rightarrow^j e'$  and  $\mathbf{irred}(e')$ , then we need to show  $e' \in \mathcal{V}_{k-j}[\mu\alpha.\tau]$ . As we have assumed that  $\mathbf{fold} \gamma(e)$  reduces to something irreducible, and the operational semantics of this language are deterministic, we know that  $\gamma(e)$  must have evaluated down to something irreducible. We therefore know that  $\gamma(e) \rightarrow^{j_1} e_1$  where  $j_1 \leq j$  and  $\mathbf{irred}(e_1)$ . Now we use our induction hypothesis:

$$\Gamma \models e : \tau[\mu\alpha.\tau/\alpha]$$

We instantiate this with  $k$  and  $\gamma \in \mathcal{G}_k[\Gamma]$  to get  $\gamma(e) \in \mathcal{E}_k[\tau[\mu\alpha.\tau/\alpha]]$ . Which we then can instantiate with  $j_1$  and  $e_1$  to get  $e_1 \in \mathcal{V}_{k-j_1}[\tau[\mu\alpha.\tau/\alpha]]$ .

Now let us take a step back and see what happened:

$$\begin{aligned} \mathbf{fold} \gamma(e) &\rightarrow^{j_1} \mathbf{fold} e_1 \\ &\equiv \mathbf{fold} v_1 \\ &\equiv e' \end{aligned}$$

We started with a `fold`  $\gamma(e)$  which took  $j_1$  steps to `fold`  $e_1$ . We have just shown that this  $e_1$  is actually a value because it is in the value interpretation of  $\mathcal{V}_{k-j_1} \llbracket \tau[\mu\alpha.\tau/\alpha] \rrbracket$ . To remind us  $e_1$  is a value let us henceforth refer to it as  $v_1$ . We further know that `fold`  $\gamma(e)$  reduces to  $e'$  in  $j$  steps and that  $e'$  is irreducible. We can further conclude that  $e' = \text{fold } v_1$  and  $j = j_1$  as the language is deterministic and `fold`  $v_1$  is irreducible (because it is a value). Our proof obligation is to show  $e' = \text{fold } v_1 \in \mathcal{V}_{k-j} \llbracket \mu\alpha.\tau \rrbracket$  to show this suppose we have  $l < k - j$  (this also gives us  $l < k - j_1$  as  $j = j_1$ ). We then need to show  $v_1 \in \mathcal{V}_l \llbracket \tau[\mu\alpha.\tau/\alpha] \rrbracket$ , we obtain this result from the monotonicity lemma using the assumption  $v_1 \in \mathcal{V}_{k-j_1} \llbracket \tau[\mu\alpha.\tau/\alpha] \rrbracket$  and  $l < k - j_1$ .  $\square$

The *list* type from Section 6.1 uses the sum type. Sums are a straight forward extension to this language. The extension of the value interpretation would be:

$$\mathcal{V}_k \llbracket \tau_1 + \tau_2 \rrbracket = \{\text{inl } v_1 \mid v_1 \in \mathcal{V}_k \llbracket \tau_1 \rrbracket\} \cup \{\text{inr } v_2 \mid v_2 \in \mathcal{V}_k \llbracket \tau_2 \rrbracket\}$$

We can use  $k$  directly or  $k$  decremented by one depending on whether casing should take up a step or not. Either way the definition is well-founded.

### 6.3 Exercises

1. Do the lambda and application case of the proof of the *Fundamental Property* (Theorem 11).
2. Try to prove the *monotonicity* lemma where the definition of the value interpretation has been adjusted with:

$$\mathcal{V}_k \llbracket \tau_1 \rightarrow \tau_2 \rrbracket = \{\lambda x : \tau_1. e \mid \forall v \in \mathcal{V}_k \llbracket \tau_1 \rrbracket. e[v/x] \in \mathcal{E}_k \llbracket \tau_2 \rrbracket\}$$

This will fail, but it is instructive to see how it fails.

## 7 References and Worlds

This section is not based on Amal Ahmed's lectures at OPLSS '15, so the reader will have to make do with these notes.

## 7.1 STLC with References

In order to add references to STLC, we add means to allocate new references, make assignments to existing references, and dereference existing references:

$$\begin{aligned}\tau &::= \dots \mid \mathit{ref} \ \tau \\ e &::= \dots \mid \mathbf{ref} \ e \mid e := e \mid !e \mid l \\ E &::= \dots \mid \mathbf{ref} \ E \mid E := e \mid v := E \mid !E \\ v &::= \dots \mid l\end{aligned}$$

The expression  $\mathbf{ref} \ e$  allocates a new cell with an initial value specified by  $e$ . The expression  $e := e$  makes an assignment to an existing reference. The expression  $!e$  dereferences an existing reference. We also add locations  $l$  to the language. Locations are references to the heap.

In order to model references, we need to add a store or a heap to our language. Our heap  $h$  is simply a finite partial map from location  $\text{Location}$  to the values of the language:

$$h : \text{Location} \xrightarrow{\text{fin}} \text{Value}$$

This is a reasonable model of a real heap: The map contains all the allocated locations. Any real program will only ever allocated a finite amount of memory, and we have infinitely many available heap cells, so we do not have to worry about running out of memory.

We also need to update the operational semantics, so it has the heap available during evaluation. We update the step relation to be a partial relation on configurations where a configuration is a pair of a heap and an expression.

$$\langle h, e \rangle \rightarrow \langle h', e' \rangle$$

The pure reductions of the language are all the reductions of STLC. These reductions are pure because they do not make changes to the heap.

$$\frac{e \rightarrow e'}{\langle h, E[e] \rangle \rightarrow \langle h, E[e'] \rangle}$$

We introduce an impure reduction, i.e. a reduction that manipulate or interacts with the heap, for each of the three new ways to interact with the heap.

$$\frac{l \notin \text{dom}(h)}{\langle h, E[\mathbf{ref} \ v] \rangle \rightarrow \langle h[l \mapsto v], E[l] \rangle} \text{E-ALLOC}$$

$$\frac{l \in \text{dom}(h)}{\langle h, E[l := v] \rangle \rightarrow \langle h[l \mapsto v], E[v] \rangle} \text{E-ASSIGN}$$

$$\frac{l \in \text{dom}(h)}{\langle h, E[!l] \rangle \rightarrow \langle h, E[h(l)] \rangle} \text{E-DEREF}$$

An allocation  $\mathbf{ref} v$  allocates a new location  $l$  on the heap with the initial value  $v$ . The location must be new in the sense that the heap did not previously use that location. Further, we do not expose the locations in the surface language (this is enforced by the type system), so allocation is the only way to obtain locations. An assignment  $l := v$  updates the heap at location  $l$  to point to  $v$ . Here the result of an assignment is the assigned value, but often it will just be unit value (we do not opt for this as we have not introduced unit value). Finally,  $!l$  dereferences location  $l$  which means that it looks up the value denoted by  $l$  in the heap.

