

Melocoton: A Program Logic for Verified Interoperability Between OCaml and C

Armaël Guéneau

Johannes Hostert

Simon Spies

Michael Sammler

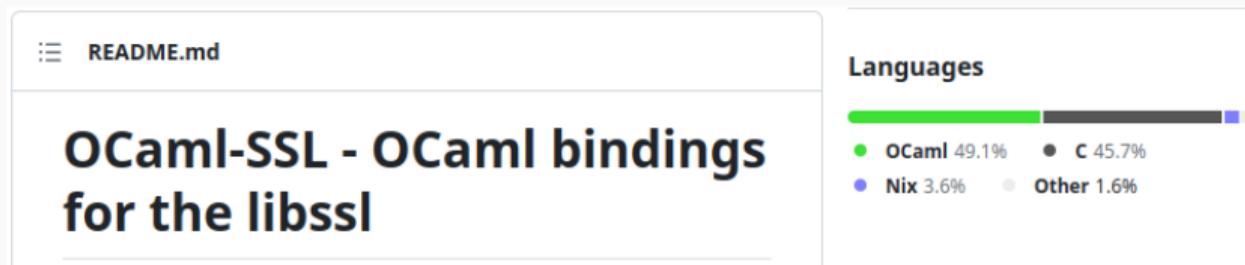
Lars Birkedal Derek Dreyer

May 24, 2023

Many Real Programs Are Multi-Language

Consider the `ocaml-ssl` library:

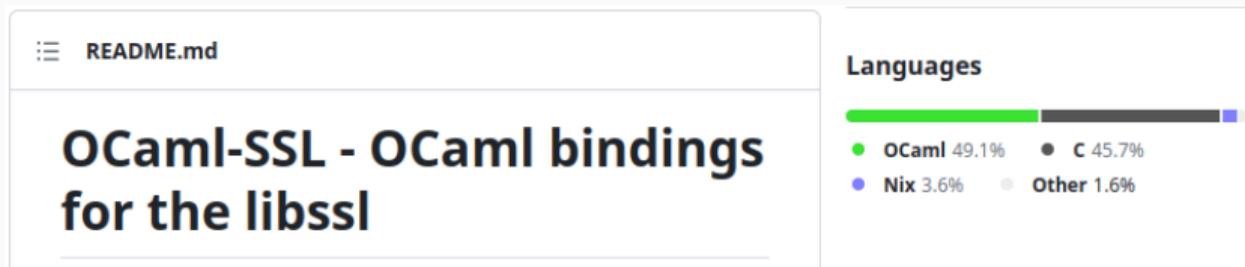
- Exposes OpenSSL (a C library) as an OCaml library
- To do so, it is implemented using a mix of *both* OCaml *and* C code:



Many Real Programs Are Multi-Language

Consider the `ocaml-ssl` library:

- Exposes OpenSSL (a C library) as an OCaml library
- To do so, it is implemented using a mix of *both* OCaml *and* C code:



How do we reason about such code (in Iris)?

Mind the gap!

OCaml

C

Mind the gap!

OCaml

Structured values

$$\boxed{\lambda_{\text{ML}}} \quad V \in \text{Val} ::= (n \in \mathbb{Z}) \mid (\ell \in \text{Loc}) \\ \mid \text{true} \mid \text{false} \\ \mid \langle \rangle \mid \langle V, V \rangle \dots$$

Garbage collection

$$\boxed{\text{Iris}_{\text{ML}}} \quad \ell \mapsto_{\text{ML}} \vec{V}$$

C

Integers and pointers

$$\boxed{\lambda_{\text{C}}} \quad w \in \text{Val} ::= (n \in \mathbb{Z}) \mid (a \in \text{Addr})$$

Manual memory management

$$\boxed{\text{Iris}_{\text{C}}} \quad a \mapsto_{\text{C}} w$$

Mind the gap!

OCaml ← OCaml FFI → C

Structured values

λ_{ML} $V \in \text{Val} ::= (n \in \mathbb{Z}) \mid (\ell \in \text{Loc})$
| true | false
| $\langle \rangle$ | $\langle V, V \rangle \dots$

Garbage collection

Iris_{ML} $\ell \mapsto_{\text{ML}} \vec{V}$

Integers and pointers

λ_{C} $w \in \text{Val} ::= (n \in \mathbb{Z}) \mid (a \in \text{Addr})$

Manual memory management

Iris_{C} $a \mapsto_{\text{C}} w$

Key Challenge

Can we build a program logic for reasoning about interoperability with an FFI,
while preserving language-local reasoning?

λ_{ML} Semantics

Iris_{ML}

Program Logic

given as black box

λ_C Semantics

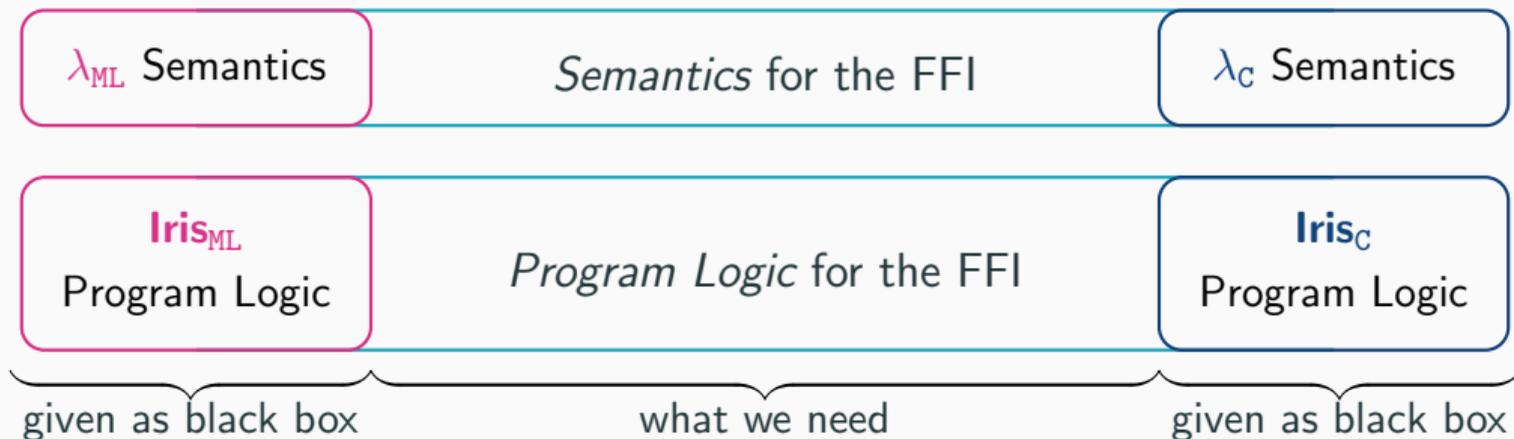
Iris_C

Program Logic

given as black box

Key Challenge

Can we build a program logic for reasoning about interoperability with an FFI,
while preserving language-local reasoning?



