

# Proofs about Mapping Polymorphism

Ryan Wisnesky and Mauricio Hernandez and Lucian Popa

September 6, 2009

## Theorem 1 - Satisfiability

This proof needs to use the full mapping language. We write  $\langle M, S, T \rangle$  to indicate that a set of mapping expressions  $M$  can be typing using two contexts,  $S$  and  $T$ , where  $S$  contains source variables and  $T$  target variables. At the start, we assume that  $M$  has only two free variables,  $Src$  on the source side and  $Dst$  on the target side. A set of mapping expressions is typechecked by typechecking each mapping expression  $m \in M$ .

*Proof.* The idea is that we go backwards on the typing derivation for  $\langle M, S, T \rangle$ , and we construct the source instance  $I$  and the target instance  $J$  that we need.

First,  $I$  is “initialized” to contain a root element  $Src$  of type  $S$ , and  $J$  is “initialized” to contain a root element  $Dst$  of type  $T$ . This is what we do for  $I$ . (A similar process applies to create the initial  $J$ .) If  $S$  is a record type, then  $Src$  must contain all the labels given by the record type. For each label  $l$ , we will create a value of the corresponding type  $t$ . If  $t$  is another record type, we recurse. If  $t$  is a SETRCD or SETCHC type, the value we create is an empty set. If  $t$  is ATOMIC, we create a new “value” (or labeled null)  $X$  of type  $a$ . This  $X$  is not a real value, we should think of it as a variable. But in the end all such “variables” will be replaced by real values.

I hope this initialization step is clear. Basically, these  $I$  and  $J$  are skeletons of instances, containing just the top record parts (all sets are empty). Now we start adding to these sets, based on the typing derivations for each  $m \in M$ .

So, for each  $m \in M$ , we do the following. When we hit the typing rule SETRCD-ELIM, we already have some set value  $X$  for  $p$ , where  $p$  is the expression that is of type SETRCD  $r$ . (In general, as we go up the type derivation tree, each path obtained from some variable that is in the context  $\Gamma$  will have some value. The initial case is when  $p$  is a path from the root (either  $Src$  or  $Dst$ ). In that case, if this is the first  $m \in M$  to type-check, the value for  $p$  is simply the empty set.) What we do is to create a record value  $R$  of type  $r$  and insert it into the set  $X$ . The way we create the value  $R$  is by applying the same process we applied for the initialization step (just fill in with labels and atomic values, and leave sets empty).

Note that it is possible to hit the SETRCD-ELIM rule again later for the *same* set value  $X$ . So, we may insert more values into  $X$ . That’s ok.

Now, for the SETCHC rule, the process is the same, except that instead of creating a record value we need to create a choice value  $C$  that we insert into  $X$ . We know which branch  $C$  will have because the rule specifies the label. The construction of what we put inside the label follows, again, the same process (fill in with labels and atomic values, and leave sets empty).

As in the case of SETRCD, it is possible to hit the same SETCHC rule again later for the *same* set value  $X$ . So, we may insert more values into  $X$ , possibly with different branches. But that’s ok, since  $X$  is a set. (When choices are not bundled together in a SETCHC, we can run into trouble – unsatisfiable cases. But not now).

Then, in both cases of SETRCD or SETCHC, once we construct the record value  $R$  or the choice value  $C$ , we “assign it” to the variable  $v$ , and then we go recursively, with the new context  $\Gamma + v$ ,

up the right branch of the derivation tree. (This “assignment” of a value to  $v$  ensures that, later on, each path reachable from some variable in the new context will have a value.)

So, this process continues until we hit the WF-EQ rules. Then  $p$  and  $p'$  are two values,  $X$  and  $X'$ . (Both could be from  $J$ , or both could be from  $I$ , or one from each.) We then identify  $X$  and  $X'$ . This identification never fails, since  $X$  and  $X'$  are not constants (like 3 and 4) but variables.

In the end, we obtain two instances  $I$  and  $J$ , where the atomic elements are variables. And some of the variables in  $I$  appear in  $J$ . We replace all variables uniformly with real values. It's easy to argue that the resulting  $I$  and  $J$  satisfy the constraints in  $M$  (they were minimally constructed to satisfy them).  $\square$

## Proposition 1 - Principal Typings are Unique

*Proof.* Let  $fv(\Gamma)$  indicate the union of the type, row, and atomic variables in the types in the range of context  $\Gamma$ .

Suppose that  $\Gamma_1$  and  $\Gamma_2$  are principal typings. As such, we know that there exists substitutions  $\phi_{1,2}$  and  $\phi_{2,1}$  such that (A)  $\phi_{1,2}\Gamma_1 = \Gamma_2$  and (B)  $\phi_{2,1}\Gamma_2 = \Gamma_1$ . Substituting for  $\Gamma_1$  and  $\Gamma_2$  into (A) and (B) yields that  $\phi_{2,1}\phi_{1,2}\Gamma_1 = \Gamma_1$  and  $\phi_{1,2}\phi_{2,1}\Gamma_2 = \Gamma_2$ . Therefore we know that (up to  $\alpha$ -equivalence) that

$$(C) \quad \forall v \in fv(\Gamma_1), \phi_{2,1}\phi_{1,2}v = v$$

and

$$(D) \quad \forall v \in fv(\Gamma_2), \phi_{1,2}\phi_{2,1}v = v$$

Our goal is to show that  $\forall v \in fv(\Gamma_1), \phi_{1,2}v = v$ , from which it follows that  $\phi_{1,2}\Gamma_1 = \Gamma_1$ , which by (A) establishes that  $\Gamma_1 = \Gamma_2$ .

So, let  $v \in fv(\Gamma_1)$ . From (C) we have that

$$(E) \quad \phi_{2,1}\phi_{1,2}v = v$$

Hence  $\phi_{1,2}v$  must be a variable in  $fv(\Gamma_2)$ , and so we can instantiate (D) with  $\phi_{1,2}v$  to get

$$\phi_{1,2}\phi_{2,1}\phi_{1,2}v = v$$

Rewriting by (E) gives  $\phi_{1,2}v = v$ , as required. □

## **Proposition 2 - Schema Unification produces MGUs**

*Proof.* Our schema unification rules are simply those of the Jone's type unification system specialized to NR schema. Hence this is a straightforward consequence of this property in that system.  $\square$

## Theorem 2 - Soundness

We start by establishing the soundness of path inference over path checking. We will need:

**Lemma 1** (Path checking respects substitution).  $\forall \Gamma p t \phi, \Gamma \vdash p : t \rightarrow \phi\Gamma \vdash p : \phi t$ .