With the heap introduced, we will also have to change some of the definitions that have been reoccurring in previous sections. One such definition is *irred*:

$$\begin{aligned} & \text{irred}(h, e) \\ & \text{iff} \\ & \nexists h', e'. \langle h, e \rangle \rightarrow \langle h', e' \rangle \end{aligned}$$

For each of the new expressions, except locations, we introduce a new typing rule to the type system:

$$\begin{array}{c} \frac{\Gamma \vdash e : \tau}{\Gamma \vdash \mathbf{ref} e : \text{ref } \tau} \text{T-ALLOC} \qquad \frac{\Gamma \vdash e_1 : \text{ref } \tau \quad \Gamma \vdash e_2 : \tau}{\Gamma \vdash e_1 := e_2 : \tau} \text{T-ASSIGN} \\ \\ \frac{\Gamma \vdash e : \text{ref } \tau}{\Gamma \vdash !e : \tau} \text{T-DEREF} \end{array}$$

We also introduce the notion of a well-types heap:

$$\Gamma \vdash h : \Sigma \quad \text{iff} \quad \begin{cases} \text{dom}(h) = \text{dom}(\Sigma) \wedge \\ \forall l \in \text{dom}(h). \Sigma; \Gamma \vdash h(l) : \Sigma(l) \end{cases}$$

## 7.2 Properties of STLC with References

With the language defined, we take a step back to consider what properties the language have.

**The operational semantics is non-deterministic:** In the reduction rule E-ALLOC, we only require  $l$  to be a new location that is not in the domain of the heap. It can, however, be any new location which is the cause of the non-determinism. By leaving allocation underspecified, our system can handle different real implementations of allocation. One drawback of non-determinism is that we cannot rely on determinism in our proofs (which we have previously done).

**Evaluation can get stuck:** The reduction rules for assignment and dereference require the location to be in the heap. If this is not the case, then the evaluation is stuck. One way to think of this is as a memory fault. No surface language for expressing locations is provided, so to obtain a reference one has to allocate it which means that there should (hopefully) be no way to obtain a reference for memory that has not been allocated.

**Only mutation on the heap:** The only place we have introduced mutation is on the heap. Variables bound with a  $\lambda$ -expression can still not be modified.

**Values in the same memory cell stay the same type:** The type system enforces the invariant that memory cells always store values of the same type.

**Recursion:** It may come as a surprise that the language has recursion as it does not have recursive types or a fixed-point operator. However, recursion can be achieved through the heap with a technique known as *Landin's knot* [Landin, 1964].

We demonstrate this technique with the following program

```
(let x = ref (\ y : int. y) in
  x := (\ n : int. (!x 0));
  !x) 0
```

which recurses through the heap to diverge. We have written the program in an ML-style, but it can be written as a well-typed expression (see Appendix A) in the language we have presented here (assuming the trivial extension with integers). The program first allocates a new reference with a dummy value as the default value. Specifically, the dummy is chosen, so it has the type of the value we actually want to store (in this case, the dummy is the identity function on integers), but otherwise the function does not matter. The reference is stored in variable  $x$ . Next, the program assigns the function we are actually interested in to  $x$ . This function dereferences  $x$  to get a function which it applies to 0. The function takes an integer as an argument in order to have the correct type. At the point when this function is actually applied, the location denoted by  $x$  will contain the function itself, so it will call itself repeatedly. To start the recursion, the program dereferences  $x$ , which yields the function we just described, and applies the result to 0.

The example should give an idea about how Landin's knot can be used to emulate a fixed-point operator which gives general recursion.

### 7.3 Logical Predicate

In this section, we define a logical relation that we use to prove type safety of STLC with references.

We will have to deal with a couple of new things, but the main thing is how to interpret locations in the value interpretation. The interpretation will of course look something like this

$$\mathcal{V}[\![ref\ \tau]\!] = \{l \mid \dots\},$$

but what should we require from the location? To answer this question, we should consider the elimination form of things of reference type which is dereference. Dereferencing a location yields the value stored in that location of the heap, so that value should be in the value interpretation, i.e.  $h(l) \in \mathcal{V}[\![\tau]\!]$ . However, the value interpretation does not consider a heap in its interpretation, so we cannot write this immediately. We could try to index the value interpretation with a heap, but we are not just interested in safety with respect to one heap. Instead we need a semantic model of the heap.

A semantic model for heaps is called a world. It specifies what values one can expect to find at a given location. We will define worlds in detail later, but for now we will just assume that a world is a function from locations to value predicates. We index our value interpretation with a world, i.e.

$$\mathcal{V}[\![ref\ \tau]\!](W) = \{l \mid \dots\},$$

Returning to the elimination form, we still cannot express that the contents of the heap at a specific location has to be in the value interpretation. We can, however, put restrictions on what the world can allow to be at a specific position. Specifically, we want to make sure that if a location  $l$  is in the value interpretation, then  $l$  can only dereference values that respect safety. This can be achieved by requiring the world to only allow values that respect safety to reside at location  $l$ . The values that respect safety are exactly what is captured by the value interpretation, so we require the predicate at location  $l$  in the world to be equal to the value interpretation.

$$\mathcal{V}[\![ref\ \tau]\!](W) = \{l \mid l \in \text{dom}(W) \wedge W(l) = \mathcal{V}[\![\tau]\!]\},$$

This is exactly the definition we would like for our value interpretation of reference types, but unfortunately we cannot use it. The problem is that the domain of the worlds cannot exist. To see why, we need to take a step back and consider what we have been doing when defining value interpretations so far. The value interpretations defined in previous sections have been elements in the space of all

predicates over values. We call such a space *the space of semantic types*. For instance in Section 3, the space of semantic types  $T$  was  $\text{Pred}(\text{Value})$ . The value interpretation we want now is world indexed, so it needs to be from a different space of semantic types, namely

$$T = \text{World} \rightarrow \text{Pred}(\text{Value})$$

A world is a function from locations to safe values. A predicate of safe values is in the space of semantic types, so the domain of worlds is

$$\text{World} = \text{Location} \xrightarrow{\text{fin}} T$$

If we inline  $T$  in the definition of World, we get

$$\text{World} = \text{Location} \xrightarrow{\text{fin}} (\text{World} \rightarrow \text{Pred}(\text{Value}))$$

which is a recursive domain equation for which no solution exists. This means that we cannot have the proposed definition of the value interpretation as the worlds we need for the definition do not exist. It may not come as a surprise that the proposed definition does not work. After all, this language is similar in its attributes to STLC with recursive types. When we defined the logical relation for that language, we had to introduce step-indexing to approximate the interpretation of values. In our proposed definition, we have nothing like step-indexing, and it would be a bit surprising if we could make do without here.

