Parametric Polymorphism and Operational Improvement

JENNIFER HACKETT and GRAHAM HUTTON, University of Nottingham, UK

Parametricity, in both operational and denotational forms, has long been a useful tool for reasoning about
program correctness. However, there is as yet no comparable technique for reasoning about program improve-
ment, that is, when one program uses fewer resources than another. Existing theories of parametricity cannot
be used to address this problem as they are agnostic with regard to resource usage. This article addresses
this problem by presenting a new operational theory of parametricity that is sensitive to time costs, which
can be used to reason about time improvement properties. We demonstrate the applicability of our theory
by showing how it can be used to prove that a number of well-known program fusion techniques are time
improvements, including fixed point fusion, map fusion and short cut fusion.

CCS Concepts: • Software and its engineering → Functional languages; Polymorphism; • Theory of
computation → Operational semantics;

ACM Reference Format:
Jennifer Hackett and Graham Hutton. 2018. Parametric Polymorphism and Operational Improvement. N/A,
ICFP (March 2018), 23 pages. https://doi.org/10.1145/nnnnnn.nnnnnnn

1 INTRODUCTION

Parametricity is everywhere. In typed functional languages, many if not most of the built-in and
user-defined functions are parametric. Because of this ubiquity, it is important that we study
carefully the properties of parametric functions. Chief among these properties is the abstraction
theorem [Reynolds 1983], which shows that any well-typed term must satisfy a property that can
be derived uniformly from its type. By instantiating this theorem, we obtain the so-called “free
theorems” of Wadler [1989], properties held by any term of a given type.

The abstraction theorem was first presented by Reynolds [1983], who proved it using the notion
of relational parametricity. In relational parametricity, one starts with a denotational semantics
based on sets (or more generally, some form of domain) and builds on top of it another denotational
semantics based on relations between those sets. It can then be shown that interpreting any term
in related contexts must produce related results. Deriving the free theorem is then a matter of
calculating the relational interpretation of the type in question.

The original denotational presentation of the abstraction theorem is promising, as it suggests
that similar theorems will exist for any System F-like language. However, it is often not obvious
what a parametric model of such a language should be. For this reason, it is helpful to investigate
more operational notions of parametricity such as the one developed by Pitts [2000], where the
relations we work with are between terms rather than between the interpretations of terms. The
result is an abstraction theorem that respects observational equivalence, provided the relations
involved satisfy the intuitive property of \( \top \top \)-closure.
In this article we develop a new operational theory of parametricity that can be used to reason about time improvement, i.e. when one term can be replaced by another without increasing time cost. Our theory is built on a standard lazy abstract machine [Sestoft 1997], making it applicable to call-by-need languages such as Haskell. Specifically, we make the following contributions:

- We show how Pitts’ notion of $\top \top$-closure can be adapted to produce a resource-aware notion that we call machine-closure. The key idea is that whereas $\top \top$-closed relations respect observational equivalence, machine-closed relations respect time improvement.
- We use the notion of machine-closure to prove an abstraction theorem for call-by-need programs that use recursion in a restricted manner. The resulting theorem can be used to reason about time improvement properties in call-by-need languages such as Haskell.
- We demonstrate the application of our abstraction theorem by justifying a number of program fusion techniques as time improvements, including short cut fusion [Gill et al. 1993].

This work has similar aims to that of Seidel and Voigtländer [2011], who investigate efficiency-based free theorems in a call-by-value language, but differs in setting and approach. Firstly, we consider call-by-need rather than call-by-value. Secondly, their work is based on a denotational semantics instrumented with costs and it is not clear how this can be applied in a call-by-need setting, so instead we use an operational semantics with an explicit stack and heap. Finally, our work builds on the call-by-need improvement theory of Moran and Sands [1999a], using an improvement relation to abstract away from explicit costs where possible.

This paper is part of a wider project to make questions of call-by-need efficiency more tractable [Hutton and Hackett 2016]. By developing new techniques for questions of improvement that are compatible with the existing techniques used to prove correctness, we seek to bring the two issues of correctness and improvement closer together, reducing the work that must be done to formally justify a particular program transformation. In this case, a technique based on parametricity for reasoning about improvement makes it easier to reason about the efficiency aspects of program transformations that rely on parametricity for their correctness properties.

2 BACKGROUND

We begin in this section with some background on the two key technical elements that underpin our work, namely parametric polymorphism and operational improvement.

2.1 Parametric Polymorphism

The viewpoint of parametric polymorphism is that a polymorphic function must do the same thing at every type. This contrasts with ad-hoc polymorphism [Strachey 2000], where it is only required that a function do something at every type. The result is that parametrically polymorphic functions are forced to respect the abstractions of the calling code, being prevented from inspecting the structure of the type at which they are called. This property was first observed by Reynolds [1983], who proved the abstraction theorem for the polymorphic $\lambda$-calculus.

Reynolds’ abstraction theorem works by first defining a set-theoretic denotational semantics, where types are interpreted as sets and terms as elements of those sets, and then building on top of this a semantics of logical relations. These relations are built by defining relation-lifted definitions of the type constructors of the language. For example, two terms at a function type are related if they take related arguments to related results:

\[(f, g) \in R \rightarrow S \iff \forall (x, y) \in R. \ (f \ x, g \ y) \in S\]
Once we have this relational interpretation of types, it can be shown by a straightforward induction on the structure of type derivations that interpreting any term in related environments will give related results. This model can be extended with extra constructs such as general recursion and sequencing, which translate to adding restrictions on the relations [Johann and Voigtländer 2004]. However, this method is limited to languages with a natural denotational semantics, and so cannot be applied when the only natural semantics is operational.

Pitts [2000] addressed this issue by presenting an operational treatment of parametricity for a language called PolyPCF, a version of Plotkin’s PCF [Plotkin 1977] extended with list types and polymorphism. The same technique of building relations from the structure of types is used, but these relations are between terms rather than elements of a denotational semantics. It is therefore necessary to require that the relations respect equivalence in the operational semantics.

To ensure that relations respect equivalence, Pitts introduced two notions for PolyPCF language. Firstly, there is the \( \top \) relation that holds between stacks (which function as term contexts) and terms whenever the stack applied to the term will produce the empty list Nil. Secondly, there is the \( \top \top \)-closure operator for term relations, so called because it involves using the \( \top \) relation twice: once to go from a relation on terms to one on stacks, and once to go back again. As a consequence of the definition, all \( \top \top \) closed relations respect equivalence.

We can summarise the notion of \( \top \top \) closure as follows. Given a relation \( R : \tau_1 \leftrightarrow \tau_2 \) between two types \( \tau_1 \) and \( \tau_2 \), its \( \top \top \)-closure \( R^{\top \top} \) is a relation of the same type. A pair \( (M_1, M_2) \) is in \( R^{\top \top} \) if:

\[
\text{for all stacks } S_1, S_2, \\
\text{if for all } (N_1, N_2) \in R, S_1 \top N_1 \leftrightarrow S_2 \top N_2 \\
\text{then } S_1 \top M_1 \leftrightarrow S_2 \top M_2
\]

Pitts shows that \( \top \top \) closed relations are closed under the relational versions of function space formation, type abstraction and list type formation, thus demonstrating that they are a suitable notion of relation on terms. It can then be shown that any closed term is related to itself, and that open terms are related to themselves provided that the relational interpretations of the free type variables are all \( \top \top \) closed. Applying this theorem is then a matter of instantiating with particular relations, provided these relations can be shown to be \( \top \top \) closed.

2.2 Operational Improvement

In order to reason about the operational efficiency of programs, we need some model of the cost of terms. For call-by-value this is straightforward: in most cases it is enough simply to count the steps taken to evaluate the term to normal form; functions are slightly more complicated as we have both the cost to evaluate the function itself and the cost to compute its result, where the latter may depend on the argument to the function [Shultis 1985]. The result is a semantics of call-by-value \( \lambda \)-terms that can be used to reason about time costs.

The situation for call-by-need is more complicated, however. In this case, a subterm is only evaluated when it is forced, i.e. when its value is required, so the cost to evaluate a term to normal form is a poor measure of efficiency. For example, the terms \( \bot \) and \( 0 \) are both in normal form, but one is potentially more efficient than the other: e.g. head \( \bot \) diverges while head \( 0 \) returns almost immediately. One might consider making this analysis recursive, comparing the cost of evaluating subterms as well, but this method will identify the terms let \( x = M \) in \( (x, x) \) and \( (M, M) \) even though one shares work in a way the other does not.

The solution of Moran and Sands [1999a] is to quantify over evaluation contexts. That is to say, a term \( M \) is improved by another term \( N \), written \( M \triangleright N \), if for all contexts \( C \), the term \( C[M] \) takes at least as many steps to evaluate as \( C[N] \). By taking into account all possible uses of the initial terms, we automatically take into account cost savings introduced by sharing as well as the cost of
computing subterms. Furthermore, this notion of efficiency is compositional by definition, making it amenable to techniques of inequational reasoning.

Moran and Sands’ improvement theory has been used to justify general-purpose program improvements, including the worker/wrapper transformation [Hackett and Hutton 2014]. This takes advantage of the parallels between the rules of call-by-need improvement and those of program equivalence, which makes it possible to adapt proofs of correctness into proofs of improvement. Subsequent work took this idea of having compatible equivalence and improvement proofs and placed it in a generalised setting of preorder-enriched categories [Hackett and Hutton 2015].

Improvement theory and related techniques have seen something of a resurgence in recent years. Breitner [2015] uses an operational approach to proving the safety of the call arity transformation based on counting the number of allocations made. Sergey et al. [2017] use a step-counting approach to show that floating let bindings into a lambda is an improvement provided the lambda expression is used only once. Schmidt-Schauß and Sabel [2015] use an operational semantics based on term rewriting rules to show that a number of local optimisations are improvements, including common subexpression elimination. Finally, Simões et al. [2012] define a cost model for Launchbury’s natural semantics [Launchbury 1993], and use it to prove soundness of a cost analysis.

3 FROM EQUIVALENCE TO IMPROVEMENT

In this section, we show how to build a theory of parametricity that can be used to reason about program improvements. Essentially we want to do for improvement what Pitts did for equivalence. Unfortunately, Pitts’ notion of \(\top\)-closure is too strong for our purposes: because \(\top\)-closed relations must respect program equivalence, they cannot be used to distinguish between observationally equivalent programs with different costs. In a sense, Pitts’ notion of observation is too narrow for our purposes. What we need is a notion of closure that forces relations to respect only cost-equivalence, i.e. when two programs are interchangeable in terms of time costs.

