An Algebraic Theory for Shared-State Concurrency

Citation for published version:

Digital Object Identifier (DOI):
10.1007/978-3-031-21037-2_1

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Programming Languages and Systems: 20th Asian Symposium, APLAS 2022, Auckland, New Zealand, December 5, 2022, Proceedings

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
An Algebraic Theory for Shared-State Concurrency*

Yotam Dvir1[0000–0002–6507–3791], Ohad Kammar2[0000–0002–2071–0929], and Ori Lahav1[0000–0003–2102–4338]

1 Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel yotamdvir@mail.tau.ac.il, orilahav@tau.ac.il
2 School of Informatics, University of Edinburgh, Edinburgh, Scotland ohad.kammar@ed.ac.uk

Keywords: shared state · concurrency · denotational semantics · monads · equational theory · program refinement · program equivalence · compiler transformations · compiler optimisations

Abstract. We present a monadic denotational semantics for a higher-order programming language with shared-state concurrency, i.e. global-state in the presence of interleaving concurrency. Central to our approach is the use of Plotkin and Power’s algebraic effect methodology: designing an equational theory that captures the intended semantics, and proving a monadic representation theorem for it. We use Hyland et al.’s equational theory of resumptions that extends non-deterministic global-state with an operator for yielding to the environment. The representation is based on Brookes-style traces. Based on this representation we define a denotational semantics that is directionally adequate with respect to a standard operational semantics. We use this semantics to justify compiler transformations of interest: redundant access eliminations, each following from a mundane algebraic calculation; while structural transformations follow from reasoning over the monad’s interface.

1 Introduction

Denotational semantics [−] associates every program M with its meaning, i.e. its denotation, [M]. A key feature of a denotational semantics is compositionality: the denotation of a program depends only on the denotations of its constituents.

As a concrete example, consider an imperative language that manipulates a memory store σ ∈ S, and denotational semantics for it that associates with each program M a denotation [M] : S → S modelling how M transforms the store. For example – denoting by σ[a → v] the store that is the same as σ but with

* Supported by the Israel Science Foundation (grant numbers 1566/18 and 814/22) and the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement no. 851811); and by a Royal Society University Research Fellowship and Enhancement Award.
its value at $a$ changed to $v$ — we have $[a := v] \sigma = \sigma[a \mapsto v]$. Compositionality manifests in the semantics of program sequencing: $[M ; N] \sigma = [[N]([M] \sigma)$. Thus $[a := 0 ; a := 1] \sigma = [a := 1] \sigma([a := 0]) \sigma = (\sigma[a \mapsto 0])(a \mapsto 1) \sigma = \sigma[a \mapsto 1]$. Incidentally, we also have $[a := 1] \sigma = \sigma[a \mapsto 1]$, and so $[a := 0 ; a := 1] = [a := 1]$.

A desirable property of a denotational semantics $[\cdot]$ is adequacy, meaning that $[M] = [N]$ implies that $M$ and $N$ are contextually equivalent: replacing $N$ with $M$ within some larger program does not affect the possible results of executing that program. Contextual equivalence is useful for optimizations: for example, $M$ could have better runtime performance than $N$. Adequate denotational semantics can justify optimizations without quantifying over all program contexts, serving in this way as a basis for validating compiler optimizations.

Returning to the example above, although $[a := 0 ; a := 1] = [a := 1]$, in the presence of concurrency $a:=0;a:=1$ and $a:=1$ are not contextually equivalent. For example, if $b:=a?$ (read from $a$ and write the result to $b$) is executed concurrently, it could write 0 to $b$ only with the first programs. Therefore, the semantics we defined is inadequate for a concurrent programming language; differentiating between these programs requires a more sophisticated denotational semantics.

Moreover, the transformation $a := 0 ; a := 1 \rightarrow a := 1$ eliminating the first, redundant memory access is valid in the presence of concurrency, even though the programs are not equivalent. Indeed, a compiler applying this simplification within a program would not introduce any additional possible results (though it may eliminate some), and in particular it would conserve the correctness of the program. We would like our semantics to be able to justify such transformations.

This leads us to the concept of directional adequacy, a useful refinement of adequacy. Given a partial order $\preceq$ on the set of denotations, the denotational semantics is directionally adequate (w.r.t. $\preceq$) if $[M] \preceq [N]$ implies that $M$ contextually refines $N$: replacing $N$ with $M$ within some larger program does not introduce new possible results of executing that program. Thus, directional adequacy can justify the transformation $N \rightarrow M$ even if it is not an equivalence.

In this paper we define directionally-adequate denotational semantics for a higher-order language, subsuming the imperative language above, that justifies the above transformation along with other standard memory access eliminations:

$$
\ell := w ; \ell := v \rightarrow \ell := v \quad \text{(write ; write)}
$$

$$
\ell := v ; \ell \,? \rightarrow \ell := v \,; v \quad \text{(write ; read)}
$$

$$
\text{let } x = \ell \,? \text{ in let } y = \ell \,? \text{ in } (x,y) \rightarrow \text{let } x = \ell \,? \text{ in } (x,x) \quad \text{(read ; read)}
$$

$$
\ell \,? ; M \rightarrow M \quad \text{(irrelevant read)}
$$

Other transformations and equivalences this semantics validates are structural ones, such as $\text{if } M \text{ then } N \text{ else } N \equiv M \,; N$; and concurrency-related ones, such as $(M \parallel N) \,; K \rightarrow M \,; N \,; K$. 

None of these transformations are novel. Rather, the contribution of this paper is in the methodology that is used to justify them, fitting shared-state concurrency semantics into a general, uniform model structure. In particular, each memory access elimination is proven correct via a mundane algebraic calculation; and the structural transformations can be justified using simple structural arguments that abstract away the details of our particular semantics.

Methodology  The use of monads to study the denotational semantics of effects [30] has proven fruitful, especially with its recent refinement with algebraic operators and equational theories [9,10,23,35,36]. We follow the algebraic approach to define denotational semantics for a simple higher-order concurrent programming language, using an equational theory extending non-deterministic global-state with a single unary algebraic operator for yielding computation to a concurrently running program [1]. We find a concrete representation of the monad this theory induces based on sets of traces [4,6], and use it to define a directionally adequate denotational semantics. From this adequacy result we deduce various program transformations via routine calculations.

The advantages of denotational semantics defined with this approach include:

Uniformity. Theories are stated using the same general framework. This uniformity means that many theoretical results can be stated in general terms, applying to all theories. Even if a theorem, like adequacy, must be proven separately for each theory, it is likely that a similar proof technique can be used, and experience can guide the construction of the proof.

Comparability. Comparing and contrasting theories is convenient due to uniformity [23]. While our language and semantics is very different from Abadi and Plotkin’s [1], the equational theory we obtain easily compares to theirs.

Modularity. Since the theories are stated algebraically, using operators and equations, they are amenable to be combined to form larger theories. Some combinations are the result of general theory-combinators, such as the theory of non-deterministic global-state resulting from combining the theory of global-state [34] with the theory of non-determinism [14]. In this combined theory, equations that are provable in each separate theory remain provable. Even if the combination is bespoke, using an algebraic theory breaks down the problem into smaller components [8].

Abstraction. The semantics we define for the fragment of our language without shared-state is identical in form to the standard semantics, by using the monad operations. Therefore, any structural transformation proven using these abstractions remains valid in the language extended with shared-state.

Implementability. Monads are ubiquitous as a computational device in functional programming languages, such as Haskell. Thus a theory based on a monad may in the future form a bridge to implementation.

Outline  The remaining sections are as follows. The next section provides background to the problem and overviews our results in a simplified manner. Then we
dive into the weeds, starting with a succinct presentation of the equational theory and related notions \(\S3\). We then work our way gradually to define the induced monad’s concrete representation \(\S4\). Next we define the denotations using this representation \(\S5\). With everything in place, we present our metatheoretical results and use them to justify program transformations and equivalences \(\S6\). We conclude with a discussion of related work and future prospects \(\S7\).

2 Overview

Our setting is a simple programming language with state. We fix a finite set of locations \(L := \{l_1, \ldots, l_n\}\) and a finite set of (storable) values \(V := \{v_1, \ldots, v_m\}\). A store \(\sigma\) is an element of \(S := L \rightarrow V\), where \(\sigma[l] \rightarrow v\) is the store that is equal to \(\sigma\) except (perhaps) at \(l\), where it takes the value \(v\). We use subscripts to apply stores to locations, i.e. we write \(\sigma[l]\) instead of \(\sigma(l)\). In examples we often assume \(L = \{a, b, c\}\) and \(V = \{0, 1\}\), and write stores in matrix notation, e.g. \((a \ b \ c \ | \ 1 \ 0 \ 1)\).

The language is based on a standard extension of Moggi’s \(\S30\) computational lambda calculus with products and variants (labelled sums), further extended with three shared-state constructs. Many other constructs are defined using syntactic sugar, such as if-statements and booleans, let-bindings, and program sequencing. The core syntax is presented below (where \(n \geq 0\)):

\[
G ::= (G_1 \ast \cdots \ast G_n) \mid \{t_1 \of G_1 \mid \cdots \mid t_n \of G_n\} \quad \text{(Ground types)}
\]

\[
A, B ::= (A_1 \ast \cdots \ast A_n) \mid \{t_1 \of A_1 \mid \cdots \mid t_n \of A_n\} \mid A \rightarrow B \quad \text{(Types)}
\]

\[
V, W ::= \langle V_1, \ldots, V_n \rangle \mid i V \mid \lambda x. M \quad \text{(Values)}
\]

\[
M, N ::= x \mid \langle M_1, \ldots, M_n \rangle \mid i M \mid \lambda x. M \mid MN \quad \text{(Terms)}
\]

\[
\text{match } M \text{ with } \langle x_1, \ldots, x_n \rangle \rightarrow N
\]

\[
\text{case } M \text{ of } \{t_1 x_1 \rightarrow N_1 \mid \cdots \mid t_n x_n \rightarrow N_n\}
\]

\[
M? \mid M := N \mid M \parallel N
\]

The typing rules for the shared-state constructs appear at the top of Figure 1, where we define \(\text{Loc} := \{l_1 \of () \mid \cdots \mid l_n \of ()\}\) and \(\text{Val} := \{v_1 \of () \mid \cdots \mid v_m \of ()\}\).

The language is equipped with a call-by-value, small-step operational semantics \(\sigma, M \rightsquigarrow \rho, N\), meaning that the program \(M\) executed from store \(\sigma\) progresses to \(N\) with store \(\rho\). The operational semantics for the shared-state constructs appear at the bottom of Figure 1. Parallel execution is interpreted via a standard interleaving semantics, ultimately returning the pair of the results of each side, and synchronizing on their completion. The reflexive-transitive closure of \(\rightsquigarrow\) is denoted by \(\rightsquigarrow^*\). The operational semantics can be seen in action in Example 2.

