Modular abstract syntax trees (MAST): substitution tensors with second-class sorts

Marcelo Fiore, Ohad Kammar, Georg Moser, and Sam Staton



 $$41^{\rm st}$$ Conference on Mathematical Foundations of Programming Semantics (MFPS XLI) 20 June 2025, University of Strathclyde, Glasgow, Scotland



Call-by-Value λ-calculus

```
A, B, C :=
             type
                                                     V.W :=
                                                                        value
                      base
                                                                           variable
   A \rightarrow B function
                                                         \lambda x : A.M function abst.
   | (C_i : A_i | i \in I) \text{ record } (I \text{ finite})
                                                         | (C_i : V_i | i \in I) record c'tor
   |\{C_i: A_i | i \in I\}\} variant (I finite)
                                                         A.C.V variant c'tor
              M.N.K.L :=
                                                         term
                     \mathsf{val}\,V
                                                            value
                  sequencing
                  M @ N
                                                            function application
                  (C_1: M_1, \ldots, C_n: M_n)
                                                            record constructor
                  | case M of (C_1x_1, \ldots, C_nx_n) \Rightarrow N
                                                            record pattern match
                  A.C_iM
                                                            variant constructor
                    case M of \{C_i x_i \Rightarrow M_i | i \in I\} N
                                                            variant pattern match
```

- Initial algebra characterisation for abstract syntax with binding-aware substitution
- Robust to extensions:
 - polymorphism
 - mechanisation
 - substructurality
- ► CBN works smoothly. Doesn't cover CBV. Technical reasons later:
 - ▶ Substitute in: values and terms
 - Substitute for variables: values only

Slogan

for substitution: values are 1st-class but terms are 2nd-class [cf. Levy's CBPV, '04]

[Fiore and Hamana'13] [Crole'11, Allais et al.'18, Fiore and Szamoszvancev'22] [Fiore and Ranchod'25]

Goal: abstract syntax with heterogeneous sorting

Sorting signature R

set sort

partitioned into

- ▶ bindable/ 1^{st} -class sorts $s \in Bind$
- ▶ non-bindable/2nd-class sorts

Example (CBV sorting signature)

- ▶ sort := $\{A, \text{comp } A | A \in \text{Type}\}$
- ▶ Bind := Type_{CBV}

Example (CBPV sorting signature)

- ▶ Bind := $\{A|A \text{ value type}\}$
- ▶ sort := Type_{CBPV}

Core contribution

 \mathbf{PSh} (sort \times sort $_{\vdash}$), \otimes monoidal product

generalise

this work (MAST)

 \mathbf{PSh} (sort \times Bind_{\vdash}), \otimes right-unital associative **skew** monoidal product

Skew monoidal heterogeneous tensor

$$(P \otimes Q) \otimes L \cong P \otimes (Q \otimes L)$$

(associative)

$$P\otimes \mathbb{I}\cong P$$

(right-unital)

$$\mathbb{I} \otimes Q \xrightarrow{\ell} Q$$

(non-invertible!)
$$(\mathbb{I} \otimes \mathbb{1})_s = \emptyset \ncong \mathbb{1}_s (s \notin \mathsf{Bind})$$

Contribution

Modular Abstract Syntax Trees (MAST)

- ► SOAS ⇒ 2nd-class sorts
 Using **skew** bicategories/monoidal categories, and:
 - Kleisli bicategories

▶ Familial theory of SOAS

[Gambino, Fiore, Hyland, and Winskel'19] [Fiore and Szamoszvancev'25]

- MAST tutorial
- Case-study: CBV semantics á la carte

(128 substitution lemmata)

WIP

▶ Idris 2 implementation of computational fragment

[cf. Fiore and Szamoszvancev'22]

► Replace skew monoidal structure and monoids with monoidal structure and actions

[cf. Fiore and Turi'01]

Talk structure

- ► Contribution
- ▶ High-level motivation
- ▶ MAST taster

High-level motivation

Initial Algebra Semantics Programme

[Goguen and Thatcher'74]

Denotational semantics á la carte

[homage to Swierstra'08, Forster and Stark'20]