3.1 Call-By-Need PolyPCF

For this work we use a \(\lambda\)-calculus similar to the PolyPCF language studied by Pitts, but with two key differences. Firstly, we add recursive let-bindings to the language and remove the fixed point combinator \(\text{fix}\). We do this because let-binding-based presentations of recursions are better at capturing sharing. Secondly, we only allow functions and constructors to be applied to variables. This latter distinction is useful because it means that all sharing is explicitly introduced in let-bindings; a similar restriction applied to earlier versions of the internal language used in the Glasgow Haskell Compiler, called GHC Core. This makes it easier to reason about call-by-need evaluation strategies. The type and term grammars are defined as follows:

\[
\begin{align*}
\alpha & \in TVar \\
x, y, xs & \in Var \\
A, B & \in Type ::= \alpha \\
& | A \to B \\
& | \text{List} A \\
& | \forall \alpha. A \\
M, N & \in Term ::= \\
x & \\
& | \lambda x. M \\
& | M \ x \\
& | \Lambda \alpha. M \\
& | M \ A \\
& | \text{let} \{ \overline{x} = \overline{M} \} \text{ in } N \\
& | \text{Nil} \\
& | x :: xs \\
& | \text{case } M \text{ of } \{ \text{Nil } \to M_1; x :: xs \to M_2 \}
\end{align*}
\]
The typing rules for the language are given in Figure 1, and comprise one rule for each language construct. Note that the typing context \( \Gamma \) contains both free type variables and type assignments for free term variables. We apply a standard well-formedness constraint to \( \Gamma \).

The operational semantics for this language is a small-step semantics based on Sestoft’s Mark 1 abstract machine [Sestoft 1997], extended to handle type abstraction and application. A machine state is given by a triple \( \langle H, M, S \rangle \) consisting of a heap \( H \) that binds term variables to terms, the current term to be evaluated \( M \), and a stack \( S \) consisting of tokens that describe the context in which the result of evaluating \( M \) is to be used. There are four kinds of stack tokens: variable updates \#x that signal when a variable is to be updated with the result of the computation; applications \( x \) that signify when the result should be given as an argument; type applications \( [A] \) that signify when the result should be given a type argument; and finally, alternatives \( \text{alts} \) that signify that the result should be pattern-matched and branched on.

A term is considered evaluated when it is in \textit{value form}, that is, when it is the empty list, a cons cell or either kind of abstraction. By convention, we denote value terms with letters such as \( V \) and \( W \). The complete set of transition rules for the semantics are given in Figure 2. We assume that all bound variables are unique in the statement of these rules, which can be achieved by \( \alpha \)-renaming to a fresh variable whenever an abstraction is opened.

Now we can define our notions of improvement and cost-equivalence. If for all heaps and stacks \( \langle H, S \rangle \) we have that if \( \langle H, M, S \rangle \) terminates after making \( n \) \textit{LOOKUP} steps then \( \langle H, N, S \rangle \) terminates after \( n \) or fewer \textit{LOOKUP} steps, we say \( M \) is \textit{improved} by \( N \), written \( M \preceq N \). We only need to count \textit{LOOKUPS} as these are an appropriate measure of the total cost of evaluating a term. That counting \textit{LOOKUP} steps is sufficient was demonstrated by Moran and Sands [1999a], who proved that the total number of steps is bound by a linear function of the number of \textit{LOOKUPS}. If \( M \) and \( N \) both improve each other we say that they are \textit{cost-equivalent}, written \( M \cong N \).

Fig. 1. The typing rules for our language
3.2 Machine-Closure for Relations

Now that we have our language, we must develop a notion of relations between terms of that language. These relations will be between terms of particular types in the language; if $R$ relates terms of type $A$ to terms of type $B$, we say that $R$ has the type $A \leftrightarrow B$. However, not all relations will have the desired properties of respecting costs. We must therefore have some notion of permissible relations that are all guaranteed to have this property.

As noted above, we cannot use Pitts’ notion of $\top \top$-closure as it identifies all equivalent terms regardless of cost. However, we can modify $\top \top$-closure to produce another notion of closure that is more suitable for our purposes. We obtain this notion by replacing Pitts’ stacks by heap and stack pairs that capture all of the state of the abstract machine besides the term itself, and adapting his $\top$ relation to take costs into account. We also follow Voigtländer and Johann [2007] and use one-directional implication in our definition, as this is more useful for reasoning about program orderings. We call the resulting notion of closure machine-closure. Given a relation $R : A \leftrightarrow B$, two terms $M_1 : A$ and $M_2 : B$ are related by the machine-closure of $R$, written $R^n$, if:

\[
\begin{align*}
\langle H, \{ x = M \}, x, S \rangle & \rightarrow \langle H, M, \#x : S \rangle \quad \{ \text{LOOKUP} \} \\
\langle H, V, \#x : S \rangle & \rightarrow \langle H, \{ x = V \}, V, S \rangle \quad \{ \text{UPDATE} \} \\
\langle H, M x, S \rangle & \rightarrow \langle H, M, x : S \rangle \quad \{ \text{UNWIND} \} \\
\langle H, \lambda x.M, y : S \rangle & \rightarrow \langle H, M \ [ y/x ], S \rangle \quad \{ \text{SUBST} \} \\
\langle H, M A, S \rangle & \rightarrow \langle H, M, [ A ] : S \rangle \quad \{ \text{TYPEUNWIND} \} \\
\langle H, \Delta \alpha.M, [ A ] : S \rangle & \rightarrow \langle H, M [ A/\alpha ], S \rangle \quad \{ \text{TYPESUBST} \} \\
\langle H, \text{case } M \text{ of } \text{alts} \rangle & \rightarrow \langle H, M, \text{alts} : S \rangle \quad \{ \text{CASE} \} \\
\langle H, \text{Nil}, \{ \text{Nil} \rightarrow N_1; x :: xs \rightarrow N_2 \} : S \rangle & \rightarrow \langle H, N_1, S \rangle \quad \{ \text{BRANCHNIL} \} \\
\langle H, y :: ys, \{ \text{Nil} \rightarrow N_1; x :: xs \rightarrow N_2 \} : S \rangle & \rightarrow \langle H, N_2 \ [ y/x, ys/xs ], S \rangle \quad \{ \text{BRANCHCONS} \} \\
\langle H, \text{let } \{ x = \tilde{M} \} \text{ in } N, S \rangle & \rightarrow \langle H, \{ x = \tilde{M} \}, N, S \rangle \quad \{ \text{LETREC} \}
\end{align*}
\]

By $\langle H, N, S \rangle \downarrow_n$, we mean that if we start the abstract machine in state $\langle H, N, S \rangle$, then it will take at most $n$ LOOKUP steps to finish evaluating. We can prove a number of useful theorems about machine-closure. Firstly, it is monotone, i.e. if $R \subseteq S$ then $R^n \subseteq S^m$. This can be shown as follows: given a pair $(M_1, M_2)$ in $R^m$, for any heap and stack pairs $(H_1, S_1)$ and $(H_2, S_2)$ such that for any $(N_1, N_2) \in R$, $(H_1, N_1, S_1) \downarrow_n \Rightarrow (H_2, N_2, S_2) \downarrow_n$, we know that $(H_1, M_1, S_1) \downarrow_n \Rightarrow (H_2, M_2, S_2) \downarrow_n$. Any heap and stack pairs that have this property for $S$ will also have the same property for $R$ (as there are more pairs in $S$), so $(M_1, M_2)$ must be in $S^m$ as well.

Secondly, we can show that machine-closure really is a notion of closure, that is that $R \subseteq S^m = (R^m)^m$. That $R$ is smaller than $R^m$ follows from the definition, as when $(M_1, M_2) \in R$ then the criterion for membership in $R^m$ becomes a tautology. Then, as any heap and stack pairs that identify pairs in $R$ will also identify pairs in $R^m$ by definition, we can conclude that any pair in $(R^m)^m$ will also be in $R^m$, so our closure operation is idempotent.

Finally, we show that machine-closed relations respect improvement. Assume we have a pair $(M_1, M_2) \in R^m$, and that we know $M'_1 \triangleright M_1$ and $M_2 \triangleright M'_2$. We can then conclude that for any heap

\[
\text{Fig. 2. The call-by-need abstract machine}
\]
and stack pairs \( \langle H_1, S_1 \rangle \) and \( \langle H_2, S_2 \rangle \) such that for any \((N_1, N_2) \in R\), we have \( \langle H_1, N_1, S_1 \rangle \downarrow_n \Rightarrow \langle H_2, N_2, S_2 \rangle \downarrow_n \), as demonstrated by the following simple calculation:

\[
\begin{align*}
\langle H_1, M'_1, S_1 \rangle \downarrow_n & \quad \Rightarrow \quad \{ \text{ improvement } \}
\langle H_1, M_1, S_1 \rangle \downarrow_n \\
\langle H_2, M_1, S_1 \rangle & \downarrow_n \\
\langle H_2, M'_2, S_2 \rangle & \downarrow_n \\
\end{align*}
\]

Therefore, we can conclude that \((M'_1, M'_2)\) is also in \(R^m\).

### 3.3 Actions on Relations

Now that we have defined an appropriate notion of relation on terms, we must in turn define how the type constructors of our language act on those relations. In particular, we need these actions to preserve the machine-closure property, in order that any relation built out of machine-closed relations using these constructors will also be machine-closed. We can define relational actions for the three type constructors of our language as follows:

- **Function space formation.** Given two relations, \( R : T_1 \leftrightarrow T'_1 \) and \( S : T_2 \leftrightarrow T'_2 \), we can define the relation \( R \rightarrow S : (T_1 \rightarrow T_2) \leftrightarrow (T'_1 \rightarrow T'_2) \). Two terms \( M : T_1 \rightarrow T_2 \) and \( M' : T'_1 \rightarrow T'_2 \) are related by \( R \rightarrow S \) iff for all collections of bindings \( B, B' \), we have the following implication:

  \[
  \begin{align*}
  & (\text{let } B \text{ in } y, \text{ let } B' \text{ in } y) \in R \\
  \Rightarrow & \quad (\text{let } B \text{ in } M y, \text{ let } B' \text{ in } M' y) \in S \\
  \end{align*}
  \]

- **List type formation.** Given \( R : T \leftrightarrow T' \), we can define the relation \( List R : List T \leftrightarrow List T' \) as the least fixed point of the following equation, for all collections of bindings \( B, B' \):

  \[
  \text{List } R = (\{(\text{Nil, Nil})\} \cup \\
  \{ (\text{let } B \text{ in } y :: ys, \text{ let } B' \text{ in } y :: ys) \\
  | (\text{let } B \text{ in } y, \text{ let } B' \text{ in } y) \in R \\
  | (\text{let } B \text{ in } ys, \text{ let } B' \text{ in } ys) \in \text{List } R \})^m
  \]

  The fixed point is guaranteed to exist because all operations in the definition are monotone and relations \( List A \leftrightarrow List A' \) form a lattice. Note that our approach here differs from Pitts, who uses greatest fixed point; we use a least fixed point as induction is useful later on.