2.1 Global-State (for Sequential Computation)

To make this exposition more accessible, we focus on sequential computation before advancing to denotational semantics of concurrent computation. Sequential computations with global state cause two kinds of side-effects: looking a value up
in the store, and updating a value in the store. Plotkin and Power \[34\] propose a corresponding equational theory with two operators:

**Lookup.** Suppose $\ell \in L$, and $(t_v)_{v \in V}$ is a $V$-indexed sequence of terms. Then $L_\ell (t_v)_{v \in V}$ is a term representing looking the value in $\ell$ up, calling it $v$, and continuing the computation with $t_v$. We write $L_\ell (v. t_v)$ instead of $L_\ell (t_v)_{v \in V}$.

**Update.** Suppose $\ell \in L$, $v \in V$, and $t$ is a term. Then $U_{\ell,v} t$ is a term representing updating the value in $\ell$ to $v$ and continuing the computation with $t$.

The equations of the theory of global-state are generated by taking the closure of the axioms – listed at the top of Figure 2 – under reflexivity, symmetry, transitivity, and substitution. The grayed-out axioms happen to be derivable, and are included for presentation’s sake. The theory of global-state can be used to define adequate denotational semantics for the sequential fragment of our language, obtained by removing concurrent execution ($\parallel$).

**Example 1.** Global-state includes the following equation (1) which, when considering sequential programs, represents the program equivalence (2):

\[
L_b (v. U_{a,b} L_c (w. U_{a,w} \langle \rangle )) = L_c (w. U_{a,w} \langle \rangle )
\]

(1)

\[
a := b? ; b := c \equiv a := c?
\]

(2)

### 2.2 Shared-State

The equivalence (2) from Example 1 fails in the concurrent setting, since the two program can be differentiated by program contexts with concurrency:
Example 2. Consider the program context \( \Xi \quad [-] = [-] \parallel a \), i.e. executing each program in parallel to a thread dereferencing the location \( a \). Then there is no execution of \( \Xi \[ a := c ? \] \) that starts with the store \( (\xi, 0, 1) \) and returns \( () , 0 \), but there is such a execution of \( \Xi \[ a := b ? ; a := c ? \] \):

\[
\begin{align*}
\langle \xi, 0, 1 \rangle, a := b ? ; a := c ? \parallel a ? \leadsto & \langle 0, 0, 1 \rangle, a := c ? \parallel a ? \leadsto \\
\langle 0, 0, 1 \rangle, a := c ? \parallel 0 \leadsto & \langle a, b \xi, () \parallel 0 \leadsto \langle a, b \xi, () , 0 \rangle
\end{align*}
\]

Therefore, the aforementioned denotational semantics defined using the theory of global-state cannot be extended with a denotation for \( () \) while preserving adequacy. More sophistication is needed in the concurrent setting: the denotations, even of sequential programs, must account for environment effects.

We thus extend the global-state theory with a single unary operator:

**Yield.** Suppose \( t \) is a term. Then \( \text{Yt} \) is a term. Its intended meaning is to let the environment read and write and then continue with the computation \( t \).

We also need to account for the non-determinism inherent in parallel executions. We do so by extending the resulting theory with operators for finite non-determinism with their equational theory, and further standard equations axiomatizing commutative interaction with the other operators [14]:

**Choice.** For every \( \alpha \in \mathbb{N} \) there is a respective choice operator. Suppose \( (t_i)_{i \in \alpha} \) is a sequence of terms \( t_0, \ldots , t_{\alpha-1} \). Then \( \text{V}_\alpha (t_i)_{i \in \alpha} \) is a term. Its intended
meaning is to choose $\iota < \alpha$ non-deterministically and continue with the computation $t_1$. We write $V_{\iota<\alpha} t_1$ instead of $V_{\alpha} (t_1)_{\iota<\alpha}$; and when $\alpha = 2$ we use infix notation, i.e. instead of $V_{\iota<2} t_1$ we may write $t_0 \vee t_1$.

The axioms of the resulting theory of resumptions $\text{Res}$ are listed in Figure 2. The novelty is not in the equational theory, but rather in the way we use it to define denotations. We compare to related work in §7.

2.3 Denotations

In §5 we define denotations of programs, but for the sake of this discussion we simplify, defining the denotation $[M]$ of a program $M$ to be an equivalence class $[t]$ of a particular term $t$ in $\text{Res}$. Our actual denotations defined in §5 will use the concrete representation of the monad (developed in §4.3), similar to how state transformers represent equivalence classes of terms of global-state.

Dereference & Assignment We use the monadic bind $\Downarrow$ (defined in §4.3), which we can think of as “chaining operators”; and possible yields $Y^2 t := \iota \vee Y t$:

\[
[M?] := [M] \Downarrow \lambda \ell. \left| L_\ell (v, Y^2 v) \right|
\]

\[
[M := N] := [M] \Downarrow \lambda \ell. \left| N \right| \Downarrow \lambda v. \left| U_\ell v Y^2 () \right|
\]

The main idea is to intersperse possible yields to block the use of global-state equations such as in (1) and to allow computations to interleave. Although equation (1) still holds in $\text{Res}$, it does not imply the program equivalence (2) because the programs in (2) do not map to the algebraic terms in (1). Rather:

**Example 3.** The denotations of the programs in **Example 1** are:

\[
[a := c \ni ?] = [L_c (v, Y^2 v)] \Downarrow \lambda v. \left| U_{a,v} Y^2 () \right| = [Y^2 a_{a,v} Y^2 ()]
\]

\[
[a := b \ni ; a := c \ni ?] = [L_b (w, Y^2 b_{a,w} Y^2 c_{a,w} Y^2 ())]
\]

So the denotation (3) looks $c$ up finding a value $v$, then possibly yields, then updates $a$ to $v$, then possibly yields, and finally returns the empty tuple. The concrete representation (Theorem 1) immediately proves that (3) and (4) are not equal, in contrast to the situation in **Example 1**.

Parallel Execution Computations interleave using the yield operator. Interleaving execution of programs suggests, for example, the following calculation:

\[
[\ell ? \parallel \ell := 0] = [L_\ell (v, Y^2 [v \parallel \ell := 0])] \vee U_{\ell,0} Y^2 [\ell ? \parallel ()]
\]

\[
= [L_\ell (v, Y^2 U_{\ell,0} Y^2 [v \parallel ()]) \vee U_{\ell,0} Y^2 L_\ell (v, Y^2 [v \parallel ()])]
\]

\[
= [L_\ell (v, Y^2 U_{\ell,0} Y^2 (v, *)) \vee U_{\ell,0} Y^2 L_\ell (v, Y^2 (v, *))]
\]

The problem with the above is that it lacks compositionality: $[\ell ? \parallel \ell := 0]$ should be defined in terms of $[\ell ?]$ and $[\ell := 0]$, without referring to the underlying programs. In §4.6 we define a function $[[\parallel]]$ such that $[M \parallel N] = [M] \parallel [N]$. This definition relies on the concrete representation from §4.3.
2.4 Program Transformations

Our main result is directional adequacy (Theorem 5). Under this simplified view it can be stated, in terms of the partial-order on our denotations generated by $\|t\| \leq \|t \lor s\|$, as follows: if $\llbracket M \rrbracket \leq \llbracket N \rrbracket$ then the transformation $N \rightarrow M$ is valid in the concurrent setting. The following example illustrates how directional-adequacy can be used to validate program transformations of interest, of the relatively few that are valid in the strong memory-model we consider here.

Example 4. We validate $(\text{write ; read})$ (see also Example 9):

$$\llbracket \ell := v ; \ell? \rrbracket = \| U_{\ell,v} Y^\ell (w. Y^w) \| \geq \| U_{\ell,v} L_{\ell} (w. Y^w) \| = \| U_{\ell,v} Y^w v \| = \llbracket \ell := v ; v \rrbracket$$

We can similarly validate the other memory access eliminations from §1. By using $Y^\ell$ rather than $Y$ in the denotations, the cases where the environment is known to not interleave are taken into accounted explicitly. The relevant global-state equation can then be exploited to eliminate the redundant memory access.

2.5 Caveats and Limitations

Our goal in this work is to fit concurrency semantics on equal footing with other semantic models of computational effects. As a consequence, the proposed model has its fair share of fine print, which we bring to the front:

Memory Model. We study a very strong memory model: sequential consistency. Modern architectures adhere to much weaker memory-models, where further program transformations are valid.

Concurrency Model. Our semantics involves a simple form of concurrency in which threads interleave their computation without restriction, acting on a shared memory store. This is in contrast to a well-established line of work in which models include a causal partial-order in which incomparable events denote “truly” parallel execution [31]. These causal models are showing promise in modelling sophisticated execution [31]. We hope further work would fit these causal models into a relatable semantic footing that easily accommodates higher-order structure.

Features. Our analysis lacks many valuable features that appear in related work, such as recursion [1], higher-order state [3], probabilities [4], and infinitely many locations/values. This simplification is intended: we took to reductionism, finding a minimal model still accounting for core features of shared state. The benefit of the algebraic approach is that this model can be modularly combined with other features, hopefully using standard technology such as sum-and-tensor [14], domain-enrichment [27], and functor categories [21,32,33,38]. For example, to support recursion our model may be integrated with one of the known powerdomain theory-combinators [42]. This requires making a semantic-design choice that is orthogonal to shared-state concurrency, each with different trade-offs. We avoid making such choices.
Semantic Precision. The equational theory and denotational semantics based on it leave much room for improvement in terms of precision and abstraction. For example, our denotational model does not support the introduction of irrelevant reads, i.e. it does not justify the valid transformation \( M \rightarrow \ell? ; M \). Indeed, taking \( M = \langle \rangle \), we have \( \llbracket \ell? ; \langle \rangle \rrbracket = \llbracket L_\ell \{ v. Y? (\langle \rangle) \} \rrbracket = \llbracket Y? (\langle \rangle) \} \rrbracket \neq \llbracket \langle \rangle \rrbracket = \llbracket \langle \rangle \rrbracket \). The problem stems from a “counting” issue: even though the value being looked-up in \( \ell \) is discarded, the additional possible-yield remains. We hope further work could address this semantic inaccuracy.