Design choice: reuse most of existing semantics/program logics;
do not drop down to a lowest-common denominator (assembly)!

Melocoton:

- Two instantiations of Iris for a ML-like and C-like language with *external calls*
- An *operational semantics* for the OCaml FFI, bridging between the two languages.
- A *separation logic* for the OCaml FFI, bridging between the two language logics.
- A number of interesting *case studies*

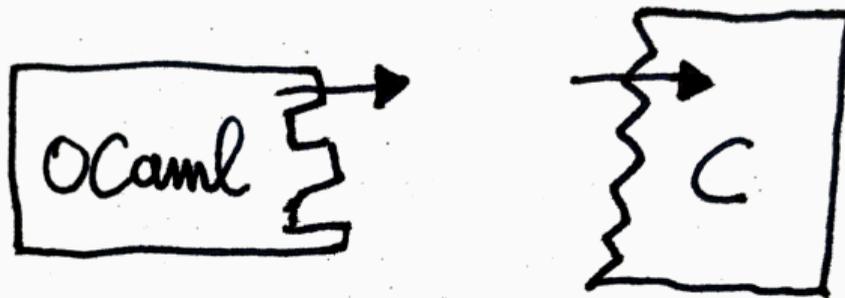
Melocoton:

- Two instantiations of Iris for a ML-like and C-like language with *external calls*
- An *operational semantics* for the OCaml FFI, bridging between the two languages.
- A *separation logic* for the OCaml FFI, bridging between the two language logics.
- A number of interesting *case studies*

Language-locality: Verification of mixed OCaml/C programs can be done *almost entirely* in logics for OCaml and C!

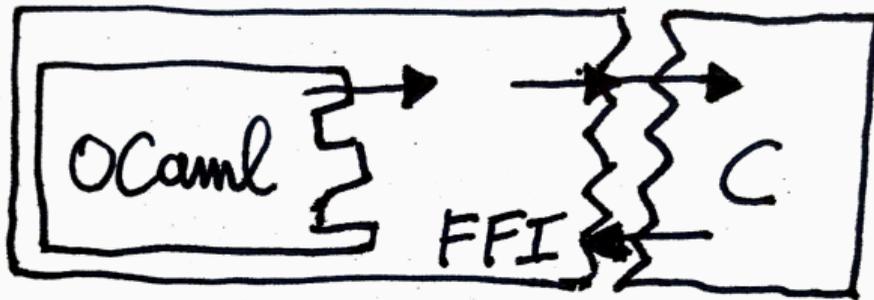
Outline

1. Language-local program logics with external calls



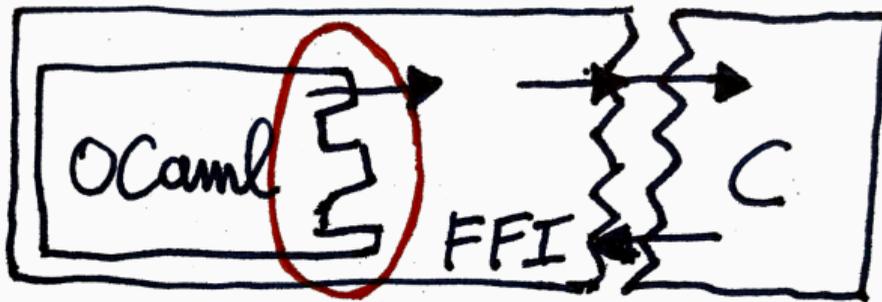
Outline

1. Language-local program logics with external calls
2. Program logic for FFI



Outline

1. Language-local program logics with external calls
2. Program logic for FFI
3. Focus: the language boundary



Example: updating an OCaml reference from C code

OCaml code:

```
let main () =  
  let r = ref 0 in  
  update_ref r; (* TODO call C code and use rand() *)  
  print_int !r
```

C code:

```
int rand(int x) { ... }
```

Example: updating an OCaml reference from C code

OCaml code:

```
external update_ref : int ref -> unit = "caml_update_ref"  
let main () =  
  let r = ref 0 in  
  update_ref r;  
  print_int !r
```

C code:

```
int rand(int x) { ... }
```

Example: updating an OCaml reference from C code

OCaml code:

```
external update_ref : int ref -> unit = "caml_update_ref"  
let main () =  
  let r = ref 0 in  
  update_ref r;  
  print_int !r
```

C code:

```
int rand(int x) { ... }
```

Glue code:

```
value caml_update_ref(value r) {  
  /* TODO */  
  int y = rand(x);  
  /* TODO */  
}
```

Example: updating an OCaml reference from C code

OCaml code:

```
external update_ref : int ref -> unit = "caml_update_ref"  
let main () =  
  let r = ref 0 in  
  update_ref r;  
  print_int !r
```

