

Use-Based Inference of Reference Polymorphism

Dave King and John Hannan

Pennsylvania State University, University Park PA 16802, USA

Abstract

Polymorphism is an important language feature, allowing generic code to operate on many different types. However, adding mutable state to a language places many constraints on how polymorphism interacts with references. Current solutions either restrict references to be used monomorphically or require explicit type declaration and thus limit the polymorphism of data retrieved from the store. We present a type system and type inference algorithm which allow the safe polymorphic use of data in a mutable store, as well as functions which take and return such references to the store. We refer to this language feature as *reference polymorphism*.

1 INTRODUCTION

Combining polymorphism with mutable state is a long-studied problem. Polymorphic code can be written generically once and then later instantiated to operate on many different types. A mutable store allows programmers to create references to values that can change throughout the execution of the program. However, the naive way to extend polymorphism to operate on reference values violates language soundness. Existing solutions restrict polymorphism on data retrieved from the store in various ways.

The first class of solutions treats references to polymorphic objects as second-class types; they allow creating a reference to a polymorphic value, but they do not allow retrieving a polymorphic value from that reference [12, 7, 14, 15, 11, 5]. The goal in this line of research has been on restoring soundness to the language rather than enabling polymorphism of the store. Wright's value restriction, which limits polymorphism to syntactic values, is a critical key in many of these systems.

The second class of solutions allows the creation of references to polymorphic values that can be used polymorphically, but require explicitly declaring the type of the reference [3]. For larger values and complicated objects, this can be prohibitive for the programmer. Additionally, the presence of the value restriction limits how polymorphic data retrieved from the store can be used. Using semi-explicit polymorphism, it is not always possible to restore polymorphism to data retrieved from the store using η -expansion. This is the case when the data retrieved is a record of methods, as in functional object encodings.

Both classes of solutions restrict references to polymorphic values in severe ways. Either we cannot treat the data in the store as polymorphic at all, or we cannot treat the reference to that polymorphic data like any other data. By focusing on inferring the types of references and adding polymorphism to the imperative subset of the language, we can gain more expressivity. In this paper, we present a type system that allows references to polymorphic values to be created, dereferenced, and updated. Polymorphic behavior of references is inferred based on how they are used; specifically, the types of values stored and retrieved in the store. Finally, we present an algorithm for automatic type inference.

Throughout the paper, we refer to a reference that can be used polymorphically as a *polymorphic reference*; this is distinct from past systems combining polymorphism with references, where the terminology

is used to describe references to polymorphic values that become monomorphic when accessed through the reference.

1.1 Motivating Examples

We will work within the framework of the ML language. Each of the existing solutions combining references and polymorphism does not completely infer the behavior of code creating and dereferencing polymorphic references. Consider the following code in a dialect of ML extended with operations for creating, dereferencing, and assigning polymorphic references.

```
let val rK = polyref (fn x => fn y => x)
in
  (polyderef rK) 3 4;
  (polyderef rK) true false;
  polyasgn rK := (fn x => fn y => y);
  (polyderef rK) [1, 3, 7] [4, 8, 6];
  (polyderef rK) (fn n => n * n) (fn n => n + 1)
end
```

The above code gives a very simple example of how a reference can be safely used polymorphically in our system. Here, the polymorphic K -combinator is stored in a cell that is bound to the variable rK . We use the syntax `polyref` to indicate that the value being created is a polymorphic reference; similarly, polymorphic assignment (`polyasgn`) and polymorphic dereference (`polyderef`) operators here work on polymorphic references¹. Next, we use rK at the types $int \rightarrow int \rightarrow int$ and $bool \rightarrow bool \rightarrow bool$. Later, rK is updated to a combinator that takes two arguments and returns the second. The value stored in the reference is then again used at the two types:

$$int\ list \rightarrow int\ list \rightarrow int\ list$$

$$(int \rightarrow int) \rightarrow (int \rightarrow int) \rightarrow (int \rightarrow int)$$

We can see that the reference which rK is bound to is used in a safe way: at any time, the types of values which are retrieved from it using dereference are instances of the types stored in it through assignment or reference creation. As our inference procedure keeps track of what polymorphic types are stored in and retrieved from reference cells, we refer to it as *use-based*.

In order to check the safety of code using polymorphic references, we check the spots at which the reference is used, separating each into two categories: storage and retrieval. If all of the types of values stored in a reference can be instantiated to all of the types of values retrieved from a reference, then that reference is used consistently. For example, if we store a value with most general type $\forall \alpha. \alpha \rightarrow \alpha$, it is consistent to later retrieve a value of type $int \rightarrow int$ from it. On the other hand, if we were to store a value of type $bool \rightarrow bool$ in a reference, it would not be consistent to later retrieve a value of type $int \rightarrow int$ from that same reference. Using this intuition, we can see that the reference rK is used consistently in the above code.

In order to keep track of polymorphic references, we type them to the distinct type $M\ pref$. Here, M is an alias by which we can keep track of how that reference is used [10]. We use capital letters M, N, S, T, \dots to indicate distinct aliases. There is a 1-1 correspondence between reference statements and aliases; each

¹If a language supports both traditional ML references and polymorphic references, it is important each have their own creation, assignment, and dereference primitives. Otherwise, the type inference algorithm is no longer syntax-directed.

new reference statement is represented by a different alias. In the above example, let rK be assigned type $M \text{ pref}$.

The way that code uses a reference with an alias type is given by constraints c . There are two types of constraints in our system: storage constraints, written $\sigma \succ M$, and retrieval constraints, written $M \succ \tau$. A storage constraint $\sigma \succ M$ indicates that a value with polymorphic type σ is stored in the reference cell with type $M \text{ pref}$. A retrieval constraint $M \succ \tau$ indicates that a value with type τ has been retrieved from a reference with type $M \text{ pref}$. The constraints on the above code are thus:

$$C = \left\{ \begin{array}{l} M \succ int \rightarrow int \rightarrow int, M \succ bool \rightarrow bool \rightarrow bool, \\ M \succ int \text{ list} \rightarrow int \text{ list} \rightarrow int \text{ list}, \\ M \succ (int \rightarrow int) \rightarrow (int \rightarrow int) \rightarrow (int \rightarrow int) \\ \forall \alpha \beta. \alpha \rightarrow \beta \rightarrow \alpha \succ M, \forall \alpha \beta. \alpha \rightarrow \beta \rightarrow \beta \succ M \end{array} \right\}$$

One can verify that each of the storage constraints on M instantiates to each of the retrieval constraints on M . If, for example, C included the constraint $int \text{ list} \succ M$, indicating that an $int \text{ list}$ was stored inside a reference with the alias M , then C would be inconsistent since we also retrieve a functional type from M . The above example would thus not type correctly. Formal definitions for the above are given in Section 2.

1.2 Alias Polymorphism

Functions which take and return references require some additional machinery in order to incorporate polymorphism over aliases. For example, consider the following code:

```
let val choose_first = polyref (fn x => fn y => x)
    val choose_second = polyref (fn x => fn y => y)
    val do_choose = fn (r, x, y) => (polyderef r) x y
in
  do_choose (choose_first, 3, 4);
  do_choose (choose_second, true, false)
end
```

In order for the function `do_choose` to be fully polymorphic, it should be able to accept either `choose_first` or `choose_second` as arguments. Therefore, we must have some way to abstract away the alias in the type of the function `do_choose`, similar to the abstraction of type variables in traditional ML let-polymorphism.

Suppose the variable r is assigned type $N \text{ pref}$. Before generalization, the function `do_choose` has the most general type: $N \text{ pref} \rightarrow \alpha \rightarrow \alpha \rightarrow \alpha$. This type is read as follows: given the reference to a polymorphic function to choose between two values of the same type, we return one of those values.

However, we cannot generalize α without also generalizing the alias N used for typing `do_choose`. In order to generalize the alias in this function, we use a form of bounded polymorphism which places the constraints on a polymorphic reference in the polymorphic type of the function. The only constraint which the function `do_choose` requires is $N \succ \alpha \rightarrow \alpha \rightarrow \alpha$, i.e. all that `do_choose` does is retrieve a value of type $\alpha \rightarrow \alpha \rightarrow \alpha$ from N . We can only abstract N and generalize α if this constraint is discharged into the polymorphic type of the function. With this idea, the function `do_choose` is given the polymorphic type

$$\forall \alpha. \forall N. \{N \succ \alpha \rightarrow \alpha \rightarrow \alpha\}. N \text{ pref} \rightarrow \alpha \rightarrow \alpha \rightarrow \alpha$$

The above polymorphic type is read as follows: for all type variables β , aliases P , and constraint sets S such that S satisfies the constraint $P \succ \beta \rightarrow \beta \rightarrow \beta$, the function `do_choose` can take on the type $P \text{ pref} \rightarrow \beta \rightarrow \beta \rightarrow \beta$.