Full Abstraction. Brookes’s seminal work [1,6] defined denotational semantics for concurrency that is fully-abstract, meaning that the converse of adequacy holds: programs that are replaceable in every context have equal denotations. Our semantics is far from being fully-abstract: there is a first-order valid equivalence, \( M \equiv \ell? ; M \), that our semantics does not support. Moreover, we do not include atomic block executions in our language as Brookes did, which was crucial for the proof of full-abstraction. However, even if our model was precise enough to capture the first-order equivalences, and even if we were to include atomic block executions, we still would not expect to obtain full-abstraction, since this result is infamously elusive for higher-order languages (see Abramsky’s recent overview on the full-abstraction problem of PCF [2]).

3 Equational Theory

At the foundation of our approach is the equational theory of resumptions \( \text{Res} [1,14,30] \) presented in §2, consisting of operators and equational axioms over them. We succinctly fill-in the formal details below, followed by related definitions.

The signature of \( \text{Res} \) consists of the following parameterized operators. The notation \( O : A(P) \) means that the arity of the operator \( O \) is the set \( A \) and it is parameterized over the set \( P \):

\[
\begin{align*}
L : \forall \langle \ell \rangle & \quad \text{lookup} \\
Y : \exists \langle \ell \rangle & \quad \text{yield} \\
U : \exists \langle \ell \times \ell \rangle & \quad \text{update} \\
\bigvee \alpha & \quad \text{non-deterministic choice for every } \alpha \in \mathbb{N}
\end{align*}
\]

From now on, whenever we refer to an operator, we mean an operator of \( \text{Res} \). We denote the set of terms freely generated by the signature over \( X \) by \( \text{Term}_X \).

Figure 2 lists the axioms of \( \text{Res} \), classified as follows: an axiomatization of the equational theory of global-state [34]; the standard axiomatization of non-determinism; and an axiomatization of the commutative interaction of non-determinism with the other operators [13] via the tensor [14].

A \textit{Res-algebra} \( \mathcal{A} \) consists of a carrier set \( \mathcal{A} \) together with interpretations \( \hat{O}^\mathcal{A} : \mathcal{A}^A \times P \rightarrow \mathcal{A} \) for each operator \( O : A(P) \). We elide the superscript if it is clear from context. For a set \( X \), a \textit{Res-algebra on} \( X \) consists of a \textit{Res-algebra} \( \mathcal{A} \) and a function \( \text{env} : X \rightarrow A \); which extends to \( \text{eval} : \text{Term}_X \rightarrow \mathcal{A} \) homomorphically along the inclusion \( X \hookrightarrow \text{Term}_X \). A \textit{Res-model on} \( X \) is a \textit{Res-algebra on} \( X \) that satisfies each axiom of \( \text{Res} \), i.e. the same element of \( \mathcal{A} \) is obtained by applying \text{eval} to either side of the axiom.
In the following, we abbreviate using \( \hat{L}(\sigma, t_\alpha) := L_1(v_1 \ldots L_{k_\alpha}(v_{k_\alpha} \ldots t_{\lambda, \alpha})) \) and \( \hat{U}_\sigma t := U_{1, \sigma_1} \ldots U_{k, \sigma_k} \alpha t, \) in addition to \( Y^2 t := t \lor Y t \) that we saw in §2.3. For example, \( \hat{L}(\sigma, \hat{U}(\frac{a}{b}, \frac{c}{d}) Y^2 t_\alpha) = L_2(v, L_3(w, L_2(u, U_{a, b} U_{c, d} (t_w \lor Y t_u)))) \).

We use similar shorthands with interpretations of operators as well.

### 4 A Monad for Shared-State

The next part in our denotational semantics is a monad whose elements represent equivalence classes of \( \text{RES} \). The monad can be obtained via a universal construction \([28]\) (by quotienting the terms by the equational theory), but a concrete representation is crucial to reason about it; for example, to show that two denotations are different.

#### 4.1 Difficulty of Term Normalization

To motivate the definitions building up to this concrete representation, we first find a representative for each equivalence class in \( \text{Term}_X/\text{RES} \), by taking an arbitrary \( t \in \text{Term}_X \) and transforming it via equations in \( \text{RES} \) to a particular form – a normal form – such that there is only one term of this form equal to \( t \).

Consider an algebraic term \( t \in \text{Term}_X \). Using \( \text{LU}-\text{noop} \) once for each location, a sequence of \( \text{LU}-\text{comm} \) and \( \text{ND-return} \) we find that \( t = \hat{L}(\sigma, \mathcal{V} \mathcal{S} \hat{U}_\sigma t) \). Note:

\[
\hat{U}_\sigma L_\ell(v, s_v) \overset{\text{RES}}{=} \hat{U}_\sigma s_{\ell v} \quad \hat{U}_\sigma U_{\ell, v}s \overset{\text{RES}}{=} \hat{U}_\sigma s_{\ell v} \quad \hat{U}_\sigma \mathcal{S} \overset{s_{\ell v}}{=} \hat{U}_\sigma \mathcal{S} \quad \hat{U}_\sigma \mathcal{V} \overset{s_{\ell v}}{=} \hat{U}_\sigma \mathcal{V} \quad \hat{U}_\sigma \mathcal{S} \overset{s_{\ell v}}{=} \hat{U}_\sigma \mathcal{S}
\]

By applying these equalities left-to-right as long as possible, and applying \( \text{ND-join} \) and \( \text{ND-epi} \) to rearrange the sums, we find that \( t \) is equal to a term of the form \( \hat{L}(\sigma, \mathcal{V} \mathcal{S} \hat{U}_\rho, \mathcal{S} s_{\ell, \rho, \sigma}) \), where \( s_{\ell, \rho, \sigma} \) is either in \( \mathcal{X} \) or is of the form \( Y s'_{\ell, \sigma} \).

For every \( \sigma \), we can rearrange the sum according to common prefixes, thus we find that \( t \) is equal to a term of the form: \( \hat{L}(\sigma, \mathcal{V} \mathcal{S} \hat{U}_\rho \mathcal{V} \mathcal{S} s_{\ell, \rho, \sigma} Y s'_{\ell, \sigma}) \), where \( s_{\ell, \rho, \sigma} \) is either in \( \mathcal{X} \) or is of the form \( Y s'_{\ell, \sigma} \) (we can take \( \alpha_{\rho, \sigma} = 0 \) when the prefix \( \hat{U}_\rho \) did not appear in \( t \)). For every \( \rho \), we can rearrange to obtain the form:

\[
\hat{L}(\sigma, \mathcal{V} \mathcal{S} \hat{U}_\rho (Y r_{\rho, \sigma} \lor \lor X x_{\rho, \sigma}))
\]  

(5)

This is not yet a normal form, which to obtain would require recursively applying this procedure to \( r_{\rho, \sigma} \) and propagating empty choice operators outward.

Were we to continue in this way to find a normal form, we would still need to prove uniqueness and completeness. One standard way to achieve this is to show that this procedure equates the sides of every axiom and respects the deduction rules of equational logic. This requires a careful proof-theoretic analysis of this normalization procedure. Instead, we take a model-theoretic approach, akin to normalization-by-evaluation, constructing for every set a concrete representation of the free \( \text{RES} \)-model over it. This representation is based on finite sets of traces.
4.2 Traces

Brookes \cite{Brookes1981} defined a trace to be a non-empty sequence of transitions, where a \textit{transition} is a pairs of stores; e.g. \((\langle 0 1 0 \rangle, \langle 0 b 0 \rangle), (\langle 0 1 1 \rangle, \langle 0 b f \rangle)\). Brookes used traces to define a denotational semantics for an imperative concurrent programming language. In Brookes’s semantics, traces denote interrupted executions, where each transition corresponds to an uninterrupted sequence of computation steps that starts with the first store and end with the second store. The breaks between transitions are where the computation yields to the environment.

The concept was adapted by many, including Benton et al. \cite{Benton1990}, to define denotational semantics for a functional language, where they have added an additional value at the end of the sequence to refer to the value the computation returns. A \textit{trace} in this paper will refer to this concept: a non-empty sequence of transitions followed by an additional return value. If we wish to specify \(X\) as the set of the return values, we will call it an \(X\)-trace. For example, if \(x \in X\) then \((\langle 0 b f \rangle, (\langle 0 1 1 \rangle, \langle 0 b f \rangle)) \ x\) is an \(X\)-trace.

For sets \(Y, Z\), we denote by \(Y^+\) the set of sequences over \(Y\), by \(Y^*\) the set of non-empty sequences over \(Y\), and \(Y \cdot Z := \{yz \mid y \in Y, z \in Z\}\). That is, \((\cdot)\) is just notation for \((\times)\) which suggests that the elements of the set are written in sequence, eliding the tuple notation. Thus \((S \times S)^+\cdot X\) is the set of \(X\)-traces.

Following our discussion in \S4.1, our representation will use finite sets of \(X\)-traces instead of the algebraic syntax. In particular, the form we have found \[\text{rep } t := \{(\sigma, \rho) \tau \mid \sigma, \rho \in S, \tau \in \text{rep } r_{\rho, \sigma}\} \cup \{(\sigma, \rho) x_{j, \rho, \sigma} \mid \sigma, \rho \in S, j < \beta_{\rho, \sigma}\} \quad (6)\]
The model-theoretic approach we use below obviates the need for the syntactic manipulation that leads to the form in \(\text{rep}\) as part of finding the representation. In the model definition, \(\text{eval}\) will play the role of \(\text{rep}\).

4.3 Model Definition

We represent elements of \(\text{Term}X/\text{Res}\) by \(\mathcal{T}X := \mathcal{P}_{\text{fin}}((S \times S)^+ \cdot X)\), i.e. finite sets of \(X\)-traces. We equip \(\mathcal{T}X\) with a \(\text{Res}\)-algebra structure \(\mathcal{F}X\):

\[
\begin{align*}
\tilde{L}_\ell(v, P_i) &:= \{(\sigma, \rho) \tau \mid (\sigma, \rho) \tau \in P_{\alpha_i}\} \\
\tilde{U}_{\ell,v} P &:= \{(\sigma, \rho) \tau \mid (\sigma[\ell \mapsto v], \rho) \tau \in P\}
\end{align*}
\]

We further equip it with \(\text{env } x := \{(\sigma, \sigma) x \mid \sigma \in S\}\) to make it a \(\text{Res}\)-algebra over \(X\). We denote \(\text{env } x\) by \(\text{return } x\), or \(\tilde{x}\) for shorthand. This \(\text{Res}\)-algebra is in fact a \(\text{Res}\)-model over \(X\) by virtue of satisfying the axioms of \(\text{Res}\):

\text{Example 5.} We verify that \(\mathcal{F}X, \text{return}\) satisfies the axiom \(\text{LU-noop}\):

\[
\text{eval}(L_\ell(v, U_{\ell,v}\tilde{x})) = \tilde{L}_\ell(v, \tilde{U}_{\ell,v}\tilde{x}) = \tilde{L}_\ell(v, \{(\sigma, \sigma[\ell \mapsto v]) x \mid \sigma \in S\}) = \{(\sigma, \sigma[\ell \mapsto \sigma]) x \mid \sigma \in S\} = \{(\sigma, \sigma) x \mid \sigma \in S\} = \tilde{x} = \text{eval } x
\]
4.4 Correspondence to Non-Deterministic Global-State

The theory of non-deterministic global-state (the fragment of RES excluding \( Y \)) admits a concrete representation using non-deterministic state transformers \( S \to \mathcal{P}_{\text{fin}}(S \times S) \) \[14\]. This representation corresponds to the one we have defined in an interesting way. Namely, there is a bijection between \( \mathcal{T} \times \sigma \) and the set of functions mapping stores to finite sets of \( X \)-traces-with-the-first-store-removed: \[
\lambda P \in \mathcal{T}, \sigma. \{ \rho \tau \in \mathcal{P}_{\text{fin}}(S \times S) \mid \langle \sigma, \rho \rangle \tau \in P \}
\]

Implicitly identifying the two, the model from §4.3 can be defined using formulas that look exactly like the non-deterministic global-state ones:

\[
\tilde{L}_e(v. P_v) \defeq \lambda \sigma. P_{\sigma v} \quad \tilde{U}_{e,v} P \defeq \lambda \sigma. P(\sigma [e \mapsto v])
\]

However, these are not the same formulas – they are defined for different elements (sets of traces as opposed to non-deterministic state transformers).