- **Type abstraction.** Given a family of functions \( R \) indexed over types \( T_1, T'_1 \) that map relations of type \( T_1 \leftrightarrow T'_1 \) to relations of type \( T_2 \ [T_1] \leftrightarrow T'_2 \ [T'_1] \), we can define the relation \( \forall r. \ R (r) \). Two terms \( M : \forall \alpha.T_2 \ [\alpha] \) and \( M' : \forall \alpha.T'_2 \ [\alpha] \) are related by \( \forall r. \ R (r) \) if and only if:

  \[
  \begin{align*}
  & \forall \alpha, T_1, T'_1 \in \text{Type, } S : T_1 \leftrightarrow T'_1 \quad . \\
  & (M T_1, M' T'_1) \in R (S)
  \end{align*}
  \]

These three definitions are based on the relational actions from Pitts [2000], adapted to our setting. We use the technique of adding bindings to get around the limitation of applying terms and constructors only to variables, as well as to take account of sharing. There is an implicit requirement in the above definitions that all the bindings are type-correct.

The next step is to prove that all three of these constructions preserve the notion of machine-closure. That this is the case for the \( List \) construction is immediate, as any fixed-point of the defining equation is by definition machine-closed. For the other two, the proof is a little more involved:
Lemma 3.1 (Function Space Preserves Machine-Closure).

(i) Given bindings \( B, B' \) and heap and stack pairs \( \langle H, S \rangle, \langle H', S' \rangle \), if we have

\[
\begin{align*}
(a) & \quad (\text{let } B \text{ in } y, \text{ let } B' \text{ in } y) \in R \\
(b) & \quad \forall (N, N') \in R'. \; (H, N, S) \downarrow_n \Rightarrow (H', N', S') \downarrow_n
\end{align*}
\]

then we also have

\[
\forall (M, M') \in R \rightarrow R'. \; \langle H + B, M, y : S \rangle \downarrow_n \Rightarrow (H' + B', M', y : S') \downarrow_n
\]

(ii) Given machine-closed relations \( R \) and \( R' \), then \( R \rightarrow R' \) is also machine-closed.

Proof. For part (i), take arbitrary \( (M, M') \in R \rightarrow R' \). We note that \( (\text{let } B \text{ in } M \; y, \text{ let } B' \text{ in } M' \; y) \in R' \) by assumption (a) and the definition of \( R \rightarrow R' \), and reason as follows:

\[
\begin{align*}
\langle H + B, M, y : S \rangle \downarrow_n \\
\quad \Leftrightarrow \quad \{ \text{Unwind and Letrec steps are free} \} \\
\quad \Leftrightarrow \quad \{ \text{Unwind and Letrec steps are free} \} \\
\quad \Rightarrow \quad \{ \text{part (i), definition of machine-closure} \} \\
\quad \Leftrightarrow \quad \{ \text{Unwind and Letrec steps are free} \} \\
\langle H', \text{ let } B' \text{ in } M' \; y, S' \rangle \downarrow_n
\end{align*}
\]

For part (ii), we assume \( (M, M') \in (R \rightarrow R')^m \) for machine-closed \( R \) and \( R' \), and aim to prove \( (M, M') \in R \rightarrow R' \). Assuming type-correct bindings \( B \) and \( B' \) such that \( (\text{let } B \text{ in } y, \text{ let } B' \text{ in } y) \in R \), we must show that \( (\text{let } B \text{ in } M \; y, \text{ let } B' \text{ in } M' \; y) \in R' \). Letting \( \langle H, S \rangle, \langle H', S' \rangle \) be arbitrary heap-stack pairs such that \( \forall (N, N') \in R' \). \( (H, N, S) \downarrow_n \Rightarrow (H', N', S') \downarrow_n \), we have:

\[
\begin{align*}
\langle H, \text{ let } B \text{ in } M \; y, S \rangle \downarrow_n \\
\quad \Leftrightarrow \quad \{ \text{Unwind and Letrec steps are free} \} \\
\quad \Leftrightarrow \quad \{ \text{Unwind and Letrec steps are free} \} \\
\quad \Rightarrow \quad \{ \text{part (i), definition of machine-closure} \} \\
\quad \Leftrightarrow \quad \{ \text{Unwind and Letrec steps are free} \} \\
\langle H', \text{ let } B' \text{ in } M' \; y, S \rangle \downarrow_n
\end{align*}
\]

Because these heap-stack pairs were arbitrary, we can conclude \( (\text{let } B \text{ in } M \; y, \text{ let } B' \text{ in } M' \; y) \in R'^m \), which implies the desired result by the assumption that \( R' \) is machine-closed.

Lemma 3.2 (Type Abstraction Preserves Machine-Closure).

(i) Given heap and stack pairs \( \langle H, S \rangle, \langle H', S' \rangle \), if we have

\[
\forall r : T \leftrightarrow T', (N, N') \in R \; (r) . \\
\langle H, N, S \rangle \downarrow_n \Rightarrow (H', N', S') \downarrow_n
\]

then we also have

\[
\forall (M, M') \in \forall r \; R \; (r) . \\
\langle H, M, [T] : S \rangle \downarrow_n \Rightarrow (H, M', [T'] : S') \downarrow_n
\]
(ii) For any types $T_2$, $T'_2$, given a family of functions $R$ indexed over types $T_1$, $T'_1$ that map relations of type $T_1 \leftrightarrow T'_1$ to relations of type $T_2 \leftrightarrow T'_2$, if all the relations $R (r)$ are machine-closed then the relation $\forall r. R (r)$ will also be machine-closed.

**Proof.** Similarly to Lemma 3.1, but with TYPEUNWIND steps rather than UNWIND/LETREC. □

### 3.4 The Logical Relation

Given the above relational actions, it is possible to define a mapping from types to relations, abstracted over the interpretations of type variables. By convention this mapping is called the logical relation, denoted $\Delta$, and is defined recursively on the structure of types as follows:

$$
\begin{align*}
\Delta (\overline{R} / \overline{\alpha}) (\alpha_i) &= R_i \\
\Delta (\overline{R} / \overline{\alpha}) (T_1 \rightarrow T_2) &= \Delta (\overline{R} / \overline{\alpha}) (T_1) \rightarrow \Delta (\overline{R} / \overline{\alpha}) (T_2) \\
\Delta (\overline{R} / \overline{\alpha}) (\forall \alpha. T) &= \forall \alpha. \Delta (r^m / \alpha, \overline{R} / \overline{\alpha}) (T) \\
\Delta (\overline{R} / \overline{\alpha}) (\text{List } T) &= \text{List } (\Delta (\overline{R} / \overline{\alpha}) (T))
\end{align*}
$$

A simple proof by induction will show that $\Delta (\overline{R} / \overline{\alpha}) (T)$ is machine-closed whenever all the $\overline{R}$ are. Furthermore, another proof by induction will show that if all of $\overline{R}$ are subsets of $\triangleright$ then $\Delta (\overline{R} / \overline{\alpha}) (T)$ will also be a subset of $\triangleright$. This corresponds to Reynolds’ identity extension lemma [Reynolds 1983].

### 3.5 The Abstraction Theorem (for Non-Recursive Programs)

Now we can prove a version of the abstraction theorem. In this section we will prove the theorem for a non-recursive version of our language, but in the next section we will show how to extend the theorem to deal with recursive programs. We proceed in this manner because dealing with recursion requires applying some limitations to our language that we do not wish to apply to non-recursive programs. For the purposes of this section, we use the following weaker version of the LET rule that disallows the use of recursion:

$$
\begin{align*}
\Gamma \vdash M_1 : T_1 & \quad \ldots \quad \Gamma \vdash M_n : T_n \\
\Gamma, x_1 : T_1, \ldots, x_n : T_n & \vdash M : T \\
\Gamma & \vdash \text{let } \{ x_1 = M_1; \ldots; x_n = M_n \} \text{ in } M : T \text{ \textsc{let'}}
\end{align*}
$$

Now we state the abstraction theorem:

**Theorem 3.3 (Abstraction for Non-Recursive Programs).** Given a closed term $M$ and closed type $A$ such that $\vdash M : A$ (using the weaker $\textsc{let'}$ rule rather than $\textsc{let}$), we have $(M, M) \in \Delta () (A)$.

To prove this theorem we must actually prove a stronger theorem, extending the logical relation $\Delta$ to terms with free variables. Assuming $\Gamma = \overline{\alpha}$, $x_1 : T_1$, $\ldots$, $x_n : T_n$, we define $\Gamma \vdash M \\Delta \ M' : T$ to mean that for all bindings $B, B'$ that close the terms $M, M'$ respectively, and machine-closed relations $\overline{R} : T_i \leftrightarrow T'_i$ (where $\text{length } \overline{R} = \text{length } \overline{\alpha}$), we have:

$$
(\forall i \in [1 \ldots n] . \\
(\text{let } B \text{ in } x_i, \quad \text{let } B' \text{ in } x_i) \in \Delta (\overline{R} / \overline{\alpha}) (T_i)) \Rightarrow \\
(\text{let } B \text{ in } M [\overline{T} / \overline{\alpha}], \quad \text{let } B' \text{ in } M' [\overline{T'} / \overline{\alpha}]) \in \Delta (\overline{R} / \overline{\alpha}) (T)
$$
Theorem 3.3 then follows immediately from the following lemma:

**Lemma 3.4.** Given a context Γ, term M and type A such that Γ ⊢ M : T (using the weaker Let’ rule instead of Let), we have that Γ ⊢ M Δ M : T. (The proof is given in the Appendix.)