C code:

```
int rand(int x) { ... }
```

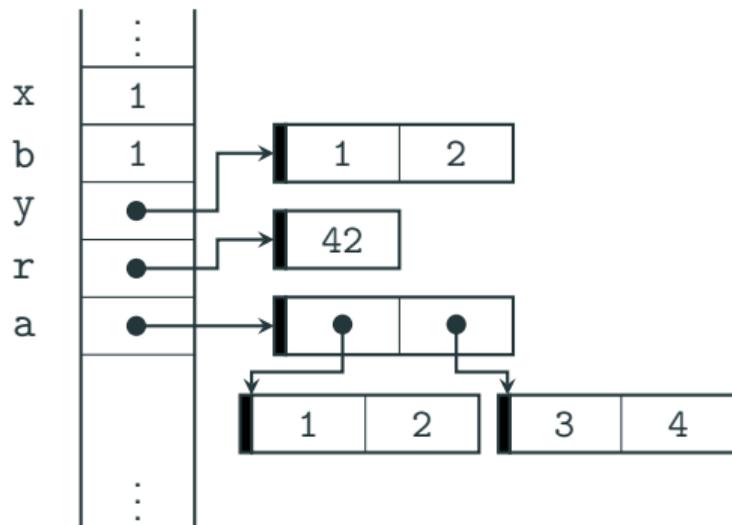
Glue code:

```
value caml_update_ref(value r) {  
  /* TODO */  
  int y = rand(x);  
  /* TODO */  
}
```

The runtime representation of OCaml values

At runtime, an OCaml value is either an integer or a pointer to a block:

```
let x = 1
let b = true
let y = (1, 2)
let r = ref 42
let a = [| (1, 2); (3, 4) |]
```



Glue code has access to this *low-level* representation of OCaml values.

Example: updating an OCaml reference from C code

OCaml code:

```
external update_ref : int ref -> unit = "caml_update_ref"  
let main () =  
  let r = ref 0 in  
  update_ref r;  
  print_int !r
```

C code:

```
int rand(int x) { ... }
```

Glue code:

```
value caml_update_ref(value r) {  
  int x = Int_val(Field(r, 0));  
  int y = rand(x);  
  Store_field(r, 0, Val_int(y));  
  return Val_int(0);  
}
```

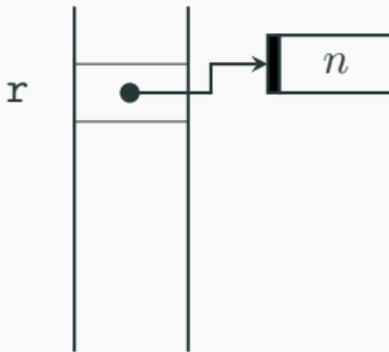
Glue code bridges between OCaml and C values by using powerful **FFI primitives**...

Writing glue code

```
value caml_update_ref(value r) {  
  int x = Int_val(Field(r, 0));    /* read the first field of the input block */  
  int y = rand(x);                 /* get a random integer */  
  Store_field(r, 0, Val_int(y));  /* store the value in the block */  
  return Val_int(0);              /* return () */  
}
```

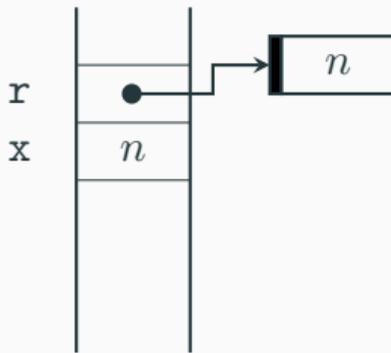
Writing glue code

```
value caml_update_ref(value r) {<--  
  int x = Int_val(Field(r, 0));      /* read the first field of the input block */  
  int y = rand(x);                   /* get a random integer */  
  Store_field(r, 0, Val_int(y));    /* store the value in the block */  
  return Val_int(0);                 /* return () */  
}
```



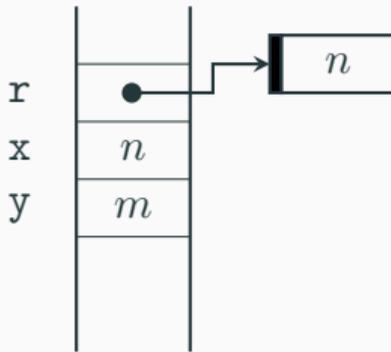
Writing glue code

```
value caml_update_ref(value r) {  
  int x = Int_val(Field(r, 0)); <-- /* read the first field of the input block */  
  int y = rand(x); /* get a random integer */  
  Store_field(r, 0, Val_int(y)); /* store the value in the block */  
  return Val_int(0); /* return () */  
}
```



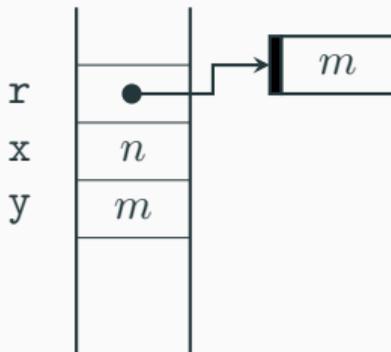
Writing glue code

```
value caml_update_ref(value r) {  
  int x = Int_val(Field(r, 0));      /* read the first field of the input block */  
  int y = rand(x);                   <-- /* get a random integer */  
  Store_field(r, 0, Val_int(y));     /* store the value in the block */  
  return Val_int(0);                 /* return () */  
}
```



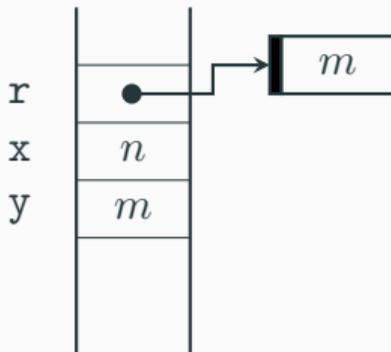
Writing glue code

```
value caml_update_ref(value r) {  
  int x = Int_val(Field(r, 0));      /* read the first field of the input block */  
  int y = rand(x);                   /* get a random integer */  
  Store_field(r, 0, Val_int(y));<-- /* store the value in the block */  
  return Val_int(0);                 /* return () */  
}
```