We encode constraint sets inside polytypes in this way in order to force the propagation of how polymorphic references are used inside polymorphic functions. For example, consider the following code:

```
let val f = fn x => polyref x
in
  f [1,2,3]; f true
end
```

At the time of generalization, `f` is assigned the polymorphic type $\forall \alpha \forall M: \{\alpha \succ M\}. \alpha \rightarrow M \text{ pref}$. In order for this polymorphic type to be instantiated to operate on an integer list and a boolean, there must be aliases O, P such that the top-level environment satisfies both $\text{int list} \succ O$ and $\text{bool} \succ P$.

To avoid inference complications, we restrict the types of values that can be stored and retrieved from polymorphic references to not contain polymorphic references of their own. So while we can create polymorphic references to the identity function and other first-class polymorphic functions, we cannot create a polymorphic reference to the polymorphic assignment function. This tradeoff simplifies the presentation of our system and prevents the system from allowing higher-order polymorphism through polymorphic references, while restricting the class of typeable programs. We believe that this encompasses most of the cases where reference polymorphism is needed.

1.3 Generalization of Values and Expressions

The value restriction is an important tool to ensure soundness in the presence of side effects. It should be no surprise that the value restriction is essential to soundness in our system as well. We are free to generalize both type variables and aliases in syntactic values, which do not cause side effects when evaluated. For other expressions, we can only generalize type variables not free in the context and constraint set; we are not allowed to generalize aliases. For example, consider the following code:

```
let a = (fn x => polyref x)
let r = a (fn y => y)
let c = (fn x => fn y => r) 3
in
  ...
end
```

In the above code, `a` is given the type $\forall \alpha. \forall M: \{\alpha \succ M\}. \alpha \rightarrow M \text{ pref}$. The variable `r` must be given a monomorphic type; otherwise, the body of the `let` expression could assign `r` to two different types, creating a heterogeneous reference at run-time. By restricting alias polymorphism to values, we avoid unsoundness. We thus give `r` the type $N \text{ pref}$ and introduce the constraint $N \succ \alpha$. Finally, because type variables not appearing free in the constraint set or context are generalizable, we can give `c` the polymorphic type $\forall \beta. \beta \rightarrow N \text{ pref}$.

1.4 Structure of the Paper

We present the formal semantics and typing rules for our system in Section 2. We outline an algorithm for automatic type inference in Section 3. Finally, we provide comments about this method's application

Locations	$l ::= l_0, l_1, \dots$
Variables	$x ::= x, y, z, \dots$
Values	$v ::= b \mid l \mid \lambda x. e \mid ()$
Expressions	$e ::= x \mid v \mid e_1 e_2 \mid$ $\text{polyref } e \mid \text{polyderef } e \mid \text{polyasgn } e_1 := e_2$

FIGURE 1. Syntax for ML extended with polymorphic references

Type Variables	$\alpha ::= \alpha, \beta, \dots$
Aliases	$M ::= M, N, S, T, \dots$
Types	$\tau ::= \alpha \mid \tau_1 \rightarrow \tau_2 \mid \text{unit} \mid M \text{ pref}$
Polymorphic Types	$\sigma ::= \tau \mid \forall \alpha. \sigma \mid \forall M: C. \sigma$
Constraints	$c ::= \sigma \succ M \mid M \succ \sigma$
Constraint Sets	$C ::= \{c_0, \dots, c_l\}$

FIGURE 2. Type Syntax

to other languages along with related and future work in Section 4.

2 TYPE SYSTEM

In this section we present the typing judgment for our system. We work within a fragment of the ML language containing let-polymorphism, function abstraction, application, and new operations to operate on references containing polymorphic values.

2.1 Abstract Syntax and Types

ML expressions and values are represented by e and v , respectively. The language syntax is given in Figure 1. Note that the expressions that manipulate references are designed to only work with references that can be used polymorphically. If traditional ML-style references are desired in the language, then to maintain the syntax-directed nature of ML inference, we need separate syntax for both sorts of references. How a reference is created thus affects how the type system treats it.

We next define the syntax for constraints, polytypes, and types. This is given in Figure 2. We use the notation $FV(\tau)$ to represent the free type variables appearing in a type. We overload this notation to represent the free variables of a polytype, a constraint, a constraint set, a context, and so on. Similarly, the notation $FA(\tau)$ represents the free aliases appearing in a type, $FA(\sigma)$ the free aliases in a polytype (that are not bound by a quantifier), and so on. In this paper, we concern ourselves with three kinds of substitutions: substitutions from type variables to types, given by φ ; substitutions from aliases to other aliases, given by ψ ; and type-and-alias substitutions, given by ϕ . When we apply an alias substitution to a constraint set, aliases are substituted both in the constraints and in the domain. For example, $[M \mapsto N](\{int \succ M\}) = \{int \succ N\}$.

In order to ensure that our constraints do not include polymorphic types that contain their own constraints, we introduce the notion of a simple type, a simple polytype, and a simple constraint. We say

$$\begin{array}{c}
\text{(IMPL-AX)} \quad \frac{c \in C}{C \supset c} \\
\text{(IMPL-TRANS)} \quad \frac{C \supset M \succ \sigma' \quad C \vdash \sigma \succ \sigma'}{C \supset M \succ \sigma} \\
\text{(INST-REFL)} \quad C \vdash \sigma \succ \sigma \\
\text{(INST-TRANS)} \quad \frac{C \vdash \sigma \succ \sigma_0 \quad C \vdash \sigma_0 \succ \sigma'}{C \vdash \sigma \succ \sigma'} \\
\text{(INST-TVAR)} \quad C \vdash \forall \alpha. \sigma \succ \sigma[\tau/\alpha] \\
\text{(INST-ALIAS)} \quad \frac{C \supset C_0[N/M]}{C \vdash \forall M: C_0. \sigma \succ \sigma[N/M]}
\end{array}$$

FIGURE 3. Instantiation and Implication Judgments

that a type is *simple* if it does not contain any types of the form $M \text{ pref}$. Similarly, a polymorphic type is simple if it contains no polytypes of form $\forall M:C. \sigma$ and its underlying type is simple. A constraint c is simple if the types and polymorphic types occurring in it are all simple. Finally, a constraint set C is simple if it contains only simple constraints. In this paper, we only consider simple constraint sets.

We have extended the notion of what a type and a polytype mean, so we must now extend the notion of what instantiation of a polytype to another polytype means. Constraints occurring inside a polymorphic types act as guards based on how they can be instantiated. Therefore, instantiation is done under a constraint context. The instantiation judgment $C \vdash \sigma \succ \sigma'$ has the following meaning: under constraint set C , polytype σ instantiates to polytype σ' . The intuition for this judgment is that constraints which are generalized in σ must be satisfied by C before instantiation. We simultaneously define the judgment $C \supset c$: under constraint set C , constraint c holds. We write $C \supset C_0$ to mean that for all $c' \in C_0$, $C \supset c'$; for every constraint $c \in C_0$, either $c \in C$ or c is implied by a more general constraint c' in C . These judgments are given in Figure 3.

In this paper, we consider an extended class of substitutions ϕ that map type variables to types and aliases to other aliases.

Lemma 2.1. *Let ϕ be a substitution on type variables and aliases.*

1. *If $C \vdash \sigma \succ \sigma'$ then $\phi(C) \vdash \phi(\sigma) \succ \phi(\sigma')$.*
2. *If $C \supset C_0$ then $\phi(C) \supset \phi(C_0)$.*

2.2 Consistency of Constraints

Type judgments for expressions are conducted under a set of constraints C , a variable context γ that maps variables to polytypes, and a location-alias mapping ρ that maps locations to aliases. It is important that the constraint set C be *consistent*: the storage and retrieval constraints must match up.