Using this identification for the yield operator, we obtained the definition \( \tilde{Y} P \defeq \lambda \sigma. \{ \sigma \tau \mid \tau \in P \} \), which we understand as “the thread does not modify the state, then allows the environment to intervene, and then continues as before.”

4.5 Representation Theorem

The model \( \langle FX, \text{return} \rangle \) defined in §4.3 represents TermX/RES because – according to the representation theorem – this is a free RES-model on \( X \), and therefore equivalent to the model of equivalence classes we used in §2 or the model of syntactic normal forms to which we have alluded in §4.1.

To prove that the model is free we first equip the family of sets \( \mathcal{T} \) with a monad structure. For every RES-model \( A \) and function \( f : X \to A \), define \( \models f : \mathcal{T} \times \sigma \to A \), the homomorphic extension of \( f \) along return, recursively; where \( R_P^{(\sigma, \rho)} \defeq \{ \tau \in (S \times S)^+ \cdot X \mid \langle \sigma, \rho \rangle \tau \in P \} \) and \( X_{P,f}^{(\sigma, \rho)} \defeq \tilde{Y}^A(\tilde{R}_P^{(\sigma, \rho)} f) \). We show it is induced by the aforementioned family of free RES-models:
Theorem 1 (Representation for shared-state). The pair $\langle F_X, \text{return} \rangle$ is a free Res-model on $X$: for every Res-model $A$ and $f : X \to A$, the function $\| f \| : T X \to A$ is the unique homomorphism $g$ satisfying $f = g \circ \text{return}$.

As a direct consequence:

Corollary 1 (Model is sound and complete). Terms over $X$ are equal in Res iff they have the same representation in $\langle F_X, \text{return} \rangle$.

4.6 Synchronization

To define the denotational semantics of $(\|)$ in §5, we will define a corresponding function $(\|\|)$ on elements of the monad. To this end we first define the trace synchronization, an inductively defined relation $\tau_1 \| \tau_2 \iff \langle \sigma, \rho \rangle \| \langle \rho, \sigma \rangle \beta y \iff \langle \sigma, \varsigma \rangle \beta (x, y)$ (Var-Left)

\[
\begin{align*}
\tau_1 \| \tau_2 \iff \langle \sigma, \rho \rangle \| \langle \rho, \varsigma \rangle \\
\langle \sigma, \rho \rangle \| \langle \rho, \varsigma \rangle \beta y \iff \langle \sigma, \varsigma \rangle \beta (x, y)
\end{align*}
\]

(Var-Right) (Brk-Left) (Seq-Left)

Symmetrically: (Brk-Right) (Seq-Right)

One way to understand these rules is to concentrate on the first transition on the left trace $\tau_1 = \langle \sigma, \rho \rangle \tau_1'$; the right-sided rules are treated symmetrically. If the first transition is also the last, i.e. $\tau_1' \in X$, then $\rho$ must be the initial store when the execution continues (recall that only a break between transitions reflects a yield to the environment). This is why $\text{Var-Left}$ combines the transitions as it does. The value in $\tau_3$ is the pair of the values in $\tau_1$ and $\tau_2$, reflecting the operational semantics of $(\|)$ returning the pair of the results. If, on the other hand, the first transition is not the last, then we may combine the transition with the continuation of the computation (Seq-Left), or we may not (Brk-Left).

The first option means the yield was used-up in this synchronization; while in the second option yield remains available to ambient synchronizations.

From this relation we derive the semantic synchronization function:

\[(\|\|) : T X \times T Y \to T (X \times Y) \quad P \|\| Q := \{ \omega \mid \exists \tau \in P, \pi \in Q. \tau \| \pi \Rightarrow \omega \}\]

Example 6. For $\sigma, \rho \in S$, we may synchronize $\langle \sigma, \rho \rangle \langle \rho, \sigma \rangle \langle \rangle$ and $\langle \rho, \rho \rangle 0$ so:

\[
\begin{align*}
\langle \rho, \sigma \rangle \langle \rangle \| \langle \rho, \rho \rangle 0 & \Rightarrow \langle \rho, \sigma \rangle \langle \rangle, 0 \text{ Var-Right} \\
\langle \sigma, \rho \rangle \langle \rho, \sigma \rangle \langle \rangle \| \langle \rho, \rho \rangle 0 & \Rightarrow \langle \sigma, \sigma \rangle \langle \rangle, 0 \text{ Seq-Left}
\end{align*}
\]
Therefore, if \( (\sigma, \rho) \{ \rho, \sigma \} \in P \) and \( (\rho, \rho) 0 \in Q \), then \( (\sigma, \sigma) \{(), 0\} \in P \parallel Q \).

The use of \texttt{SEQ-LEFT} was possible since the stores happen to match, resulting in a trace that does not allow the environment to interfere. By using \texttt{BRK-LEFT} we could find a different synchronization, one that does yield to the environment.

5 Denotational Semantics

With the monad in place, denotations of types and contexts are standard [30]:

\[
\begin{align*}
[(A_1 \ast \cdots \ast A_n)] & := [A_1] \times \cdots \times [A_n] & & [(A \rightarrow B)] := [A] \rightarrow T[B] \\
\{\ell_1 \text{ of } A_1 | \cdots | \ell_n \text{ of } A_n\} & := \bigcup_{i=1}^{n} \{\ell_i\} \times [A_i] & & [\Gamma] := \Pi(x:A)\ell[\Gamma][A]
\end{align*}
\]

Define the extension of \( \gamma \in [\Gamma] \) to \( \gamma[\sigma \rightarrow y] \in [\Gamma, \sigma : A] \) by \( \gamma[\sigma \rightarrow y]x := y \).

On the above we base two kinds of denotations for programs \( \sigma \vdash \ell : A \):

**Commutational.** \([M]^c : [\Gamma] \rightarrow T[\ell] \). When \( \Gamma \) is empty we may write \([M]^c\) instead of \([M]^c()\). We write \([M]^c \subseteq [N]^c\) for \( \forall \gamma \in [\Gamma] . \, [M]^c \gamma \subseteq [N]^c \gamma \).

**Valuational.** \([V]^v : [\Gamma] \rightarrow [A]\) defined solely for values, and satisfying \([V]^v \gamma = \text{return}([V]^v \gamma)\). When \( \Gamma \) is empty we may write \([V]^v\) instead of \([V]^v()\); and if furthermore \( A \) is a ground type, we may write \( V \) instead of \([V]^v\), noting that the restriction of \([-]^v\) to closed programs of ground type is a bijection.

Most denotations of programs are standard as well, such as:

\[
\begin{align*}
[()]^v \gamma & := () & \lambda x . [M]^v \gamma & := \lambda y . [M]^c \gamma[\sigma \rightarrow y] \\
[x]^v \gamma & := \gamma x & [N M]^c \gamma & := [N]^c \gamma [\ell \gamma] = f
\end{align*}
\]

The denotations of the state effects allow the environment to intervene:

\[
\begin{align*}
[M?]^c \gamma & := [M]^c \gamma \triangleright \ell. \operatorname{Brk-Left}(v . \check{Y}^v \check{v}) \\
[M := N]^c \gamma & := [M]^c \gamma \triangleright \ell . [N]^c \gamma \triangleright \ell . \check{U} . \check{Y}^v \check{v}(\check{v}) \\
[M \parallel N]^c \gamma & := [M]^c \gamma \parallel [N]^c \gamma
\end{align*}
\]

**Example 7.** With the definitions above, we can state the denotations from \texttt{Example 3} precisely. For instance, \([\ell] \) becomes:

\[
[a := b?; a := c]?^c = \operatorname{Brk-Left}(w . \check{Y}^v \check{Y}^v \check{U} . a, w . \check{Y}^v \check{U} . a, \check{Y}^v (\check{v}))
\]

**Example 8.** We can explain the execution of \( a := b?; a := c? \parallel a? \) from \texttt{Example 2} in denotational terms. First we find traces to synchronize:

\[
\langle (a \ b \ b \ b), (a \ b \ b \ b) \rangle \langle (a \ b \ b \ b), (a \ b \ b \ b) \rangle \langle (a \ b \ b \ b), (a \ b \ b \ b) \rangle
\]

Following from the derivation in \texttt{Example 6} with \( \sigma = (a \ b \ b \ b) \) and \( \rho = (a \ b \ b \ b) \):

\[
\langle (a \ b \ b \ b), (a \ b \ b \ b) \rangle \langle (a \ b \ b \ b), (a \ b \ b \ b) \rangle \langle (a \ b \ b \ b), (a \ b \ b \ b) \rangle
\]

This trace corresponds to the (uninterrupted) execution presented in \texttt{Example 2}.
6 Metatheoretical Results

First we find that the single-transition traces in the denotation of a program account for the possible executions of that program:

**Theorem 2 (Soundness).** If \( \sigma, M \rightarrow* \rho, V \) then \( \langle \sigma, \rho \rangle \}[V]^v \in [M]^v \).

For the proof, omitted for brevity, we instrument the operational-semantics with actions, elements of \( \{ U_{k,v}, L_t, \varepsilon \} \) signifying the effect caused by the step, to analyse the change to the denotation of the program as it runs.