The solution to circularity in the domain equation seems intuitively similar to the one for recursive types. Rather than using an exact solution to the recursive domain equation, we use an approximate solution. Developing such a solution is beyond the scope of this note, so we only present the definitions necessary to understand the rest. The following is the approximate solution:

$$\xi : \hat{T} \cong \blacktriangleright ((\text{Location} \xrightarrow{\text{fin}} \hat{T}) \xrightarrow{\text{mon}} \text{UPred}(\text{Value}))$$

(ignore the  $\blacktriangleright$  for now) which allows us to define the domain of worlds

$$\text{World} = \text{Location} \xrightarrow{\text{fin}} \hat{T}$$

and finally to define the space of semantic types as

$$T = \text{World} \xrightarrow{\text{mon}} \text{UPred}(\text{Value})$$

In order to have an approximation, we approximate over a metric. In this case, the metric is steps. This means that our predicates should be step-indexed with a

natural number which is where UPred comes into the picture. UPred defines the step-indexed, downwards closed predicates over some domain<sup>16</sup>:

$$\text{UPred}(\text{Value}) = \{A \subseteq \mathbb{N} \times \text{Value} \mid \forall(n, v) \in A. \forall m \leq n. (m, v) \in A\}$$

Because of the approximation, equalities only hold true as long as we have steps left. When we run out of steps, we can no longer distinguish elements from each other and all bets are off. Such an equality is an  $k$ -equality:  $\stackrel{k}{=}$ . An  $k$ -equality has properties reminiscent of step-indexed logical relation for STLC with recursive types ( $\mu\text{LR}$ ): In  $\mu\text{LR}$  the step-index was the amount of steps we could take without observing that a value was not of a certain type. This meant that when we hit zero steps, everything would be in  $\mu\text{LR}$ . Similarly, for an  $k$ -equality  $\stackrel{0}{=}$  is the total relation. For  $\mu\text{LR}$  we had a lemma that stated that the value interpretation was downwards closed. Here we require that the  $k$ -equalities are downwards closed, i.e.  $\stackrel{k}{=} \subseteq \stackrel{k+1}{=}$ .

We could not solve the original domain equation due to cardinality issues related to sets. In some sense,  $\mathbb{N} \xrightarrow{\text{fin}} T$  was simply too big. In fact in order to solve the recursive domain equation, we move away from the familiar space of sets into a different space. This space can be thought of as the space of sets with additional structure. The structure is related to the step-indices, and everything in the space must preserve this structure. We mention this because all the definitions that constitute our logical predicate are in this new space which means that they must preserve the added structure of the spaces. In our proofs, we will also have to make sure this structure is preserved which we sometimes do without mention. The proofs also use some of this structure without mention. We will not prove that our definitions have the necessary structure, but we will leave it as an exercise in the end of this section.

There are still a couple of things we have not mentioned in the domain equation. The first is that the space of semantic types are monotone functions with respect to the world. Later in this section, we present the preorder the functions are monotone with respect to along with some intuition for why we want them to be monotone. We will state this as a lemma later, but we will also note that all the predicates mapped to by a world have to be monotone with respect to worlds. The second is the fact that the solution to the recursive domain equation is an isomorphism  $\xi$ . This means that our definitions will use  $\xi$  to move between the two sides. This is also a technicality, but we include it so our definitions are correct. The third and last thing is the black triangle, known as *later*. The purpose of

---

<sup>16</sup>UPred(Value) was the space of semantic types from which we picked the value interpretation for the logical predicate for STLC with recursive types

the later in the recursive domain equation is to make that space of structured sets under it small enough for a solution to exist. In practice, later makes all  $k$ -equalities under it go one step down. The later is important for the construction of the solution for the recursive domain equation, but that is beyond the scope of this section (we refer the interested reader to Birkedal and Bizjak [2014]).

We can now define the  $k$ -equality for UPred.

**Definition 3.** For  $B \in \text{UPred}(A)$  and  $k \in \mathbb{N}$

$$\lfloor B \rfloor_k \stackrel{\text{def}}{=} \{(n, a) \in B \mid n < k\} \quad \blacksquare$$

**Definition 4.** For  $B, C \in \text{UPred}(A)$  and  $k \in \mathbb{N}$

$$B \stackrel{k}{=} C \quad \text{iff} \quad \lfloor B \rfloor_k = \lfloor C \rfloor_k \quad \blacksquare$$

That is, two predicates are  $k$ -equal if they are equal on all elements with smaller steps than  $k$ .

We can now state the value interpretation for the reference type:

$$\mathcal{V}[\![\text{ref } \tau]\!](W) = \{(k, l) \mid l \in \text{dom}(W) \wedge \xi(W(l)) \stackrel{k}{=} \mathcal{V}[\![\tau]\!]\},$$

This is almost the same as the definition we proposed first. The only difference is that the predicate is step indexed (here we use an explicit step, but it is essentially the same as what we had for recursive types), and it uses a  $k$ -equality rather than a normal equality.

We also need to look at the remainder of the value interpretation. As in all the other logical predicates and relations, the boolean values are always safe, so we just need to add a step index for it to conform with the remaining definitions

$$\mathcal{V}[\![\text{bool}]\!](W) = \{(k, \text{true}), (k, \text{false}) \mid k \in \mathbb{N}\}$$

For the function type, we can start with a similar definition to what we had for recursive types:

$$\mathcal{V}[\![\tau_1 \rightarrow \tau_2]\!](W) = \{(k, \lambda x. e) \mid \forall j \leq k. \forall (j, v) \in \mathcal{V}[\![\tau_1]\!](W). (j, e[v/x]) \in \mathcal{E}[\![\tau_2]\!](W)\}$$

Recall that we consider all  $j \leq k$  because an application of a lambda-abstraction may not happen immediately. Specifically, the argument may be an expression that needs to evaluate to a value before the application can take place. In this language, evaluating an expression may have side effects. It may assign new values to certain locations, which is fine as long as they are in the value interpretation, or it may also allocate new references which corresponds to adding new locations to the heap.

Because of the non-determinism of location allocation, we cannot statically (that is before the execution) say what type of values will be stored where on the heap, so we the world needs to be updated dynamically with this information. To this end, we construct a future world relation that captures allocation of new locations.

**Definition 5** (Future worlds). *For worlds  $W$  and  $W'$ :*

$$W' \sqsupseteq W \quad \text{iff} \quad \text{dom}(W') \supseteq \text{dom}(W) \wedge \forall l \in \text{dom}(W). W'(l) = W(l)$$

For  $W' \sqsupseteq W$ , we say  $W'$  is a future world of  $W$ . ■

The future world relation is extensional in the sense that a future world remains the same as the past world with respect to everything but possible extensions. We use this to finish our value interpretation by saying that the argument should be valid in a future world which means that it may have allocated new locations.

$$\mathcal{V}[\tau_1 \rightarrow \tau_2](W) = \left\{ (k, \lambda x. e) \left| \begin{array}{l} \forall W' \sqsupseteq W, j \leq k, (j, v) \in \mathcal{V}[\tau_1](W'). \\ (j, e[v/x]) \in \mathcal{E}[\tau_2](W') \end{array} \right. \right\}$$

Generally speaking, if a value is safe with respect to a world, then it should remain that even if the presence of changes to the heap. For instance, if we have a safe location  $l$ , i.e. a location that is in the value interpretation, then its safety should not be affected by allocation of new locations. We should still only be able to use  $l$  to dereference safe values. As the future world relation models changes to the heap, we want our value interpretation to be monotone with respect to future worlds, so safe values are resilient to allocation of new references.

The expression interpretation is also going to look like the one from the logical predicate for STLC with recursive types. That is, if an expression reduces to an irreducible expression, then it must be in the value interpretation. However in this language, it is configurations and not expressions that take evaluation steps, so we need to specify with respect to what heap the expression evaluates in. To this end, we simply take any heap that satisfies the world. For a heap to satisfy a world, it should have the locations specified by the world; and for all the locations in the heap, it should contain a value permitted by the world.