Definition 2.2 (Consistency). *A constraint set C is consistent if for all pairs of storage and retrieval constraints $(\sigma \succ M, M \succ \sigma')$ such that $C \supset \sigma \succ M$ and $C \supset M \succ \sigma'$, then $C \vdash \sigma \succ \sigma'$.*

Intuitively, C is consistent if the uses of C behave as if there is some polymorphic type σ_M for each alias M and everywhere that M is used, σ_M could have been used instead². A technical note is that we do not enforce consistency on the storage constraint level, i.e. the constraint set $\{int \succ M, bool \succ M\}$ is consistent under this definition. Type soundness is still preserved under this weaker consistency condition and it simplifies the task of type inference slightly.

We state two lemmas relating to consistent constraint sets that will be useful later.

Lemma 2.3 (Implication and Consistency). *Suppose C is consistent and $C \supset C_0$. Then C_0 is consistent.*

Lemma 2.4 (Constraint Set Consistency Under Substitutions). *Suppose C is consistent. Then if ϕ is a substitution on type variables, then $\phi(C)$ is consistent.*

Proof. This follows from invariance of instantiation under type substitutions. □

2.3 Type System and Theorems

We give the typing rules for our system in Figure 4. The location-label mapping ρ provides a map from locations l to aliases M , and allows us to type locations. The constraint set C and the location label map ρ fulfill the role of the typically monomorphic store typing λ from the literature.

The standard properties all apply to type judgments in our system: it is preserved under weakening, substitution on types and aliases, and substitution of values for variables.

Lemma 2.5 (Weakening). *Suppose $C; \gamma; \rho \vdash e : \sigma$. If $C \subseteq C'$ and $\gamma \subseteq \gamma'$, then $C'; \gamma'; \rho \vdash e : \sigma$.*

Lemma 2.6 (Type Substitution). *Let ϕ be a substitution on type variables and aliases and suppose $C; \gamma; \rho \vdash e : \sigma$. Then $\phi(C); \phi(\gamma); \phi(\rho) \vdash e : \phi(\sigma)$.*

Proof. This proof is straightforward. The only technical difficulty encountered is during the (TP-GEN-TVAR) and (TP-GEN-ALIAS) steps, where it is necessary to introduce a renaming substitution before applying the ϕ of the lemma. □

Lemma 2.7 (Value Substitution). *Suppose $C; \gamma[x:\sigma_0]; \rho \vdash e : \sigma$ and $C; \gamma; \rho \vdash v : \sigma_0$. Then $C; \gamma; \rho \vdash e[v/x] : \sigma$.*

2.4 Soundness

The operational semantics for polymorphic references are given in Figure 5; they are identical to the operational semantics for normal references. Here a store μ represents a mapping of locations l to values v , and the evaluation judgment $\mu \vdash e \hookrightarrow v, \mu'$ reads: “under store μ , expression e evaluates to value v with modified store μ' ”.

We now connect the type judgment for our language to the operational semantics. We first need a notion saying that a store and a constraint set are related: the facts in the constraint set should correspond to the values contained in the store. Soundness of any language with references depends on the type

²The reason that we choose a constraint-based approach rather than a direct mapping from aliases to polytypes is to ensure that functions such as $\lambda x. \text{polyderef } x$ have a principal type. It is not possible to know what a reference’s type should ultimately be at reference creation. Constraints avoid this difficulty by only keeping track of what types are put in or taken out of an alias.

(TP-UNIT)	$C; \gamma; \rho \vdash () : \text{unit}$
(TP-VAR)	$\frac{\gamma(x) = \sigma}{C; \gamma; \rho \vdash x : \sigma}$
(TP-LOC)	$\frac{l \in \text{dom}(\rho)}{C; \gamma; \rho \vdash l : \rho(l) \text{ pref}}$
(TP-LAM)	$\frac{C; \gamma[x : \tau_1]; \rho \vdash e : \tau_2}{C; \gamma; \rho \vdash \lambda x. e : \tau_1 \rightarrow \tau_2}$
(TP-APP)	$\frac{C; \gamma; \rho \vdash e_1 : \tau_1 \rightarrow \tau \quad C; \gamma; \rho \vdash e_2 : \tau_1}{C; \gamma; \rho \vdash e_1 e_2 : \tau}$
(TP-LET)	$\frac{C; \gamma; \rho \vdash e_1 : \sigma_1 \quad C; \gamma[x : \sigma_1]; \rho \vdash e_2 : \sigma}{C; \gamma; \rho \vdash \text{let } x = e_1 \text{ in } e_2 : \sigma}$
(TP-POLYREF)	$\frac{C; \gamma; \rho \vdash e : \sigma \quad C \supset \sigma \succ M}{C; \gamma; \rho \vdash \text{polyref } e : M \text{ pref}}$
(TP-POLYASGN)	$\frac{C; \gamma; \rho \vdash e_1 : M \text{ pref} \quad C; \gamma; \rho \vdash e_2 : \sigma \quad C \supset \sigma \succ M}{C; \gamma; \rho \vdash \text{polyasgn } e_1 := e_2 : \text{unit}}$
(TP-POLYDEREF)	$\frac{C; \gamma; \rho \vdash e : M \text{ pref} \quad C \supset M \succ \sigma}{C; \gamma; \rho \vdash \text{polyderef } e : \sigma}$
(TP-GEN-TVAR)	$\frac{C; \gamma; \rho \vdash e : \sigma \quad \alpha \notin \text{FV}(\gamma) \cup \text{FV}(C)}{C; \gamma; \rho \vdash e : \forall \alpha. \sigma}$
(TP-GEN-ALIAS)	$\frac{C \cup C_0; \gamma; \rho \vdash v : \sigma \quad M \notin \text{dom}(C) \cup \text{FA}(\gamma) \cup \text{range}(\rho) \quad \text{dom}(C_0) = \{M\}}{C; \gamma; \rho \vdash v : \forall M : C_0. \sigma}$
(TP-INST)	$\frac{C; \gamma; \rho \vdash v : \sigma \quad C \vdash \sigma \succ \sigma'}{C; \gamma; \rho \vdash v : \sigma'}$

FIGURE 4. Typing Rules for an extension of ML with Reference Polymorphism

system accurately capturing what values in the store can do once they. To this end, we need the retrieval constraints to match up with the values in the store μ .

Definition 2.8 (Typing a Store). *We say that under alias-location mapping ρ , store μ satisfies constraint set C , written $\rho \vdash \mu : C$, if:*

1. $\text{dom}(\rho) = \text{dom}(\mu)$
2. For each retrieval constraint $M \succ \sigma \in C$, then for all l such that $\rho(l) = M$, $C; \rho \vdash \mu(l) : \sigma$.

We are now ready to state the semantic consistency result, which connects the operational semantics with the typing judgment. It requires that there be a typing of the expression e under a consistent constraint set C .

Theorem 2.9 (Semantic Consistency). *Let C be consistent and suppose $C; \rho \vdash e : \sigma$, $\mu \vdash e \hookrightarrow v, \mu'$, and $\rho \vdash \mu : C$. Then there exists $\rho' \supseteq \rho$ such that $C; \rho' \vdash v : \sigma$ and $\rho' \vdash \mu' : C$.*

Proof. Proof follows from induction on the structure of the typing judgment $C; \rho \vdash e : \sigma$. The cases (TP-UNIT), (TP-LOC), (TP-VAR), and (TP-LAM) are all straightforward, as they are values and so their corresponding evaluation rules do not modify the store. For these we simply take $\rho' = \rho$ and the result follows from the typing of the value.

$$\begin{array}{c}
\text{(EV-LOC)} \quad \mu \vdash l \hookrightarrow l, \mu \\
\text{(EV-LAM)} \quad \mu \vdash \lambda x. e \hookrightarrow \lambda x. e, \mu \\
\text{(EV-APP)} \quad \frac{\mu \vdash e_1 \hookrightarrow \lambda x. e_0, \mu_1 \quad \mu_1 \vdash e_2 \hookrightarrow v_0, \mu_2 \quad \mu_2 \vdash e_0[v_0/x] \hookrightarrow v, \mu'}{\mu \vdash e_1 e_2 \hookrightarrow v, \mu'} \\
\text{(EV-LET)} \quad \frac{\mu \vdash e_1 \hookrightarrow v_0, \mu_1 \quad \mu_1 \vdash e_2[v_0/x] \hookrightarrow v, \mu'}{\mu \vdash \text{let } x = e_1 \text{ in } e_2 \hookrightarrow v, \mu'} \\
\text{(EV-REF)} \quad \frac{\mu \vdash e \hookrightarrow v, \mu' \quad l \notin \text{dom}(\mu')}{\mu \vdash \text{polyref } e \hookrightarrow l, \mu'[l : v]} \\
\text{(EV-ASGN)} \quad \frac{\mu \vdash e_1 \hookrightarrow l, \mu_1 \quad \mu_1 \vdash e_2 \hookrightarrow v, \mu'}{\mu \vdash \text{polyasgn } e_1 := e_2 \hookrightarrow (), \mu'[l : v]} \\
\text{(EV-DEREF)} \quad \frac{\mu \vdash e \hookrightarrow l, \mu' \quad \mu'(l) = v}{\mu \vdash \text{polyderef } e \hookrightarrow v, \mu'}
\end{array}$$