*Proof.* Introduce  $\Gamma$  and proceed by induction on  $p$ .

- In the case where  $p$  is a variable,  $(v, t) \in \Gamma$  and so  $(v, \phi t) \in \phi\Gamma$  and the result holds by VAR.
- In the case where we have a path  $p.l$ , the inductive hypothesis is that

$$\forall t \phi, \Gamma \vdash p : t \rightarrow \phi\Gamma \vdash p : \phi t$$

Introduce  $\Gamma$  and  $t$  and  $\phi$  and assume that  $\Gamma \vdash p.l : t$ . By inversion there is some  $r$  such that  $\Gamma \vdash p : \text{RCD} (l : t, r)$ . Our goal is that  $\phi\Gamma \vdash p.l : \phi t$ . Apply the inductive hypothesis to get  $\phi\Gamma \vdash p : \phi\text{RCD} (l : t, r)$ . The result then follows by the RCD-ELIM rule.  $\square$

Also recall the definition of a unifier: if  $a \stackrel{\phi}{\sim} b$ , then  $\phi a = \phi b$ . Carrying on then, we have:

**Lemma 2** (Soundness of Path Inference).  $\forall \Gamma p \varphi \tau, \varphi\Gamma \Vdash p : \tau \rightarrow \varphi\Gamma \vdash p : \tau$ .

*Proof.* Introduce  $\Gamma$  and proceed by induction on  $p$ .

The base case is that  $p$  is a variable. In this case, the checking and inference rules are identical. (The substitution returned from the inference algorithm is the identity function.)

The inductive step is that  $p$  is a projection. The inductive hypothesis is  $\forall \varphi \tau, \varphi\Gamma \Vdash p : \tau \rightarrow \varphi\Gamma \vdash p : \tau$ . We must show that  $\forall \varphi \tau, \varphi\Gamma \Vdash p.l : \tau \rightarrow \varphi\Gamma \vdash p.l : \tau$ . Introduce  $\varphi$  and  $\tau$  and assume that  $\varphi\Gamma \Vdash p.l : \tau$ . By inversion, we know that there exists  $\phi, \psi, t, \sigma, \rho$  such that  $\varphi = \psi \circ \phi$  and  $\tau = \psi\sigma$  and  $\phi\Gamma \Vdash p :: t$  and  $\text{RCD} (l : \sigma, \rho) \stackrel{\psi}{\sim} t$  with  $\sigma, \rho$  fresh. Substituting for  $\tau$  and  $\varphi$  gives us a goal of  $\psi\phi\Gamma \vdash p.l : \psi\sigma$ .

We now apply the RCD-ELIM rule, setting its  $\Gamma$  to be  $\psi\phi\Gamma$ , and its  $t$  and  $r$  to be  $\psi\sigma$  and  $\psi\rho$ , respectively. This yields a new goal of  $\psi\phi\Gamma \vdash p : \text{RCD} (l : \psi\sigma, \psi\rho)$ . Using our inductive hypothesis with  $\phi$  and  $t$  and  $\phi\Gamma \Vdash p :: t$  gives us that  $\phi\Gamma \vdash p : t$ . We can then apply  $\psi$  to both sides (by Lemma 1) to get that  $\psi\phi\Gamma \vdash p : \psi t$ . By the mgu property we know that  $\psi \text{RCD} (l : \sigma, \rho) = \psi t$ , and we're done.  $\square$

Similarly to the case for paths, we need that

**Lemma 3** (Type checking respects substitution).  $\forall m \Gamma \phi, \Gamma \vdash m \rightarrow \phi\Gamma \vdash m$ .

*Proof.* By induction on  $m$ .

- The TRUE case is trivial, as any context will work with the TRUE rule.
- For the  $m_1$  AND  $m_2$  case, we have two inductive hypotheses

$$\forall \Gamma \phi, \Gamma \vdash m_1 \rightarrow \phi\Gamma \vdash m_1$$

and

$$\forall \Gamma \phi, \Gamma \vdash m_2 \rightarrow \phi\Gamma \vdash m_2$$

Introduce  $\Gamma$  and  $\phi$  and assume that  $\Gamma \vdash m_1$  AND  $m_2$ . We want to prove that  $\phi\Gamma \vdash m_1$  AND  $m_2$ . By inversion, we have that  $\Gamma \vdash m_1$  and  $\Gamma \vdash m_2$ . Apply these with the inductive hypothesis gives  $\phi\Gamma \vdash m_1$  and  $\phi\Gamma \vdash m_2$ , and the result follows from the AND rule.

- For the  $p_1 \text{ EQ } p_2$  case, we have no inductive hypothesis. Introduce  $\Gamma$  and  $\phi$  and assume that  $\Gamma \vdash p_1 \text{ EQ } p_2$ . By inversion, there is some  $a$  such that  $\Gamma \vdash p_1 :: \text{ATOMIC } a$  and  $\Gamma \vdash p_2 :: \text{ATOMIC } a$ . Because path checking respects substitution, we have that  $\phi\Gamma \vdash p_1 :: \text{ATOMIC } \phi a$  and  $\phi\Gamma \vdash p_2 :: \text{ATOMIC } \phi a$ . We can then apply the EQ rule using  $\phi a$  and  $\phi\Gamma$  to obtain our goal that  $\phi\Gamma \vdash p_1 \text{ EQ } p_2$ .
- For the  $v \text{ IN } p.m$  case, the inductive hypothesis is that

$$\forall \Gamma \phi, \Gamma \vdash m \rightarrow \phi\Gamma \vdash m$$

Introduce  $\Gamma$  and  $\phi$  and assume that  $\Gamma \vdash v \text{ IN } p.m$ . We want to prove that  $\phi\Gamma \vdash v \text{ IN } p.m$ . By inversion we know that there is some  $r$  such that  $\Gamma \vdash p :: \text{SETRCD } r$  and  $(v, \text{RCD } r); \Gamma \vdash m$ . Apply the inductive hypothesis to get  $\phi(v, \text{RCD } r); \Gamma \vdash m$ . By the SETRCD-ELIM rule, we thus just need to obtain  $\phi\Gamma \vdash p :: \phi\text{SETRCD } r$ , which follows from the soundness of path checking.