**Definition 6** (Heap satisfaction). *For a heap  $h$  and world  $W$ , we have*

$$h :_k W \quad \text{iff} \quad \text{dom}(h) = \text{dom}(W) \wedge \forall l \in \text{dom}(h). (k, h(l)) \in \xi(W(l))(W)$$

When  $h :_k W$  we say:  $h$  ( $k$ -)satisfies  $W$ . ■

The future world relation is a partial order (exercise). With this definition, we are ready to state the expression interpretation:

$$(k, e) \in \mathcal{E}[\tau](W) \text{ iff } \begin{cases} \forall j < k, h, h', e'. \\ h :_k W \wedge \langle h, e \rangle \rightarrow^j \langle h', e' \rangle \wedge \text{irred}(h', e') \\ \implies \\ \exists W' \sqsupseteq W. h' :_{k-j} W \wedge (k-j, e') \in \mathcal{V}[\tau](W') \end{cases}$$

Note that for the result heap  $h'$  there should exist some future world of  $W$  such that  $h'$  satisfies that world. This means that during the evaluation the expression can only have made allocations consistent with the heap it started evaluation in.

As per usual, we need to have an interpretation of typing contexts. It is defined in a straight forward manner by lifting the simple definition with worlds and step-indexing:

$$\begin{aligned} \mathcal{G}[\emptyset](W) &::= \{(k, \emptyset)\} \\ \mathcal{G}[\Gamma, x : \tau](W) &::= \{(k, \gamma[x \mapsto v]) \mid (k, v) \in \mathcal{V}[\tau](W) \wedge (k, \gamma) \in \mathcal{G}[\Gamma](W)\} \end{aligned}$$

Finally, we can state semantic type safety:

$$\Gamma \models e : \tau \quad \text{iff} \quad \forall W, k \geq 0, (k, \gamma) \in \mathcal{G}[\Gamma](W). (k, \gamma(e)) \in \mathcal{E}[\tau](W)$$

## 7.4 Safety Proof

The safety proof has the same structure as in the previous sections where we proved safety. We do need a number of lemmas for the proof. We leave the proofs of these lemmas as exercises.

First of all, we need our logical predicate to be monotone with respect to worlds.

**Lemma 13** (World monotonicity). *For  $W' \sqsupseteq W$  we have*

- If  $(k, v) \in \mathcal{V}[\tau](W)$ , then  $(k, v) \in \mathcal{V}[\tau](W')$
- If  $(k, e) \in \mathcal{E}[\tau](W)$ , then  $(k, e) \in \mathcal{E}[\tau](W')$
- If  $(k, \gamma) \in \mathcal{G}[\Gamma](W)$ , then  $(k, \gamma) \in \mathcal{G}[\Gamma](W')$  ■

Note that it is not all of our definitions that are monotone with respect to future worlds. Specifically, heap satisfaction is not monotone because it would be

nonsensical. If heap satisfaction was monotone, then a heap would have to be able to handle that worlds that requires more locations allocated than it has.

We also need to make sure that our definitions are downwards closed. This is reminiscent of Lemma 12 from Section 6 (we do not call it monotonicity here to not confuse it with world monotonicity). For instance, if a value is in the value interpretation at index  $k$ , then it should also be in there for any smaller step.

**Lemma 14** (Downwards closure). *For  $j \leq k$*

- *If  $(k, v) \in \mathcal{V}[\tau](W)$ , then  $(j, v) \in \mathcal{V}[\tau](W)$*
- *If  $(k, e) \in \mathcal{E}[\tau](W)$ , then  $(j, e) \in \mathcal{E}[\tau](W)$*
- *If  $(k, \gamma) \in \mathcal{G}[\Gamma](W)$ , then  $(j, \gamma) \in \mathcal{G}[\Gamma](W)$*
- *If  $h :_k W$ , then  $h :_j W$*  ■

Finally, we need a substitution lemma for the **T-Abs** case for the Fundamental Property proof.

**Lemma 15.** *Let  $e$  be syntactically well-formed,  $v$  a closed value,  $\gamma$  a substitution of closed values where  $x$  is not mapped*

$$\gamma(e[e^v/x]) \equiv \gamma[x \mapsto v](e)$$
■

We can now state and prove the fundamental property.

**Theorem 12** (Fundamental Property).

$$\Gamma \vdash e : \tau \implies \Gamma \models e : \tau$$
■

*Fundamental Property, Theorem 12.* Proof by induction over the typing derivation. In all cases, the case for  $k = 0$  is vacuously true as no  $j < 0$  exists.

**Case T-TRUE,** Let  $W$ ,  $k > 0$ , and  $(k, \gamma) \in \mathcal{G}[\Gamma](W)$  be given and show:

$$(k, \mathbf{true}) \in \mathcal{E}[\mathit{bool}](W)$$

To this end let  $j < k$ ,  $h$ ,  $h'$ , and  $e'$  be given such that  $h :_j W$ ,  $\langle h, \mathbf{true} \rangle \rightarrow^j \langle h', e' \rangle$ , and irred  $h', e'$ . By the evaluation relation, it must be the case that  $j = 0$ ,  $h = h'$  and  $\mathbf{true} = e'$ . Now pick  $W' = W$  which means that we by assumption have  $h :_j W$ . Finally, we need to argue  $(k, \mathbf{true}) \in \mathcal{E}[\mathit{bool}](W)$  which follows from the definition of the value interpretation.

**Case T-FALSE**, Analogous to the previous case.

**Case T-VAR**, Also analogous to the above, but to argue  $(k, \gamma(x)) \in \mathcal{E}[\![\tau]\!](W)$  we use assumptions  $(k, \gamma) \in \mathcal{G}[\![\Gamma]\!](W)$  and  $\Gamma \vdash x : \tau$ .

**Case T-IF**, Let  $W, k > 0$ , and  $(k, \gamma) \in \mathcal{G}[\![\Gamma]\!](W)$  be given and show:

$$(k, \text{if } \gamma(e) \text{ then } \gamma(e_1) \text{ else } \gamma(e_2)) \in \mathcal{E}[\![\tau]\!](W)$$

To this end let  $j < k$ ,  $h, h'$ , and  $e'$  be given such that  $h :_j W$  and  $\text{irred } h', e'$  and

$$\langle h, \text{if } \gamma(e) \text{ then } \gamma(e_1) \text{ else } \gamma(e_2) \rangle \rightarrow^j \langle h', e' \rangle.$$

By the operational semantics one of the following must constitute the first part of the assumed evaluation:

1.  $\langle h, \gamma(e) \rangle \rightarrow^j \langle h'', e'' \rangle$  and  $\text{irred}(h'', e'')$  for some  $h''$  and non-value  $e''$ .
2.  $\langle h, \gamma(e) \rangle \rightarrow^i \langle h'', v \rangle$  and for some  $h''$ , value  $v$ , and  $i < j$ .

In case 1, we use our induction hypothesis:  $\Gamma \models e : \text{bool}$ . This gives us that  $(k-j, e'') \in \mathcal{V}[\![\text{bool}]\!](W')$  for some  $W'$  which means that  $e''$  is a value contradicting that it is not.