FIGURE 5. Operational Semantics for ML with References

Case (TP-APP):

By inversion on the typing judgment we have

$$C; \rho \vdash e_1 : \tau_1 \rightarrow \tau \quad C; \rho \vdash e_2 : \tau_1$$

By inversion on the evaluation we have

$$\mu \vdash e_1 \hookrightarrow \lambda x. e, \mu_1 \quad \mu_1 \vdash e_1 \hookrightarrow v_0, \mu_2 \quad \mu_2 \vdash e[v_0/x] \hookrightarrow v, \mu'$$

We apply induction to the first pair of evaluation deductions to receive a $\rho_1 \supseteq \rho$ such that $\rho_1 \vdash \mu_1 : C$ and by inversion $C; \{x \mapsto \tau_1\}; \rho_1 \vdash e : \tau_1$. By weakening we can replace ρ with ρ_1 in the second type judgment and so we have $\rho' \supseteq \rho_1$ such that $\rho' \vdash \mu' : C$ and $C; \rho' \vdash v_0 : \tau_1$. By weakening and value substitution, we have $C; \rho' \vdash v : \tau$, as required.

Case (TP-LET):

This case is straightforward and similar to the case for (TP-APP).

Case (TP-POLYREF):

By inversion on the typing judgment, we have

$$C; \rho \vdash e : \sigma \quad C \supset \sigma \succ M$$

By inversion on the evaluation judgment, we have

$$\mu \vdash e \hookrightarrow v, \mu_1$$

We apply induction to receive a ρ_1 such that $\rho_1 \vdash \mu_1 : C$ and $C; \rho_1 \vdash v : \sigma$. Let $\mu' = \mu_1[l : v]$, where l is the result of the evaluation of the polyref expression. We then let $\rho' = \rho_1[l : M]$. From this definition it is clear we have $C; \rho' \vdash l : M \text{ pref}$ by the (TP-LOC) typing rule.

We now show $\rho' \vdash \mu' : C$: it remains to show that for any $M \succ \sigma'$ such that $C \supset M \succ \sigma'$, then $C; \rho' \vdash v : \sigma'$.

We have $C; \rho' \vdash v : \sigma$. Suppose $C \supset M \succ \sigma'$; since $C \supset \sigma \succ M$ and C is consistent, then $C \vdash \sigma \succ \sigma'$. By the (TP-INST) typing rule, we thus have $C; \rho' \vdash v : \sigma'$ as required. Therefore $\rho' \vdash \mu' : C$.

Case (TP-POLYASGN):

By inversion on the typing judgment, we have

$$C; \rho \vdash e_1 : M \text{ pref} \quad C; \rho \vdash e_2 : \sigma \quad C \supset \sigma \succ M$$

By inversion on the evaluation, we have

$$\mu \vdash e_1 \hookrightarrow l, \mu_1 \quad \mu_1 \vdash e_2 \hookrightarrow v, \mu_2$$

By induction we have $\rho_1 \supseteq \rho$ such that $\rho_1 \vdash \mu_1 : C$ and $C; \rho_1 \vdash l : M \text{ pref}$ (thus $\rho_1(l) = M$). We apply induction a second time to receive $\rho_2 \supseteq \rho$ such that $\rho_2 \vdash \mu_2 : C$ and $C; \rho_2 \vdash v : \sigma$.

Let $\mu' = \mu_2[l : v]$. We claim that $\rho_2 \vdash \mu' : C$. We must check that $C; \rho_2 \vdash \mu'(l) : \sigma'$ for each τ such that $C \supset \rho_2(l) \succ \sigma'$; this reduces to checking $C; \rho_2 \vdash v : \sigma'$.

We have $C; \rho_2 \vdash v : \sigma$. Since C is consistent and $C \supset \sigma \succ M$, we have $C \vdash \sigma \succ \sigma'$. By the (TP-INST) typing rule, we thus have $C; \rho_2 \vdash v : \sigma'$ as required. Trivially we have $C; \rho_2 \vdash () : \text{unit}$. This is the required result.

Case (TP-POLYDEREF):

By inversion on the typing judgment we have

$$C; \rho \vdash e : M \text{ pref} \quad C \supset M \succ \tau$$

By inversion on the evaluation judgment we have

$$\mu \vdash e \hookrightarrow l, \mu' \quad v = \mu'(l)$$

By induction we have ρ' such that $\rho' \vdash \mu' : C$ and $C; \rho' \vdash l : M \text{ pref}$; thus $\rho'(l) = M$. Since C is consistent and $C \supset M \succ \sigma$, we have $C; \rho' \vdash \mu'(l) : \sigma$. As $\mu'(l) = v$, this is the required type derivation.

Case (TP-GEN-TVAR):

By inversion on the type judgment we have

$$C; \rho \vdash e : \sigma \quad \alpha \notin FV(C)$$

We apply induction to the type judgment to receive ρ' such that $C; \rho' \vdash v : \sigma$. By the rule (TP-GEN-TVAR) and since α is still not free in C , we thus have $C; \rho' \vdash v : \forall \alpha. \sigma$ as required.

Case (TP-GEN-ALIAS):

Since there is no evaluation judgment here (because the generalization is performed on a syntactic value), this case is trivial; let $\rho' = \rho$. Note that this case would not hold if the alias generalization was being performed on an expression; new locations might be created during its evaluation that would need to show up in the range of ρ' , which could not be generalized. \square

Soundness of polymorphic references is a direct corollary of the Theorem 2.9.

3 TYPE INFERENCE

We give a brief outline of the important components of automatic type inference before presenting their details.

Computational Generalization: We introduce a generalization function GEN , which takes arguments τ , C , γ , and e . If e is a syntactic value, then GEN generalizes all of the aliases and type variables in the type of τ which are not free in γ and returns a polytype and a reduced constraint set. Otherwise, GEN only generalizes type variables which are not free in γ and C .

Instantiation: In order to type variables which map to polytypes, we define an instantiation function $\text{INSTANTIATE}(\sigma)$. The instantiation function substitutes new type variables and aliases for generalized ones in σ , returning the substituted type, an empty substitution, and the constraints on the newly introduced aliases.

Unification: The core difficulty of type inference with polymorphic references is ensuring that different parts of code use the same mutable data consistency. For example, in the application expression, we infer information about how e_1 and e_2 both use polymorphic references. If $\gamma(x) = M \text{ pref}$, we must ensure that both e_1 and e_2 use x (through the alias M) in a consistent way; we cannot have e_1 assigning x to an integer and e_2 retrieving a boolean from x . Of course, we must still perform the normal checks of the \mathcal{W} algorithm such as ensuring that function applications are being performed with function types.

Our inference algorithm uses a unification algorithm as in traditional type inference presentations along with two inference algorithms for constraint resolution. The unification function $\text{UNIFY}(\tau_1, \tau_2, C)$ returns a substitution of type variables and aliases ϕ such that $\phi(\tau_1) = \phi(\tau_2)$ and $\phi(C)$ is also consistent. The function $\text{UNIFYCONSTRAINTS}(C)$ returns a substitution on type variables φ such that $\varphi(C)$ is consistent, while the function $\text{UNIFYINST}(\sigma, \tau)$ returns a substitution φ such that $\vdash \varphi\sigma \succ \varphi\tau$; this is a non-structural subtyping inference algorithm [9]. The UNIFYINST function ensures that the storage and retrieval constraints for a single alias are consistent.

Type Inference: The type inference algorithm calls each of the above functions. Its primary goal is to perform the actions of the traditional \mathcal{W} algorithm along with the appropriate calls to UnifyConstraints . Our inference algorithm is given in Figure 6.

We now give the formal details of the algorithms described above.

In order to construct constraint sets, we need some notation regarding more direct operations on them. The domain of a constraint set, written $\text{dom}(C)$, contains all of the aliases which are mentioned in it; formally $\text{dom}(C) = \{M \mid \sigma \succ M \in C \text{ or } M \succ \sigma \in C\}$. We use the notation $C[M]$ to represent all of the constraints on an alias M . Formally, $C[M] = \{c \in C \mid \text{ and } c \text{ is a constraint on } M\}$.