- For the  $v \text{ OF } l \text{ FROM } p.m$  case, the inductive hypothesis is that

$$\forall \Gamma \phi, \Gamma \vdash m \rightarrow \phi\Gamma \vdash m$$

Introduce  $\Gamma$  and  $\phi$  and assume that  $\Gamma \vdash v \text{ OF } l \text{ FROM } p.m$ . We want to prove that  $\phi\Gamma \vdash v \text{ OF } l \text{ FROM } p.m$ . By inversion we know that there is some  $r$  and  $t$  such that  $\Gamma \vdash p :: \text{SETCHC } (l : t, r)$  and  $(v, t); \Gamma \vdash m$ . Apply the inductive hypothesis to get  $\phi(v, t); \Gamma \vdash m$ . By the SETRCD-ELIM rule, we thus just need to obtain  $\phi\Gamma \vdash p :: \phi\text{SETCHC } (l : t, r)$ , which follows from the soundness of path checking.

□

The main result:

**Lemma 4** (Soundness of Type Inference).  $\forall m \Gamma \varphi, \varphi\Gamma \Vdash m \rightarrow \varphi\Gamma \vdash m$ .

*Proof.* The proof is by induction on  $m$ .

- The TRUE case is immediate because the checking and inference rules are the same; the identity substitution is returned from the inference algorithm.
- For the  $m \text{ AND } m'$  case, we are given two inductive hypotheses,  $\forall \Gamma \varphi, \varphi\Gamma \Vdash m \rightarrow \varphi\Gamma \vdash m$  and  $\forall \Gamma \varphi, \varphi\Gamma \Vdash m' \rightarrow \varphi\Gamma \vdash m'$ . We must show that  $\forall \Gamma \varphi, \varphi\Gamma \Vdash m \text{ AND } m' \rightarrow \varphi\Gamma \vdash m \text{ AND } m'$ . Introduce  $\Gamma$  and  $\varphi$  and assume that  $\varphi\Gamma \Vdash m \text{ AND } m'$ .

By inversion, we know that there exists  $\phi_1$  and  $\phi_2$  such that  $\varphi = \phi_2 \circ \phi_1$  and  $\phi_1\Gamma \Vdash m$  and  $\phi_2\phi_1\Gamma \Vdash m'$ . Substitution yields a new goal of  $\phi_2\phi_1\Gamma \vdash m \text{ AND } m'$ . We can apply the AND-CHECK rule with  $\phi_2\phi_1\Gamma$  as its  $\Gamma$ . This yields the two goals of  $\phi_2\phi_1\Gamma \vdash m$  and  $\phi_2\phi_1\Gamma \vdash m'$ .

To solve the first goal, by the fact that typechecking respects substitution (Lemma 3), we need to only solve  $\phi_1\Gamma \vdash m$ . We can then apply the inductive hypothesis using  $\Gamma$  and  $\phi_1$  to get a new goal of  $\phi_1\Gamma \Vdash m$ , we we assumed.

To solve the second goal, we apply the inductive hypothesis using  $\Gamma$  and  $\phi_2 \circ \phi_1$  to get a new goal of  $\phi_2\phi_1\Gamma \Vdash m'$ , which we assumed.

- For the EQ case, we have no inductive hypotheses. Assume that  $\varphi\Gamma \Vdash p \text{ EQ } p'$ ; we must show that  $\varphi\Gamma \vdash p \text{ EQ } p'$ . By inversion we know that there exists  $\phi_4, \phi_3, \phi_2, \phi_1, t, t', \alpha$  such that  $\varphi = \phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1$  and  $\phi_1\Gamma \Vdash p : t$  and  $\phi_2\phi_1\Gamma \Vdash p' : t'$  and  $\phi_2 t \stackrel{\phi_3}{\approx} t'$  and  $\alpha$  fresh and  $\phi_3 t' \stackrel{\phi_4}{\approx} \text{ATOMIC } \alpha$ . Substituting gives us a new goal of  $\phi_4\phi_3\phi_2\phi_1\Gamma \vdash p \text{ EQ } p'$ .

Using the soundness of inference for paths (Lemma 2) with  $\phi_1\Gamma \Vdash p : t$  and  $\phi_2\phi_1\Gamma \Vdash p' : t'$  yields that  $\phi_1\Gamma \vdash p : t$  and  $\phi_2\phi_1\Gamma \vdash p' : t'$ . We can then repeatedly apply that path checking respects substitution (Lemma 1), to get (1)  $\phi_4\phi_3\phi_2\phi_1\Gamma \vdash p : \phi_4\phi_3\phi_2t$  and (2)  $\phi_4\phi_3\phi_2\phi_1\Gamma \vdash p' : \phi_4\phi_3t'$ .

The mgu property gives us that (a)  $\phi_4\phi_3t' = \text{ATOMIC } \phi_4\alpha$  and (b)  $\phi_3\phi_2t = \phi_3t'$ . Rewriting by (a) in (2) gives  $\phi_4\phi_3\phi_2\phi_1\Gamma \vdash p' : \text{ATOMIC } \phi_4\alpha$ . Rewriting by (b) in (1) gives  $\phi_4\phi_3\phi_2\phi_1\Gamma \vdash p : \phi_4\phi_3t'$  and then by (a) gives  $\phi_4\phi_3\phi_2\phi_1\Gamma \vdash p : \text{ATOMIC } \phi_4\alpha$ . We can thus apply rule EQ-CHECKS.