4 DEALING WITH RECURSIVE BINDINGS

In order to extend our treatment to deal with the full Let rule, we must have some way to reason about the behaviour of recursive bindings. The usual technique, as used by Pitts [2000] and Moran and Sands [1999a], is to define some notion of unwinding and to consider a fixed point as a limit to the sequence of increasingly deep unwindings. For example, the meaning of \( \text{let } x = M \ [x] \text{ in } N \ [x] \) is generally taken to be the limit of the sequence:

\[
\begin{align*}
\text{let } x_0 &= ⊥ \text{ in } N \ [x_0] \\
\text{let } x_1 &= M \ [x_0]; x_0 = ⊥ \text{ in } N \ [x_1] \\
\text{let } x_2 &= M \ [x_1]; x_1 = M \ [x_0]; x_0 = ⊥ \text{ in } N \ [x_2] \\
\end{align*}
\]

However, this technique cannot be applied in a call-by-need setting, as there is work shared in the recursive definition that is not shared in any of the unwindings. We follow the approach taken by Moran and Sands [1999a] and restrict ourselves to cases where the right-hand side is a value, in which case the problem does not arise. We return to this assumption in the concluding section.

First of all, we prove the following lemma, known as unwinding.

**Lemma 4.1 (Unwinding).**

(i) A machine state \( \langle H \ (x = V), N, S \rangle \) terminates in \( k \) steps iff there is some non-negative integer \( n \) such that the machine state \( \langle H \ (x =^n V), N, S \rangle \) terminates in \( k \) steps.

(ii) Given a pair of closed terms \( \text{let } x = V \text{ in } N \) and let \( x = V' \text{ in } N' \) and a machine-closed relation \( R \), the membership relation (let \( x = V \text{ in } N \), let \( x = V' \text{ in } N' \)) ∈ \( R \) holds iff the membership relation (let \( x =^n V \text{ in } N \), let \( x =^n V' \text{ in } N' \)) ∈ \( R \) holds for all non-negative \( n \).

**Proof.** Part (i) follows easily from the fact that a binding can only recurce finitely many times in a given execution. For part (ii), the \( ⇒ \) direction follows from (i) and the definition of machine-closure. For the \( ⇐ \) direction, suppose we have heap and stack pairs \( \langle H_1, S_1 \rangle \) and \( \langle H_2, S_2 \rangle \) such that for any \( (P, P') \in R \), we have \( \langle H_1, P, S_1 \rangle \downarrow_n ⇒ \langle H_2, P', S_2 \rangle \downarrow_n \). We prove that \( \langle H_1, \text{let } x = V \text{ in } N, S_1 \rangle \downarrow_n ⇒ \langle H_2, \text{let } x = V' \text{ in } N', S_2 \rangle \downarrow_n \) by case analysis on the termination behaviour of the left-hand side of this result, considering the two cases in turn:

- **Termination.** If \( \langle H_1, \text{let } x = V \text{ in } N, S_1 \rangle \) terminates in \( k \) steps, then by part (i) there must be some non-negative integer \( n \) such that \( \langle H_1, \text{let } x =^n V \text{ in } N, S_1 \rangle \) terminates in \( k \) steps, in which case we also have that \( \langle H_2, \text{let } x =^n V' \text{ in } N', S_2 \rangle \) terminates in \( k \) steps, and then by part (i) we can conclude that \( \langle H_2, \text{let } x = V' \text{ in } N', S_2 \rangle \) terminates in \( k \) steps.

- **Non-termination.** If \( \langle H_1, \text{let } x = V \text{ in } N, S_1 \rangle \) does not terminate, neither can \( \langle H_2, \text{let } x = V' \text{ in } N', S_2 \rangle \), because if it did the reasoning above could be reversed. Hence, \( \langle H_1, \text{let } x = V \text{ in } N, S_1 \rangle \downarrow_n ⇒ \langle H_2, \text{let } x = V' \text{ in } N', S_2 \rangle \), and because this reasoning is generic in \( H_1, H_2, S_1, S_2 \), we have \( \text{let } x = V \text{ in } N \), let \( x = V' \text{ in } N' \) ∈ \( R \) by machine-closure of \( R \).

Now we can extend our proof of Lemma 3.4 to deal with recursive bindings. Only the Let case differs. For simplicity, we illustrate the case for a single recursive binding, but the result can be generalised easily to the case of multiple recursive bindings.

J. Hackett and G. Hutton
**Lemma 4.2 (Recursive bindings).** If \( \Gamma, x : T_1 \vdash V \Delta V : T_1 \) and \( \Gamma, x : T_1 \vdash N \Delta N : T_2 \) both hold, then \( \Gamma \vdash \text{let } x = V \text{ in } N \Delta \text{ let } x = V \text{ in } N : T_2 \) will also hold.

**Proof.** Let \( B \) and \( B' \) be bindings that close \( \text{let } x = M \text{ in } N \) (but don’t contain \( x \); we can ensure this with alpha-renaming) and let \( \bar{R} : \bar{T} \leftrightarrow \bar{T}' \) be a list of machine-closed relations with length equal to the number of free type variables in \( \Gamma \). Assume that for any \( y : A \in \Gamma \), we have \( \text{(let } B \text{ in } y, \text{ let } B' \text{ in } y') \in \Delta (\bar{R}/\bar{a}) \) (A). We aim to prove \( \text{(let } B \text{ in } x = V [\bar{T}/\alpha] \text{ in } N [\bar{T}/\alpha], \text{ let } B' \text{ in } x = V [\bar{T'}/\alpha] \text{ in } N [\bar{T'}/\alpha]) \in \Delta (\bar{R}/\bar{a}) \) (T2).

It suffices to prove \( \text{(let } B; x = V [\bar{T}/\alpha] \text{ in } x, \text{ let } B'; x = V [\bar{T'}/\alpha] \text{ in } x) \in \Delta (\bar{R}/\bar{a}) \) (T1), which implies our goal by the assumption \( \Gamma, x : T_1 \vdash \text{let } x = V [\bar{T}/\alpha] \text{ in } N [\bar{T}/\alpha], \text{ let } B' \text{ in } x = V [\bar{T'}/\alpha] \text{ in } N [\bar{T'}/\alpha] \in \Delta (\bar{R}/\bar{a}) \) (T1), First, we prove that for all \( n \), we have \( \text{let } B; x =^n V [\bar{T}/\alpha] \text{ in } x, \text{ let } B'; x =^n V [\bar{T'}/\alpha] \text{ in } x \in \Delta (\bar{R}/\bar{a}) \) (T1). We do this by induction on \( n \). The base case is simple, as both sides diverge and all machine-closed relations relate diverging terms to each other. For the inductive case, we reason as follows. Our induction hypothesis is that \( \text{let } B; x =^n V [\bar{T}/\alpha] \text{ in } x, \text{ let } B'; x =^n V [\bar{T'}/\alpha] \text{ in } x \in \Delta (\bar{R}/\bar{a}) \) (T1), so the assumption \( \Gamma, x : T_1 \vdash N \Delta V : T_1 \) we know that \( \text{let } B; x =^n V [\bar{T}/\alpha] \text{ in } \bar{V}, \text{ let } B'; x =^n V [\bar{T'}/\alpha] \text{ in } \bar{V} \in \Delta (\bar{R}/\bar{a}) \) (T1). But these terms are cost-equivalent to \( \text{let } B; x =^{n+1} V [\bar{T}/\alpha] \text{ in } \bar{V}, \text{ let } B'; x =^{n+1} V [\bar{T'}/\alpha] \text{ in } \bar{V} \), so we are done.

Finally, we apply Lemma 4.1 to give \( \text{let } B; x = V [\bar{T}/\alpha] \text{ in } x, \text{ let } B'; x = V [\bar{T'}/\alpha] \text{ in } x) \in \Delta (\bar{R}/\bar{a}) \) (T1), which completes the proof of this lemma.

**5 APPLICATIONS**

In this section, we apply our abstraction theorem to a range of examples to derive *free theorems for improvement*: theorems about the operational time efficiency of terms that rely only on their polymorphic types. Before we do this, however, we need ways to construct interesting machine-closed relations. We start by introducing a tick token \( \checkmark \) to our stacks: it is safe to do this because our reasoning was generic in the possible stack items. We introduce the following transition rule for \( \checkmark \) to the abstract machine semantics that we defined in Figure 2

\[ \langle H, M, \checkmark : S \rangle \rightarrow \langle H, M, S \rangle \{ \text{Tick} \} \]

and count these steps as well in our cost semantics, so the total cost of evaluation is now given by the number of *Lookup* and *Tick* steps. We can then formulate the following theorem:

**Theorem 5.1.** Given a context \( C \) such that \( C [\bot] \models \bot \), the following relations are machine-closed:

(i) \( \{ (M, M') | C [M] \models M' \} \)

(ii) \( \{ (M, M') | M \models C [M'] \} \)

**Proof.** We take an arbitrary pair \( (M, M') \) in the machine-closure of the relation and show it is in the original relation. For (i) we must show \( C [M] \models M' \), which by definition is equivalent to:

\[ \forall (H, S), \langle H, C [M], S \rangle \downarrow_n \Rightarrow \langle H, M', S \rangle \downarrow_n \]

Take an arbitrary \( \langle H, S \rangle \). Starting in the state \( \langle H, C [x], S \rangle \) where \( x \) is fresh, the machine will either evaluate to \( \langle H', x, s' \rangle \) in \( k \) steps, or it will diverge, because otherwise it would violate the assumption \( C [\bot] \models \bot \). If \( \langle H, C [x], S \rangle \) diverges then so does \( \langle H, C [M], S \rangle \), so \( \langle H, C [M], S \rangle \downarrow_n \Rightarrow \langle H, M', S \rangle \downarrow_n \) is vacuously true and we are done. Otherwise, take an arbitrary \( (N, N') \) such that \( C [N] \models N' \). We know that \( \langle H, C [N], S \rangle \) evaluates to \( \langle H', N, S' \rangle \) in \( k \) steps by construction, so we can conclude \( \langle H', N, k \checkmark : S' \rangle \downarrow_n \Rightarrow \langle H, C [N], S \rangle \downarrow_n \) (where \( k \checkmark \) represents \( k \) copies of \( \checkmark \)). From this and the assumption \( C [N] \models N' \) we can conclude \( \langle H', N, k \checkmark : S' \rangle \downarrow_n \Rightarrow \langle H, N', S \rangle \downarrow_n \).


Because $N$ and $N'$ were arbitrary, by the definition of machine-closure this implies $\langle H', M, k \triangleright : S' \rangle \downarrow \Rightarrow \langle H, M', S \rangle \downarrow \Rightarrow$, which by construction implies $\langle H, C [M], S \rangle \downarrow \Rightarrow \langle H, M', S \rangle \downarrow \Rightarrow$.