3.1 Computational Generalization

We now detail an automatic generalization operator similar to the traditional presentation of the Close operator for implementing a let-polymorphism algorithm.

As we have seen in the previous section, allowing type variable generalization of expressions is safe, while we can generalize aliases only for syntactic values. Our generalization function Gen takes four arguments: a type to generalize, a constraint set, a context, and an expression. Because alias generalization can discharge constraints, the generalization function returns two arguments: the generalized type and the resulting constraints from generalization. If the expression argument is a syntactic value, then both aliases and type variables are generalized in the resulting polymorphic type.

Definition 3.1 (Generalizing a Type Variable). *We define the alias abstraction operation $\text{GENALIASES}(C, \gamma, \sigma)$*

by

$$\begin{aligned} \text{GENALIASES}(C, \gamma, \sigma) &= \text{GENALIASES}(C \setminus C[M], \gamma, \forall M:C[M].\sigma) \text{ for } M \text{ such that } M \notin \text{FA}(\gamma) \text{ and} \\ &\quad \text{there does not exist } \alpha_i \text{ such that } \alpha_i \in \text{FV}(C([M])) \text{ and } \alpha_i \in \text{FV}(\gamma) \\ &= (C, \sigma) \text{ if no such } M \text{ exists} \end{aligned}$$

Definition 3.2 (Generalization Function). *Let the generalization function Gen be defined by:*

$$\begin{aligned} \text{GEN}(\tau, C, \gamma, v) &= \text{let } (C', \sigma) = \text{GENALIASES}(C, \gamma, \tau) \\ &\quad \{\alpha_1, \dots, \alpha_m\} = \text{FV}(\sigma) \setminus (\text{FV}(\gamma) \cup \text{FV}(C')) \\ &\quad \text{in } (C', \forall \alpha_1 \dots \alpha_m. \sigma) \\ \text{GEN}(\tau, C, \gamma, e) &= (C, \forall \alpha_1 \dots \alpha_n. \tau) \text{ if } e \neq v \text{ and where } \{\alpha_1, \dots, \alpha_n\} = \\ &\quad \text{FV}(\tau) \setminus (\text{FV}(\gamma) \cup \text{FV}(C)) \end{aligned}$$

Lemma 3.3 (Abstraction Correctness). *Suppose $\text{GENALIASES}(C, \gamma, \sigma) = (C', \sigma')$ and $C; \gamma \vdash v : \sigma$. Then $C'; \gamma \vdash v : \sigma'$.*

Proof. It is safe to generalize each alias M from σ since each of them does not appear in $\text{FA}(\gamma)$, and each of the constraint sets removed does not contain a type variable occurring free in $\text{FV}(\gamma)$. \square

Lemma 3.4 (Generalization Correctness). *Suppose $\text{GEN}(\tau, C, \gamma, e) = (C', \sigma)$ and $C; \gamma \vdash e : \tau$. Then $C'; \gamma \vdash e : \sigma$.*

Proof. There are two cases, depending on whether e is a value or not. If e is not a value, $\sigma \equiv \forall \alpha_1 \dots \alpha_n. \tau$, where $\{\alpha_1, \dots, \alpha_n\} \cap (\text{FV}(C) \cup \text{FV}(\gamma)) = \emptyset$. We can then abstract each α_i from τ using the (TP-GEN-TVAR) rule, resulting in the required derivation.

If e is a value, then $\sigma \equiv \forall \alpha_1 \dots \alpha_n \forall M_1:C_1 \dots M_k:C_k. \tau$. By the previous lemma, we can use (TP-GEN-ALIAS) to abstract the aliases M_1, \dots, M_k in τ . Afterwards, we can use (TP-GEN-TVAR) to generalize the remaining aliases. \square

3.2 Instantiation

We provide here the instantiation algorithm required to type variables along with a proof of its correctness.

$$\begin{aligned} \text{INSTANTIATE}(\sigma) &= \text{let } \sigma \equiv \forall \alpha_1 \dots \alpha_n \forall M_1:C_1 \dots M_k:C_k. \tau \\ &\quad \text{let } \{\beta_1, \dots, \beta_n\} \text{ be fresh type variables} \\ &\quad \text{let } \{N_1, \dots, N_k\} \text{ be fresh aliases} \\ &\quad \Psi = \{\alpha_1 \mapsto \beta_1, \dots, \alpha_n \mapsto \beta_n, M_1 \mapsto N_1, \dots, M_k \mapsto N_k\} \\ &\quad \text{return } (\Psi(\tau), \emptyset, \Psi(C_1 \cup \dots \cup C_k)) \end{aligned}$$

We call a polytype σ consistent if $\text{INSTANTIATE}(\sigma) = (\tau, \emptyset, C)$ and C is consistent. We extend this definition to contexts γ in the usual way: γ is consistent if for all $x \in \text{dom}(\gamma)$, $\gamma(x)$ is consistent. Note that consistency of polytypes and contexts is invariant under both type and alias substitutions.

Theorem 3.5 (Correctness of Instantiate). *If $\text{INSTANTIATE}(\sigma) = (\tau, \emptyset, C)$, then $C \vdash \sigma \succ \tau$.*

Proof. Proof follows easily from the definition of $C \vdash \sigma \succ \tau$. \square

3.3 Unification

In our type inference algorithm, we need to have three related unification algorithms. The first, `UNIFY`, returns a substitution that makes two types equal, similar to the normal type unification algorithm. A second unification algorithm, `UNIFYCONSTRAINTS` will make an inconsistent constraint set into a consistent one. This algorithm will depend on a third unification algorithm, `UNIFYINST`, which will make a polytype instantiate to a type under a constraint set. We do not provide `UNIFYINST`: it is a non-structural subtyping algorithm for ensuring that a polytype instantiates to a type.

$$\begin{aligned}
\text{UNIFY}(\tau, \tau, C) &= \emptyset \\
\text{UNIFY}(\alpha, \tau_2, C) &= \text{if } \alpha \notin \text{FV}(\tau) \cup \{\text{FV}(C[M]) \mid M \in \text{FA}(\tau)\} \text{ then} \\
&\quad \{\alpha \mapsto \tau_2\} \\
\text{UNIFY}(\tau_1, \alpha, C) &= \text{UNIFY}(\alpha, \tau_1, C) \\
\text{UNIFY}(\tau_1 \rightarrow \tau_2, \tau'_1 \rightarrow \tau'_2, C) &= \text{let } \phi_1 = \text{UNIFY}(\tau_1, \tau'_1, C) \\
&\quad \phi_2 = \text{UNIFY}(\phi_1(\tau_2), \phi_1(\tau'_2), \phi_1(C)) \\
&\quad \text{in } \phi_2 \circ \phi_1 \\
\text{UNIFY}(M \text{ pref}, N \text{ pref}, C) &= \text{let } \phi = \{N \mapsto M\} \\
&\quad \varphi_C = \text{UNIFYCONSTRAINTS}(C[M] \cup \phi(C[N])) \\
&\quad \text{in } \varphi_C \circ \phi \\
\text{UNIFY}(\tau_1, \tau_2,) &= \text{raise exception}
\end{aligned}$$

The `UNIFYCONSTRAINTS` function returns a type substitution φ such that a possibly inconsistent constraint set C is made consistent by φ ; i.e. $\varphi(C)$ is consistent.

$$\begin{aligned}
\text{UNIFYCONSTRAINTS}(C) &= \text{let } \varphi = \emptyset \\
&\quad \text{for each } M \in \text{dom}(C) \\
&\quad \quad \text{for each pair } (\sigma \succ M, M \succ \tau) \in C[M] \times C[M] \\
&\quad \quad \quad \varphi_0 = \text{UNIFYINST}(\varphi(\sigma), \varphi(\tau)) \\
&\quad \quad \quad \varphi = \varphi_0 \circ \varphi \\
&\quad \text{return } \varphi
\end{aligned}$$

The function `UNIFYINST` makes a polytype instantiate to a type. It is a non-structural subtyping algorithm [9]. It takes arguments σ, τ and returns a type substitution φ such that $\vdash \varphi(\sigma) \succ \varphi(\tau)$.