- For the IN case, the inductive hypothesis is that  $\forall\Gamma\varphi, \varphi\Gamma \Vdash m \rightarrow \varphi\Gamma \vdash m$ . We must show that  $\forall\Gamma\varphi, \varphi\Gamma \Vdash v \text{ IN } p.m \rightarrow \varphi\Gamma \vdash v \text{ IN } p.m$ . To that end, introduce  $\Gamma$  and  $\varphi$  and assume that  $\varphi\Gamma \Vdash v \text{ IN } p.m$ . By inversion, we know there exists  $\phi_3, \phi_2, \phi_1, t, \rho$  with  $\rho$  fresh such that  $\varphi = \phi_3 \circ \phi_2 \circ \phi_1$  and  $\phi_1\Gamma \Vdash p : t$  and  $\text{SETRCD } \rho \stackrel{\phi_2}{\approx} t$  and  $\phi_3\phi_2((v, \text{RCD } \rho); \phi_1\Gamma) \Vdash m$ . We can then apply the IN-CHECK rule, taking its  $\Gamma$  to be  $\phi_3\phi_2\phi_1\Gamma$  and its  $r$  to be  $\phi_3\phi_2\rho$ . We are thus left with two new goals:  $\phi_3\phi_2\phi_1\Gamma \vdash p : \text{SETRCD } \phi_3\phi_2\rho$  and  $(v, \text{RCD } \phi_3\phi_2\rho); \phi_3\phi_2\phi_1\Gamma \vdash m$ .

To solve the first, note that we have  $\phi_1\Gamma \Vdash p : t$  and thus by soundness of path inference we have that  $\phi_1\Gamma \vdash p : t$ , and thus also that (1)  $\phi_2\phi_1\Gamma \vdash p : \phi_2t$  by Lemma todo. By the mgu property, we have  $\text{SETRCD } \phi_2\rho = \phi_2t$ , and so we can rewrite (1) to obtain  $\phi_2\phi_1\Gamma \vdash p : \text{SETRCD } \phi_2\rho$ , and then apply todo with  $\phi_3$  to get  $\phi_3\phi_2\phi_1\Gamma \vdash p : \text{SETRCD } \phi_3\phi_2\rho$ , as required.

To solve the second, apply the inductive hypothesis with its  $\Gamma$  as  $(v, \text{SETRCD } \rho); \phi_1\Gamma$  and its  $\varphi$  as  $\phi_3 \circ \phi_2$  to yield a new goal of  $(v, \text{RCD } \phi_3\phi_2\rho); \phi_3\phi_2\phi_1\Gamma \Vdash m$ , which we already assumed.

- For the OF case, the inductive hypothesis is that  $\forall\Gamma\varphi, \varphi\Gamma \Vdash m \rightarrow \varphi\Gamma \vdash m$ . We must show that  $\forall\Gamma\varphi, \varphi\Gamma \Vdash v \text{ OF } l \text{ FROM } p.m \rightarrow \varphi\Gamma \vdash v \text{ IN } p.m$ . To that end, introduce  $\Gamma$  and  $\varphi$  and assume that  $\varphi\Gamma \Vdash v \text{ OF } l \text{ FROM } p.m$ . By inversion, we know there exists  $\phi_3, \phi_2, \phi_1, t, \rho, \sigma$  with  $\rho, \sigma$  fresh such that  $\varphi = \phi_3 \circ \phi_2 \circ \phi_1$  and  $\phi_1\Gamma \Vdash p : t$  and  $\text{SETCHC } (l : \sigma, \rho) \stackrel{\phi_2}{\approx} t$  and  $\phi_3\phi_2((v, \sigma); \phi_1\Gamma) \Vdash m$ . We can then apply the OF-CHECK rule, taking its  $\Gamma$  to be  $\phi_3\phi_2\phi_1\Gamma$  and its  $r$  to be  $\phi_3\phi_2\rho$  and its  $t$  to be  $\phi_3\phi_2\sigma$ . We are thus left with two new goals:  $\phi_3\phi_2\phi_1\Gamma \vdash p : \text{SETCHC } (l : \phi_3\phi_2\sigma, \phi_3\phi_2\rho)$  and  $(v, \text{RCD } \phi_3\phi_2\rho); \phi_3\phi_2\phi_1\Gamma \vdash m$ .

To solve the first, note that we have  $\phi_1\Gamma \Vdash p : t$  and thus by soundness of path inference we have that  $\phi_1\Gamma \vdash p : t$ , and thus also that (1)  $\phi_2\phi_1\Gamma \vdash p : \phi_2t$  because typing respect substitution. By the mgu property, we have  $\text{SETCHC } (l : \phi_2\sigma, \phi_2\rho) = \phi_2t$ , and so we can rewrite (1) to obtain  $\phi_2\phi_1\Gamma \vdash p : \text{SETCHC } (l : \phi_2\sigma, \phi_2\rho)$ , and then apply todo with  $\phi_3$  to get  $\phi_3\phi_2\phi_1\Gamma \vdash p : \text{SETCHC } (l : \phi_3\phi_2\sigma, \phi_3\phi_2\rho)$ , as required.

To solve the second, apply the inductive hypothesis with its  $\Gamma$  as  $(v, \sigma); \phi_1\Gamma$  and its  $\varphi$  as  $\phi_3 \circ \phi_2$  to yield a new goal of  $(v, \phi_3\phi_2\sigma); \phi_3\phi_2\phi_1\Gamma \Vdash m$ , which we already assumed.

□

## Theorem 2 - Completeness

Intuitively, completeness (and soundness) holds because we are simply doing iterated unification in a way similar to Hindley-Milner or the Jones system.

We start by establishing the completeness of path inference over path checking. Let  $fv(\Gamma)$  indicate the union of the type, row, and atomic variables in the types in the range of context  $\Gamma$ .

**Lemma 5** (Completeness of Path Inference).

$$\forall \Gamma \ p \ \tau \ \varphi, \ \varphi\Gamma \vdash p : \tau \rightarrow \exists S \ T \ s, \ S\Gamma \Vdash p : T \wedge \tau = sT \wedge \forall v \in fv(\Gamma), \varphi v = sSv$$

*Proof.* Introduce  $\Gamma$  and proceed by induction on  $p$ .