Working our way up to the fundamental lemma, we define a unary logical relation: functions \( \mathcal{V}(\cdot) \) and \( \mathcal{E}(\cdot) \) from types to sets of closed programs by mutual recursion. Specifically, \( \mathcal{V}(A) \) is a set of closed values of type \( A \), and \( \mathcal{E}(A) \) is a set of closed programs of type \( A \). The definition of \( \mathcal{V}(\cdot) \) is standard:

\[
\mathcal{V}(A \rightarrow B) := \{ \lambda x. M \mid \forall V \in \mathcal{V}(A). M[V/x] \in \mathcal{E}(B) \}
\]

\[
\mathcal{V}(\{A_1 \cdots \cdots A_n\}) := \{(V_1, \ldots, V_n) \mid \forall i. V_i \in \mathcal{V}(A_i)\}
\]

\[
\mathcal{V}(\{t_1 \text{ of } A_1 | \cdots | t_n \text{ of } A_n\}) := \cup_i \{t_i V \mid V \in \mathcal{V}(A_i)\}
\]

The definition of \( \mathcal{E}(\cdot) \) is also standard in that it ensures programs in \( \mathcal{E}(A) \) compute to values in \( \mathcal{V}(A) \), but bespoke in its requirement about how they compute. This requirement is based on the way traces specify interrupted executions, a notion we have discussed in §4.2 and now make precise. For a non-empty sequence of transitions \( \alpha = \langle \sigma_1, \rho_1 \rangle \cdots \langle \sigma_m, \rho_m \rangle \) we write \( M \xrightarrow{\alpha} N \) when there exist \( M = M_1, M_2 \ldots M_m, M_{m+1} = N \) such that \( \sigma_i, M_i \rightarrow* \rho_i, M_{i+1} \) for all \( i \in \{1, \ldots, m\} \). We write \( M \xrightarrow{\alpha} V \) when \( M \xrightarrow{\alpha} V \) and \( [V]^v = x \). We now define:

\[
\mathcal{E}(A) := \{M \in \vdash A \mid \forall \tau \in [M]^v \exists V \in \mathcal{V}(A). M \xrightarrow{\tau} V\}
\]

The last component needed is the function \( \mathcal{G}(\cdot) \) from typing contexts to sets of program substitutions: \( \mathcal{G}(\Gamma) := \{\Theta \mid \forall (x : A) \in \Gamma. \Theta x \in \mathcal{V}(A)\} \). The semantic typing judgment \( \Gamma \vdash M : A \) is then defined as: \( \forall \Theta \in \mathcal{G}(\Gamma). \Theta M \in \mathcal{E}(A) \).

**Theorem 3 (Fundamental Lemma).** If \( \Gamma \vdash M : A \) then \( \Gamma \vdash M : A \).

This brings us one step closer to proving the theorem of directional adequacy. One piece is still missing: since the theorem assumes set inclusion of denotations rather than equality, we will need a different form of compositionality of the denotations than the one that holds by definition.

To state this form of compositionality we first define the standard notion of a program with holes. A function \( \Xi[-] : \Gamma \vdash A \rightarrow A \rightarrow B \) is a program context (or context for short) if, in the language extended with a program \( \bullet \) and additional axioms \( \Gamma' \vdash \bullet : A \) for all \( \Gamma' \geq \Gamma \), we have \( \Delta \vdash \Xi[\bullet] : B \); and if \( \Gamma \vdash M : A \), then \( \Xi[M] \) is obtained from \( \Xi[\bullet] \) by replacing every occurrence of \( \bullet \) with \( M \).

**Theorem 4 (Compositionality).** Let \( \Xi[-] : \Gamma \vdash A \rightarrow \bullet \rightarrow G \) be a context for ground \( G \), and \( M, N \in \Gamma \vdash A \). If \( [M]^c \subseteq [N]^c \) then \( [\Xi[M]]^c \subseteq [\Xi[N]]^c \).
The condition that the context be closed and ground is necessary, so an attempt to prove directly by induction on the structure of the context fails. The proof, omitted for brevity, instead uses a binary logical relation approximating containment that identifies with it on ground types; the main ingredient being:

$$E^g(A) := \{ \langle P, Q \rangle \in T [A] \times T [A] \mid \forall \alpha x \in P \exists \beta y \in Q. \alpha = \beta \land \langle x, y \rangle \in V^g(A) \}$$

With this compositionality in hand we are finally ready to prove the main result of this paper, that we will then use to justify program transformation. To state it we first spell out the standard definition of contextual refinement.

Suppose that $$M, N \in N \vdash A$$. We say that $$M$$ refines $$N$$, and write $$M \subseteq N$$, if $$\sigma, \Xi [M] \Rightarrow \rho, V$$ implies $$\sigma, \Xi [N] \Rightarrow \rho, V$$ whenever $$\Xi [-] : \Gamma \vdash A \rightarrow \cdot \vdash G$$ is a context for ground $$G$$. This justifies the transformation $$N \Rightarrow M$$, since replacing $$N$$ with $$M$$ within a larger program introduces no additional behaviours.

**Theorem 5 (Directional Adequacy).** If $$[[M]]^c \subseteq [[N]]^c$$ then $$M \subseteq N$$.

**Proof.** Let $$\Xi [-] : \Gamma \vdash A \rightarrow \cdot \vdash G$$ be a program context for some ground $$G$$, and assume $$\sigma, \Xi [M] \Rightarrow \rho, V$$ for some $$V$$. By soundness $$\langle \sigma, \rho \rangle [V]^* \subseteq \Xi [[N]]^c$$. Using compositionality by assumption $$\langle \sigma, \rho \rangle [V]^* \subseteq \Xi [[N]]^c$$. By the fundamental lemma $$\Xi [[N]]^c \Rightarrow W$$ for some $$W$$ such that $$[[W]]^* = [[V]]^*$$. They are of ground type, so $$W = V$$. Therefore, $$\sigma, \Xi [N] \Rightarrow \rho, V$$.

### 6.1 Example Transformations

Thanks to **directional adequacy** we can now justify various transformations and equivalences using rather mundane calculations, requiring no reasoning about the context in which these transformations are to take place.

**Example 9.** We make the reasoning from **Example 4** precise.

Denote $$\Gamma := x : \text{Loc}, y : \text{Val}$$. We have $$\Gamma \vdash x := y; x ? : \text{Val}$$ and $$\Gamma \vdash x := y; y : \text{Val}$$. Let $$\gamma \in [[\Gamma]]$$, and denote $$\ell := \gamma x$$ and $$v := \gamma y$$. Calculating, we have:

$$[[x := y; x ?]]^c \gamma = \bar{U}_{\ell,v} \bar{Y}^2 \bar{L}_v(\bar{w} \cdot \bar{Y}^2 \bar{W}^\beta \bar{v} \in [[x := y; y]]^c \gamma$$

By **directional adequacy** $$x := y ; y \in x := y ; x ?$$.

**Example 10.** We validate **elimination of irrelevant reads** i.e. $$M \subseteq x ? ; M$$:

$$[[x ? ; M]]^c \gamma = [[(\lambda_1. x) ?] x ?] \gamma = [[x ?]]^c \gamma \gamma = \lambda v. [[M]]^c \gamma = \bar{Y}^2 ([[M]]^c \gamma)$$

As mentioned in [2,3], the semantics does not validate **introduction** of irrelevant reads, i.e. we have $$[[x ? ; M]]^c \notin [[M]]^c$$ even though $$x ? ; M \subseteq M$$.

**Example 11.** Thanks to our use of standard monad-based semantics, structural transformations and equivalences follow from structural reasoning, avoiding considerations relating to shared-state. For instance:

$$[[\text{if} \ y \ \text{then} \ \lambda z. K_{\text{true}} \ \text{else} \ \lambda z. K_{\text{false}}] \gamma] = [[\lambda z. K_{\gamma}]] (\gamma [x \mapsto z]) = [[x. \text{if} \ y \ \text{then} \ K_{\text{true}} \ \text{else} \ K_{\text{false}}] \gamma$$

Therefore, if $$y \ \text{then} \ \lambda x. K_{\text{true}} \ \text{else} \ \lambda z. K_{\text{false}} \ \equiv \ \lambda y. \ \text{if} \ y \ \text{then} \ K_{\text{true}} \ \text{else} \ K_{\text{false}}$$.
Finally, adequacy can help validate expected transformations involving $(∥)$:

**Example 12.** Defining map $ψP := \{α(ψx) \mid αx ∈ P\}$ we have:

$⟦⟨M, N⟩⟩^c γ ⊆ ⟦M ∥ N⟧^c γ$ (Sequencing)

$⟦M ∥ V⟧^c γ = \text{map}(λx. ⟨x, ⟦V⟧ v⟩)(⟦M⟧^c γ)$ (Neutrality)

$⟦M ∥ N⟧^c γ = \text{map}(λ⟨y, x⟩. ⟨x, y⟩)(⟦N ∥ M⟧^c γ)$ (Symm.)

$⟦(M ∥ N) ∥ K⟧^c γ = \text{map}(λ(x, ⟨y, z⟩). ⟨(x, y), z⟩)(⟦M ∥ (N ∥ K)⟧^c γ)$ (Assoc.)

Unlike the previous examples, proving the above involves careful reasoning at the level of the traces. We still gain the benefit of justifying equivalences and transformations of programs – even open ones – without resorting to analysis under arbitrary program contexts and substitutions:

$(M, N) ⊑ M ∥ N$ (Sequencing)

$(M, V) ≅ M ∥ V$ (Neutrality)

$M ∥ N ≅ \text{match } N \mid M \text{ with } ⟨y, x⟩ → ⟨x, y⟩$ (Symm.)

$(M ∥ N) ∥ K ≅ \text{match } M \mid (N ∥ K) \text{ with } ⟨⟨x, ⟨y, z⟩⟩⟩ → ⟨x, y⟩, z$ (Assoc.)

Coordinating the returned values make these somewhat awkward. More convenient but less informative forms are derivable, such as $M ; N ; K ⊑ (M ∥ N) ; K$ (mentioned as a transformation in §1) which is a consequence of (Sequencing).

### 7 Conclusion, Related Work, and Future Work

We have defined a monad-based denotational semantics for a language for shared-state providing standard higher-order semantics supporting standard meta-theoretic development. This monad is a representation of the one induced by the equational theory of resumptions, which extends non-deterministic global-state with a delaying/yielding operator [14].

Abadi and Plotkin [1] design a modification for the theory of resumptions to define a denotational semantics for a concurrent imperative programming language with cooperative asynchronous threads. We have shown that the theory of resumptions can be used as-is to define denotational semantics for concurrency, albeit of a different kind. It is interesting to note that they interpret the unary operator analogously to our interpretation of $Y^?$, rather than $Y$. By decomposing into a sum we were able to validate transformations that are not equivalences.