In case 2, we use the same induction hypothesis to get  $h'' :_{k-i} W''$  and  $(k-i, v) \in \mathcal{V}[\![\text{bool}]\!](W'')$  for some  $W'' \sqsupseteq W$ . From this we conclude that either  $v = \text{true}$  or  $v = \text{false}$ . W.l.o.g. assume  $v = \text{true}$  which means that

$$\langle h'', \text{if } v \text{ then } \gamma(e_1) \text{ else } \gamma(e_2) \rangle \rightarrow \langle h'', \gamma(e_1) \rangle$$

By the evaluation assumption, we know that  $\langle h'', \gamma(e_1) \rangle \rightarrow^{j-i-1} \langle h', e' \rangle$ . At this point, we use another induction hypothesis:  $\Gamma \models e_1 : \tau$ . From this we get  $W' \sqsupseteq W''$  such that  $h :_{k-(j-i-1)} W'$  and  $(k-(j-i-1), e') \in \mathcal{V}[\![\tau]\!](W')$  from which we get  $h :_{k-j} W'$  and  $(k-j, e') \in \mathcal{V}[\![\tau]\!](W')$  using Lemma 14. From  $\sqsupseteq$  being a partial order,  $W' \sqsupseteq W''$  and  $W'' \sqsupseteq W$ , we get  $W' \sqsupseteq W$ , so  $W'$  is exactly the world we needed to provide.

**Case T-ABS**, Let  $W, k > 0$ , and  $(k, \gamma) \in \mathcal{G}[\![\Gamma]\!](W)$  be given and show:

$$(k, \lambda x : \tau_1. \gamma(e)) \in \mathcal{E}[\![\tau_1 \rightarrow \tau_2]\!](W)$$

To this end let  $j < k$ ,  $h, h'$ , and  $e'$  be given such that  $h :_j W$  and  $\langle h, \lambda x : \tau_1. \gamma(e) \rangle \rightarrow^j \langle h', e' \rangle$  and  $\text{irred } h', e'$ . By the evaluation relation we know that  $j = 0$   $e' = \lambda x : \tau_1. \gamma(e)$  and  $h = h'$  as  $\lambda x : \tau_1. \gamma(e)$  is a value. We pick  $W$  and need to show

$$(k, \lambda x : \tau_1. \gamma(e)) \in \mathcal{V}[\![\tau_1 \rightarrow \tau_2]\!](W)$$

To this end let  $W' \sqsubseteq W$ ,  $i \leq k$ , and  $(i, v) \in \mathcal{V}[\tau_1]$  be given and show

$$(i, \gamma(e)[v/x]) \in \mathcal{E}[\tau_2](W')$$

By Lemma 15 it suffices to show  $(i, \gamma[x \mapsto v](e)) \in \mathcal{E}[\tau_2](W')$  By the induction hypothesis we have  $\Gamma, x : \tau_1 \models e : \tau_2$ . Using assumption  $(i, v) \in \mathcal{V}[\tau_1]$ ,  $(k, \gamma) \in \mathcal{G}[\Gamma](W)$ , and Lemma 13 and 14, we get  $\gamma[x \mapsto v] \in \mathcal{G}[\Gamma, x : \tau_1](W')$  which allows us to use the induction hypothesis to conclude

$$(i, \gamma[x \mapsto v](e)) \in \mathcal{E}[\tau_2](W')$$

**Case T-APP**, Let  $W$ ,  $k > 0$ , and  $(k, \gamma) \in \mathcal{G}[\Gamma](W)$  be given and show:

$$(k, \gamma(e_1) \gamma(e_2)) \in \mathcal{E}[\tau](W)$$

To this end let  $j < k$ ,  $h, h'$ , and  $e'$  be given such that  $h :_j W$  and  $\langle h, \gamma(e_1) \gamma(e_2) \rangle \rightarrow^j \langle h', e' \rangle$  and irred  $h', e'$ . From this evaluation and the operational semantics, we conclude that one of the following evaluations of  $e_1$  must take place:

1.  $\langle h, \gamma(e_1) \rangle \rightarrow^j \langle h'', e'' \rangle$  and irred  $(h'', e'')$  for some  $h''$  and non-value  $e''$ .
2.  $\langle h, \gamma(e_1) \rangle \rightarrow^i \langle h'', v_2 \rangle$  and for some  $h''$ , value  $v_2$ , and  $i < j$ .

In case 1, we use our induction hypothesis:  $\Gamma \models e_1 : \tau_2 \rightarrow \tau$ . This gives us that  $(k - j, e'') \in \mathcal{V}[\tau_2 \rightarrow \tau](W')$  for some  $W'$  which means that  $e''$  is a value contradicting that it is not.

In case 2, we use the same induction hypothesis to get  $h'' :_{k-j} W''$  and  $(k - i, v_2) \in \mathcal{V}[\tau_2 \rightarrow \tau](W'')$  for some  $W'' \sqsupseteq W$ . From this we conclude that either  $v_2 = \lambda \tau_2 : x. e$ . Now we do the same thing with the argument evaluation. By the operational semantics one of the following must be the case:

1.  $\langle h'', \gamma(e_2) \rangle \rightarrow^{j-i} \langle h''', e''' \rangle$  and irred  $(h''', e''')$  for some  $h'''$  and non-value  $e'''$ .
2.  $\langle h'', \gamma(e_2) \rangle \rightarrow^n \langle h''', v_2 \rangle$  and for some  $h'''$ , value  $v_2$ , and  $n < j - i$ .

In case 1, we use our induction hypothesis:  $\Gamma \models e_2 : \tau_2$ . This gives us that  $(k - j - i, e''') \in \mathcal{V}[\tau](W')$  for some  $W'$  which means that  $e'''$  is a value contradicting that it is not.

In case 2, we use the same induction hypothesis to get  $h''' :_{k-i-n} W'''$  and  $(k - i - n, v_2) \in \mathcal{V}[\tau_2](W''')$  for some  $W''' \sqsupseteq W''$ . At this point, we have seen the following part of the assumed evaluation:

$$\begin{aligned} \langle h, \gamma(e_1) \gamma(e_2) \rangle &\rightarrow^i \langle h'', \lambda \tau_2 : x. e \gamma(e_2) \rangle \\ &\rightarrow^n \langle h''', \lambda \tau_2 : x. e v_2 \rangle \\ &\rightarrow \langle h''', e[v_2/x] \rangle \end{aligned}$$

where the last step follows from the operational semantics. At this point we use the assumption  $(k - i, v_2) \in \mathcal{V}[\![\tau_2 \rightarrow \tau]\!](W'')$  along with  $W''' \sqsupseteq W''$  and  $(k - i - n, v_2) \in \mathcal{V}[\![\tau_2]\!](W''')$  to conclude  $(k - i - n, e^{[v_2/x]}) \in \mathcal{E}[\![\tau]\!](W''')$ . By the evaluation assumption we had initially we know that  $\langle h''', e^{[v_2/x]} \rangle \rightarrow^{j-k-i-n-1} \langle h, e' \rangle$ . We also know that  $h''' :_{k-i-n} W'''$ , so  $(k - i - n, e^{[v_2/x]}) \in \mathcal{E}[\![\tau]\!](W''')$  gives us  $W'$  such that  $h' :_{k-j+1} W'$  and  $(k - j + 1, e') \in \text{vpred}\tau(W')$ . Using Lemma 14, this is exactly what we needed to show.