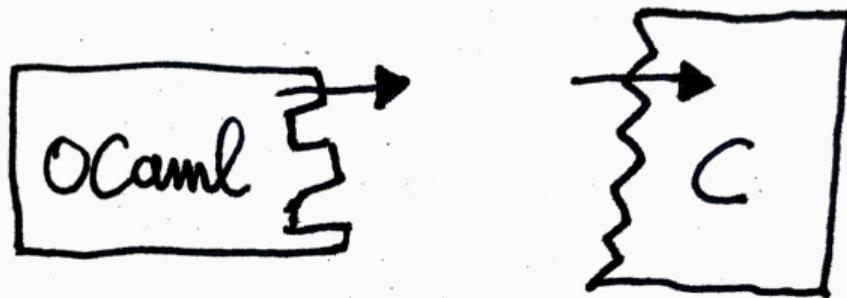
Writing glue code

```
value caml_update_ref(value r) {  
  int x = Int_val(Field(r, 0));      /* read the first field of the input block */  
  int y = rand(x);                  /* get a random integer */  
  Store_field(r, 0, Val_int(y));    /* store the value in the block */  
  return Val_int(0);                <-- /* return () */  
}
```



Outline: Language-local reasoning

1. Language-local program logics with external calls



Language-local reasoning

We reuse:

λ_{ML} Semantics

Iris_{ML}
Program Logic

λ_C Semantics

Iris_C
Program Logic

The one change: a minimal extension allowing **external calls**.

Modeling External Calls

```
external update_ref : int ref -> unit = "caml_update_ref"  
  
let main () :=  
  let r = ref 0 in  
  update_ref r;  
  print_int !r
```

Modeling External Calls

```
external update_ref : int ref -> unit = "caml_update_ref"  
  
let main () :=  
  let r = ref 0 in  
  update_ref r;  
  print_int !r
```

We model external calls as a new syntactic construct (inlining the declaration):

$$e \in Expr ::= \dots \mid \text{call } fn \vec{e}$$

Modeling External Calls

```
external update_ref : int ref -> unit = "caml_update_ref"  
  
let main () :=  
  let r = ref 0 in  
  update_ref r;  
  print_int !r
```

We model external calls as a new syntactic construct (inlining the declaration):

$$e \in Expr ::= \dots \mid \text{call } fn \vec{e}$$

We assign **no semantics** to external calls: they are simply stuck!

Interface Specifications

We still want to *reason* about calls to `caml_update_ref`, as if it had the specification:

$$\forall l n. \{l \mapsto_{\text{ML}} n\} \text{ call caml_update_ref } [l] \{V'. \exists m. V' = \langle \rangle * l \mapsto_{\text{ML}} m\}_{\text{ML}}$$

Interface Specifications

We still want to *reason* about calls to `caml_update_ref`, as if it had the specification:

$$\forall l n. \{l \mapsto_{\text{ML}} n\} \text{call caml_update_ref } [l] \{V'. \exists m. V' = \langle \rangle * l \mapsto_{\text{ML}} m\}_{\text{ML}}$$

To do so, we introduce **interfaces** Ψ , and weakest preconditions $\text{wp } e @ \Psi \{v. Q\}$ that verify programs against them. For example, for `caml_update_ref`, we assume:

$$\forall l n. \langle l \mapsto_{\text{ML}} n \rangle \text{ caml_update_ref } [l] \langle V'. \exists m. V' = \langle \rangle * l \mapsto_{\text{ML}} m \rangle \quad \sqsubseteq \Psi$$

Interface Specifications

We still want to *reason* about calls to `caml_update_ref`, as if it had the specification:

$$\forall l n. \{l \mapsto_{\text{ML}} n\} \text{ call caml_update_ref } [l] \{V'. \exists m. V' = \langle \rangle * l \mapsto_{\text{ML}} m\}_{\text{ML}}$$

To do so, we introduce **interfaces** Ψ , and weakest preconditions $\text{wp } e @ \Psi \{v. Q\}$ that verify programs against them. For example, for `caml_update_ref`, we assume:

$$\forall l n. \langle l \mapsto_{\text{ML}} n \rangle \text{ caml_update_ref } [l] \langle V'. \exists m. V' = \langle \rangle * l \mapsto_{\text{ML}} m \rangle \sqsubseteq \Psi$$



This is an assumption, not a (atomic) Hoare triple



Desugaring To Predicate Transformers

Implement interface triples as a predicate transformer Ψ :

$$\Psi : \underbrace{FnName}_{\text{Name}} \rightarrow \underbrace{\vec{Val}}_{\text{Args}} \rightarrow \underbrace{(Val \rightarrow iProp)}_{\text{Postcondition}} \rightarrow \underbrace{iProp}_{\text{Precondition}}$$

Desugaring To Predicate Transformers

Implement interface triples as a predicate transformer Ψ :

$$\Psi : \underbrace{FnName}_{\text{Name}} \rightarrow \underbrace{\vec{Val}}_{\text{Args}} \rightarrow \underbrace{(Val \rightarrow iProp)}_{\text{Postcondition}} \rightarrow \underbrace{iProp}_{\text{Precondition}}$$

We desugar

$$\forall \ell n. \langle \ell \mapsto_{ML} n \rangle \text{ caml_update_ref } [\ell] \\ \langle V'. \exists m. V' = \langle \rangle * \ell \mapsto_{ML} m \rangle$$

Desugaring To Predicate Transformers

Implement interface triples as a predicate transformer Ψ :

$$\Psi : \underbrace{FnName}_{\text{Name}} \rightarrow \underbrace{\vec{Val}}_{\text{Args}} \rightarrow \underbrace{(Val \rightarrow iProp)}_{\text{Postcondition}} \rightarrow \underbrace{iProp}_{\text{Precondition}}$$

We desugar

$$\forall l n. \langle l \mapsto_{ML} n \rangle \text{caml_update_ref } [l] \\ \langle V'. \exists m. V' = \langle \rangle * l \mapsto_{ML} m \rangle$$

as follows:

$$\Psi_{upd} fn \vec{V} \Phi := \exists l n. l \mapsto_{ML} n * fn = \text{caml_update_ref} * \vec{V} = [l] \\ * (\forall V' m. V' = \langle \rangle * l \mapsto_{ML} m \multimap \Phi(V'))$$