Benton et al. [4] also define a monad for higher-order shared-state, with additional features such as recursion and abstract locations, using Brookes’s style of semantics. Contrasting, the monad we defined is presented algebraically, and has finite sets of traces, whereas Benton et al.’s denotations are infinite even for recursion-free programs. Although this finiteness makes our definition simpler, we saw in Example 10 that it leads to a resumption-counting issue, thus less abstract semantics. It would be interesting to analyse their semantic model from the algebraic perspective as it may lead to more abstract semantics.
Like in previous work, including those mentioned above, our semantics is based on the sets of traces, originally used by Brookes [6] to define denotational semantics for an imperative concurrent language. Brookes proved that this semantics is not only directionally adequate, but also fully abstract. The proof makes crucial use of atomic execution blocks which we have not included.

Birkedal et al. [5] provide an interesting related model, given by logical relations (step-indexed, Kripke, etc.) over syntactic terms as semantics. Their language is substantially more expressive including higher-order local store, and accounts for a type-and-effect system semantics. A more precise model could lead to a monadic account that reproduces these results less syntactically.

Also of note are process calculi and algebraic laws concerning the structure of programs. Hoare and van Staden [12] give such an account for concurrent programs, unifying previous work. Their laws are much more general, parameterizing over the notions of sequencing programs and running programs in parallel. It would be interesting to discover if and how our semantics is an instance of theirs. There is also a lot of work on semantics of “while” languages where all information flows through the state, which support more advanced features such as probabilistic choice [3,11,41]. Others approach the study of concurrency through game semantics, such as Jaber and Murawski’s [15] study of the semantics of a higher-order call-by-value concurrent language. Trace semantics features in their study too, though their traces are quite different, being sequences of player/opponent actions that incrementally transform configurations.

In the future we plan to refine the type system into a type-and-effect system [18,20,22,29,39,40], by annotating the typing judgments with the allowed effects. The denotations then depend on the effect annotations, with each annotation having its own associated equational theory. This may allow additional transformations that are currently beyond this model’s reach. For example, the converse of (Sequencing) under certain syntactic and static guarantees would enable compiler parallelism.

Atomic constructs that disallow interference from the environment are a common feature of concurrent languages. Adding such constructs may be a simple matter, since we have a dedicated operator, yield, for allowing interference. Nevertheless, in the spirit of reductionism, we leave this investigation to future work.

We would also like to see how well our approach extends to weak-memory models. In particular, we believe that the timestamp-based operational semantics of the release-acquire memory model [19,24,26,37] is amenable to a similar treatment by using more sophisticated traces.

**Acknowledgments** We thank Andrés Goens for providing his perspective on a previous version of this paper, and the anonymous APLAS reviewers for their helpful feedback.
References


A Programming Language

We define a simple idealized call-by-value language for shared-state. The language is parameterized by the number of locations \( \bar{n} \) and the number of values \( \bar{m} \). The technical development is mostly standard, included here for the sake of completeness, while highlighting key points along the way.

A.1 Syntax

The language is higher order, extending Moggi’s \([30]\) computational lambda calculus with products and sums; and three shared-state constructs:

\[ M, N ::= \ldots \mid M? \mid M := N \mid M \parallel N \]

Figure 3 presents the full syntax; Figure 4 presents syntactic sugar that makes the programs written in this language more familiar and readable.

A.2 Typing

We define a typing context as a finitely supported partial map \( \Gamma \) from variable names \( x \) to types \( A \). We write \( \cdot \) for the typing context with empty domain. We will often omit the empty typing context \( \cdot \) from typing judgments. A typing context \( \Delta \) extends a typing context \( \Gamma \), written \( \Gamma \preceq \Delta \), when \( \Delta \) extends \( \Gamma \) as a
$G ::= \begin{align*}
& (G_1 \ldots \ast G_n) \quad \text{product} \\
| \{ t_1 \; \text{of} \; G_1 \} \ldots \quad \text{variant} \\
| t_n \; \text{of} \; G_n \} \quad \text{variant} \\
A, B ::= \begin{align*}
& (A_1 \ldots \ast A_n) \quad \text{product} \\
| \{ t_1 \; \text{of} \; A_1 \} \ldots \quad \text{variant} \\
| t_n \; \text{of} \; A_n \} \quad \text{variant} \\
& A \rightarrow B \quad \text{function} \\
\end{align*}
\end{align*}$

$M, N ::= \begin{align*}
& x \quad \text{variable} \\
& (M_1, \ldots, M_n) \quad \text{tuple} \\
& \lambda x. M \quad \text{constructor} \\
& \text{case} \; M \; \text{of} \; \{ \\
& \; \; \text{t}_1 \; \text{x}_1 \; \rightarrow \; N_1 \; \} \quad \text{case split} \\
& \text{let} \; x = M \; \text{in} \; N \quad \text{assignment} \\
& \_ \quad \text{par. exec.} \\
\end{align*}$

Fig. 3. The language's terms and types
partial map: \( \text{dom} \, \Gamma \subseteq \text{dom} \, \Delta \), and \( \forall x \in \text{dom} \, \Gamma. \, \Gamma x = \Delta x \). We write \( \Gamma, x : A \) for the shadowing extension of \( \Gamma \) by the type assignment \( x : A \):

\[
(\Gamma, x : A)y := \begin{cases} 
A & y = x \\
\Gamma y & \text{otherwise}
\end{cases}
\]

When shadowing the empty typing context we will simply omit it, i.e. we write \( x : A \) instead of \( \cdot, x : A \). We also write \( (x : A) \in \Gamma \) for \( \Gamma x = A \).

Figure 5 presents the typing judgment \( \Gamma \vdash M : A \), an inductively defined relation asserting that \( M \) is a well-formed term of type \( A \), under typing context \( \Gamma \). When \( \vdash M : A \) we say that \( M \) is closed. We define \( \Gamma \vdash A \) to be the set of all programs \( M \) such that \( \Gamma \vdash M : A \).

\[
\frac{\Gamma \vdash M : A}{\Gamma \vdash x : A} \quad \frac{\Gamma, x : A \vdash M : B}{\Gamma \vdash M : A \rightarrow B} \quad \frac{\Gamma \vdash M : A \quad \Gamma \vdash N : A \rightarrow B}{\Gamma \vdash \text{match} \; M \; \text{with} \; (x_1, \ldots, x_n) \rightarrow N : A}
\]

\[
\frac{\forall i \vdash M_i : A_i}{\Gamma \vdash \{M_1, \ldots, M_n\} : (A_1 \times \cdot \cdot \cdot \times A_n)} \quad \frac{\Gamma \vdash M : A_i \quad \{t_1 \; \text{of} \; A_1 \mid \cdots \mid t_n \; \text{of} \; A_n\}}{\Gamma \vdash \{t_1 \mid \cdots \mid t_n\} : A}
\]

\[
\frac{\Gamma \vdash M : (A_1 \times \cdot \cdot \cdot \times A_n) \quad \Gamma, x_1 : A_1, \ldots, x_n : A_n \vdash N : A}{\Gamma \vdash \text{case} \; M \; \text{of} \; \{t_1 \; x_1 \rightarrow N_1 \mid \cdots \mid t_n \; x_n \rightarrow N_n\} : A}
\]

\[
\frac{\Gamma \vdash M : \text{Loc}}{\Gamma \vdash M : \text{Val}} \quad \frac{\Gamma \vdash N : \text{Val}}{\Gamma \vdash M \; \text{=} \; N : 1} \quad \frac{\Gamma \vdash M : A \quad \Gamma \vdash N : B}{\Gamma \vdash M \; \parallel \; N : (A \times B)}
\]

Fig. 5. The language’s type system

A.3 Substitutions

A program substitution (or just substitution) is a partial function from program variables to closed values. We write \( M[V_1/x_1 \ldots V_m/x_m] \) for the application of the substitution that maps \( x_i \mapsto V_i \) on \( M \). For a substitution \( \Theta \) and a set of program variables \( X \), we obtain \( \Theta|_{\Theta} \) by removing \( X \) from \( \Theta \)'s domain. A substitution \( \Theta \) extends to a function defined on all programs: first by acting as the identity on all other variables; and then by recursively applying to subprograms, removing variables from the domain of the substitution when going under binders, e.g.
\[ \Theta(\text{match } M \text{ with } (x_1, \ldots, x_n) \rightarrow N) \]
\[ := \text{match } \Theta M \text{ with } (x_1, \ldots, x_n) \rightarrow \Theta_{\{x_1, \ldots, x_n\}} N \]

We relate substitutions to denotations in the following lemma.

**Lemma 1 (Substitution).** Let \( \Gamma \) and \( \Delta \) be typing contexts such that \( \Gamma \leq \Delta \). Let \( \Theta \) be a substitution such that \( \text{dom } \Theta = \text{dom } \Delta \setminus \text{dom } \Gamma \) and \( \forall x \in \text{dom } \Theta. \, x \mapsto \Delta x \). Let \( \gamma \in \llbracket \Gamma \rrbracket \), and define \( \gamma^\Theta \in \llbracket \Delta \rrbracket \) by \( \forall x \in \text{dom } \Theta. \, \gamma^\Theta x := \llbracket \Theta x \rrbracket^\gamma \). Assume \( \Delta \vdash M : A \) for some \( A \). Then \( \llbracket M \rrbracket^\gamma \Theta = \llbracket \Theta M \rrbracket^\gamma \).

### A.4 Operational Semantics

The reduction step \( \sigma, M \xrightarrow{a} \sigma', M' \) is an inductively defined relation asserting that, with initial store \( \sigma \), the closed term \( M \) reduces in a single computation step — via action \( a \in \{U_{\ell,v}, L_{\ell}, \epsilon\} \) — to the closed term \( M' \), and the store changes to \( \sigma' \). Figure 6 presents the inference rules for the shared-state constructs. The other rules reflect standard small-step rules, where the action is \( \epsilon \) in axioms, and the action carries over in unary rules; e.g. \((\begin{array}{ccc} a & b & c \\ 1 & 0 & 1 \end{array})\), \( b := a ? \) \xrightarrow{L} \((\begin{array}{ccc} a & b & c \\ 1 & 0 & 1 \end{array})\), \( b := 1 \).

We write \( \xrightarrow{*} \) for its reflexive-transitive closure; e.g. \((\begin{array}{ccc} a & b & c \\ 1 & 0 & 1 \end{array})\), \( b := a ? \xrightarrow{*} \((\begin{array}{ccc} a & b & c \\ 1 & 1 & 1 \end{array})\), \( \langle \rangle \).

### B Proofs

The proofs below were omitted from the body of the paper for the sake of brevity.