The base case is that  $p$  is a variable  $v$ , and we are to prove that

$$\forall \tau \ \varphi, \ \varphi\Gamma \vdash v : \tau \rightarrow \exists S \ T \ s, \ S\Gamma \Vdash v : T \wedge \tau = sT \wedge \forall v \in fv(\Gamma), \varphi v = sSv$$

Introduce  $\varphi$  and  $\tau$  and assume  $\varphi\Gamma \vdash v : \tau$ . By inversion, there exists some  $T$  such that  $(v, T) \in \Gamma$ ; witness  $S$  with the identity substitution,  $s$  with  $\varphi$ , and  $T$  with  $T$ . The goal is then

$$\Gamma \Vdash v : T \wedge \tau = \varphi T \wedge \forall v \in fv(\Gamma), \varphi v = \varphi v$$

The leftmost conjunct follows from the VAR-INF rule and that we have  $(v, T) \in \Gamma$ . The rightmost conjunct is trivial. The middle conjunct,  $\tau = \varphi T$ , follows because we have  $(v, T) \in \Gamma$  and hence that  $(v, \varphi T) \in \varphi\Gamma$ . Because  $\varphi\Gamma \vdash v : \tau$ , it must be the case that  $\tau = \varphi T$ .

The inductive case is that  $p$  is a projection  $p.l$ , and we are to prove that

$$\forall \tau \varphi, \ \varphi\Gamma \vdash p.l : \tau \rightarrow \exists S \ T \ s, \ S\Gamma \Vdash p.l : T \wedge \tau = sT \wedge \forall v \in fv(\Gamma), \varphi v = sSv$$

The inductive hypothesis is that

$$\forall \tau \varphi, \ \varphi\Gamma \vdash p : \tau \rightarrow \exists S \ T \ s, \ S\Gamma \Vdash p : T \wedge \tau = sT \wedge \forall v \in fv(\Gamma), \varphi v = sSv$$

Introduce  $\varphi$  and  $\tau$  and assume that  $\varphi\Gamma \vdash p.l : \tau$ . By inversion, we know that there exists  $r$  such that  $\varphi\Gamma \vdash p : \text{RCD} \langle l : \tau, r \rangle$ . We can then apply the inductive hypothesis taking its  $\tau$  to be  $\text{RCD} \langle l : \tau, r \rangle$  and its  $\varphi$  to be  $\varphi$ , to obtain that there exists  $S$  and  $T$  and  $s$  such that

$$S\Gamma \Vdash p : T \wedge \text{RCD} \langle l : \tau, r \rangle = sT \wedge \forall v \in fv(\Gamma), \varphi v = sSv \tag{1}$$

Doing some renaming of bound variables to avoid confusion, the goal is

$$\exists S' \ T' \ s', \ S'\Gamma \Vdash p.l : T' \wedge \tau = s'T' \wedge \forall v \in fv(\Gamma), \varphi v = s'S'v$$

We can choose fresh  $\rho, \sigma \notin fv(\Gamma)$  and we can specialize the goal's existentials, taking  $S' = \psi \circ \phi$  and  $T' = \psi\sigma$  to obtain a goal of

$$\exists \psi \ \phi \ s', \ \psi\phi\Gamma \Vdash p.l : \psi\sigma \wedge s', \tau = s'\psi\sigma \wedge \forall v \in fv(\Gamma), \varphi v = s'\psi\phi v$$

By the RCD-ELIM-INF rule, and because  $\rho, \sigma$  are fresh this goal becomes

$$\exists \psi \ \phi \ s' \ t, \ \phi\Gamma \Vdash p : t \wedge \text{RCD} \langle l : \sigma, \rho \rangle \stackrel{\psi}{\sim} t \wedge \tau = s'\psi\sigma \wedge \forall v \in fv(\Gamma), \varphi v = s'\psi\phi v$$

We can then instantiate  $t = T$  and  $\phi = S$ , to obtain a goal of

$$\exists \psi \ s', \ S\Gamma \Vdash p : T \wedge \text{RCD} \langle l : \sigma, \rho \rangle \stackrel{\psi}{\sim} T \wedge \tau = s'\psi\sigma \wedge \forall v \in fv(\Gamma), \varphi v = s'\psi S v$$

The leftmost conjunct we have in (1), and we can substitute  $\varphi = s \circ S$  by (1) to obtain a goal of

$$\exists \psi s', \text{RCD} \langle l : \sigma, \rho \rangle \stackrel{\psi}{\sim} T \wedge \tau = s' \psi \sigma \wedge \forall v \in fv(\Gamma), sSv = s' \psi Sv \quad (2)$$

From (1), we know that  $\text{RCD} \langle l : \tau, r \rangle = sT$ , which means that  $T$  must have one of three forms:

- $T = u$ , for some type variable  $u$ .
- $T = \text{RCD} \langle r', u \rangle$ , for some row variable  $u$  and some row  $r'$  that does not contain label  $l$ .
- $T = \text{RCD} \langle l : \tau', r' \rangle$  for some row  $r'$  that does not contain label  $l$ .

In all cases, we can compute a most general unifying substitution  $\psi$ , which means that

$$\text{RCD} \langle l : \psi \sigma, \psi \rho \rangle = \psi T \wedge \forall U', U' T = \text{RCD} \langle l : U' \sigma, U' \rho \rangle \rightarrow \exists s', U' = s' \circ \psi \quad (3)$$

Instantiating (3) with  $U' = (\sigma \mapsto \tau, \rho \mapsto r) \circ s$  gives

$$\begin{aligned} ((\sigma \mapsto \tau, \rho \mapsto r) \circ s) T &= \text{RCD} \langle l : ((\sigma \mapsto \tau, \rho \mapsto r) \circ s) \sigma, ((\sigma \mapsto \tau, \rho \mapsto r) \circ s) \rho \rangle \\ &\rightarrow \exists s', ((\sigma \mapsto \tau, \rho \mapsto r) \circ s) = s' \circ \psi \end{aligned}$$

Because  $\sigma, \rho$  do not appear in  $T$  (or  $s$ ), this becomes

$$sT = \text{RCD} \langle l : \tau, r \rangle \rightarrow \exists s', (\sigma \mapsto \tau, \rho \mapsto r) \circ s = s' \circ \psi$$