The second case (ii) follows the same pattern. □

5.1 Example: Fixed Point Fusion

Consider a term $\text{fix} : \forall \alpha. (\alpha \rightarrow \alpha) \rightarrow \alpha$. The abstraction theorem implies the following:

for any bindings $B_2, B'_2$ and machine-closed relation $R : T \leftrightarrow T'$,
if for any bindings $B_1, B'_1$ we have
$(\text{let } B_1 \text{ in } x, \text{ let } B'_1 \text{ in } x) \in R \Rightarrow$
$(\text{let } B_1 \text{ in } (\text{let } B_2 \text{ in } f) x, \text{ let } B'_1 \text{ in } (\text{let } B'_2 \text{ in } f) x) \in R$
then $(\text{let } B_2 \text{ in } \text{fix } T f, \text{ let } B'_2 \text{ in } \text{fix } T' f) \in R$

Simplifying, by letting $B_2 = \{ f = M_1 \}, B'_2 = \{ f = M_2 \}$, we obtain:

for any machine-closed relation $R$,
if for any bindings $B_1, B'_1$ we have
$(\text{let } B_1 \text{ in } x, \text{ let } B'_1 \text{ in } x) \in R \Rightarrow (\text{let } B_1 \text{ in } M_1 x, \text{ let } B'_1 \text{ in } M_2 x) \in R$
then $(\text{let } f = M_1 \text{ in } \text{fix } T f, \text{ let } g = M_2 \text{ in } \text{fix } T' g) \in R$

Now if we let $R = \{ (N, N') | C [N] \triangleright N' \}$, we obtain:

for any context $C$ such that $C [\bot] \triangleright \bot$,
if for any bindings $B_1, B'_1$ we have
$\text{let } B_1 \text{ in } C [x] \triangleright \text{let } B'_1 \text{ in } x \Rightarrow \text{let } B_1 \text{ in } C [M_1 x] \triangleright \text{let } B'_1 \text{ in } M_2 x$
then $\text{let } f = M_1 \text{ in } C [\text{fix } T f] \triangleright \text{let } g = M_2 \text{ in } \text{fix } T' g$

Finally, we observe that the precondition here is implied by $\text{let } B_1 \text{ in } C [M_1 x] \triangleright \text{let } B_1, y = C [x] \text{ in } M_2 y$, so we obtain the following result:

for any context $C$ such that $C [\bot] \triangleright \bot$,
if for any bindings $B_1, B'_1$ we have
$\text{let } B_1 \text{ in } C [M_1 x] \triangleright \text{let } B_1, y = C [x] \text{ in } M_2 y$
then $\text{let } f = M_1 \text{ in } C [\text{fix } T f] \triangleright \text{let } g = M_2 \text{ in } \text{fix } T' g$

Thus we see that our abstraction theorem implies an improving version of the standard fixed point fusion rule. We can also take $R = \{ (N, N') | N \triangleright C [N'] \}$ to obtain a version of this rule where the improvement goes in the other direction:

for any context $C$ such that $C [\bot] \triangleright \bot$,
if for any bindings $B_1, B'_1$ we have
$\text{let } B_1, y = C [x] \text{ in } M_1 y \triangleright \text{let } B_1 \text{ in } C [M_2 x]$
then $\text{let } f = M_1 \text{ in } \text{fix } T f \triangleright \text{let } g = M_2 \text{ in } C [\text{fix } T' g]$

The first rule tells us that fusion is an improvement when the fusion precondition is an improvement in one direction. The second tells us that reverse fusion (“fission”) is an improvement when the fusion precondition is an improvement in the other direction. Together, these imply that if the fusion precondition is a cost equivalence, then fusion itself is a cost equivalence.

5.2 Example: Map Fusion

For our second example, consider the $\text{map}$ function on lists:

$\text{map} : \forall \alpha \beta. (\alpha \rightarrow \beta) \rightarrow \text{List } \alpha \rightarrow \text{List } \beta$

$\text{map} = \Lambda \alpha \beta. \lambda f \text{ xs } \rightarrow \text{case } \text{xs } \text{ of}$
We proceed by case analysis on the elements of this relation. First of all, the pair \( X \) is empty, so is clearly included on the right hand side. For the inductive step, we assume \( X \subseteq \{(L_1, L_2) \mid \text{let } \text{map} = \ldots ; f = \lambda x.\Gamma [x]; l = L_1 \text{ in } \text{map} \ T_1, T_2 \ f \ l \triangleright L_2\} \) and try to prove the same for the right hand side of the recursive equation. Because both sides of our inclusion are machine-closed, it suffices to show that the following relation is included in the right hand side:

\[
\{ (\text{Nil} , \text{Nil}) \} \cup \\
\{ \text{let } B \text{ in } y :: ys, \text{let } B' \text{ in } y :: ys \} \\
\{ \text{let } B \text{ in } \Gamma [y] \triangleright B' \text{ in } y, \} \\
\{ \text{let } B \text{ in } ys, \text{let } B' \text{ in } ys \in X \}
\]

We proceed by fixed point induction on \( X \). In the base case, \( X \) is empty, so is clearly included on the right hand side. For the inductive step, we assume \( \Gamma \vdash M : T_1 \Rightarrow \Gamma + \Gamma [M] : T_2 \). However, it is not clear that the \( \text{List} \) action applied to a relation based on a context will still be a relation based on a context. To proceed, we need the following result:

**Theorem 5.2.** Given a context \( \Gamma \) such that \( \Gamma [\bot] \not\models \bot \) and \( \Gamma + M : T_1 \Rightarrow \Gamma + \Gamma [M] : T_2 \),

\[
\text{List} \; \{ (M_1, M_2) \mid \Gamma [M_1] \triangleright M_2 \} \\
\subseteq \\
\{ (L_1, L_2) \mid \text{let } \text{map} = \ldots ; f = \lambda x.\Gamma [x]; l = L_1 \text{ in } \text{map} \; T_1, T_2 \ f \ l \triangleright L_2\}
\]

**Proof.** From the definition of \( \text{List} \), we know that \( \text{List} \; \{ (M_1, M_2) \; \Gamma[M1] \; M2 \} \) is the greatest solution to the following equation:

\[
X = \{ (\text{Nil} , \text{Nil}) \} \cup \\
\{ (\text{let } B \text{ in } y :: ys, \text{let } B' \text{ in } y :: ys \} \\
\{ \text{let } B \text{ in } \Gamma [y] \triangleright B' \text{ in } y, \} \\
\{ \text{let } B \text{ in } ys, \text{let } B' \text{ in } ys \in X \}
\]

We proceed by fixing point induction on \( X \). In the base case, \( X \) is empty, so is clearly included on the right hand side. For the inductive step, we assume \( X \subseteq \{ (L_1, L_2) \mid \text{let } \text{map} = \ldots ; f = \lambda x.\Gamma [x]; l = L_1 \text{ in } \text{map} \; T_1, T_2 \ f \ l \triangleright L_2\} \) and try to prove the same for the right hand side of the recursive equation. Because both sides of our inclusion are machine-closed, it suffices to show that the following relation is included in the right hand side:

\[
\{ (\text{Nil} , \text{Nil}) \} \cup \\
\{ (\text{let } B \text{ in } y :: ys, \text{let } B' \text{ in } y :: ys \} \\
\{ \text{let } B \text{ in } \Gamma [y] \triangleright B' \text{ in } y, \} \\
\{ \text{let } B \text{ in } ys, \text{let } B' \text{ in } ys \in X \}
\]

We proceed by case analysis on the elements of this relation. First of all, the pair \( (\text{Nil} , \text{Nil}) \) is included in \( \{ (L_1, L_2) \mid \text{let } \text{map} = \ldots ; f = \lambda x.\Gamma [x]; l = L_1 \text{ in } \text{map} \; T_1, T_2 \ f \ l \triangleright L_2\} \) simply by applying the definition of \( \text{map} \). In turn, for \( \text{let } B \text{ in } y :: ys, \text{let } B' \text{ in } y :: ys \) we know the following two properties from the way in which the relation is constructed:

\[
\text{let } B \text{ in } \Gamma [y] \triangleright \text{let } B' \text{ in } y \\
\text{let } B \text{ in } ys, \text{let } B' \text{ in } ys \in X
\]

From the inductive hypothesis, the second of these two properties implies:
We then reason as follows:

\[
\text{let } \text{map} = \ldots; f = \lambda x. C \ [x]; B\ in\ \text{map}\ T_1\ T_2\ f\ ys \mapsto \text{let } B'\ in\ ys
\]

We note that

\[
\text{This result states that if the function } B\ \text{for any bindings } B_2, B'_2,\ \text{we have}
\]

\[
\text{let } B_1\ in\ C\ [(\text{let } B_2\ in\ f)\ x] \mapsto \text{let } B_1\ in\ (\text{let } B'_2\ in\ f)\ x
\]

\[
\text{and (let } B_2\ in\ xs, \text{let } B'_2\ in\ xs) \in \text{List } R_1
\]

\[
\text{then (let } B_2\ in\ \text{map } T_1\ T_2\ f\ xs, \text{let } B'_2\ in\ \text{map } T_1\ T'_2\ f\ xs) \in \text{List } R_2
\]

Hence this pair is also included in the right-hand side.