#### B.1 Representation Theorem

**Proof (Theorem 1).** We will use the following notation for conditionals:

\[ C ? t : f := \begin{cases} t & C \\ f & \text{ else} \end{cases} \]

and a notation that helps with decomposing sets of traces by their first transition:

\[ \llbracket \tau \rrbracket := (S \times S) \times T X \rightarrow P_{\text{fin}} \left( (S \times S)^* \cdot X \right) \]

In the notations of §4.5, \( R^P_{\sigma, \rho} \) is \( (\sigma, \rho) / P \setminus X \) and \( X^P_{\sigma, \rho} \). First show that \( f = (\langle \rangle = f) \circ \text{return} \) by using the equations of RES:

\[ \text{return } x \rangle = f = \tilde{\Lambda}^A \left( \sigma, \tilde{\Lambda}^A \tilde{\eta}^A \left( \tilde{Y}^A (\emptyset) = f \right) \tilde{\Lambda}^A \right) \]

\[ \text{where } x \in X \text{ and } A = \{z \in X \mid \sigma = \rho \wedge z = x \} \]
Fig. 6. Operational semantics for shared-state constructs
Next, we show that \((\models f)\) is a homomorphism case-by-case:

**Lookup.** Since \(L_\ell(v, \tilde{L}(\sigma, x_{\sigma, v})) \equiv \tilde{L}(\sigma, x_{\sigma, \sigma})\), the equation holds in any Res-model \(A\). Note also that \((\sigma, \rho)\tilde{L}_\ell(v, P_v) = (\sigma, \rho)/P_{\sigma, \ell}\). So:

\[
\tilde{L}_\ell(v, P_v \equiv f) = \tilde{L}_\ell(v, \tilde{L}(\sigma, x_{\sigma, v})) = \tilde{L}(\sigma, x_{\sigma, \sigma}) \equiv f \quad \text{follows from} \quad U_{\sigma, v} \tilde{L}(\sigma, \sqrt{\rho \sigma}) \equiv (\sigma, \rho)\tilde{L}_\ell(v, P_v) \equiv (\sigma, \rho)/P_{\sigma, \ell} \equiv P_{\sigma, \ell} \bigtimes X.
\]

**Update.** \(\tilde{U}_{\ell, v}(P \equiv f) = \tilde{U}_\ell(v) \equiv f\) follows from \(U_{\ell, v} \tilde{L}(\sigma, \sqrt{\rho \sigma}) \equiv (\sigma, \rho)/P_{\sigma, \ell} = (\sigma, \rho)\tilde{L}_\ell(v, P_v) \equiv (\sigma, \rho)/P_{\sigma, \ell} \equiv P_{\sigma, \ell} \bigtimes X\). Thus the right-hand-side of the \(\tilde{\nu}_A\) vanishes, and more specifically we have:

\[
\tilde{Y}_A \big((\sigma, \rho)/\tilde{Y}_A \big \times X \equiv f\big) \equiv \tilde{Y}_A \big((\sigma, \rho)/\tilde{Y}_A \big \times X \equiv f\big) = \tilde{Y}_A \big(\sigma \rho \big \times X \equiv f\big) = \tilde{Y}_A \big(\sigma \rho \big \times X \equiv f\big).
\]

The desired \(\tilde{Y}_A(P \equiv f) = \tilde{Y}_A \equiv f\) now follows from:

\[
\tilde{L}(\sigma, \sqrt{\rho \sigma}) \equiv \sigma \rho \big \times X \equiv f\big) = \tilde{Y}_A \big(\sigma \rho \big \times X \equiv f\big).
\]

**Choice.** By induction on the length of the longest trace in \(\tilde{V}_{\sigma, \ell} P_v\). We have:

\[
(\sigma, \rho)/\tilde{V}_{\sigma, \ell} P_v \cap T = (\sigma, \rho)/\tilde{V}_{\sigma, \ell} P_v \cap T = (\sigma, \rho)/\tilde{V}_{\sigma, \ell} P_v \cap T
\]
\[
= \bigcup_{i < \alpha} \left( (\sigma, \rho)P_i \right) \cap T = \check{\check{\nu}} \left( (\sigma, \rho)P_i \right) \cap T \quad (\ast)
\]

In particular, \( (\sigma, \rho)\check{\check{\nu}}_{i < \alpha} P_i \setminus X = \check{\check{\nu}}_{i < \alpha} \left( (\sigma, \rho)P_i \right) \setminus X \). The longest trace here is shorter than in \( \check{\check{\nu}}_{i < \alpha} P_i \). By induction we have

\[
\left( \check{\check{\nu}} \left( (\sigma, \rho)P_i \right) \setminus X \right) \models f = \check{\check{\nu}}\check{A} \left( (\sigma, \rho)P_i \right) \setminus X \models f
\]

Also by Lemma 2, we have \( (\sigma, \rho)\check{\check{\nu}}_{i < \alpha} P_i \setminus X = \bigcup_{i < \alpha} (\sigma, \rho)P_i \setminus X \). By ND-join we have \( \check{\check{\nu}}_{i < \alpha} (\sigma, \rho)P_i \setminus X f \check{f} = \check{\check{\nu}}_{i < \alpha} \check{\check{\nu}}_{i < \alpha} (\sigma, \rho)P_i \setminus X \check{f} \check{f} \). By further rearranging \( \check{\check{\nu}}\check{A} \), the desired \( \check{\check{\nu}}_{i < \alpha} \check{A} P \models f = \check{\check{\nu}}\check{A} P \models f \) follows.

Finally, to show that \( \models f \) uniquely satisfies the above, suppose \( \mathcal{G} \) is a homomorphism satisfying \( f = \mathcal{G} \circ \text{return} \). For any \( P \in \mathcal{T}X \) and \( h : X \to A \), by induction on the length of the longest trace in \( P \), we have \( \mathcal{G}(P \models h) = P \models \mathcal{G} \circ h \); since \( \models \mathcal{G} \) is defined in terms of the operators and \( \mathcal{G} \) is a homomorphism. Furthermore, \( (\models \mathcal{G}) \text{return} \) is the identity (easily verified using the simplified definition of \( \models \mathcal{G} \)), and therefore \( \mathcal{G} P = \mathcal{G}(P \models \text{return}) = P \models \mathcal{G} \circ \text{return} = P \models f \).

**B.2 Metatheoretical Results**

### Soundness

**Lemma 2.** If \( P \subseteq Q \) then \( P \models f \subseteq Q \models f \).

**Proof.** By induction on the length of the longest trace in \( Q \).

To show **soundness**, we first show that a single step respects the denotations:

**Lemma 3.** If \( \sigma, M \models^\alpha \sigma', M' \) then \( a^1 \models^\epsilon [M']^\sigma \subseteq [M]^\sigma \), where:

\[
U_{\ell, v} P := \check{\check{\nu}}_{\ell, v} Y^\epsilon P \\
L_{\ell} P := \check{Y}^\epsilon P \\
\epsilon^1 P := P
\]

For example, \((\frac{a}{b} \frac{\bar{a}}{\bar{b}}), a \text{?} \parallel a := 0 U_{\bar{a}, \bar{b}} \models^\epsilon (\frac{a}{b} \frac{\bar{a}}{\bar{b}}), a \text{?} \parallel \langle \rangle \), and indeed one can calculate that:

\[
U_{\bar{a}, \bar{b}} [a \text{?} \parallel \langle \rangle]^\epsilon (\frac{a}{b} \frac{\bar{a}}{\bar{b}}) \subseteq [a \text{?} \parallel a := 0]^\epsilon (\frac{\bar{a}}{\bar{b}})
\]

**Proof.** By induction on the reduction step’s derivation:

\( \sigma, (\lambda x.M) \models^\xi \sigma, M[V/x] \). By **Lemma 1**

\( \sigma, V := N \models^\alpha \sigma', V := N' \). We calculate:

\[
[V := N]^\sigma = \left( [V]^\xi \models = \lambda \ell. [N]^\xi \models = \lambda v. \check{\check{\nu}}_{\ell, v} Y^\epsilon \langle \ell \rangle \right) \sigma \\
= \left( \text{return} [V]^\nu \models = \lambda \ell. [N]^\xi \models = \lambda v. \check{\check{\nu}}_{\ell, v} Y^\epsilon \langle \ell \rangle \right) \sigma
\]
Proof (Soundness). By induction on the number of steps, in each step using

\[ a^t [N]^c \rightarrow \lambda \nu. \tilde{U}_{V, \nu} \tilde{Y}^2 (\tilde{Y}) \sigma \]

Other cases are similar to one of the above.

Lemma 4. Assume \( \tau \parallel \pi \Rightarrow \omega \). If \( M \Rightarrow V \) and \( N \Rightarrow W \), then \( M \parallel N \Rightarrow \{V, W\} \).
Proof. By induction on the assumed synchronization.

**VAR-LEFT.** There exist $\sigma, \rho, x, \varsigma, \alpha, y$ such that:
1. $\tau = \langle \sigma, \rho \rangle x$
2. $\pi = \langle \rho, \varsigma \rangle \alpha y$
3. $\omega = \langle \sigma, \varsigma \rangle \alpha \langle x, y \rangle$

By $\text{#1} \& \text{#2}$

$$M \parallel N \xrightarrow{(\sigma, \rho)} V \parallel N \xrightarrow{(\rho, \varsigma)\alpha} V \parallel W \xrightarrow{\varsigma} \{V, W\}$$

Combining the executions and using transitivity, by $\text{#3}$ we are done.

**BRK-LEFT.** There exist $\sigma, \rho, \alpha, \beta, \eta, x, y$ such that:
1. $\tau = \langle \sigma, \rho \rangle \alpha x$
2. $\pi = \beta y$
3. $\omega = \langle \sigma, \rho \rangle \eta \langle x, y \rangle$
4. $\alpha x \parallel \beta y \implies \eta \langle x, y \rangle$

By $\text{#1} \& \text{#2}$ there exists $K$ such that:

$$M \xrightarrow{(\sigma, \rho)} K \xrightarrow{\alpha x} V$$
$$N \xrightarrow{\beta y} W$$

Thus by $\text{#4}$ using the induction hypothesis:

$$M \parallel N \xrightarrow{(\sigma, \rho)} K \parallel N \xrightarrow{\eta(\langle x, y \rangle)} \{V, W\}$$

By $\text{#3}$ we are done.