We have the antecedent in (1), so we know that there exists some  $s'$  such that

$$(\sigma \mapsto \tau, \rho \mapsto r) \circ s = s' \circ \psi \quad (4)$$

Witnessing our goal (2) with  $s'$  and  $\psi$  gives a new goal of

$$\text{RCD} \langle l : \sigma, \rho \rangle \stackrel{\psi}{\sim} T \wedge \tau = s' \psi \sigma \wedge \forall v \in fv(\Gamma), sSv = s' \psi Sv$$

The leftmost of these is solved by (3). Since we are only interested in the free variables in  $\Gamma$ , and  $\sigma, \rho$  do not appear in  $\Gamma$ , the goal is the same as

$$\tau = s' \psi \sigma \wedge \forall v \in fv(\Gamma), ((\sigma \mapsto \tau, \rho \mapsto r) \circ s) Sv = s' \psi Sv$$

Rewrite the goal by (4) to get

$$\tau = s' \psi \sigma \wedge \forall v \in fv(\Gamma), s' \psi Sv = s' \psi Sv$$

Now the rightmost conjunct is trivial, leaving the goal as

$$\tau = s' \psi \sigma$$

We can un-rewrite the goal by (4) to get a new goal

$$\tau = (\sigma \mapsto \tau, \rho \mapsto r) \circ s \sigma$$

Because  $\sigma$  is fresh for  $s$ ,  $s$  has no effect on  $\sigma$  and this reduces to  $\tau = \tau$ , and we're done. □

The main theorem is

$$\forall m \Gamma \varphi, \varphi \Gamma \vdash m \rightarrow \exists S s, S \Gamma \Vdash m \wedge \forall v \in fv(\Gamma), \varphi v = sSv$$

*Proof.* By induction on  $m$ .

- For the case where  $m = \text{TRUE}$ , witness  $S$  as the identity substitution and  $s$  as  $\varphi$ . This yields a goal of  $\Gamma \Vdash \text{TRUE} \wedge \varphi v = \varphi v$ , which is trivially true.
- For the  $m_1 \text{ AND } m_2$  case, we have two inductive hypotheses:

$$\forall \Gamma \varphi, \varphi \Gamma \vdash m_1 \rightarrow \exists S s, S \Gamma \Vdash m_1 \wedge \forall v \in fv(\Gamma), \varphi v = s S v$$

and

$$\forall \Gamma \varphi, \varphi \Gamma \vdash m_2 \rightarrow \exists S s, S \Gamma \Vdash m_2 \wedge \forall v \in fv(\Gamma), \varphi v = s S v$$

Introduce  $\Gamma$  and  $\varphi$  and assume that  $\varphi \Gamma \vdash m_1 \text{ AND } m_2$ . By inversion, we know that  $\varphi \Gamma \vdash m_1$  and that  $\varphi \Gamma \vdash m_2$ . We want to prove that

$$\exists S s, S \Gamma \Vdash m_1 \text{ AND } m_2 \wedge \forall v \in fv(\Gamma), \varphi v = s S v$$

Applying the AND-INF rule gives a new goal of

$$\exists \phi_1 \phi_2 s, \phi_2 \phi_1 \Gamma \Vdash m_1 \text{ AND } m_2 \wedge \forall v \in fv(\Gamma), \varphi v = s \phi_2 \phi_1 v$$

and thus a new goal of

$$\exists \phi_1 \phi_2 s, \phi_1 \Gamma \Vdash m_1 \wedge \phi_2 \phi_1 \Gamma \Vdash m_2 \wedge \forall v \in fv(\Gamma), \varphi v = s \phi_2 \phi_1 v$$

Applying the first inductive hypothesis gives us a  $\phi_1$  and  $s_1$  such that  $\phi_1 \Gamma \Vdash m_1 \wedge \forall v \in fv(\Gamma), \varphi v = s_1 \phi_1 v$ . We can then witness the goal and substitute for  $\varphi$  to get a new goal of

$$\exists \phi_2 s, \phi_2 \phi_1 \Gamma \Vdash m_2 \wedge \forall v \in fv(\Gamma), s_1 \phi_1 v = s \phi_2 \phi_1 v$$

Substituting for  $\varphi$  gives us  $s_1 \phi_1 \Gamma \vdash m_2$ . We can then apply the second inductive hypothesis (choosing its  $\Gamma$  to be  $\phi_1 \Gamma$  and its  $\varphi$  to be  $s_1$ , to obtain that there exists some  $\phi_2$  and  $s_2$  such that  $\phi_2 \phi_1 \Gamma \Vdash m_2 \wedge \forall v \in fv(\Gamma), s_1 v = s_2 \phi_2 v$ . The goal's leftmost conjunct is then immediate, and the rightmost conjunct follows by inserting  $\phi_1$  on both sides of the substitution equality.

- For the  $p_1 \text{ EQ } p_2$  case, we have no inductive hypothesis. Introduce  $\Gamma$  and  $\varphi$  and assume that  $\varphi \Gamma \vdash p_1 \text{ EQ } p_2$ . By inversion we know that there is some  $a$  such that  $\varphi \Gamma \vdash p_1 :: \text{ATOMIC } a$  and  $\varphi \Gamma \vdash p_2 :: \text{ATOMIC } a$ . We want to prove that

$$\exists S s, S \Vdash p_1 \text{ EQ } p_2 \wedge \forall v \in fv(\Gamma), \varphi v = s S v$$

Applying the EQ-INF rule gives a new goal of

$$\exists \phi_1 \phi_2 \phi_3 \phi_4 s, \phi_4 \phi_3 \phi_2 \phi_1 \Vdash p_1 \text{ EQ } p_2 \wedge \forall v \in fv(\Gamma), \varphi v = s \phi_4 \phi_3 \phi_2 \phi_1 v$$

and then by choosing a fresh  $\alpha$ , a new goal of

$$\exists \phi_1 \phi_2 \phi_3 \phi_4 s t_1 t_2, \forall v \in fv(\Gamma), \varphi v = s \phi_4 \phi_3 \phi_2 \phi_1 v \wedge$$

$$\phi_1 \Gamma \Vdash p_1 :: t_1 \wedge \phi_2 \phi_1 \Gamma \Vdash p_2 :: t_2 \wedge \phi_2 t_1 \stackrel{\phi_3}{\approx} t_2 \wedge \phi_3 t_2 \stackrel{\phi_4}{\approx} \text{ATOMIC } \alpha$$