Implementing Interface Triples

$$\Psi : \underbrace{FnName}_{\text{Name}} \rightarrow \underbrace{\vec{Val}}_{\text{Args}} \rightarrow \underbrace{(Val \rightarrow iProp)}_{\text{Postcondition}} \rightarrow \underbrace{iProp}_{\text{Precondition}}$$

Parameterize weakest pre by Ψ (inspired by de Vilhena and Pottier [2021]):

$$\text{wp } e @ \Psi \{ \Phi \} := \begin{cases} \Phi(v) & e = v \\ \forall e', (e \rightarrow e') \Rightarrow \text{wp } e' @ \Psi \{ \Phi \} & e \text{ reducible} \\ \Psi \text{ fn } \vec{V} \underbrace{(\lambda V'. \text{wp } K[V'] @ \Psi \{ \Phi \})}_{\text{Postcondition}} & e = K[\text{call fn } \vec{V}] \end{cases}$$

Implementing Interface Triples

$$\Psi : \underbrace{FnName}_{\text{Name}} \rightarrow \underbrace{\vec{Val}}_{\text{Args}} \rightarrow \underbrace{(Val \rightarrow iProp)}_{\text{Postcondition}} \rightarrow \underbrace{iProp}_{\text{Precondition}}$$

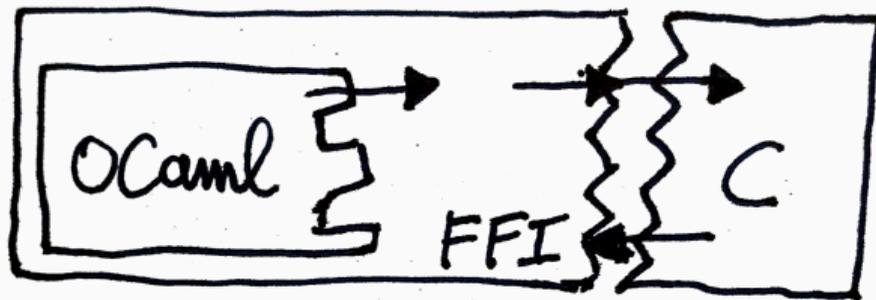
Parameterize weakest pre by Ψ (inspired by de Vilhena and Pottier [2021]):

$$\text{wp } e @ \Psi \{ \Phi \} := \begin{cases} \Phi(v) & e = v \\ \forall e', (e \rightarrow e') \Rightarrow \text{wp } e' @ \Psi \{ \Phi \} & e \text{ reducible} \\ \Psi \text{ fn } \vec{V} \underbrace{(\lambda V'. \text{wp } K[V'] @ \Psi \{ \Phi \})}_{\text{Postcondition}} & e = K[\text{call fn } \vec{V}] \end{cases}$$

Note: In a OCaml-and-C program (after linking), adequacy holds for $\Psi \text{ fn } \vec{V} \Phi := \perp$

Outline: The OCaml FFI

1. Language-local program logics with external calls
2. Glue code and program logic for FFI



External Calls in Glue Code

In glue code we treat operations of the OCaml FFI as **external functions**.

```
value caml_update_ref(value r) {  
  int x = Int_val(Field(r, 0));  
  int y = rand(x);  
  Store_field(r, 0, Val_int(y));  
  return Val_int(0);  
}
```

Glue code is verified using the program logic for C, but additionally **assuming an interface** Ψ_{FFI} for the OCaml FFI primitives, with resources e.g. $\gamma \mapsto_{\text{blk}[t|m]} \vec{v}$.

External Calls in Glue Code

In glue code we treat operations of the OCaml FFI as **external functions**.

```
value caml_update_ref(value r) {  
  int x = Int_val(Field(r, 0));  
  int y = rand(x);  
  Store_field(r, 0, Val_int(y));  
  return Val_int(0);  
}
```

Glue code is verified using the program logic for C, but additionally **assuming an interface** Ψ_{FFI} for the OCaml FFI primitives, with resources e.g. $\gamma \mapsto_{\text{blk}[t|m]} \vec{v}$.

$$\begin{aligned} & \langle \text{GC}(\theta) * \gamma \mapsto_{\text{blk}[0|\text{mut}]} \vec{v} * \gamma \sim_{\text{C}}^{\theta} w * v' \sim_{\text{C}}^{\theta} w' \rangle \\ & \quad \text{Store_field}(w, i, w') \quad \sqsubseteq \Psi_{\text{FFI}} \\ & \langle \text{GC}(\theta) * \gamma \mapsto_{\text{blk}[0|\text{mut}]} \vec{v}[i := v'] \rangle \end{aligned}$$

External Calls in Glue Code

In glue code we treat operations of the OCaml FFI as **external functions**.

```
value caml_update_ref(value r) {  
  int x = Int_val(Field(r, 0));  
  int y = rand(x);  
  Store_field(r, 0, Val_int(y));  
  return Val_int(0);  
}
```

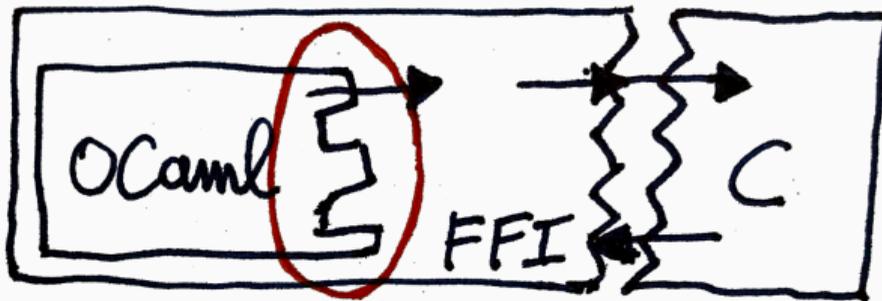
$$\{GC(\theta) * \gamma \mapsto_{\text{blk}[0|\text{mut}]} [n] * \gamma \sim_C^\theta w\}$$
$$\text{call caml_update_ref } [w] @ \Psi_{\text{FFI}}$$
$$\{w'. \exists m. GC(\theta) * \gamma \mapsto_{\text{blk}[0|\text{mut}]} [m] * w' \sim_C^\theta 0\}$$