Now we can specialise our free theorem for \text{map} to something more immediately useful. Letting \(T_1 = T'_1, R_1 = \{(M_1, M_2) \mid M_1 \preceq M_2\}\) and \(R_2 = \{(M_1, M_2) \mid C \ [M_1] \preceq M_2\}\), we obtain the following:

\[
\text{for any bindings } B_2, B'_2,
\]

\[
\text{if for any bindings } B_1, B'_1\ \text{we have}
\]

\[
\text{let } B_1\ in\ C\ [(\text{let } B_2\ in\ f)\ x] \mapsto \text{let } B_1\ in\ (\text{let } B'_2\ in\ f)\ x
\]

\[
\text{and (let } B_2\ in\ xs, \text{let } B'_2\ in\ xs) \in \text{List } R_1
\]

\[
\text{then (let } B_2\ in\ \text{map } T_1\ T_2\ f\ xs, \text{let } B'_2\ in\ \text{map } T_1\ T'_2\ f\ xs) \in \text{List } R_2
\]

We note that \text{List } R_1\ contains the improvement relation, and simplify:

\[
\text{for any bindings } B_2, B'_2,
\]

\[
\text{if for any bindings } B_1, B'_1\ \text{we have}
\]

\[
\text{let } B_1\ in\ C\ [(\text{let } B_2\ in\ f)\ x] \mapsto \text{let } B_1\ in\ (\text{let } B'_2\ in\ f)\ x
\]

\[
\text{and (let } B_2\ in\ xs, \text{let } B'_2\ in\ xs) \in \text{List } R_1
\]

\[
\text{then let } h = \lambda x. C\ [x]; l = \text{let } B_2\ in\ \text{map } T_1\ T_2\ f\ xs\ \text{in } \text{map } T_2\ T'_2\ h\ l \mapsto
\]

\[
\text{let } B'_2\ in\ \text{map } T_1\ T'_2\ f\ xs
\]

Finally, we let \(B_2 = \{f = P, xs = M\}, B'_2 = \{f = Q, xs = M\}\) and obtain:

\[
\text{if for any bindings } B_1, B'_1\ \text{we have}
\]

\[
\text{let } B_1\ in\ C\ [P\ x] \mapsto \text{let } B_1\ in\ Q\ x
\]

\[
\text{then let } h = \lambda x. C\ [x]; f = P; xs = M; l = \text{map } T_1\ T_2\ f\ xs\ \text{in } \text{map } T_2\ T'_2\ h\ l \mapsto
\]

\[
\text{let } g = Q; xs = M\ \text{in } \text{map } T_1\ T'_2\ g\ M
\]

This result states that if the function \(P\) can be fused with the context \(C\) to make an improved function \(Q\), then the combination of \text{map} (\lambda x. C\ [x]) with \text{map} \(P\) is improved by \text{map} \(Q\), which establishes that map fusion is indeed an optimisation.

### 5.3 Example: Fold Fusion

Now consider the fold function on lists, with the following type:

\[
\text{fold} : \forall \alpha. (\alpha \rightarrow \beta \rightarrow \beta) \rightarrow \beta \rightarrow \text{List } \alpha \rightarrow \beta
\]

Applying our abstraction theorem, we obtain:

\[
\text{for any bindings } B_2, B'_2\ and\ machine-closed\ relations\ R_1 : T_1 \leftrightarrow T'_1, R_2 : T_2 \leftrightarrow T'_2,
\]

\[
\text{if for any bindings } B_1, B'_1\ \text{we have}
\]

\[
(\text{let } B_1\ in\ x, \text{let } B'_1\ in\ x) \in R_1\ \text{and (let } B_1\ in\ y, \text{let } B'_1\ in\ y) \in R_2
\]

\[
\text{together imply (let } B_1; B_2\ in\ h\ x\ y, \text{let } B'_1; B'_2\ in\ h'\ x\ y) \in R_2
\]

\[
\text{and (let } B_2\ in\ n, \text{let } B'_2\ in\ n') \in R_2
\]
and \( (\text{let } B_2 \text{ in } l, \text{ let } B'_2 \text{ in } l') \in \text{List } R_1 \)

then \( (\text{let } B_2 \text{ in } \text{fold } T_1 \text{ } T_2 \text{ } h \text{ } n \text{ } l, \text{ let } B'_2 \text{ in } \text{fold } T'_1 \text{ } T'_2 \text{ } h' \text{ } n' \text{ } l') \in R_2 \)

Letting \( T_1 = T'_1, R_1 = \{ (M_1, M_2) \mid M_1 \triangleright M_2 \}, R_2 = \{ (M_1, M_2) \mid C [M_1] \triangleright M_2 \} \), we obtain:

for any bindings \( B_2, B'_2 \),

if for any bindings \( B_1, B'_1 \) we have

\[
\text{let } B_1 \text{ in } x \triangleright \text{ let } B'_1 \text{ in } x \text{ and let } B_1 \text{ in } C [y] \triangleright \text{ let } B'_1 \text{ in } y
\]

then \( (\text{let } B_2 \text{ in } \text{fold } T_1 \text{ } T_2 \text{ } h \text{ } n \text{ } l, \text{ let } B'_2 \text{ in } \text{fold } T'_1 \text{ } T'_2 \text{ } h' \text{ } n' \text{ } l) \in R_2 \)

Simplifying further, we obtain:

for any bindings \( B_2, B'_2 \),

if for any bindings \( B_1 \) we have

\[
\text{let } B_1; B_2 \text{ in } C [h \times y] \triangleright \text{ let } B_1; B'_2 \text{ in } h' \times y'
\]

then \( (\text{let } B_2 \text{ in } C [\text{fold } T_1 \text{ } T_2 \text{ } h \text{ } n \text{ } l] \triangleright \text{ let } B'_2 \text{ in } \text{fold } T'_1 \text{ } T'_2 \text{ } h' \text{ } n' \text{ } l) \in R_2 \)

This is an improving version of the usual fusion rule for \text{fold}, much like the improving version of \text{fix} fusion above, telling us that when the fusion precondition is an improvement in one direction then fusion itself is also an improvement.

### 5.4 Example: Short Cut Fusion

Short cut fusion, as introduced by Gill et al. [1993], is a general purpose transformation that eliminates the use of intermediate lists in functional programs. The transformation itself is based around \text{fold} function from the previous example, coupled with a new function called \text{build}:

\[
\text{build} : \forall \alpha. (\forall \beta. (\alpha \rightarrow \beta \rightarrow \beta) \rightarrow \beta \rightarrow \beta) \rightarrow \text{List } \alpha
\]

\[
\text{build} = \lambda \alpha. \lambda g \rightarrow g (\text{List } \alpha) (\lambda x \text{ } x \rightarrow x :: xs) \text{ Nil}
\]

Essentially, \text{build} is used to produce lists by abstracting over the two list constructors, while \text{fold} is used to consume the resulting lists in a structured manner by replacing the two constructors.

These two functions are linked by the following equation:

\[
\text{fold } h \text{ } n \text{ (build } g) = g \text{ } h \text{ } n
\]

Short cut fusion involves applying this equation anywhere the left-hand side appears, thereby removing the intermediate list. The strength of this technique comes from the generality of \text{build}. In particular, many library functions that produce lists can be implemented using \text{build}, and any function implemented in such a way then becomes a candidate for short cut fusion.

Unlike the previous examples, which could in theory be proved from the structure of the function in question, the structure of \( g \) is not known ahead of time. As such, parametricity is essential in the proof of correctness of short cut fusion. This suggests that we may be able to prove a result about its efficiency properties using our new abstraction theorem. First of all, instantiating the abstraction theorem for the type of our function \( g \) gives:

for any bindings \( B_2, B'_2 \) and machine-closed relation \( R : T_1 \leftrightarrow T'_1 \),

if for any bindings \( B_1, B'_1 \) we have

\[
(\text{let } B_1 \text{ in } x, \text{ let } B'_1 \text{ in } x) \in \Delta () (A) \text{ and } (\text{let } B_1 \text{ in } y, \text{ let } B'_1 \text{ in } y) \in R
\]

\[
(\text{let } B_1; B_2 \text{ in } f \times y, \text{ let } B'_1; B'_2 \text{ in } f \times y) \in R
\]

then \( (\text{let } B_2 \text{ in } g \text{ } f \text{ } c, \text{ let } B'_2 \text{ in } g \text{ } f \text{ } c) \in R \)

In this case, we let $T_1 = \text{List } A$, $R = \{ (M_1, M_2) \mid \text{let } x y = M_1 \text{ in } \text{fold } h f n x y \supset M_2 \}$, $B_2 = \{ f = \lambda x y. x :: y ; c = \text{Nil} \}$ and $B'_2 = \{ f = h ; c = n \}$. The fact that $R$ is machine-closed follows from the strictness of $\text{fold}$. We obtain:

if for any bindings $B_1, B'_1$ we have

$$(\text{let } B_1 \text{ in } x, \text{let } B'_1 \text{ in } x) \in \Delta () (\text{A})$$ and let $B_1 \text{ in } \text{fold } h n y \supset \text{let } B'_1 \text{ in } y$

... together imply let $B_1; x y = \text{fold } h n x y \supset \text{let } B'_1 \text{ in } h x y$

and $\text{fold } h n \text{ Nil } \supset n$

then let $f = \lambda x. x :: x ; c = \text{Nil} \text{ in } \text{fold } h n (g f c) \supset \text{let } f = h ; c = n \text{ in } g f c$.

We now aim to prove the precondition. We assume $(\text{let } B_1 \text{ in } x, \text{let } B'_1 \text{ in } x) \in \Delta () (\text{A})$ and let $B_1 \text{ in } \text{fold } h n y \supset \text{let } B'_1 \text{ in } y$, and attempt to prove let $B_1; x y = \text{fold } h n x y \supset \text{let } B'_1 \text{ in } h x y$. Note that by our equivalent of the identity extension lemma the first assumption implies let $B_1 \text{ in } x \supset \text{let } B'_1 \text{ in } x$. We reason as follows:

Let $B_1; x y = \text{fold } h n x y$

$\supset$ {definition of $\text{fold}$}

Let $B_1 \text{ in } h x (\text{fold } h n y)$

$\supset$ {assumptions}

Let $B'_1 \text{ in } h x y$

We can therefore conclude that let $f = \lambda x. x :: x ; c = \text{Nil} \text{ in } \text{fold } h n (g f c) \supset g h n$, which establishes that applying the short cut fusion technique results in an improvement.

6 CONCLUSION AND FURTHER WORK

We have shown that a parametricity result holds for a call-by-need operational semantics with a notion of program cost, extending the work of Pitts [2000]. The resulting abstraction theorem can be used to derive free theorems that give results about program efficiency, giving conditions under which program transformations are guaranteed to maintain or improve the time performance of programs. We have applied our theorem to a range of examples, including the well-known short cut fusion of Gill et al. [1993], showing that these particular transformations are all safe, in the sense that they do not degrade performance. This is an important step towards making formal reasoning about the performance of call-by-need programs tractable.

This work considers the time performance of call-by-need programs. However, space behaviour of call-by-need programs can also be counterintuitive, and this raises the question of whether a similar parametricity result will hold. This work was based on call-by-need time improvement as developed by [Moran and Sands 1999a]; corresponding work for space costs could be based on the theory of [Gustavsson and Sands 1999, 2001]. Ultimately, it would be best to have a unified theory for both space and time based on some abstract notion of resource, so that this asymmetry can be avoided. The work of Sands [1997] may offer a way forward here, and we are also in the process of developing a generic foundation for program improvement [Hackett and Hutton 2017], based upon the use of metric spaces to abstract over various aspects including the cost model.