From completeness of path inference, we know that there exists some  $\phi_1$  and some  $t_1$  and some  $s_1$  such that  $\phi_1 \Gamma \Vdash p_1 :: t_1 \wedge \text{ATOMIC } a = s_1 t_1 \wedge \forall v \in fv(\Gamma), \varphi v = s_1 \phi_1 v$ . Witnessing the goal and substituting for  $\varphi$  gives the new goal

$$\exists \phi_2 \phi_3 \phi_4 s t_2, \forall v \in fv(\Gamma), s_1 \phi_1 v = s \phi_4 \phi_3 \phi_2 \phi_1 v \wedge$$

$$\phi_2\phi_1\Gamma \vdash p_2 :: t_2 \wedge \phi_2t_1 \stackrel{\phi_3}{\sim} t_2 \wedge \phi_3t_2 \stackrel{\phi_4}{\sim} \text{ATOMIC } \alpha$$

Substituting for  $\varphi$  into  $\varphi\Gamma \vdash p_2 :: \text{ATOMIC } a$  gives that  $s_1\phi_1\Gamma \vdash p_2 :: \text{ATOMIC } a$ . Hence by completeness of path inference (taking its  $\Gamma$  to be  $\phi_1\Gamma$  and its  $\varphi$  to be  $s_1$ ) there is some  $\phi_2$  and  $t_2$  and  $s_2$  such that  $\phi_2\phi_1\Gamma \vdash p_2 :: t_2 \wedge \text{ATOMIC } a = s_2t_2 \wedge \forall v \in fv(\Gamma), s_1v = s_2\phi_2v$ . The goal now becomes

$$\begin{aligned} \exists \phi_3\phi_4s, \forall v \in fv(\Gamma), s_1\phi_1v = s\phi_4\phi_3\phi_2\phi_1v \wedge \\ \phi_2t_1 \stackrel{\phi_3}{\sim} t_2 \wedge \phi_3t_2 \stackrel{\phi_4}{\sim} \text{ATOMIC } \alpha \end{aligned}$$

We know the unifiers  $\phi_3$  and  $\phi_4$  must exist, because  $t_2$  is either a typevariable or itself atomic; moreover,  $\phi_3$  and  $\phi_4$  must simply send one atomic typevariable to another. As such, our goal is simply

$$\exists s, \forall v \in fv(\Gamma), s_1\phi_1v = s\phi_2\phi_1v$$

Using  $s_2$  does the trick; we get a goal of  $s_1\phi_1v = s_2\phi_2\phi_1v$ , which follows because  $s_1 = s_2 \circ \phi_2$

- The IN and OF cases are similar

□

## Theorem 4 - Mapping Subtyping

We want to prove that

$$\forall m \Gamma' \Gamma, \Gamma' \leq \Gamma \wedge \Gamma \vdash m \rightarrow \Gamma' \vdash m$$

First, an auxillary lemma about path checking.

**Lemma 6.**  $\forall p \Gamma' \Gamma t, \Gamma' \leq \Gamma \wedge \Gamma \vdash p :: t \rightarrow \exists t', \Gamma' \vdash p :: t' \wedge t' \leq t.$

*Proof.* Proceed by induction on  $p$ .

- The base case is that  $p$  is a variable  $v$ . Introduce  $\Gamma'$  and  $\Gamma$  and  $t$  and assume that  $\Gamma' \leq \Gamma$  and  $\Gamma \vdash v :: t$ . We want to show that  $\exists t', \Gamma' \vdash v :: t' \wedge t' \leq t$ . By inversion we know that  $(v, t) \in \Gamma$  and because  $\Gamma' \leq \Gamma$  that there exists some  $t'$  such that (A)  $t' \leq t$  and (B)  $(v, t') \in \Gamma'$ . Witness the goal with  $t'$ , so we must show that  $t' \leq t$  and  $\Gamma' \vdash v :: t'$ . The first of these we already have in (A), and the second follows by applying the VAR rule with (B).
- The inductive case is that  $p$  is a projection  $p.l$ . The inductive hypothesis is that

$$\forall \Gamma' \Gamma t, \Gamma' \leq \Gamma \wedge \Gamma \vdash p :: t \rightarrow \exists t', \Gamma' \vdash p :: t' \wedge t' \leq t$$

Introduce  $\Gamma'$  and  $\Gamma$  and  $t$ , and assume that  $\Gamma' \leq \Gamma$  and  $\Gamma \vdash p.l :: t$ . We are to prove that

$$\exists t', \Gamma' \vdash p.l :: t' \wedge t' \leq t$$

By inversion, we know that there exists  $r$  such that  $\Gamma \vdash p :: \text{RCD } (l : t, r)$ . Apply the inductive hypothesis to get that there exists  $t'$  such that

$$(C) \quad \Gamma' \vdash p :: t' \wedge t' \leq \text{RCD } (l : t, r)$$

From the right conjunct we know that there exists  $T$  and  $R$  such that

$$(D) \quad t' = \text{RCD } (l : T, R) \wedge T \leq t \wedge R \leq r$$

Witness our goal with  $T$  so that we need to prove

$$\Gamma' \vdash p.l :: T \wedge T \leq t$$

The rightmost conjunct follows from (D). We can rewrite (C) by (D) to obtain

$$\Gamma' \vdash p :: \text{RCD } (l : T, R)$$

from which our goal follows by the RCD-ELIM rule. □

Now the main result.

*Proof.* Proceed by induction on  $m$ .