Glue code is verified using the program logic for C, but additionally **assuming an interface** Ψ_{FFI} for the OCaml FFI primitives, with resources e.g. $\gamma \mapsto_{\text{blk}[t|m]} \vec{v}$.

$$\langle GC(\theta) * \gamma \mapsto_{\text{blk}[0|\text{mut}]} \vec{v} * \gamma \sim_C^\theta w * v' \sim_C^\theta w' \rangle$$
$$\text{Store_field}(w, i, w') \quad \sqsubseteq \Psi_{\text{FFI}}$$
$$\langle GC(\theta) * \gamma \mapsto_{\text{blk}[0|\text{mut}]} \vec{v}[i := v'] \rangle$$

Outline: The OCaml-FFI boundary

1. Language-local program logics with external calls
2. Glue code and program logic for FFI
3. Focus: the OCaml-FFI boundary



View Reconciliation

We assumed an interface for `caml_update_ref` that uses ML points-tos:

$$\forall \ell n. \langle \ell \mapsto_{\text{ML}} n \rangle \text{ caml_update_ref } [\ell] \langle V'. \exists m. V' = \langle \rangle * \ell \mapsto_{\text{ML}} m \rangle$$

Meanwhile, we proved the following specification for `caml_update_ref` using Ψ_{FFI} :

$$\begin{aligned} & \{ \text{GC}(\theta) * \gamma \mapsto_{\text{blk}[0|\text{mut}]} [n] * \gamma \sim_{\text{C}}^{\theta} w \} \\ & \quad \text{call caml_update_ref } [w] @ \Psi_{\text{FFI}} \\ & \{ w'. \exists m. \text{GC}(\theta) * w' \sim_{\text{C}}^{\theta} 0 * \gamma \mapsto_{\text{blk}[0|\text{mut}]} [m] \} \end{aligned}$$

These express two different views about the **same piece of state!**

View Reconciliation: Update Rules

Idea:

- make $l \mapsto_{\text{ML}} \vec{V}$ and $\gamma \mapsto_{\text{blk}[0|\text{mut}]} \vec{v}$ mutually exclusive (for related l and γ)
- have *view reconciliation* rules to switch between the two representations

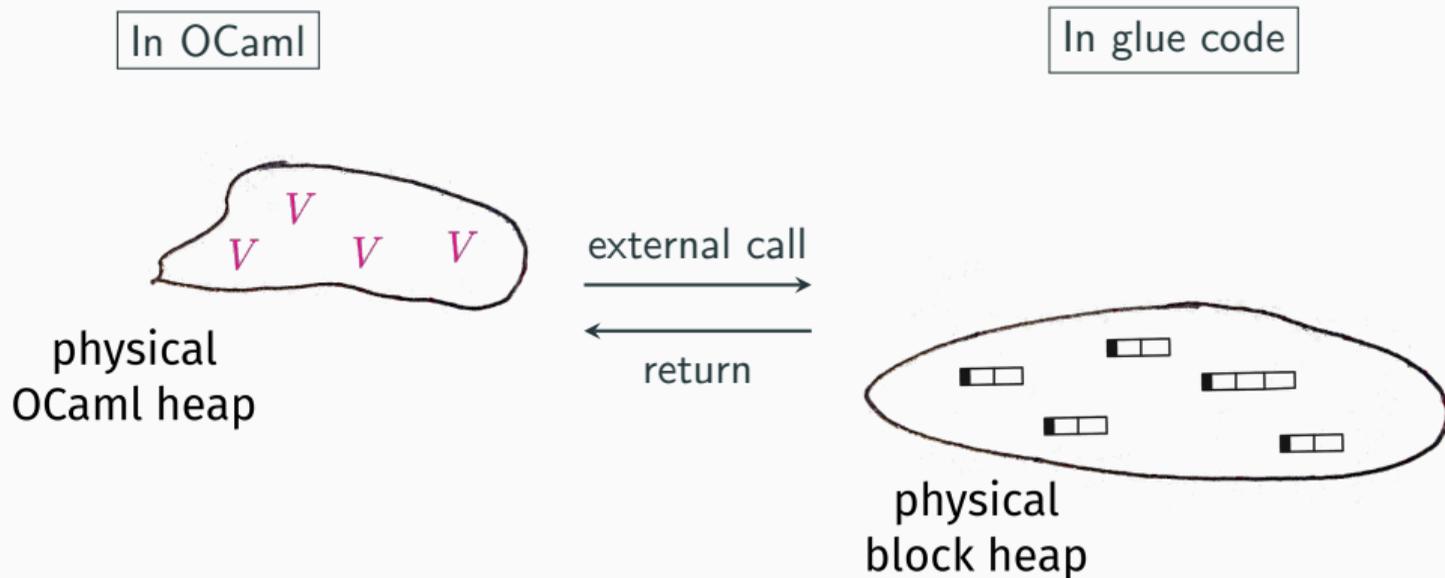
$$\text{GC}(\theta) * l \mapsto_{\text{ML}} \vec{V} \equiv * \exists \vec{v}, \gamma. \text{GC}(\theta) * \gamma \mapsto_{\text{blk}[0|\text{mut}]} \vec{v} * l \sim_{\text{ML}} \gamma * \vec{V} \sim_{\text{ML}} \vec{v} \quad (\text{ML-TO-FFI})$$

$$\text{GC}(\theta) * \gamma \mapsto_{\text{blk}[0|\text{mut}]} \vec{v} * \vec{V} \sim_{\text{ML}} \vec{v} \equiv * \exists l. \text{GC}(\theta) * l \mapsto_{\text{ML}} \vec{V} * l \sim_{\text{ML}} \gamma \quad (\text{FFI-TO-ML})$$