Another way to extend this work would be to consider selective strictness, where an operator such as Haskell’s $\text{seq}$ can be used to force evaluation of subterms. Parametricity in the presence of $\text{seq}$ has already been investigated both denotationally [Johann and Voigtländer 2004] and operationally in terms of program equivalence and partial equivalence [Voigtländer and Johann 2007], so it would be a natural next step to attempt to do the same for program improvement.

Improvement-theory based techniques can show that a transformation does not have a negative impact on performance, but there are other questions we may wish to ask about our optimisations. In particular, we would like to quantify how much performance is improved, and whether this
change is merely a constant factor or an actual asymptotic improvement. This would also allow us to verify the properties of optimisations that are not simple improvements, such as those that improve performance in all but a few pathological cases.

This work inherits the limitation of Moran and Sands [1999a] in that it can only be used to reason about certain forms of recursive definition, i.e. those where the right-hand side is in a value form. This covers the usual case of recursive functions, because the right-hand side is then a lambda-term and hence in value form, but it fails to address a number of programs written in the so-called "knot-tying" style. For example, the function cycle that takes a list and produces the list of infinite repetitions of that list, can naturally be written in two ways:

(i) \( \text{cycle } xs = xs + \text{cycle } xs \)

(ii) \( \text{cycle } xs = \text{let } r = xs + r \text{ in } r \)

Our theory allows us to reason about the first definition, but not the second. The problem is the unwinding lemma (4.1) fails when the right-hand side of the definition is unrestricted, as when \( x = M [x] \) is unwound to the infinite sequence \( x_0 = x; x_1 = M [x_0]; x_2 = M [x_1] \ldots \) we lose sharing between the different levels of recursion. We believe this limitation can be removed by considering bindings with fuel instead of unwound bindings, where \( x =^n M \) is treated as a binding that allows \( x \) to be looked up at most \( n \) times. However, while the unwinding lemma holds for this form of binding, we have yet to prove the corresponding Let case for the abstraction theorem.

Deriving free theorems can be time consuming. Generation of standard free theorems has been mechanised for a sublanguage of Haskell [Bartsch 2017], so the work of this article could potentially also be mechanised. This would benefit from further work on machine-closed relations, so as to increase the number of ways we can specialise these theorems. In particular, it would be useful to know whether the inclusion from Theorem 5.2 can be strengthened to an equality.

Acknowledgements
We would like to thank Martin Handley for many useful comments and suggestions. This work was funded by the Engineering and Physical Sciences Research Council (EPSRC) grant EP/P00587X/1, Mind the Gap: Unified Reasoning About Program Correctness and Efficiency.

REFERENCES
Extended version of [Moran and Sands 1999b], available at http://tinyurl.com/ohuv8ox.
A PROOF OF LEMMA 3.4

We proceed by induction on the derivation of the typing judgement $\Gamma \vdash M : T$ according to the
trules in Figure 1, considering each of the nine cases in turn.

VAR Follows immediately from the definition of $\Delta$.

ABS In this case, $M$ is of the form $\lambda y. M'$ and $T$ is of the form $T_1 \rightarrow T_2$. The inductive hypothesis
is that $\Gamma, y : T_1 \vdash M' : T_2$. We must prove that $\Gamma \vdash \lambda y. M' \Delta \lambda y. M' : T_1 \rightarrow T_2$.

We let $B_1, B_2$ be bindings that close $\lambda y. M'$ and let $\bar{R} : \bar{T} \leftrightarrow \bar{T}'$ be a list of machine-closed
relations with length equal to the number of free type variables in $\Gamma$. Assume that for all $x : A \in \Gamma$, (let $B_1$ in $x$, let $B_2$ in $x$) $\in \Delta (\bar{R}/\bar{a}) (A)$, and let $B'_1$ and $B'_2$ be extensions of $B_1$ and $B_2$ such that (let $B'_1$ in $y$, let $B'_2$ in $y$) $\in \Delta (\bar{R}/\bar{a}) (T_1)$ also. We reason as follows:

\[
\{ \text{ induction hypothesis } \} \\
\Rightarrow \\
\{ \text{ definition of } \rightarrow \text{ action on relations } \} \\
\{ \text{ removing extra bindings } \} \\
\{ \text{ type}\beta\text{-reduction, machine-closure } \}
\]

The removing bindings step is valid because let $B_1$ in $(\lambda y. M')$ and let $B_2$ in $(\lambda y. M')$ were
already closed, so the extra bindings in the extended versions cannot affect evaluation. Because all relations are machine-closed, this means that the terms before and after removing bindings must be treated equally. As this reasoning was generic in $\bar{R}$, $B_1$ and $B_2$, we can conclude that $\Gamma \vdash \lambda y. M' \Delta \lambda y. M' : T_1 \rightarrow T_2$, as required.

APP Follows straightforwardly from the definitions of $\Delta$ and the relational action of $\rightarrow$.

TABS In this case, $M$ is of the form $\Lambda \alpha. M'$ and $T$ is of the form $\forall \alpha'. T'$. The inductive hypothesis is
that $\Gamma, \alpha \vdash M' : T'$. We must prove that $\Gamma \vdash \Lambda \alpha. M' \Delta \Lambda \alpha. M' : \forall \alpha. T'$.

We let $B_1, B_2$ be bindings that close $M'$ and let $\bar{R} : \bar{T} \leftrightarrow \bar{T}'$ be a list of machine-closed
relations with length equal to the number of free type variables in $\Gamma$. Assume that for all $x : A \in \Gamma$, (let $B_1$ in $x$, let $B_2$ in $x$) $\in \Delta (\bar{R}/\bar{a}) (A)$, and let $X, X'$ be arbitrary types with an
arbitrary machine-closed relation $R : X \leftrightarrow X'$. We reason as follows:

\[
\{ \text{ induction hypothesis } \} \\
\Rightarrow \\
\{ \text{ type}\beta\text{-reduction, machine-closure } \} \\
\]


(let B₁ in ((Λα.M') X) [T/α₁],
let B₂ in ((Λα.M') X) [T'/α₆]) ∈ Δ (R/α₁, R/α₆) (T')

Because this reasoning is generic in X, X' and R : X ↔ X', and because R is machine-closed, we can conclude:

(let B₁ in (Λα.M') [T/α],
let B₂ in (Λα.M') [T'/α₆]) ∈ Δ (R/α₁) (∀ α₆.T')

Finally, because all of the above reasoning was generic in R, B₁ and B₂, we can conclude that Γ ⊢ Λα.M' Δ Λα.M' : ∀ α₆.T', as required.

**TAPP** Follows straightforwardly from the definitions of Δ and the relational action of ∀.

**NIL** Using the fact that let B in Nil ⊢ Nil, this case follows from the definition of the logical relation Δ and the relational action of List.

**CONS** Follows immediately from the definition of Δ and the relational action of List.

**CASE** In this case, we know that M is of the form case M' of {Nil → M₁; x :: xs → M₂}. The induction hypotheses are that Γ ⊢ M' Δ M' : List T', Γ ⊢ M₁ Δ M₁ : T, Γ, x : T', xs : List T' ⊢ M₂ : T for some type T'.

We assume that x, xs are fresh, which can be ensured by alpha-renaming. We let B₁, B₂ be bindings that close M and let R : T ↔ T' be a list of machine-closed relations with length equal to the number of free type variables in Γ.

To prove Γ ⊢ case M' of { ... } Δ case M' of { ... } : T we assume that for all y : A ∈ Γ we have

(let B₁ in y,
let B₂ in y) ∈ Δ (R/α₁) (A)

and try to prove

(let B₁ in case M' [T/α₁] of { Nil → M₁ [T/α₁]; x :: xs → M₂ [T/α₁]},
let B₂ in case M' [T'/α₆] of { Nil → M₁ [T'/α₆]; x :: xs → M₂ [T'/α₆]})

∈ Δ (R/α₁) (T)

To prove this, we need the following sub-lemma:

**LEMMA A.1.** Consider machine-closed relations R₁ R₂, heap and stack pairs ⟨H₁, S₁⟩, ⟨H₂, S₂⟩, bindings B₁, B₂ and terms M₁, M₁', M₂, M₂' that satisfy the following assumptions:

(i) (let B₁ in M₁, let B₂ in M₁') ∈ R₂

(ii) For all bindings B₁', B₂' and variables y, ys, if

(let B₁' in y, let B₂' in y) ∈ R₁

and

(let B₁' in ys, let B₂' in ys) ∈ List R₁

then it follows that

(let B₁' in M₂ [y/x, ys/xs], let B₂' in M₂' [y/x, ys/xs]) ∈ R₂

(iii) For any pair of terms (N, N') ∈ R₂, ⟨H₁, N, S₁⟩ ↓ₙ ⇒ ⟨H₂, N', S₂⟩ ↓ₙ

Then for any pair of terms (L, L') ∈ List R₂, we have:
Parametric Polymorphism and Operational Improvement: 21

Now we can prove this case of the main lemma. Let \( H_1 + B_1, L, \{ \text{Nil} \to M_1; x :: xs \to M_2 \} : S_1 \) \( \downarrow_n \) =>
\( H_2 + B_2, L', \{ \text{Nil} \to M'_1; x :: xs \to M'_2 \} : S_2 \) \( \downarrow_n \)

Proof. Because List \( R_1 \) is the machine-closure of Rel = \{ Nil, Nil \} \cup \{ (let B_1 \ in \ y :: ys, let B_2 \ in \ y :: ys) \} (let B_1 \ in \ y, let B_2 \ in \ y) \in R_1, (let B_1 \ in \ y, let B_2 \ in \ ys) \in \text{List } R_1 \}, the definition of machine-closure implies that our result holds if \( H_1 + B_1, L, \{ \text{Nil} \to M_1; x :: xs \to M_2 \} : S_1 \) \( \downarrow_n \) => \( H_2 + B_2, L', \{ \text{Nil} \to M'_1; x :: xs \to M'_2 \} : S_2 \) \( \downarrow_n \) for any \( (L, L') \) in the underlying set Rel. This \( (L, L') \) will be of one of two forms.