**Case T-DEREF**, Let  $W$ ,  $k > 0$ , and  $(k, \gamma) \in \mathcal{G}[\![\Gamma]\!](W)$  be given and show:

$$(k, !\gamma(e)) \in \mathcal{E}[\![\tau]\!](W)$$

To this end let  $j < k$ ,  $h$ ,  $h'$ , and  $e'$  be given such that  $h :_j W$  and  $\langle h, !\gamma(e) \rangle \rightarrow^j \langle h', e' \rangle$  and irred  $h', e'$ . By the assumed evaluation one of the following must be the case

1.  $\langle h, \gamma(e) \rangle \rightarrow^j \langle h'', e'' \rangle$  and irred  $(h'', e'')$  for some  $h''$  and non-value  $e''$ .
2.  $\langle h, \gamma(e) \rangle \rightarrow^i \langle h'', v \rangle$  and for some  $h''$ , value  $v$ , and  $i < j$ .

In case 1, we use our induction hypothesis:  $\Gamma \models e : \text{ref } \tau$ . This gives us that  $(k - j, e'') \in \mathcal{V}[\![\text{ref } \tau]\!](W')$  for some  $W'$  which means that  $e''$  is a value contradicting that it is not.

In case 2, we use the same induction hypothesis to get  $h'' :_{k-i} W''$  and  $(k - i, v) \in \mathcal{V}[\![\text{ref } \tau]\!](W')$  for some  $W'' \sqsubseteq W$ . By the definition of the value interpretation this means that  $v = l$  for some location  $l$ . Now the dereference can happen, i.e.

$$\langle h'', l \rangle \rightarrow \langle h'', h''(l) \rangle$$

as we have been considering the assumed evaluation this must mean that  $h' = h''$  and  $e' = h''(l)$ , so we pick  $W'$  as the we need to pick and get  $h' :_{k-j} W''$  by assumption and Lemma 14. This means that it remains to show  $(k - j) \in \mathcal{V}[\![\tau]\!](W')$ . To show this we need the following:

- From  $h' :_{k-i} W''$ , we get that  $(k - i, h''(l)) \in \xi W(l)$ .
- From  $(k - i, l) \in \mathcal{V}[\![\text{ref } \tau]\!](W')$ , we get  $\xi(W(l)) \stackrel{k-i}{=} \mathcal{V}[\![\tau]\!]$ .

From this and Lemma 14, we conclude  $(k - i - 1, h''(l)) \in \mathcal{V}[\![\tau]\!](W)$  (remember that the definition of  $k$ -equality is defined in terms of  $k$ -cut which takes away a step). We know that  $i < j$ , so  $k - j \leq k - i - 1$ , so the desired result follows from Lemma 14.

**Case T-ALLOC**, Let  $W$ ,  $k > 0$ , and  $(k, \gamma) \in \mathcal{G}[\![\Gamma]\!](W)$  be given and show:

$$(k, \text{ref } \gamma(e)) \in \mathcal{E}[\![\text{ref } \tau]\!](W)$$

To this end let  $j < k$ ,  $h$ ,  $h'$ , and  $e'$  be given such that  $h :_j W$  and  $\langle h, !\gamma(e) \rangle \rightarrow^j \langle h', e' \rangle$  and irred  $h', e'$ . By the assumed evaluation one of the following must be the case

1.  $\langle h, \gamma(e) \rangle \rightarrow^j \langle h'', e'' \rangle$  and irred  $(h'', e'')$  for some  $h''$  and non-value  $e''$ .
2.  $\langle h, \gamma(e) \rangle \rightarrow^i \langle h'', v \rangle$  and for some  $h''$ , value  $v$ , and  $i < j$ .

In case 1, we use our induction hypothesis:  $\Gamma \models e : \tau$ . This gives us that  $(k - j, e'') \in \mathcal{V}[\![\tau]\!](W')$  for some  $W'$  which means that  $e''$  is a value contradicting that it is not.

In case 2, we use the same induction hypothesis to get  $h'' :_{k-i} W''$  and  $(k-i, v) \in \mathcal{V}[\![ref \ \tau]\!](W')$  for some  $W'' \sqsubseteq W$ . Now that  $e$  has been evaluated to a value, the allocation can happen, so the next evaluation step is

$$\langle h'', v \rangle \rightarrow \langle h''[l \mapsto v], l \rangle \text{ for } l \notin \text{dom}(h'')$$

As we have just been looking at the assumed evaluation and this expression is irreducible, it must be the case that  $e' = l$  and  $h' = h''[l \mapsto v]$ . Now we need to pick a world we want to use. We pick  $W' = W''[l \mapsto \xi^{-1}(\mathcal{V}[\![\tau]\!])]$  i.e. we allow anything from the value interpretation of  $\tau$  to reside at location  $l$  (note that we need to apply  $\xi^{-1}$  as is from the wrong domain). This means that we need to show the following

- $h''[l \mapsto v] :_{k-i-1} W'$

The fact that  $\text{dom}(h''[l \mapsto v]) = \text{dom}(W')$  follows from  $h'' :_{k-i} W''$  and  $W' \sqsubseteq W''$ .

Take  $l' \in \text{dom}(W')$ . We need to consider two cases:

- $l' = l$ : In this case we must show  $(k-i-1, h''[l \mapsto v](l)) \in \xi(W(l))(W)$ , i.e.  $(k-i-1, v) \in \mathcal{V}[\![\tau]\!](W)$ , which follows by assumption and Lemma 14.
- $l' \neq l$ : This follows by  $h'' :_{k-i} W''$  and the fact that  $\xi(W''(l))$  is monotone with respect to the world (you can see this if you take a look at the recursive domain equation).

- $(k-i-1, l) \in \mathcal{V}[\![ref \ \tau]\!](W')$

This amounts to showing  $\xi(W(l)) = \xi(\xi^{-1}(\mathcal{V}[\![\tau]\!])) \stackrel{k-i-1}{=} \mathcal{V}[\![\tau]\!]$  which is true as if they are equal, then they are also equal if we limit them to everything of index less than  $k-i-1$ .

**Case T-ASSIGN**, Let  $W$ ,  $k > 0$ , and  $(k, \gamma) \in \mathcal{G}[\Gamma](W)$  be given and show:

$$(k, \gamma(e_1) := \gamma(e_2)) \in \mathcal{E}[\text{ref } \tau](W)$$

To this end let  $j < k$ ,  $h, h'$ , and  $e'$  be given such that  $h :_j W$  and  $\langle h, \gamma(e_1) := \gamma(e_2) \rangle \rightarrow^j \langle h', e' \rangle$  and irred  $h', e'$ .

1.  $\langle h, \gamma(e_1) \rangle \rightarrow^j \langle h'', e'' \rangle$  and irred  $(h'', e'')$  for some  $h''$  and non-value  $e''$ .
2.  $\langle h, \gamma(e_1) \rangle \rightarrow^i \langle h'', v_1 \rangle$  and for some  $h''$ , value  $v_1$ , and  $i < j$ .

In case 1, we use our induction hypothesis:  $\Gamma \models e_1 : \text{ref } \tau$ . This gives us that  $(k-j, e'') \in \mathcal{V}[\text{ref } \tau](W')$  for some  $W'$  which means that  $e''$  is a value contradicting that it is not.