CBV customisation menu

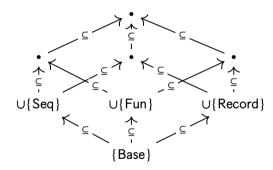
fragment	syntactic constructs	types	semantics
base	returning a value: val		strong monad over a Cartesian category
			Cartesian Category
sequential	sequencing: let		
functions	abst., app.	function	Kleisli exponentials
	$(\lambda x. : A), (@)$	(\rightarrow)	
variants	c'tors, pattern match	variant	distributive category
	$A.C_i$ -, case - of	$\{\![C_i:- i\in I]\!\}$	
	$\{C_i x_i \Rightarrow - i \in I\}$		
:			

Dream

Iterative semantic development

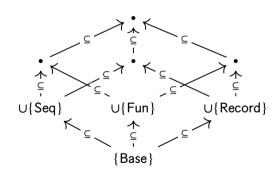
- ► Add syntax
- Add semantics

▶ Profit!



Iterative semantic development

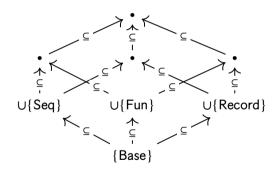
- Add syntax
- Add semantics
- ▶ Develop meta-theory:
 - Substitution lemma
 - Compositionality
 - Soundness
 - Adequacy
- Profit!



Dream vs. Bleak Reality

Iterative semantic development

- Add syntax
- Add semantics
- ▶ Develop meta-theory:
 - Substitution lemma Tedious and boring
 - Compositionality
 Tedious and boring
 - Soundness
 - Adequacy
- Profit!



Meta-theory: the tedious parts

Lemma (substitution)

Syntactic substitution corresponds to semantic composition:

$$\llbracket M \left[\theta\right] \rrbracket = \llbracket M \rrbracket \circ \llbracket \theta \rrbracket$$

Lemma (compositionality)

$$[C[M]] = plug([C[-]], [M])$$

Meta-theory: the tedious parts

Lemma (substitution)

Syntactic substitution corresponds to semantic composition:

Proof.

$$\llbracket M \left[\theta\right] \rrbracket = \llbracket M \rrbracket \circ \llbracket \theta \rrbracket$$

Presupposes a syntactic substitution lemma. Typically several inductions over all constructs.

Lemma (compositionality)

Composite semantics is independent of component syntax:

Proof.

$$[C[M]] = plug([C[-]], [M])$$

Tediously define terms with holes, plugging holes syntactically, carefully capturing some variables but not others. Then induction over semantics.

Dream

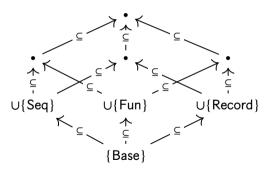
It would be nice if tedious bits were... ... free

Dream vs. Reality

It would be nice if tedious bits were...

... free

... syntactically scaleable: additive syntactic work per new feature



Capstone: abstract syntax and substitution universality

Thm (representation)

abstract syntax with operators in **O** and holes in **H**amounts to
free substitution **O**-monoid over **H**:

$$\begin{array}{c} \mathbf{H} \\ \downarrow ?-[\mathsf{id}] \\ \$\mathbf{H} \otimes \$\mathbf{H} \xrightarrow{-[-]} & \$\mathbf{H} & \longleftrightarrow \\ \llbracket - \rrbracket \uparrow \\ \mathbf{O}(\$\mathbf{H}) \end{array}$$

Plugging holes/metavariable substitution

Kleisli extension (≽) for **O**-monoid monad.

Capstone: semantics

Key propaganda

compositional, binding-respecting denotational semantics amounts to substitution **O**-monoid:

$$\begin{array}{ccc} \mathbf{M} \otimes \mathbf{M} \xrightarrow{-[-]} & \mathbf{M} & \xleftarrow{\mathsf{var}} \mathbb{I} \\ & & & & & \\ \mathbf{OM} & & & & \end{array}$$

The denotational semantics for terms with holes in ${\bf H}$ is the unique substitution ${\bf O}$ -monoid homomorphism over ${\bf H}$:

$$\left(\$\mathbf{H},-[-],\mathsf{var},[\![-]\!]\;,?-[\mathsf{id}]\right) \xrightarrow{[\![-]\!]} \left(\mathbf{M},-[-],\mathsf{var},[\![-]\!]\;,\mathsf{menv}\right) \qquad \qquad \left(\mathbf{H} \xrightarrow{\mathsf{menv}} \mathbf{M}\right)$$