Firstly, we consider the case of \( \{ \text{Nil}, \text{Nil} \} \). In this case, we reason as follows:
\( \langle H_1 + B_1, \text{Nil}, \{ \text{Nil} \to M_1; x :: xs \to M_2 \} : S_1 \rangle \downarrow_n \)
\( \Rightarrow \{ \text{BranchNil} \} \)
\( \langle H_1 + B_1, M_1, S_1 \rangle \downarrow_n \)
\( \Rightarrow \{ \text{conditions (i) and (iii)} \} \)
\( \langle H_2 + B_2, M_2, S_2 \rangle \downarrow_n \)
\( \Rightarrow \{ \text{BranchNil} \} \)
\( \langle H_2 + B_2, \text{Nil}, \{ \text{Nil} \to M'_1; x :: xs \to M'_2 \} : S_2 \rangle \downarrow_n \)

Next, we consider the case of \( \langle \text{let } B'_1 \ \text{in } y :: ys, \ \text{let } B'_2 \ \text{in } y :: ys \rangle \). In this case, we know that \( \langle \text{let } B'_1 \ \text{in } y, \ \text{let } B'_2 \ \text{in } y \rangle \in R_1 \) and \( \langle \text{let } B'_1 \ \text{in } y, \ \text{let } B'_2 \ \text{in } ys \rangle \in \text{List } R_1 \). We reason as follows:
\( \langle H_1 + B_1, \text{let } B'_1 \ \text{in } y :: ys, \ \{ \text{Nil} \to M_1; x :: xs \to M_2 \} : S_1 \rangle \downarrow_n \)
\( \Rightarrow \{ \text{Letrec} \} \)
\( \langle H_1 + B_1 + B'_1, y :: ys, \ \{ \text{Nil} \to M_1; x :: xs \to M_2 \} : S_1 \rangle \downarrow_n \)
\( \Rightarrow \{ \text{BranchCons} \} \)
\( \langle H_1 + B_1 + B'_1, M_2 [y/x, ys/xs], S_1 \rangle \downarrow_n \)
\( \Rightarrow \{ \text{Letrec} \} \)
\( \langle H_1 + B_1, \text{let } B'_1 \ \text{in } M_2 [y/x, ys/xs], S_1 \rangle \downarrow_n \)
\( \Rightarrow \{ \text{conditions (ii) and (iii)} \} \)
\( \langle H_2 + B_2, \text{let } B'_2 \ \text{in } M'_2 [y/x, ys/xs], S_2 \rangle \downarrow_n \)
\( \Rightarrow \{ \text{Letrec} \} \)
\( \langle H_2 + B_2 + B'_2, y :: ys, \ \{ \text{Nil} \to M'_1; x :: xs \to M'_2 \} : S_2 \rangle \downarrow_n \)
\( \Rightarrow \{ \text{BranchCons} \} \)
\( \langle H_2 + B_2 + B'_2, y :: ys, \ \{ \text{Nil} \to M'_1; x :: xs \to M'_2 \} : S_2 \rangle \downarrow_n \)
\( \Rightarrow \{ \text{Letrec} \} \)
\( \langle H_2 + B_2, \text{let } B'_2 \ \text{in } y :: ys, \ \{ \text{Nil} \to M'_1; x :: xs \to M'_2 \} : S_2 \rangle \downarrow_n \)

□

Now we can prove this case of the main lemma. Let \( \langle H_1, S_1 \rangle \) and \( \langle H_2, S_2 \rangle \) be heap and stack pairs such that for any pair of terms \( (N, N') \in \Delta (\overline{R}/\overline{a}) (T), \) we have \( \langle H_1, N, S_1 \rangle \downarrow_n \Rightarrow \langle H_2, N', S_2 \rangle \downarrow_n \). We then reason as follows:
\( \langle H_1, \text{let } B_1 \ \text{in } \text{case } M' [\overline{T}/\overline{a}] \text{ of } \{ \text{Nil} \to M_1 [\overline{T}/\overline{a}]; x :: xs \to M_2 [\overline{T}/\overline{a}] \}, S_1 \rangle \downarrow_n \)
\( \Rightarrow \{ \text{Letrec} \} \)
\( \langle H_1 + B_1, \text{case } M' [\overline{T}/\overline{a}] \text{ of } \{ \text{Nil} \to M_1 [\overline{T}/\overline{a}]; x :: xs \to M_2 [\overline{T}/\overline{a}] \}, S_1 \rangle \downarrow_n \)
\( \Rightarrow \{ \text{Case} \} \)
\( \langle H_1 + B_1, M' [\overline{T}/\overline{a}], \ {\text{Nil} \to M_1 [\overline{T}/\overline{a}]; x :: xs \to M_2 [\overline{T}/\overline{a}] \} : S_1 \rangle \downarrow_n \)
\( \Rightarrow \{ \text{lemma; conditions (i) and (ii) follow from IHs, (iii) holds by assumption} \} \)
\[ (H_2 + B_2, M' [\vec{\alpha} / \vec{\alpha}], \{ \text{Nil} \to M_1 [\vec{\alpha} / \vec{\alpha}]; x :: xs \to M_2 [\vec{\alpha} / \vec{\alpha}] \} : S_2) \downarrow_n \]

\[ \Leftrightarrow \quad \{ \text{CASE} \} \]

\[ (H_2 + B_2, \text{case } M' [\vec{\alpha} / \vec{\alpha}] \text{ of } \{ \text{Nil} \to M_1 [\vec{\alpha} / \vec{\alpha}]; x :: xs \to M_2 [\vec{\alpha} / \vec{\alpha}] \}, S_2) \downarrow_n \]

\[ \Leftrightarrow \quad \{ \text{LETREC} \} \]

\[ (H_2, \text{let } B_2 \text{ in case } M' [\vec{\alpha} / \vec{\alpha}] \text{ of } \{ \text{Nil} \to M_1 [\vec{\alpha} / \vec{\alpha}]; x :: xs \to M_2 [\vec{\alpha} / \vec{\alpha}] \}) \]

But this reasoning was generic in \( \langle H_1, S_1 \rangle, \langle H_2, S_2 \rangle \), so by machine-closure we have:

(\text{let } B_1 \text{ in case } M' [\vec{\alpha} / \vec{\alpha}] \text{ of } \{ \text{Nil} \to M_1 [\vec{\alpha} / \vec{\alpha}]; x :: xs \to M_2 [\vec{\alpha} / \vec{\alpha}] \},

\text{let } B_2 \text{ in case } M' [\vec{\alpha} / \vec{\alpha}] \text{ of } \{ \text{Nil} \to M_1 [\vec{\alpha} / \vec{\alpha}]; x :: xs \to M_2 [\vec{\alpha} / \vec{\alpha}] \}) \in \Delta (\bar{R}/\bar{\alpha}) (T)

Finally, since this reasoning was generic in \( \bar{R}, B_1 \) and \( B_2 \), we can then conclude that \( \Gamma \vdash \text{case } M' \text{ of } \{ \ldots \} \Delta \text{ case } M' \text{ of } \{ \ldots \} : T \), as required.

\textbf{LET'} In this case, \( M \) is of the form \( \bar{x} = \bar{M} \) in \( M' \). The induction hypotheses are that \( \Gamma \vdash M_i \Delta M_i : T_i \) for all \( i \) and \( \Gamma \), \( x_1 : T_1, \ldots, x_n : T_n \vdash M' \Delta M' : T \).

We let \( B_1, B_2 \) be bindings that close \( \bar{x} = \bar{M} \) in \( M' \) and let \( \bar{R} : \bar{T} 

\leftrightarrow \bar{T}' \) be a list of machine-closed relations with length equal to the number of free type variables in \( \Gamma \).

To prove \( \Gamma \vdash \text{let } \bar{x} = \bar{M} \text{ in } M' : T \), we must assume that for all \( y : A \in \Gamma \) it is the case that:

(\text{let } B_1 \text{ in } y,

\text{let } B_2 \text{ in } y) \in \Delta (\bar{R}/\bar{\alpha}) (A)

and try to prove:

(\text{let } B_1 \text{ in let } \bar{x} = \bar{M} [\vec{\alpha} / \vec{\alpha}] \text{ in } M' [\vec{\alpha} / \vec{\alpha}],

\text{let } B_2 \text{ in let } \bar{x} = \bar{M} [\vec{\alpha} / \vec{\alpha}] \text{ in } M' [\vec{\alpha} / \vec{\alpha}]) \in \Delta (\bar{R}/\bar{\alpha}) (T)

We note that the induction hypotheses for the \( M_i \) imply:

(\text{let } B_1 \text{ in } M_i [\vec{\alpha} / \vec{\alpha}],

\text{let } B_2 \text{ in } M_i [\vec{\alpha} / \vec{\alpha}]) \in \Delta (\bar{R}/\bar{\alpha}) (T_i)

In turn, by machine closure, this implies that

(\text{let } B_1; \bar{x} = \bar{M} [\vec{\alpha} / \vec{\alpha}] \text{ in } x_i,

\text{let } B_2; \bar{x} = \bar{M} [\vec{\alpha} / \vec{\alpha}] \text{ in } x_i) \in \Delta (\bar{R}/\bar{\alpha}) (T_i)

because each term takes exactly one \text{LOOKUP} step to replace \( x_i \) with \( M_i [\vec{\alpha} / \vec{\alpha}] \), and the extra bindings serve no purpose beyond this. These, combined with the inductive hypothesis of \( \Gamma, x_1 : T_1, \ldots, x_n : T_n \vdash M' \Delta M' : T \), imply

(\text{let } B_1; \bar{x} = \bar{M} [\vec{\alpha} / \vec{\alpha}] \text{ in } M' [\vec{\alpha} / \vec{\alpha}],

\text{let } B_2; \bar{x} = \bar{M} [\vec{\alpha} / \vec{\alpha}] \text{ in } M' [\vec{\alpha} / \vec{\alpha}]) \in \Delta (\bar{R}/\bar{\alpha}) (T)

which implies

(\text{let } B_1 \text{ in let } \bar{x} = \bar{M} [\vec{\alpha} / \vec{\alpha}] \text{ in } M' [\vec{\alpha} / \vec{\alpha}],

\text{let } B_2 \text{ in let } \bar{x} = \bar{M} [\vec{\alpha} / \vec{\alpha}] \text{ in } M' [\vec{\alpha} / \vec{\alpha}]) \in \Delta (\bar{R}/\bar{\alpha}) (T)
Because these are each cost-equivalent to the terms in the previous relation and machine-closed relations respect cost-equivalence. Finally, we note that this reasoning was generic in $\tilde{R}$, $B_1$ and $B_2$, so we can conclude $\Gamma \vdash \text{let } \tilde{x} = \tilde{M} \text{ in } \Delta \text{ let } \tilde{x} = \tilde{M} \text{ in } \Delta' : T$.

This completes the proof. □