Theorem 3.6 (Unify Correctness). *Let C be simple. Then the following hold:*

- (1) *If $\text{UNIFY}(\tau_1, \tau_2, C) = \phi$ and C is consistent, then $\phi(\tau_1) = \phi(\tau_2)$ and $\phi(C)$ is consistent.*
- (2) *If $\text{UNIFYCONSTRAINTS}(C) = \varphi$ and C is consistent, then $\varphi(C)$ is consistent.*
- (3) *If $\text{UNIFYINST}(\sigma, \tau) = \varphi$, then φ is a substitution on type variables and $\vdash \varphi(\sigma) \succ \varphi(\tau)$.*

Proof. By the correctness of the non-structural subtyping algorithm, (3) holds. We now show that (2) and (1) hold.

Case (2) \implies (1):

Assume (2) holds. Suppose $\text{UNIFY}(\tau_1, \tau_2, C) = \phi$ and C is consistent.

If $\tau_1 \equiv \alpha$ or $\tau_2 \equiv \alpha$, then ϕ is a substitution on type variables and so by Lemma 2.4, $\phi(C)$ is consistent.

If $\tau_1 \equiv \tau'_1 \rightarrow \tau'_2$ and $\tau_2 \equiv \tau_1^* \rightarrow \tau_2^*$, then by induction $\phi_1(C)$ is consistent and $\phi_2(\phi_1(C)) = \phi'(C)$ is consistent.

If $\tau_1 \equiv M \text{ pref}$ and $\tau_2 \equiv N \text{ pref}$, then by (2), we have $\varphi_C(C[M] \cup \phi(C[N]))$ consistent. Let $C' = C \setminus (C[M] \cup C[N])$. By the simplicity of C , we have the equality

$$\begin{aligned} \phi'(C) = \varphi_C(\phi(C)) &= \varphi_C(\phi(C') \cup \phi(C[M]) \cup \phi(C[N])) \\ &= \varphi_C(C' \cup C[M] \cup \phi(C[N])) \end{aligned}$$

Here C' and $C[M] \cup \phi(C[N])$ are both consistent; since their domains do not overlap, we thus have that their union is consistent and so applying a type substitution to the resulting constraint set is also consistent.

Case (3) \implies (2):

Assume (3) holds, i.e. $\text{UNIFYINST}(\sigma, \tau) = \varphi$ implies that $\vdash \varphi(\sigma) \succ \varphi(\tau)$ for all σ and τ mentioned in C .

We prove the following claim that immediately implies the result: after each M has been considered by the UNIFYCONSTRAINTS algorithm, $\varphi(C[M])$ is consistent. Suppose $\varphi(C[M])$ is not consistent; i.e. there is a pair of constraints $(\varphi(\sigma) \succ M, M \succ \varphi(\tau)) \in \varphi(C[M] \times C[M])$ such that $\vdash \varphi(\sigma) \not\succeq \varphi(\tau)$. Let $\varphi = \varphi_0 \circ \dots \circ \varphi_k$; each of these substitutions was added to φ by being the return result from UNIFYINST for some pair of constraints $(\sigma' \succ M, M \succ \tau')$.

Let ϕ_j be the substitution such that the pair $(\phi_1 \circ \dots \circ \phi_{j-1}(\sigma) \succ M, M \succ \phi_1 \circ \dots \circ \phi_{j-1}(\tau))$ was considered by the algorithm. By (3), $\vdash \phi_1 \circ \dots \circ \phi_j(\sigma) \succ \phi_1 \circ \dots \circ \phi_j(\tau)$. By 2.1 we can apply the rest of the φ_i and so we have $\vdash \varphi(\sigma) \succ \varphi(\tau)$. We thus have a contradiction and so $\varphi(C[M])$ is consistent. Since for all $M \in \text{dom}(\varphi(C))$, $C[M]$ is consistent, $\varphi(C)$ is therefore consistent. □

Our inference algorithm is given in Figure 6. The key difficulty in our use-based inference approach is in combining constraints from multiple branches of the inference algorithm. For example, in the application expression, we infer information about how e_1 and e_2 both use polymorphic references; for example, maybe $\gamma(x) = M \text{ pref}$. We must ensure that both e_1 and e_2 use x (through the alias M) in a consistent way; we cannot have e_1 assigning it to an integer and e_2 reading a boolean. In this case, we call the inference algorithm recursively on both e_1 and e_2 and then use UNIFYCONSTRAINTS to combine the constraint sets returned by these recursive calls. Extra work is necessary in order to perform the traditional actions of the \mathcal{W} inference algorithm, i.e. ensuring that e_1 is a function type that matches up with e_2 .

Lemma 3.7 (Substitution Idempotency). *If $\text{INFER}(e, \gamma) = (\tau, \phi, C)$, then $\phi(C) = C$, $\phi(\phi(\gamma)) = \phi(\gamma)$, and $\phi(\tau) = \tau$.*

Theorem 3.8 (Soundness of Infer). *If $\text{INFER}(e, \gamma) = (\tau, \phi, C)$ and γ is consistent, then $C; \phi(\gamma) \vdash e : \tau$ and C is consistent.*

Proof. We prove the theorem by induction on e , using Theorem 3.6 as necessary.

Case $e \equiv b$:

Straightforward.

Case $e \equiv x$:

The consistency of C is guaranteed by the consistency of γ . The typability of x to τ is given by Theorem 3.5.

Case $e \equiv \lambda x.e$:

This case is easy, as we do not modify the constraint set.

Case $e \equiv e_1 e_2$:

By induction we have

$$C_1; \phi_1(\gamma) \vdash e : \tau_1 \quad C_2; \phi_2(\phi_1(\gamma)) \vdash e : \tau_2$$

and C_1, C_2 consistent.

By Theorem 3.6 and $\text{UNIFY}(\phi'(\tau_1), \phi'(\tau_2 \rightarrow \beta), C) \circ \phi'$, we have

$$\phi_u(\tau_1) = \phi_u(\tau_2 \rightarrow \beta)$$

Also by Theorem 3.6 and, we have $\text{UNIFYCONSTRAINTS}(\phi_u(C)) \circ \phi_u = \phi'$ that $\phi'(C)$ is consistent; therefore, the returned constraint set is consistent. This satisfies the second requirement of the theorem. It still remains to show the first.

Applying substitutions to the above type derivations and the above equality, we have

$$\phi_u(C_1); \phi_u(\gamma) \vdash e : \phi_u(\tau_2) \rightarrow \phi_u(\beta) \quad \phi_u(C_2); \phi_u(\gamma) \vdash e : \phi_u(\tau_2)$$

Here $\phi_u(C_1) \subseteq \phi_u(C)$ and $\phi_u(C_2) \subseteq \phi_u(C)$, so by Lemma 2.5 we can replace them in the deduction by $\phi_u(C)$.

$$\phi_u(C); \phi_u(\gamma) \vdash e : \phi_u(\tau_2) \rightarrow \phi_u(\beta) \quad \phi_u(C); \phi_u(\gamma) \vdash e : \phi_u(\tau_2)$$

Lastly, we apply ϕ' to gain the final two preconditions for our application deduction.

$$\phi'(C); \phi'(\gamma) \vdash e : \phi'(\tau_2) \rightarrow \phi'(\beta) \quad \phi'(C); \phi'(\gamma) \vdash e : \phi'(\tau_2)$$

We thus can construct the required application type deduction using the (TP-APP) typing rule.

$$\frac{\phi'(C); \phi'(\gamma) \vdash e : \phi'(\tau_2) \rightarrow \phi'(\beta) \quad \phi'(C); \phi'(\gamma) \vdash e : \phi'(\tau_2)}{\phi'(C); \phi'(\gamma) \vdash e : \phi'(\beta)}$$

Case $e \equiv \text{polyref } e$:

By induction we have the type judgement

$$C; \phi(\gamma) \vdash e : \tau$$

By Theorem 3.4, we have

$$C'; \phi(\gamma) \vdash e : \sigma$$

Since M is a new alias, $C' \cup \{\sigma \succ M\}$ is consistent. We can thus construct the required type deduction using the (TP-POLYREF) rule, we have $C'; \phi(\gamma) \vdash \text{polyref } e : M \text{ pref}$.

Case $e \equiv \text{polyderef } e$:

By induction we have the type judgement

$$C; \phi(\gamma) \vdash e : \tau$$

By the correctness of UNIFY, we have

$$C; \phi(\gamma) \vdash e : N \text{ pref}$$

Here $N = \phi_u(M)$. By the correctness of UNIFYCONSTRAINTS, we have that $\phi_c(C')$ is consistent and $\phi_c(C') \vdash \phi_c(\phi_u(\alpha)) \succ$.