Meta-theory in one line

Lemma (substitution)

Syntactic substitution corresponds to semantic composition:

$$\llbracket M \left[\theta\right] \rrbracket = \llbracket M \rrbracket \circ \llbracket \theta \rrbracket$$

Lemma (compositionality)

Meta-theory in one line

Lemma (substitution)

Syntactic substitution corresponds to semantic composition:

substitution monoid homomorphism

$$\begin{bmatrix} M \left[\theta\right] \end{bmatrix} = \begin{bmatrix} -[-] \left[M, \theta\right] \end{bmatrix} = -[-] \left[\begin{bmatrix} M \end{bmatrix}, \begin{bmatrix} \theta \end{bmatrix} \right] \coloneqq \begin{bmatrix} M \end{bmatrix} \circ \begin{bmatrix} \theta \end{bmatrix}$$

Lemma (compositionality)

$$[\![C[M]]\!] =$$

$$plug([\![C[-]]\!], [\![M]\!])$$

Meta-theory in one line

Lemma (substitution)

Syntactic substitution corresponds to semantic composition:

substitution monoid homomorphism

$$\llbracket M \left[\theta\right] \rrbracket = \llbracket -[-] \left[M, \theta\right] \rrbracket = -[-] \left[\llbracket M \rrbracket, \llbracket \theta \rrbracket \right] \coloneqq \llbracket M \rrbracket \circ \llbracket \theta \rrbracket$$

Lemma (compositionality)

Talk structure

- ► Contribution
- ► High-level motivation
- ▶ MAST taster

MAST taster: semantic domain for syntax and semantics

MAST provides ($\mathbf{R} = (\text{sort}, \text{Bind}) \text{ sorting system}$)

- Contexts
- Renamings
- **R**-structures:
- Variables structure:
- substitution tensors:

$$(P \otimes Q)_s \Gamma \ni [p,\theta]_\Delta:$$

```
\operatorname{sort}_{\sqsubseteq} \ni \Gamma := [x_1 : s_1, \dots, x_n : s_n]
                                                                  sort_{\perp}(\Gamma, \Delta) \ni \Gamma \vdash \rho : \Delta
         \mathbf{PSh}(\mathsf{sort} \times \mathsf{Bind}_{\vdash}) \ni P : \mathsf{sort} \times \mathsf{Bind}_{\vdash}^{\mathsf{op}} \to \mathbf{Set}
P_s\Gamma \ni p: sort s element with variables in \Gamma
                                    R-Struct \ni \mathbb{I}_s \Gamma \coloneqq \{x | (x : s) \in \Gamma\}
                                        R-Struct \ni P \otimes Q, P \otimes_{\bullet} \left( \text{var} \downarrow_{A}^{\mathbb{I}} \right)

P-element: p \in P_{\bullet}\Gamma
                                             Q-closure : \theta \in \prod_{(v:r) \in \Lambda} Q_r \Gamma
                   identifying, e.g.:
                          \begin{aligned} &[p[\mathsf{weaken}], \theta]_{\Delta_1 + \Delta_2} &= [p, \theta \circ \rho]_{\Delta_1} \\ &[p[x', x'' \mapsto x]_{x \in \Lambda}, \theta]_{\Delta} &= [p, \theta + \theta]_{\Delta + \Delta} \end{aligned}
```

MAST taster: semantic domain for syntax and semantics

MAST provides ($\mathbf{R} = (\text{sort}, \text{Bind}) \text{ sorting system}$)

- Contexts
- Renamings
- **R**-structures:

- $\operatorname{sort}_{\vdash} \ni \Gamma \coloneqq [x_1 : s_1, \dots, x_n : s_n]$ $\operatorname{sort}_{\vdash}(\Gamma, \Delta) \ni \Gamma \vdash \rho : \Delta$
- **PSh**(sort × Bind_⊢) ∋ P : sort × Bind_⊢^{op} → **Set** $P_{\circ}\Gamma$ ∋ p: sort s element with variables in Γ