In case 2, we also use the induction hypothesis to get  $h'' :_{k-i} W''$  and  $(k-i, v_1) \in \mathcal{V}[\text{ref } \tau](W'')$  for some  $W'' \sqsubseteq W$ . From this we can conclude  $v_1 = l$  for some location. At this point, the left side of the assignment is fully evaluated, so the right side starts evaluating. The assumed evaluation does one of the following two things

1.  $\langle h'', \gamma(e_2) \rangle \rightarrow^{j-i} \langle h''', e''' \rangle$  and irred  $(h''', e''')$  for some  $h'''$  and non-value  $e'''$ .
2.  $\langle h'', \gamma(e_2) \rangle \rightarrow^{j-1} \langle h''', v_2 \rangle$  and for some  $h'''$ , value  $v_2$ .

In case 1, we use our induction hypothesis:  $\Gamma \models e_2 : \tau$ . This gives us that  $(k-j, e''') \in \mathcal{V}[\tau](W')$  for some  $W'$  which means that  $e'''$  is a value contradicting that it is not.

In case 2, we use the same induction hypothesis to conclude  $h''' :_{k-j+1} W'''$  and  $(k-j+1, v_2) \in \mathcal{V}[\tau](W''')$  for some  $W''' \sqsubseteq W''$ . At this point both left and right side of the assignment has been evaluate, so the last step of the assumed evaluation is the assignment itself:

$$\langle h''', l := v_2 \rangle \rightarrow \langle h'''[l \mapsto v_2], v_2 \rangle$$

Note that since we have just considered the actual steps the computation takes, we know that  $e' = v_2$  and  $h' = h'''[l \mapsto v_2]$ . Now pick  $W' = W'''$  and show

- $(k-j, v_2) \in \mathcal{V}[\tau](W')$

This follows from  $(k-j+1, v_2) \in \mathcal{V}[\tau](W''')$ , and Lemma 14.

- $h' :_{k-j} W'$

Seeing as the assignment did not get stuck, it must be the case that  $l \in \text{dom}(h''')$ . This means that  $\text{dom}(h') = \text{dom}(W')$  follows from  $h''' :_{k-j+1} W'''$ . Further, for  $l' \neq l$  and  $l' \in \text{dom}(h')$  we have  $(k-j, h'(l')) \in \xi(W'(l))(W')$  from the same assumption and the fact that  $\xi(W'(l))(W')$  must be downwards-closed. Finally, we need to show for  $l' = l$ . From  $(k-i, l) \in \mathcal{V}[\![\text{ref } \tau]\!](W'')$ , we know that  $\xi(W''(l)) \stackrel{k-i}{=} \mathcal{V}[\![\tau]\!]$ . Further by assumption, we have  $(k-j+1, v_2) \in \mathcal{V}[\![\tau]\!](W''')$  which gives us  $(k-j, v_2) \in \xi(W''(l))$  - exactly what we need to show. □

The safety predicate we used in previous sections needs to change a bit in this setting, so it takes the heap into account

$$\begin{aligned} & \text{safe}(e) \\ & \text{iff} \\ & \forall h', e'. \langle \emptyset, e \rangle \rightarrow^* \langle h', e' \rangle \implies \text{Val}(e') \vee \exists h'', e''. \langle h', e' \rangle \rightarrow \langle h'', e'' \rangle \end{aligned}$$

**Lemma 16.**

$$\emptyset \models e : \tau \implies \text{safe}(e) \quad \blacksquare$$

*Proof.* Assume  $\emptyset \models e : \tau$  and let  $h, h', e'$  be given such that  $\langle h, e \rangle \rightarrow^* \langle h', e' \rangle$ . Proceed by case on  $\text{irred}(h', e')$ .

**Case**  $\neg \text{irred}(h', e')$ ,: In this case, it follows by definition of  $\text{irred}$  that there exists  $h''$  and  $e''$  such that  $\langle h', e' \rangle \rightarrow \langle h'', e'' \rangle$ .

**Case**  $\text{irred}(h', e')$ ,: Say the evaluation takes  $k$  steps to do the assumed evaluation, i.e.  $\langle \emptyset, e \rangle \rightarrow^k \langle h', e' \rangle$ . Now use assumption  $\emptyset \models e : \tau$  to get  $(k+1, e) \in \mathcal{E}[\![\tau]\!](\emptyset)$ . If we use this with the assumed evaluation,  $\text{irred}(h', e')$ , and the fact that  $\emptyset :_k \emptyset$  (trivially true), then we get  $W' \sqsubseteq \emptyset$  such that  $h' :_1 W'$  and  $(1, e) \in \mathcal{V}[\![\tau]\!](W')$ . This means that  $e$  is indeed a value. □

**Theorem 13** (STLC with references is type safe). *If  $\emptyset \vdash e : \tau$ , then  $\text{safe}(e)$*  ■

*Proof.* Follows from Theorem 12 and Lemma ?? □

## 7.5 Exercises

1. Verify that the  $k$ -equality on  $\text{UPred}(A)$  satisfy the necessary properties. That is

- $\stackrel{0}{=}$  is the total relation, i.e.  $\forall a, a' \in \text{UPred}(A). a \stackrel{0}{=} a'$ .
- $\stackrel{k+1}{=} \subseteq \stackrel{k}{=}$ , i.e.  $\forall a, a' \in \text{UPred}(A). a \stackrel{k+1}{=} a' \implies a \stackrel{k}{=} a'$

A  $k$ -equality must also satisfy the following property omitted from the above presentation

$$\forall a, a' \in \text{UPred}(A). (\forall k \in \mathbb{N} a \stackrel{k}{=} a') \implies a = a'$$

That is, if two elements approximates each other for any index, then they should in fact be equal.

2. Verify that the future world relation is a partial order. That is for worlds  $W, W'$  and  $W''$  verify that it satisfies the following three properties
  - Reflexivity:  $W \sqsubseteq W$
  - Antisymmetry: If  $W' \sqsubseteq W$  and  $W \sqsubseteq W'$ , then  $W = W'$
  - Transitivity: If  $W' \sqsubseteq W$  and  $W'' \sqsubseteq W'$ , then  $W'' \sqsubseteq W$
3. Prove the future world monotonicity lemma, i.e. Lemma 13.
4. Prove the downwards closure lemma, i.e. Lemma 14.
5. Prove the substitution lemma, i.e. Lemma 15.
6. In the space where the recursive domain equation is solved, functions must preserve the added structure, i.e. step indices. Specifically, all functions must be non-expansive. A function is non-expansive if it preserves  $k$ -equalities, i.e.  $f : X \rightarrow Y$  is non-expansive if the following holds true

$$\forall x, x' \in X. \forall k \in \mathbb{N}. x \stackrel{k}{=} x' \implies f(x) \stackrel{k}{=} f(x')$$

Note that the first  $k$ -equality is defined for elements in  $X$  and the second  $k$ -equality is defined for elements in  $Y$ . It is these  $k$ -equalities that must be preserved. The value interpretation is a function of the space in which the recursive domain equation is solved. Prove that this is the case.

7. Prove the remaining cases of Theorem 12.

## 7.6 Further Reading

We refer the interested reader to the note *Logical Relations and References* [Skorstengaard, 2015] for further reading. It contains: important details that we omitted here for the sake of presentation, a more expressive language which allows for interesting examples but also calls for a different kind of world, and a logical relation rather than a predicate. There is some overlap between this section and the note in question.