**SEQ-LEFT.** There exist $\sigma, \rho, \varsigma, \alpha, \beta, \eta, x, y$ such that:
1. $\tau = \langle \sigma, \rho \rangle \alpha x$
2. $\pi = \beta y$
3. $\omega = \langle \sigma, \varsigma \rangle \eta \langle x, y \rangle$
4. $\alpha x \parallel \beta y \implies \langle \rho, \varsigma \rangle \eta \langle x, y \rangle$

By $\text{#1} \& \text{#2}$ there exists $K$ such that:

$$M \xrightarrow{(\sigma, \rho)} K \xrightarrow{\alpha x} V$$
$$N \xrightarrow{\beta y} W$$

Thus by $\text{#4}$ using the induction hypothesis:

$$M \parallel N \xrightarrow{(\sigma, \rho)} K \parallel N \xrightarrow{(\rho, \varsigma)\eta(xy)} \{V, W\}$$

By $\text{#3}$ we are done.

Other cases are symmetric.

We find the following concept relevant for the proof of the **Fundamental Lemma**: we say a trace is *preserving* if it consists solely of transitions in which the stores are equal; e.g. $\llbracket \ell ? \rrbracket^c$ consists solely of preserving traces.
Proof (Fundamental Lemma). By induction on the typing judgment.

\[ \Gamma \vdash NM : B. \] Let \( \Theta \in G(\Gamma) \), and let \( \tau \in \sem{\Theta(NM)} = \sem{\Theta N} \) \( \Rightarrow \lambda f. \sem{\Theta M} \Rightarrow f \). Therefore, there exist \( \alpha, \sigma, \rho, f, \varsigma, \pi \) such that:
1. \( \alpha(\sigma, \rho) f \in \sem{\Theta N} \)
2. \( \langle \rho, \varsigma \rangle \pi \in \sem{\Theta M} \Rightarrow f \)
3. \( \tau = \alpha(\sigma, \varsigma) \pi \)

By \#2, there exist \( \eta, \varrho, \theta, x, \mu, \omega \) such that:
4. \( \eta(\varrho, \theta) x \in \sem{\Theta M} \)
5. \( \langle \theta, \mu \rangle \omega \in fx \)
6. \( \langle \rho, \varsigma \rangle \pi = \eta(\varrho, \mu) \omega \)

By \#6, there exists \( \beta \) such that:
7. \( \eta = \langle \rho, \varsigma \rangle \beta \)
8. \( \pi = \beta(\varrho, \mu) \omega \)

By \#1 \#4 \& \#7 using the induction hypothesis, we have:

\[ \exists W \in V(A \rightarrow B). \Theta N \xrightarrow{\alpha(\sigma, \rho) f} W \]
\[ \exists V \in V(A). \Theta M \xrightarrow{\langle \rho, \varsigma \rangle \beta(\varrho, \theta)x} V \]

In particular, there exist \( x, K \) such that \( W = \lambda x. K \) and \( K[V/x] \in \E(B) \). So:

\[ \Theta(NM) \xrightarrow{\alpha(\sigma, \rho)} (\lambda x. K) (\Theta M) \xrightarrow{\langle \rho, \varsigma \rangle \beta(\varrho, \theta)} (\lambda x. K) V \xrightarrow{\langle \theta, \mu \rangle} K[V/x] \]

Furthermore, we have \( \sem{\lambda x.K}^\gamma = f \) and \( \sem{V}^\gamma = x \). By Lemma 1

\[ \sem{K[V/x]}^\gamma = \sem{K}^\gamma [x \mapsto \sem{V}^\gamma] = \sem{\lambda x. K}^\gamma [V] = fx \]

By \#5

\[ \exists U \in V(B). K[V/x] \xrightarrow{\langle \theta, \mu \rangle \omega} U \]

Combining the executions and using transitivity:

\[ \exists U \in V(B). \Theta(NM) \xrightarrow{\alpha(\sigma, \varsigma) \beta(\varrho, \mu) \omega} U \]

By \#3 \& \#8 \( \alpha(\sigma, \varsigma) \beta(\varrho, \mu) \omega = \tau \), so we are done.

\[ \Gamma \vdash M ? \Theta. Val \] Let \( \Theta \in G(\Gamma) \), and let \( \tau \in \sem{\Theta(M?)} = \sem{\Theta M} \Rightarrow \lambda \ell. \tilde{L}_\ell(v. \tilde{Y}^2 \tilde{v}) \).

Therefore, there exist \( \alpha, \sigma, \rho, \ell, \varsigma, \pi \) such that:
1. \( \alpha(\sigma, \rho) \ell \in \sem{\Theta M} \)
2. \( \langle \rho, \varsigma \rangle \pi \in \tilde{L}_\ell(v. \tilde{Y}^2 \tilde{v}) \)
3. \( \tau = \alpha(\sigma, \varsigma) \pi \)

By \#2 \( \langle \rho, \varsigma \rangle \pi \in \tilde{Y}^2 \rho \). In particular, \( \langle \rho, \varsigma \rangle \pi \) is preserving. By \#1 using the induction hypothesis, we have:

\[ \exists V \in V(\Theta. \Theta M) \xrightarrow{\alpha(\sigma, \rho) \ell} V \]
Since \textbf{Loc} is a ground type, \( V = \ell \); and since \( \langle \rho, \varsigma \rangle \pi \) is preserving:

\[
\Theta (M?) \xrightarrow{\alpha(\rho,\rho)} \ell? \xrightarrow{(\rho,\rho)} \rho_\ell \xrightarrow{(\rho,\varsigma)\pi} \rho_\ell
\]

Combining the executions and using transitivity:

\[
\Theta (M?) \xrightarrow{\alpha(\rho,\varsigma)\pi} \rho_\ell
\]

Since \( \rho_\ell \in \mathcal{V}(\mathcal{Val}) \), by \ref{eq:extensional} we are done.

\( \Gamma \vdash M \parallel N : (A \star B) \). Let \( \Theta \in \mathcal{G}(\Gamma) \), and let \( \omega \in [\Theta (M \parallel N)]^c = [\Theta M]^c \parallel [\Theta N]^c \). So there exist \( \tau \in [\Theta M]^c \) and \( \pi \in [\Theta N]^c \) such that \( \tau \parallel \pi \Rightarrow \omega \).

By the induction hypothesis it follows that:

\[
\exists V \in \mathcal{V}(A), \Theta M \xrightarrow{\tau} V
\]

\[
\exists W \in \mathcal{V}(B), \Theta N \xrightarrow{\tau} W
\]

By Lemma 4 \( \Theta (M \parallel N) \xrightarrow{\omega} (V, W) \in \mathcal{V}((A \star B)) \).

Other cases are similar.

\textbf{Compositionality} Proving the theorem directly by induction on the structure of the program context fails, in particular when attempting to handle the abstraction (\( \lambda \)) case. The condition in the theorem that the context be closed and ground is necessary, so this failure is expected. To surpass this hurdle we define a “contains up-to extensionality” logical relation.

We define functions \( \mathcal{V}^\circ(\cdot) \) and \( \mathcal{E}^\circ(\cdot) \) from types to binary relations by mutual recursion.

Specifically, \( \mathcal{V}^\circ(A) \) is a relation on \([A]\), and \( \mathcal{E}^\circ(A) \) on \( T[A] \).

\[
\mathcal{V}^\circ(A \to B) := \{(f, g) \mid \forall (x, y) \in \mathcal{V}^\circ(A). \ (f_x, g_y) \in \mathcal{E}^\circ(B)\}
\]

\[
\mathcal{V}^\circ((A_1 \star \cdots \star A_n)) := \{((x_1, \ldots, x_n), (y_1, \ldots, y_n)) \mid \forall i. (x_i, y_i) \in \mathcal{V}^\circ(A_i)\}
\]

\[
\mathcal{V}^\circ(\{\ell_1 \text{ of } A_1 | \cdots | \ell_n \text{ of } A_n\}) := \bigcup\{\{\ell_1 x, \ell_i y \mid (x, y) \in \mathcal{V}^\circ(A_i)\}
\]

\[
\mathcal{E}^\circ(A) := \{(P, Q) \mid \forall x \in P \exists y \in Q. \ x = y \wedge (x, y) \in \mathcal{V}^\circ(A)\}
\]

\textbf{Lemma 5.} \( \mathcal{V}^\circ(A) \) and \( \mathcal{E}^\circ(A) \) are reflexive for every \( A \).

\textbf{Proof.} By induction on \( A \).

\textbf{Corollary 2.} If \( P \subseteq Q \) then \( (P, Q) \in \mathcal{E}^\circ(A) \).

\textbf{Lemma 6.} \( \mathcal{V}^\circ(A) \) and \( \mathcal{E}^\circ(A) \) are transitive for every \( A \).

\textbf{Proof.} By induction on \( A \).

For every typing context \( \Gamma \) we define a binary relation \( \mathcal{G}^\circ(\Gamma) \) over \([\Gamma]\):

\[
\mathcal{G}^\circ(\Gamma) := \{(\gamma, \delta) \mid \forall (x : A) \in \Gamma. \ (\gamma x, \delta x) \in \mathcal{V}^\circ(A)\}
\]

Define the judgment \( \Gamma \vdash M \leq N : A \) as follows:

\[
\forall (\gamma, \delta) \in \mathcal{G}^\circ(\Gamma). \ [M]^c \gamma, [N]^c \delta \in \mathcal{E}^\circ(A)
\]


Lemma 7. $\Gamma \vdash N \leq N : A$ for every $N \in \Gamma \vdash A$.

Proof. By induction on $\Gamma \vdash N : A$.

Lemma 8. For $M, N \in \Gamma \vdash A$, if $[[M]]^c \subseteq [[N]]^c$ then $\Gamma \vdash M \leq N : A$.

Proof. Let $(\gamma, \delta) \in G^*(\Gamma)$. By Corollary 2, $[[M]]^c \gamma, [[N]]^c \gamma \in E^*(A)$. So by Lemma 7, $[[N]]^c \gamma, [[N]]^c \delta \in E^*(A)$. By Lemma 6, $[[M]]^c \gamma, [[N]]^c \delta \in E^*(A)$.

Lemma 9. $\vdash$ is closed under typing rules. That is, for $\Xi[\cdot] : \Gamma \vdash \Delta \vdash B$, if $\Gamma \vdash M \leq N : A$ then $\Delta \vdash \Xi[M] \leq \Xi[N] : B$

Proof. By induction on the typing of $\Xi[\cdot]$.

Lemma 10. $V^*(G)$ is equality for every ground $G$.

Proof. By induction on $G$.

Corollary 3. $E^*(G)$ is containment for ground $G$.

Proof (Theorem 4). Assume $[[M]]^c \subseteq [[N]]^c$. By Lemma 8, $\Gamma \vdash M \leq N : A$. By Lemma 9, $\Xi[M] \leq \Xi[N] : G$; i.e. $[[\Xi[M]]^c, [[\Xi[N]]^c] \in E^*(G)$. By Corollary 3, $[[\Xi[M]]^c \subseteq [[\Xi[N]]^c]$. 