- Variables structure:
- substitution tensors:

$$(P \otimes Q)_s \Gamma \ni [p, \theta]_{\Delta}$$
:

R-Struct $\ni \mathbb{I}_s \Gamma := \{x | (x : s) \in \Gamma\}$

R-Struct
$$\ni P \otimes Q, P \otimes_{\bullet} \left(\text{var} \downarrow_{A}^{\mathbb{I}} \right)$$

$$P\text{-element:} \quad p \in P_{s} \Gamma$$

$$Q\text{-closure:} \quad \theta \in \prod_{(v:r) \in A} Q_{r} \Gamma$$

Scope-change as tensorial strength identifying, e.g.:

$$\operatorname{str}^{\mathbf{O}}: (\mathbf{O}P) \otimes_{\bullet} \left(\operatorname{var} \downarrow_{A}^{\mathbb{I}}\right) \to \mathbf{O}\left(P \otimes_{\bullet} \left(\operatorname{var} \downarrow_{A}^{\mathbb{I}}\right)\right)$$

$$\begin{split} [p[\mathsf{weaken}], \theta]_{\Delta_1 + \Delta_2} &= [p, \theta \circ \rho]_{\Delta_1} \\ \big[p[x', x'' \mapsto x\big]_{x \in \Delta}, \theta]_{\Delta} &= [p, \theta + \theta]_{\Delta + \Delta} \end{split}$$

MAST taster: semantic domain for syntax and semantics

MAST provides ($\mathbf{R} = (\text{sort}, \text{Bind}) \text{ sorting system}$)

- Contexts
- Renamings
- R-structures:

- $\operatorname{sort}_{\sqsubseteq} \ni \Gamma := [x_1 : s_1, \dots, x_n : s_n]$ $sort_{\perp}(\Gamma, \Delta) \ni \Gamma \vdash \rho : \Delta$
- $\mathbf{PSh}(\mathsf{sort} \times \mathsf{Bind}_{\vdash}) \ni P : \mathsf{sort} \times \mathsf{Bind}_{\vdash}^{\mathsf{op}} \to \mathbf{Set}$ $P_s\Gamma \ni p$: sort s element with variables in Γ

- Variables structure:
- substitution tensors:

$$(P \otimes Q)_s \Gamma \ni [p, \theta]_{\Delta}$$
:

R-Struct $\ni \mathbb{I}_s \Gamma \coloneqq \{x | (x : s) \in \Gamma\}$

R-Struct
$$\ni P \otimes Q, P \otimes_{\bullet} \left(\text{var} \downarrow_{A}^{\mathbb{I}} \right)$$

$$P\text{-element:} \quad p \in P_{s}\Gamma$$

$$Q\text{-closure:} \quad \theta \in \prod_{(v:r) \in A} Q_{r}\Gamma$$

identifying, e.g.:

Scope-change as tensorial strength
$$\operatorname{str}^{\mathbf{O}}: (\mathbf{O}P) \otimes_{\bullet} \left(\operatorname{var} \downarrow_{A}^{\mathbb{I}}\right) \to \mathbf{O}\left(P \otimes_{\bullet} \left(\operatorname{var} \downarrow_{A}^{\mathbb{I}}\right)\right)$$

Substitution monoids

$$\otimes \mathbf{M} \xrightarrow{-[-]} \mathbf{M} \overset{\mathsf{var}}{\longleftarrow} \mathbb{I}$$

$$\begin{split} [p[\mathsf{weaken}], \theta]_{\Delta_1 +\!\!\!- \Delta_2} &= [p, \theta \! \circ \! \rho]_{\Delta_1} \\ \big[p[x', x'' \mapsto x\big]_{x \in \Delta}, \theta]_{\Delta} &= [p, \theta +\!\!\!- \theta]_{\Delta +\!\!\!- \Delta} \end{split}$$

What breaks the unitor?