By substitution and weakening, we have the judgement

$$\phi'(C'); \phi'(\gamma) \vdash e : \phi'(M \text{ pref})$$

Since α was new, by the idempotency of substitutions, we have $\phi'(C') \vdash \phi'(M) \succ \phi'(\alpha)$. Using the (TP-POLYDEREF) rule, we have the deduction

$$\frac{\phi'(C'); \phi'(\gamma) \vdash e : \phi'(M \text{ pref}) \quad \phi'(C') \supset \phi'(M) \succ \phi'(\alpha)}{\phi'(C'); \phi'(\gamma) \vdash \text{polyderef } e : \phi'(\alpha)}$$

This proves the required result.

Case $e \equiv \text{polyasgn } e_1 := e_2$:

By induction we have the type judgements

$$C_1; \phi_1(\gamma) \vdash e_1 : \tau_1 \quad C_2; \phi_2(\phi_u(\phi_1(\gamma))) \vdash e_2 : \tau_2$$

By the correctness of unification, we have

$$\phi_u(C_1); \phi_u(\phi_1(\gamma)) \vdash e_1 : \phi_u(M \text{ pref})$$

By Theorem 3.4, we have

$$C_g; \phi_2(\phi_u(\phi_1(\gamma))) \vdash e_2 : \sigma$$

By Theorem 3.6, the larger constraint set $\phi_c(C')$ containing $\phi_c(C_g)$ is consistent, so we have (through substitution and weakening),

$$\phi_c(C'); \phi_c(\phi'(\gamma)) \vdash e_2 : \phi_c(\sigma)$$

Through application of substitution and weakning to the first deduction, we also have

$$\phi_c(C'); \phi_c(\phi'(\gamma)) \vdash e_1 : \phi_c(M \text{ pref})$$

We also have $\phi_c(C') \supset \phi_c(\sigma) \succ \phi'(M)$. This follows from the definition of C' and the application of the type substitution ϕ_c to the polytype σ .

We thus can construct the required deduction using the (TP-POLYASGN) type rule.

$$\frac{\phi_c(C'); \phi_c(\phi'(\gamma)) \vdash e_1 : \phi_c(M \text{ pref}) \quad \phi_c(C'); \phi_c(\phi'(\gamma)) \vdash e_2 : \phi_c(\sigma) \quad \phi_c(C') \supset \phi_c(\sigma) \succ \phi'(M)}{\phi_c(C'); \phi_c(\phi'(\gamma)) \vdash \text{polyasgn } e_1 := e_2 : \text{unit}}$$

Case $e \equiv \text{let } x = e_1 \text{ in } e_2$:

By induction on the first recursive call, we have the type deduction:

$$C_1; \phi_1(\gamma) \vdash e_1 : \tau_1$$

By applying Lemma 3.4, we have

$$C_g; \phi_1(\gamma) \vdash e_1 : \sigma$$

By induction on the second recursive call, we thus have the type deduction:

$$C_2; \phi_2(\phi_1(\gamma))[x:\phi_2(\sigma)] \vdash e_2 : \tau$$

By the correctness of UNIFYCONSTRAINTS, we have that $\phi_c(C') = \phi_c(C_2) \cup \phi_c(\phi_2(C_g))$ is consistent. By substitution and weakening, we thus have the deductions

$$\phi'(C'); \phi'(\gamma) \vdash e_1 : \phi'(\sigma) \quad \phi'(C'); \phi'(\gamma)[x:\phi'(\sigma)] \vdash e_2 : \phi'(\tau)$$

We can thus use the inference rule (TP-LET) to construct the final required let derivation.

$$\frac{\phi'(C'); \phi'(\gamma) \vdash e_1 : \phi'(\sigma) \quad \phi'(C'); \phi'(\gamma)[x:\phi'(\sigma)] \vdash e_2 : \phi'(\tau)}{\phi'(C'); \phi'(\gamma); \text{let } x = e_1 \text{ in } e_2 \vdash \phi'(\tau) :}$$

This is the required result. □

We call ϕ simple if it maps type variables to simple types. The INFER algorithm does not introduce non-simple types to the constraint set, summarized in the following lemma.

Lemma 3.9 (Simplicity of C). *Let γ be a simple context. If $\text{INFER}(e, \gamma) = (C, \phi, \tau)$, then ϕ is simple and C is simple.*

Proof. There is no way that a non-simple type could have been added to ϕ or C , since γ is simple and we throw exceptions on adding non-simple types to C . □

By Theorem 3.8, we have that the INFER algorithm produces a consistent constraint set and a type derivation. By the results from the previous section, a typing under a consistent constraint set ensures that there are no runtime type errors.

Theorem 3.10 (Soundness of Infer). *If $\text{INFER}(e, \gamma) = (C, \phi, \tau)$, then the evaluation of e will not produce a type error.*

Proof. The result follows from the correctness of the INFER algorithm and the soundness of the type judgement. □

We are still in the process of showing that this inference algorithm produces a principal type.

4 DISCUSSION

4.1 Application to Other Languages

Languages extended with a use-based inference system allow more flexibility when dealing with data structures that store and retrieve polymorphic objects. For example, in Java, a number of containers implement the `List` interface. In Java 1.5, programmers can parameterize Lists with type arguments, such as `List<String>`. However, it is not permissible to use subtyping within the type argument. Suppose $A \succ B$ and consider the following code examples in a Java-like language:

```

List<A> aList = { list of A's }
List<B> bList = { list of B's }
bList = aList // dangerous
B b = bList.getFirst() // bList might now hold A's
// that are not B's

List<A> aList' = { list of A's }
List<B> bList' = { list of B's }
aList' = bList'
A a = aList.getFirst() // still safe

```

The traditional Java type system prevents both of the above approaches. With a use-based approach to inference, we can safely add the inference rule $\vdash \text{List} \langle A \rangle \succ \text{List} \langle B \rangle$. The first list reassignment in the example code is not allowed, since we cannot restrict the polymorphic type stored in a reference cell. However, the second reassignment of `aList'` to `bList'` is permitted.

For completely use-based approach, we could design a type system where a programmer could just write code like the following:

```

List bList = { list of B's }
A a = l.getFirst();
B b = l.getFirst();

```

The inference engine would then attempt to infer the type parameter for the list `bList`. In order to ensure that this is being done safely, we will need to guard the internal state of the `List` object just as we guard other references; specifically, that what we are putting into that reference has the same type as what we are taking out. A polymorphic `List` class would then have an extended (and inferred) constraint-based signature:

```

T PolyList :
  state : M
  T getFirst() : (M > T)
  void putFirst(T) : (T > M)

```

In the above, in order to retrieve a value of type `T` from a polymorphic list, we need the internal state (maybe an array) to be able to retrieve a value of type `T` from it. Similarly, to add an item to a polymorphic list, we need the internal state to be able to store a value of type `T`.

This type system would require annotating each of the user libraries with these usage constraints. Type inference would then involve propagating these constraints throughout a program to ensure that each polymorphic reference cell is used safely. The results would be similar to variance-based subtyping for parameteric types [6], except with more of an inference flavor to the language.

4.2 Verifying Code Security

Security-typed languages ensure that programs do not violate an explicitly written security policy [13]. We can decorate traditional polytypes with security levels l ; for example, an integer represents a user's

bank account might have type $int\{\text{secret}\}$. Privacy levels can be safely raised, so all public data can be classified upward to secret data. We can represent this in our type system by extending the $C \vdash \sigma \succ \tau$ relation to include security types; i.e. $C \vdash \sigma\{\text{public}\} \succ \sigma\{\text{secret}\}$ and $C \vdash \sigma\{l\} \succ \sigma'\{l\}$ assuming $C \vdash \sigma \succ \sigma'$. This subtyping relation ensures that we cannot convert secret data into public data by laundering it through the store. This extension to the type system is not fine-grained enough to prevent implicit information flows without additional program decoration.

4.3 Related Work

As previously mentioned, combining polymorphism with mutable state is a large problem; we do not give a complete survey of the work done in this field. Earlier work has focused on finding typing semantics for references that were sound [8, 1, 12, 5], and later work has worked to increase the amount that polymorphism can be combined with mutable state [14, 4, 11]. With the adoption of the value restriction, Standard ML avoids more complicated type systems in favor of simpler error messages and easier type-checking.