## Acknowledgements

It is established practice for authors to accept responsibility for any and all mistakes in documents like this. I, however, do not. If you find anything amiss, please let me know so I can figure out who of the following are to blame: Amal Ahmed, Mathias Høier, Morten Krogh-Jespersen, Kent Grigo, and Kristoffer Just Andersen.

## References

- Benjamin C. Pierce. *Types and Programming Languages*. The MIT Press, 1st edition, 2002. ISBN 0262162091, 9780262162098.
- Philip Wadler. Theorems for free! In *Proceedings of the Fourth International Conference on Functional Programming Languages and Computer Architecture*, FPCA '89, pages 347–359, New York, NY, USA, 1989. ACM. ISBN 0-89791-328-0. doi: 10.1145/99370.99404. URL <http://doi.acm.org/10.1145/99370.99404>.
- John C. Mitchell and Gordon D. Plotkin. Abstract types have existential type. *ACM Trans. Program. Lang. Syst.*, 10(3):470–502, July 1988. ISSN 0164-0925. doi: 10.1145/44501.45065. URL <http://doi.acm.org/10.1145/44501.45065>.
- Peter J. Landin. The mechanical evaluation of expressions. *The Computer Journal*, 6(4):308–320, 1964.
- Lars Birkedal and Aleš Bizjak. A Taste of Categorical Logic — tutorial notes. <http://cs.au.dk/~birke/modures/tutorial/categorical-logic-tutorial-notes.pdf>, 2014.

## A Landin's Knot

In this section, we show an implementation of

```
(let x = ref (\ y : int. y) in
  x := (\ n : int. (!x 0));
  !x) 0
```

in the language with references presented in Section 7. We also demonstrate how this causes a recursion.

First we define the following gadgets:

$$\begin{aligned} dummy &= \lambda x : int. x \\ Landin's &= \lambda y : int \rightarrow int. !x \\ knot &= x := (\lambda n : int. (!x 0)) \end{aligned}$$

The implementation of the program is

$$((\lambda x : ref (\tau). Landin's\ knot) (\mathbf{ref}\ dummy)) 0$$

and it does indeed type check which the type derivation tree on the next page demonstrates:

Let  $\tau = \text{int} \rightarrow \text{int}$

$$\begin{array}{c}
\frac{x : \text{ref } \tau, n : \text{int}; \bullet \vdash x : \text{ref } \tau}{x : \text{ref } \tau, n : \text{int}; \bullet \vdash !x : \tau} \quad \frac{x : \text{ref } \tau, n : \text{int}; \bullet \vdash 0 : \text{int}}{x : \text{ref } \tau, n : \text{int}; \bullet \vdash !x 0 : \text{int}} \\
\frac{x : \text{ref } \tau, n : \text{int}; \bullet \vdash !x 0 : \text{int}}{x : \text{ref } \tau; \bullet \vdash \lambda n : \text{int}. (!x 0) : \tau} \\
\frac{x : \text{ref } \tau; \bullet \vdash x : \text{ref } \tau}{x : \text{ref } \tau; \bullet \vdash x := (\lambda n : \text{int}. (!x 0)) : \tau} \\
\frac{x : \text{ref } \tau; \bullet \vdash \text{knot} : \tau}{x : \text{ref } \tau, y : \tau; \bullet \vdash x : \text{ref } \tau} \\
\frac{x : \text{ref } \tau, y : \tau; \bullet \vdash !x : \tau}{x : \text{ref } \tau; \bullet \vdash \text{Landin's knot} : \tau} \quad \vdots \\
\frac{x : \text{ref } \tau; \bullet \vdash \text{Landin's knot} : \tau}{\bullet; \bullet \vdash (\lambda x : \text{ref } \tau. \text{Landin's knot}) : \text{ref } \tau \rightarrow \tau} \quad \frac{x : \text{int}; \bullet \vdash x : \text{int}}{\bullet; \bullet \vdash \text{dummy} : \text{int} \rightarrow \text{int}} \\
\frac{\bullet; \bullet \vdash (\lambda x : \text{ref } \tau. \text{Landin's knot}) (\text{ref } \tau) \rightarrow \tau}{\bullet; \bullet \vdash (\lambda x : \text{ref } \tau. \text{Landin's knot}) (\text{ref dummy}) : \tau} \\
\vdots \\
\frac{\bullet; \bullet \vdash (\lambda x : \text{ref } \tau. \text{Landin's knot}) (\text{ref dummy}) : \tau}{\bullet; \bullet \vdash ((\lambda x : \text{ref } \tau). \text{Landin's knot}) (\text{ref dummy}) 0 : \text{int}}
\end{array}$$

In order to see how the expression diverges, we first consider the left side of the application in which the knot is prepared:

$$\begin{aligned}
& \langle h, (\lambda x : \text{ref } (\tau). \text{Landin's knot}) (\mathbf{ref } \text{dummy}) \rangle \\
& \rightarrow \langle h[l \mapsto \text{dummy}], (\lambda x : \text{ref } (\tau). \text{Landin's knot}) l \rangle \\
& \rightarrow \langle h[l \mapsto \text{dummy}], (\text{Landin's knot})[l/x] \rangle \\
& \equiv \langle h[l \mapsto \text{dummy}], \text{Landin's}[l/x] (l := (\lambda n : \text{int.} (!l 0))) \rangle \\
& \rightarrow \langle h[l \mapsto \lambda n : \text{int.} (!l 0)], \text{Landin's}[l/x] (\lambda n : \text{int.} (!l 0)) \rangle \\
& \rightarrow \langle h[l \mapsto \lambda n : \text{int.} (!l 0)], !l \rangle \\
& \rightarrow \langle h[l \mapsto \lambda n : \text{int.} (!l 0)], \lambda n : \text{int.} (!l 0) \rangle
\end{aligned}$$

With the not prepared, 0 is applied to activate it

$$\begin{aligned}
& \langle h, ((\lambda x : \text{ref } (\tau). \text{Landin's knot}) (\mathbf{ref } \text{dummy}) 0) \rangle \\
& \rightarrow^* \langle h[l \mapsto \lambda n : \text{int.} (!l 0)], (\lambda n : \text{int.} (!l 0)) 0 \rangle \\
& \rightarrow \langle h[l \mapsto \lambda n : \text{int.} (!l 0)], (!l 0) \rangle \\
& \rightarrow \langle h[l \mapsto \lambda n : \text{int.} (!l 0)], ((\lambda n : \text{int.} (!l 0)) 0) \rangle \\
& \rightarrow \langle h[l \mapsto \lambda n : \text{int.} (!l 0)], (!l 0) \rangle \\
& \rightarrow \langle h[l \mapsto \lambda n : \text{int.} (!l 0)], ((\lambda n : \text{int.} (!l 0)) 0) \rangle \\
& \rightarrow \langle h[l \mapsto \lambda n : \text{int.} (!l 0)], (!l 0) \rangle \\
& \rightarrow \langle h[l \mapsto \lambda n : \text{int.} (!l 0)], ((\lambda n : \text{int.} (!l 0)) 0) \rangle \\
& \rightarrow \dots
\end{aligned}$$

the execution keeps alternating between two configurations, so it diverges.