Substitution tensor

$$(P \otimes Q)_s \Gamma \coloneqq \int^{\Delta} P_s \Delta \times \prod_{(v:r) \in \Delta} Q_r \Gamma$$

for $s \notin Bind$, Q = 1, $\mathbb{I}_s \Delta = \emptyset$:

$$(\mathbb{I} \otimes Q)_s \Gamma \coloneqq \int^{\Delta} \underbrace{\emptyset} \times \prod_{(v:r) \in \Delta} Q_r \Gamma = \int^{\Delta} \emptyset = \emptyset \neq \mathbb{1} = Q_s \Gamma$$

MAST taster: semantic domain for syntax

Signature functors

Scope-change as tensorial strength

$$\mathsf{str}^\mathbf{O}\,:\, (\mathbf{O}P) \otimes_{\:\raisebox{1pt}{\text{\circle*{1.5}}}} \left(\mathsf{var} \,\,{}^{\mathbb{I}}_{\underset{A}{\downarrow}}\right) \to \mathbf{O}\left(P \otimes_{\:\raisebox{1pt}{\text{\circle*{1.5}}}} \left(\mathsf{var} \,\,{}^{\mathbb{I}}_{\underset{A}{\downarrow}}\right)\right)$$

Example

Sequential fragment signature functor:

$$(\operatorname{Seq} X)_{\operatorname{comp} B} \Gamma \coloneqq \coprod_{A \in \operatorname{Type}} \begin{pmatrix} (\operatorname{let} x : A = \underline{\quad in \quad}) : \\ \left(X_{\operatorname{comp} A} \Gamma \times X_{\operatorname{comp} B} (\Gamma, x : A)\right) \end{pmatrix}$$

$$(\operatorname{Seq} X)_A \Gamma \coloneqq \emptyset$$

$$\mathsf{str}^\mathsf{Seq} \left[\mathsf{let} \ x : A = (p \in P_{\mathsf{comp} \, A} \Delta) \ \mathsf{in} \ (q \in P_{\mathsf{comp} \, B} (\Delta, x : A)), \theta \right]_{\Delta}$$

$$\coloneqq \left(\mathsf{let} \ x : A = [p, \theta]_{\Delta} \ \mathsf{in} \ [q, (\theta, x : \mathsf{var} \, x)]_{\Delta, x : A} \right)$$

Each syntactic construct defines its own binding, renaming, and substitution structure

MAST taster: semantic domain for syntax

Abstract syntax: inductive representation

Every initial algebra:

$$\mathbb{S}^{\mathbf{O}}\mathbf{H} \coloneqq \mu X.(\mathbf{O}X) \coprod \mathbb{I} \coprod \mathbf{H} \otimes X$$

Supports standard definitions:

$$\begin{array}{c} \mathbf{H} \\ \downarrow ?-[-] \\ \mathbb{S}\mathbf{H} \otimes \mathbb{S}\mathbf{H} \xrightarrow{-[-]} & \mathbb{S}\mathbf{H} & \stackrel{\mathsf{var}}{\longleftarrow} \mathbb{I} \\ \boxed{[-]] \uparrow} \\ \mathbf{O}(\mathbb{S}\mathbf{H}) \end{array}$$

Independently of concrete representation, e.g.,:

- ▶ De-Bruijn
- Nominal

- ► Locally nameless
- ► Co-de Bruijn

Graphical

Example

$$\mathbf{M} = (C, T, \text{return}, \gg, [-])$$
:

- ▶ C: Cartesian category with chosen finite products
- \blacktriangleright (T, return, \gg) strong monad over C
- ▶ [-]: Type $\rightarrow C$ type interpretation

induces:

► A CBV-structure:

$$CBV\text{-}\mathbf{Struct}\ni \mathbf{M}_s\Gamma\coloneqq\mathcal{C}(\llbracket\Gamma\rrbracket\;,\llbracket s\rrbracket)$$

▶ Standard interpretation of contexts, computations, renaming:

$$C\ni \llbracket\Gamma\rrbracket\coloneqq \prod_{(x:A)\in\Gamma}\llbracket A\rrbracket \qquad C\ni \llbracket \mathsf{comp}\, A\rrbracket\coloneqq \mathsf{T}\,\llbracket A\rrbracket$$

$$\llbracket\rho\rrbracket\colon \llbracket\Gamma\rrbracket \xrightarrow{(\pi_{x[\rho]}\colon \llbracket\Gamma\rrbracket\to \llbracket A\rrbracket)_{(x:A)\in\Delta}} \prod_{(x:A)\in\Delta}\llbracket A\rrbracket = \llbracket\Delta\rrbracket$$