View Reconciliation: Challenge

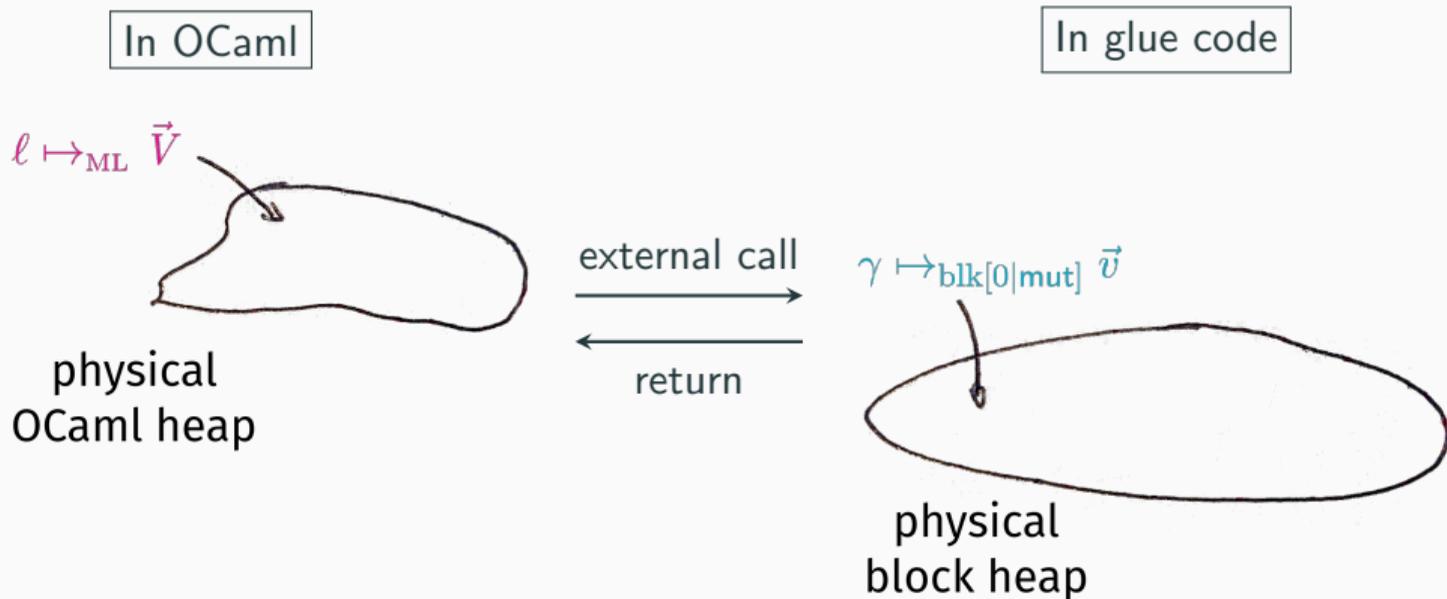
In **operational semantics**, there is *only one simultaneous view* of the OCaml state.

But resources do not reflect that!



View Reconciliation: Challenge (2) and Solution

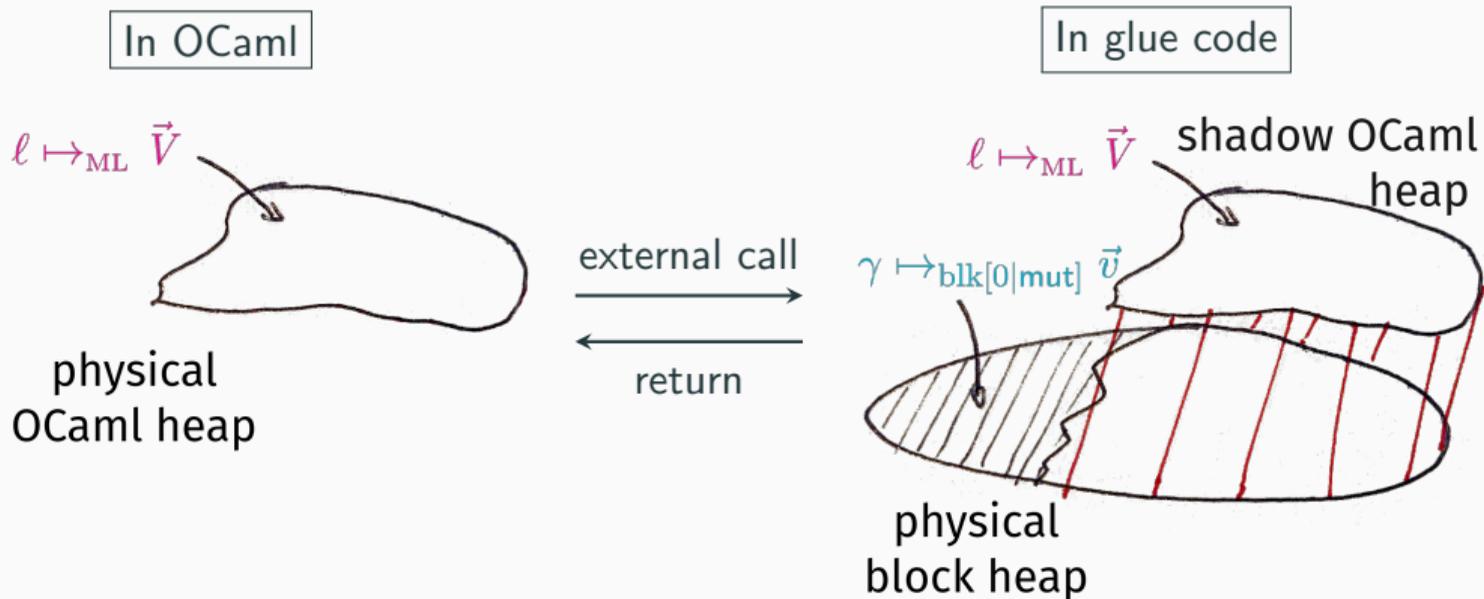
In **ghost state**: what happens to OCaml points-to?



View Reconciliation: Challenge (2) and Solution

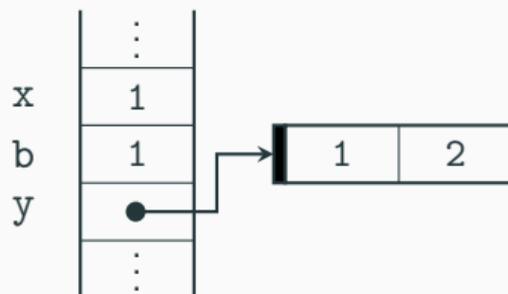
In **ghost state**: what happens to OCaml points-to?