- For the case where  $m = \text{TRUE}$ , note that  $\text{TRUE}$  typechecks in any context, and we're done.
- For the case where  $m = m_1 \text{ AND } m_2$ , we have two inductive hypothesis, that

$$\forall \Gamma' \Gamma, \Gamma' \leq \Gamma \wedge \Gamma \vdash m_1 \rightarrow \Gamma' \vdash m_1$$

and

$$\forall \Gamma' \Gamma, \Gamma' \leq \Gamma \wedge \Gamma \vdash m_2 \rightarrow \Gamma' \vdash m_2$$

Introduce  $\Gamma'$  and  $\Gamma$  and assume that  $\Gamma' \leq \Gamma$  and  $\Gamma \vdash m_1 \text{ AND } m_2$ . We want to prove that

$$\Gamma' \vdash m_1 \text{ AND } m_2$$

By inversion we know that  $\Gamma \vdash m_1$  and  $\Gamma \vdash m_2$ , which we can use with the inductive hypothesis to obtain that  $\Gamma' \vdash m_1$  and  $\Gamma' \vdash m_2$ . Our goal follows by the AND rule.

- For the case where  $m = p_1 \text{ EQ } p_2$ , we have no inductive hypothesis. Introduce  $\Gamma'$  and  $\Gamma$  and assume that  $\Gamma' \leq \Gamma$  and  $\Gamma \vdash p_1 \text{ EQ } p_2$ . We want to prove that

$$\Gamma' \vdash p_1 \text{ EQ } p_2$$

By inversion we know that there is some  $a$  such that  $\Gamma \vdash p_1 :: \text{ATOMIC } a$  and  $\Gamma \vdash p_2 :: \text{ATOMIC } a$ . Using these with the lemma we just proved gives us that there exists  $T_1 \leq \text{ATOMIC } a$  and  $T_2 \leq \text{ATOMIC } a$  such that  $\Gamma' \vdash p_1 :: T_1$  and  $\Gamma' \vdash p_2 :: T_2$ . However,  $T_1$  and  $T_2$  must be equal to  $\text{ATOMIC } a$ , because reflexivity is the only subtyping rule that applies to atomic types. The goal then follows from the EQ rule.

- For the  $v \text{ IN } p. m$  case, the inductive hypothesis is

$$\forall \Gamma' \Gamma, \Gamma' \leq \Gamma \wedge \Gamma \vdash m \rightarrow \Gamma' \vdash m$$

Introduce  $\Gamma'$  and  $\Gamma$  and assume that  $\Gamma' \leq \Gamma$  and that  $\Gamma \vdash v \text{ IN } p. m$ . We want to show that

$$\Gamma' \vdash v \text{ IN } p. m$$

By inversion, we know that there is some  $r$  such that  $(v, \text{RCD } r); \Gamma \vdash m$  and  $\Gamma \vdash p :: \text{SETRCD } r$ . From the lemma we just proved we have that there is some  $R \preceq r$  such that (A)  $\Gamma' \vdash p :: \text{SETRCD } R$ , and so we can apply the inductive hypothesis taking its  $\Gamma$  to be  $(v, \text{RCD } r)$  and its  $\Gamma'$  to be  $(v, \text{RCD } R)$  to obtain

$$(B) \quad (v, \text{RCD } R); \Gamma' \vdash m$$

From which the result follows by the SETRCD-ELIM rule applied with (A) and (B).

- For the  $v \text{ OF } l \text{ FROM } p. m$  case, the inductive hypothesis is

$$\forall \Gamma' \Gamma, \Gamma' \leq \Gamma \wedge \Gamma \vdash m \rightarrow \Gamma' \vdash m$$

Introduce  $\Gamma'$  and  $\Gamma$  and assume that  $\Gamma' \leq \Gamma$  and that  $\Gamma \vdash v \text{ OF } l \text{ FROM } p. m$ . The goal is

$$\Gamma' \vdash v \text{ IN } p. m$$

By inversion, there is some  $t$  and  $r$  such that  $(v, t); \Gamma \vdash m$  and  $\Gamma \vdash p :: \text{SETCHC } (l : t, r)$ . From the lemma we just proved we have that there is some  $R \preceq r$  and  $T \leq t$  such that (A)  $\Gamma' \vdash p :: \text{SETCHC } (l : T, R)$ , and so we can apply the inductive hypothesis taking its  $\Gamma$  to be  $(v, t)$  and its  $\Gamma'$  to be  $(v, T)$  to obtain

$$(B) \quad (v, T); \Gamma' \vdash m$$

From which the result follows by the SETCHC-ELIM rule applied with (A) and (B).

□

### Proposition 3

Suppose we have that  $X' \leq X$  and  $I \in \llbracket X' \rrbracket$ . Intuitively, the derivation used for *erase* does not matter because the fields removed from the data instance are exactly those found in  $X'$  but not in  $X$  – which does not depend on the order in which we determine these fields.

*Proof.* More formally, let the rank of a schema be the maximum number of *Row* nestings, so that NR schema are thought of as trees. The proof is by induction on the rank of  $X'$ . Rank induction says that for a predicate  $P(X')$  of a schema  $X'$ , if we can establish  $P(X')$  for all schema of rank 0, and assuming that  $P(X')$  holds for all schema of rank less than  $n$ , that  $P(X')$  holds of all schema of rank  $n$ , then we may conclude that  $P(X')$  holds for all schema. The predicate we want to use is

$$P(X') = \forall X I (pf \ pf' : X' \leq X), \text{erase}(pf)(I) = \text{erase}(pf')(I)$$

For the base case, when the rank is zero,  $X'$  is an **ATOMIC**, and the subtyping derivations are both necessarily uses of reflexivity, which removes no data from  $I$ .

For the inductive step, suppose the rank of  $X'$  is  $n$  and the inductive hypothesis is that we've established the irrelevance of the subtyping derivation for all ranks less than  $n$ .  $X'$  is of the form  $Cr'$  for some row  $r'$ , and  $X$  has the form  $Cr$  for some  $r$  such that  $r' \preceq r$ .

Note that for each  $(l : t) \in r', r$ , that the rank of  $t$  is less than  $n$ . The effect of  $\text{erase}(pf)$  and  $\text{erase}(pf')$  must necessarily consist of a sequence of width-removals and depth-removals to the labels in  $I$ . For each  $l$  that occurs in  $X'$  but not in  $X$ , the final erasure to apply to that label must be width; hence, we just need to guarantee that the order in which the depth erasures were applied to labels that do appear in the output does not matter. In this case, it must be the case that the final erasure to apply to a label erases into the label's type in  $X$ ; as such, the inductive hypothesis guarantees that the particular derivation used to do this erasure does not matter. □