MAST taster: common structure for substitution

Syntactic substitution monoid

$$\mathbb{S}^{O}H \otimes \mathbb{S}^{O}H \xrightarrow{-[-]} \mathbb{S}^{O}H \xleftarrow{\text{var}} \mathbb{I}$$

Monoid axioms amounts to syntactic substitution lemma

Example

Semantic substitution monoid:

$$M \otimes M \xrightarrow{-[-]} M \xleftarrow{\text{var}} \mathbb{I}$$

▶ Substitution via composition:

$$\left(\left[\left[\Delta \right] \right] \xrightarrow{f} \left[\left[s \right] \right] \right) \left[\left[\left[\Gamma \right] \xrightarrow{\theta} \left[\left[\Delta \right] \right] \right] : \left[\left[\Gamma \right] \xrightarrow{\theta} \left[\left[\Delta \right] \right] \xrightarrow{f} \left[\left[s \right] \right] \right]$$

 Variables: (1st-class sorts only)

$$\operatorname{var}:\left((x:A)\in\Gamma\mapsto\left(\llbracket\Gamma\rrbracket\xrightarrow{\pi_x}\llbracket A\rrbracket\right)\right)$$

MAST taster: compatibility

Substitution-compatible algebra

$$\begin{array}{c} [\![-]\!] : \mathbf{OM} \to \mathbf{M}: \\ & \text{str} \underbrace{\mathbf{O}(\underline{\mathbf{M}} \otimes \underline{\mathbf{M}})}_{\text{compatibility}} \\ \underbrace{(\mathbf{OM}) \otimes_{\bullet} \operatorname{env}^{\mathbf{M}}}_{\text{compatibility}} \\ = \underbrace{\mathbf{OM}}_{\text{[}-]\!]} \otimes \operatorname{id} \times \\ & \underbrace{\mathbf{M}} \otimes \underline{\mathbf{M}} \\ & \underbrace{-[-]_{\mathbf{M}}} \times \underline{\mathbf{M}} \\ \end{array}$$

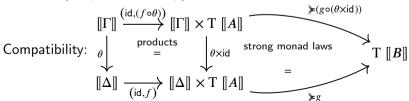
Example (Seq-compatibility)

Example (Seq-algebra)
$$\begin{bmatrix} let x : A = (\llbracket \Gamma \rrbracket \xrightarrow{f} T \llbracket A \rrbracket) \\ in (\llbracket \Gamma \rrbracket \times \llbracket A \rrbracket \xrightarrow{g} T \llbracket B \rrbracket) \end{bmatrix} :$$

$$\llbracket \Gamma \rrbracket \xrightarrow{(id,f)} \llbracket \Gamma \rrbracket \times T \llbracket A \rrbracket \xrightarrow{kg} T \llbracket B \rrbracket$$
Takeausy

Takeaway

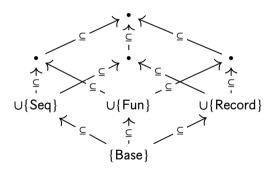
Equip each semantic interpretation with its compatibility proof



MAST: modularity and scalability

Substitution O-monoid

Substitution monoid with compatible O-algebra structure



Want more?

In the paper:

- ▶ All the details
- ▶ A CBV case-study (128 substitution lemmata)



In the future:

▶ Idris 2 implementation of computational fragment

[cf. Fiore and Szamoszvancev'22]

► Replace skew monoidal structure and monoids with monoidal structure and actions

Contribution

Modular Abstract Syntax Trees (MAST)

- ► SOAS ^{generalise} 2nd-class sorts
 Using **skew** bicategories/monoidal categories, and:
 - Kleisli bicategories
 - ► Familial theory of SOAS

- [Gambino, Fiore, Hyland, and Winskel'19]
 - [Fiore and Szamoszvancev'25]

- MAST tutorial
- Case-study: CBV semantics á la carte

(128 substitution lemmata)

WIP

- ▶ Idris 2 implementation of computational fragment
 - [cf. Fiore and Szamoszvancev'22]
- ► Replace skew monoidal structure and monoids with monoidal structure and actions

[cf. Fiore and Turi'01]