Semi-explicit polymorphism allows for programmers to explicitly annotate polymorphic types on functions, that are then propagated through the rest of the program using inference methods [2]. Our system differs from theirs with respect to references in two key ways: firstly, we do not require explicit polymorphic type declarations. Secondly, the value restriction prevents the results of a dereference from being used polymorphically. The standard technique of η -expansion restores polymorphism in some cases, but not in others. In our generalization schema, we are able to safely generalize type variables (but not aliases) in the types of expressions.

4.4 Future Work

We are working on a prototype implementation of our system within Standard ML and hope to extend this implementation to an object-oriented language such as Java. We additionally are investigating modifying this system into a linear type system, wherein references to polymorphic objects can be created, assigned to a different type, and then used at that new type. Such a type system would allow for more safety in dynamic programming languages such as scripting languages.

4.5 Conclusion

In this paper we have investigated reference polymorphism, a language feature where polymorphic data in the store can be created, updated, and then dereferenced and used polymorphically in that context. We have given a proof of the type soundness of our approach in a superset of ML extended with new language features to manipulate polymorphic references and presented a type inference algorithm for automatic type reconstruction.

REFERENCES

- [1] DAMAS, L., AND MILNER, R. Principal type-schemes for functional programs. In *POPL '82: Proceedings of the 9th ACM SIGPLAN-SIGACT symposium on Principles of programming languages* (New York, NY, USA, 1982), ACM Press, pp. 207–212.

- [2] GARRIGUE, J., AND REMY, D. Extending ML with semi-explicit higher-order polymorphism. In *Theoretical Aspects of Computer Software* (1997), pp. 20–46.
- [3] GARRIGUE, J., AND REMY, D. Semi-explicit first-class polymorphism for ML. *Information and Computation* 155, 1-2 (1999), 134–169.
- [4] GREINER, J. Weak polymorphism can be sound. *Journal of Functional Programming* 6, 1 (1996), 111–141.
- [5] HARPER, R. A simplified account of polymorphic references. *Information Processing Letters* 51, 4 (1994), 201–206.
- [6] IGARASHI, A., AND VIROLI, M. Variant parametric types: A flexible subtyping scheme for generics. *ACM Trans. Program. Lang. Syst.* 28, 5 (2006), 795–847.
- [7] LEROY, X., AND WEIS, P. Polymorphic type inference and assignment. In *18th symposium Principles of Programming Languages* (1991), ACM Press, pp. 291–302.
- [8] MILNER, R. A theory of type polymorphism in programming. *Journal of Computer and System Sciences* 17 (1978), 348–375.
- [9] PALSBERG, J., WAND, M., AND O’KEEFE, P. Type inference with non-structural subtyping. *Formal Aspects of Computing* 9, 1 (1997), 49–67.
- [10] SMITH, F., WALKER, D., AND MORRISSETT, J. G. Alias types. In *ESOP ’00, 9th European Symposium on Programming, Berlin, Germany, March 25–April 2* (2000), Springer-Verlag, NY, pp. 366–381.
- [11] SMITH, G., AND VOLPANO, D. Polymorphic typing of variables and references. *ACM Transactions on Programming Languages and Systems* 18, 3 (May 1996), 254–267.
- [12] TOFTE, M. Type inference for polymorphic references. *Information and Computation* 89, 1 (1990), 1–34.
- [13] VOLPANO, D., SMITH, G., AND IRVINE, C. A sound type system for secure flow analysis. *Journal of Computer Security* 4, 3 (1996), 167–187.
- [14] WRIGHT, A. K. Typing references by effect inference. In *ESOP ’92, 4th European Symposium on Programming, Rennes, France, February 1992, Proceedings*, B. Krieg-Bruckner, Ed., vol. 582. Springer-Verlag, New York, N.Y., 1992, pp. 473–491.
- [15] WRIGHT, A. K. Simple imperative polymorphism. *Lisp and Symbolic Computation* 8, 4 (1995), 343–355.

$$\begin{aligned}
\text{INFER}(b, \gamma) &= (\text{TYPEOF}(b), \emptyset, \emptyset) \\
\text{INFER}(x, \gamma) &= \text{INSTANTIATE}(\gamma(x)) \\
\text{INFER}(\lambda x. e, \gamma) &= \text{let } \alpha \text{ be a fresh type variable} \\
&\quad (\tau, \phi, C) = \text{INFER}(e, \gamma[x:\alpha]) \\
&\quad \text{in } (\phi(\alpha) \rightarrow \tau, \phi, C) \\
\text{INFER}(e_1 \ e_2, \gamma) &= \text{let } (\tau_1, \phi_1, C_1) = \text{INFER}(e_1, \gamma) \\
&\quad (\tau_2, \phi_2, C_2) = \text{INFER}(e_2, \phi_1(\gamma)) \\
&\quad C = \phi_2(C_1) \cup C_2 \\
&\quad \text{let } \beta \text{ be a fresh type variable and } \phi_3 = \phi_2 \circ \phi_1 \\
&\quad \phi_u = \text{UNIFY}(\phi_3(\tau_1), \phi_3(\tau_2 \rightarrow \beta), C) \circ \phi_3 \\
&\quad \phi' = \text{UNIFYCONSTRAINTS}(\phi_u(C)) \circ \phi_u \\
&\quad \text{in } (\phi'(\beta), \phi', \phi'(C)) \\
\text{INFER}(\text{polyref } e, \gamma) &= \text{let } (\tau, \phi, C) = \text{INFER}(e, \gamma) \\
&\quad \text{let } M \text{ be a new alias} \\
&\quad (C', \sigma) = \text{GEN}(\tau, C, \gamma, e) \\
&\quad \text{if } \sigma \text{ is not simple then raise exception} \\
&\quad \text{in } (M \text{ pref}, \phi, C' \cup \{\sigma \succ M\}) \\
\text{INFER}(\text{polyderef } e, \gamma) &= \text{let } (\tau, \phi, C) = \text{INFER}(e, \gamma) \\
&\quad \text{let } M \text{ be a new alias and } \alpha \text{ a new type variable} \\
&\quad \phi_u = \text{UNIFY}(\tau, M \text{ pref}, C) \\
&\quad \text{if } \phi_u(\alpha) \text{ is not simple then raise exception} \\
&\quad C' = \phi_u(C) \cup \{\phi_u(M) \succ \phi_u(\alpha)\} \\
&\quad \phi_c = \text{UNIFYCONSTRAINTS}(C') \\
&\quad \phi' = \phi_c \circ \phi_u \circ \phi \\
&\quad \text{in } (\phi'(\alpha), \phi', \phi'(C')) \\
\text{INFER}(\text{polyasgn } e_1 := e_2, \gamma) &= \text{let } (\tau_1, \phi_1, C_1) = \text{INFER}(e_1, \gamma) \\
&\quad \text{let } M \text{ be a new alias} \\
&\quad \phi_u = \text{UNIFY}(\tau_1, M \text{ pref}, C_1) \\
&\quad (\tau_2, \phi_2, C_2) = \text{INFER}(e_2, \phi_u(\phi_1(\gamma))) \\
&\quad (C_g, \sigma) = \text{GEN}(\tau_2, C_2, \phi_u(\phi_1(\gamma)), e) \\
&\quad \phi' = \phi_u \circ \phi_2 \circ \phi_1 \\
&\quad \text{if } \sigma \text{ is not simple then raise exception} \\
&\quad C' = \phi'(C_1) \cup C_g \cup \{\sigma \succ \phi'(M)\} \\
&\quad \phi_c = \text{UNIFYCONSTRAINTS}(C') \\
&\quad \text{in } (\text{unit}, \phi_c \circ \phi', \phi_c(C')) \\
\text{INFER}(\text{let } x = e_1 \text{ in } e_2, \gamma) &= \text{let } (\tau_1, \phi_1, C_1) = \text{INFER}(e_1, \gamma) \\
&\quad (C_g, \sigma) = \text{GEN}(\tau, C_1, \phi_1(\gamma), e_1) \\
&\quad (\tau, \phi_2, C_2) = \text{INFER}(e_2, \gamma[x:\phi_1(\sigma)]) \\
&\quad C' = C_2 \cup \phi_2(C_g) \\
&\quad \phi_c = \text{UNIFYCONSTRAINTS}(C') \\
&\quad \phi' = \phi_c \circ \phi_2 \circ \phi_1 \\
&\quad \text{in } (\phi'(\tau), \phi', \phi'(C'))
\end{aligned}$$

FIGURE 6. The INFER Algorithm for Type Inference