Solution: track *both* views of the state in ghost state



Changing The Representation: Making Difficult Choices

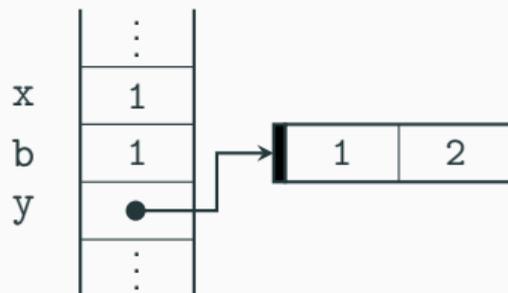
```
let x = ?  
let b = ?  
let y = (?, ?)
```



Quiz Time: What are the OCaml values of x, b, and y?

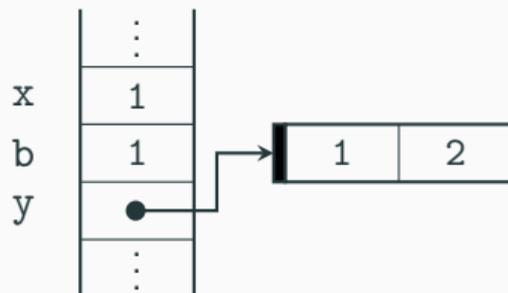
Changing The Representation: Making Difficult Choices

```
let x = 1
let b = true
let y = (1, 2)
```



Changing The Representation: Making Difficult Choices

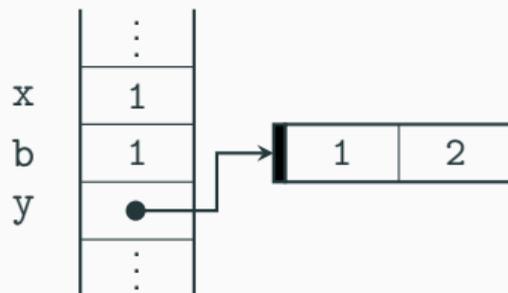
```
let x = 1
let b = true
let y = (1, 2)
```



High-level representation is **not unique!**

Changing The Representation: Making Difficult Choices

```
let x = 1
let b = true
let y = (1, 2)
```



High-level representation is **not unique!**

How does Operational Semantics choose the right value when switching to ML values?

Angelic Non-Determinism And The Weakest Pre

We use angelic nondeterminism, based on multi-relations (see DimSum, CCR)!

Angelic Non-Determinism And The Weakest Pre

We use angelic nondeterminism, based on multi-relations (see DimSum, CCR)!

$\text{wp } e \{ \Phi \} : \hat{=} \dots \vee (e \text{ reducible} * \forall e'. e \rightarrow e' \multimap \text{wp } e' \{ \Phi \})$ usual Iris

$\text{wp } e \{ \Phi \} : \hat{=} \dots \vee (\exists X. e \twoheadrightarrow X * \forall e'. e' \in X \multimap \text{wp } e' \{ \Phi \})$ multi-relations

Regular C and ML, not having angelic non-determinism, retain usual SOS

Angelic Non-Determinism And The Weakest Pre

We use angelic nondeterminism, based on multi-relations (see DimSum, CCR)!

$\text{wp } e \{ \Phi \} : \hat{=} \dots \vee (e \text{ reducible} * \forall e'. e \rightarrow e' \multimap \text{wp } e' \{ \Phi \})$ usual Iris

$\text{wp } e \{ \Phi \} : \hat{=} \dots \vee (\exists X. e \rightarrow X * \forall e'. e' \in X \multimap \text{wp } e' \{ \Phi \})$ multi-relations

Regular C and ML, not having angelic non-determinism, retain usual SOS

For adequacy, **existential** needs to be extracted \Rightarrow *transfinite Iris*

Contribution: An Iris for toy C+ML+FFI, emphasizing **language-local reasoning**.

Contribution: An Iris for toy C+ML+FFI, emphasizing **language-local reasoning**.

We give a **general recipe** for merging two languages:

1. Abstract over “the other side” using **interfaces and external calls**
2. Formalize the **semantics of the FFI** (memory model and primitives)
3. Bridge between memory models using **view reconciliation**

Contribution: An Iris for toy C+ML+FFI, emphasizing **language-local reasoning**.

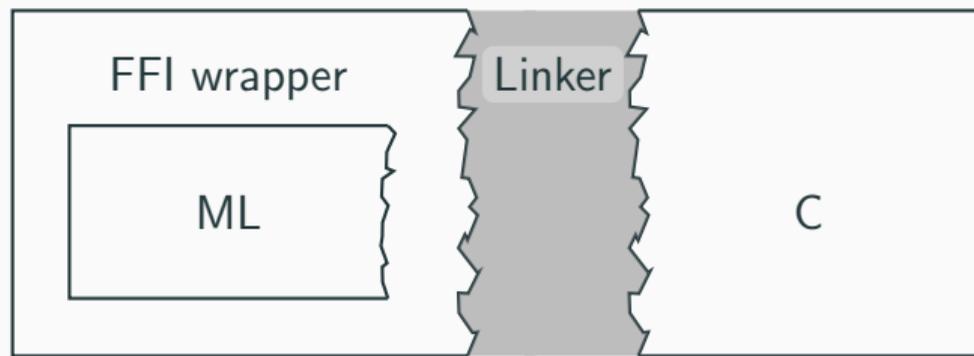
We give a **general recipe** for merging two languages:

1. Abstract over “the other side” using **interfaces and external calls**
2. Formalize the **semantics of the FFI** (memory model and primitives)
3. Bridge between memory models using **view reconciliation**

More in the paper:

- **more detailed FFI**: callbacks, custom blocks, GC interaction
- **logical relation** for semantic typing of external functions

bonus slides



The FFI wrapper

- Convert ML values to block-level
- Provide FFI: a C calling convention for ML

The Linker

- Link programs using the same calling convention
- Resolve external calls