

**Abstract**

We present the Flyweight Object-Oriented (FOO) calculus for the modeling of object-oriented languages. FOO is a simple, minimal class-based calculus, modeling only essential computational aspects and emphasizing larger-scale features (e.g., inheritance and generics). FOO is motivated by the observation that recent language design work focuses on elements not well-captured either by traditional object calculi or by language-specific modeling efforts, such as Featherweight Java. FOO integrates seamlessly both nominal and structural subtyping ideas, leveraging the latter to eliminate the need for modeling object fields and constructors. Comparing to recent formalization efforts in the literature, FOO is more compact, yet versatile enough to be usable in multiple settings modeling Java, C#, or Scala extensions.

**Categories and Subject Descriptors** D.3.1 [Programming Languages]: Formal Definitions and Theory—Semantics; D.3.3 [Programming Languages]: Language Constructs and Features—Inheritance, Polymorphism

**General Terms** Languages

**Keywords** object-oriented programming, formal semantics, type system, structural types, nominal types

1. **Introduction and Motivation**

Modeling programming languages via concise formalisms has a long history. Languages are complex artifacts whose informal specifications often weigh in at many hundreds of pages, rivaling in length and topping in complexity all but few human-produced texts. Yet, a diminutive formalism, often under a couple of pages in length, can capture key insights on the language’s design. By establishing properties of the formalism, researchers can reason about the correctness of important language elements. Formal modeling has often helped identify design errors, has been used to settle algorithmic questions (e.g., decidability of core typing), and has aided the community’s understanding of language features. Researchers routinely leverage a formalism in order to propose language extensions in their purest form, and to quickly test new ideas and their interaction with a core model.

The literature landscape of language formalisms contains several core object-oriented calculi [1, 4, 6] as well as more targeted language modeling efforts [10, 15, 21]. The core calculi are typically object-based and attempt to capture elements of language design that have gradually become less studied in the past 15 years. Language-specific modeling efforts are often non-minimal and carry significant baggage that slows down further formal development. It is telling, for instance, that the reference model for Scala [21] contains elements such as dependent types, and has an undecidable type system.

In this work, we present the Flyweight OO (FOO) calculus for modeling object-oriented languages. FOO is inspired by Featherweight Java (FJ) [10]: it is also a class-based formalism and emphasizes inheritance. Indeed, the motivation for FOO stems from our own past language modeling efforts [2, 7–9] which were based on FJ. We found that the language features that are most pertinent to our language extensions had little to do with key elements of Featherweight Java, such as casts, fields, or constructors. (The modeling of casts, including the hallmark “stupid cast” problem of FJ, has, to our knowledge, rarely arisen in the literature subsequent to the original Featherweight Java publications.) The same observation holds regarding the work of others. Recent language models in the literature focus on high-level features, and not on the structure of expressions or low-level computation in general. Such high-level features include mixins and traits [3, 17, 18], polymorphism and gradual typing [12, 22], modules [11, 14], rich type constraints [23], interactions between different kinds of subtyping [16], and domain-specific extensions [5, 19].

Therefore, we believe there is a need for a minimal calculus that abstracts away low-level computation to its essence (much like a foundational calculus) yet fully supports high-level typing elements (e.g., nominal and structural subtyping, classes, generics). The FOO calculus attempts to strike such a balance. FOO models nominal class-based inheritance, as well as anonymous classes. The latter enable emulations of fields and constructors as well as of (breadth) structural subtyping. FOO tries to be language agnostic, however FOO programs directly map to Scala programs (modulo simple, local syntax transformations). Furthermore, FOO has a straightforward runtime semantics and a simple type-system that makes its algorithmic properties (sound and complete subtyping algorithm) elementary, without a need for external assumptions (e.g., hierarchy cycle checking). Informal inspection of the recent literature suggests that FOO could be leveraged for a large number of language modeling efforts that include a formalism, in the context of either Java, C#, or Scala, and would yield consistent conciseness benefits.

In this short paper, we present our language design informally, via examples (Section 2) and detail a formalism that captures the essence of our approach (Section 3).
2. **F00, Informally**

Before presenting our formalism, we illustrate its syntactic features in a more palatable form, with the help of some syntax sugar. All examples are valid Scala code, but map straightforwardly to concepts in our calculus.

F00 is a class-based calculus. Type expressions are hybrid, consisting of a nominal part and an anonymous set of method signatures. This is quite similar to a feature of the Scala language [20], allowing on-the-fly extension of an existing named class with an anonymous part.

**Example 1 (On-the-fly and anonymous classes).** An existing Scala class, `Employee`, can be extended with extra functionality by adding method `extra`.

```scala
(new Employee { def extra() = println("add-on") }).extra;
```

In F00 all types are of the above hybrid form. The anonymous part of a type can be empty, resulting in purely nominal typing, while the nominal part of the type can be `Object` (the root of all class hierarchies) allowing subtyping to be determined structurally, by the contents of anonymous method sets.

**Example 2 (Structural types for anonymous classes).** The following example illustrates subtyping based on the structure of anonymous classes.

```scala
def fun1(e : { def extra() }) = e.extra
...
fun1(new Object { def extra() = println("subtyping") })
```

Indeed, inheritance itself (i.e., the extension of a class by another) can be viewed as merely the naming of a hybrid type, consisting of a nominal part (the named superclass) and the anonymous extension part.

**Example 3 (Inheritance defined in terms of anonymous classes).** We see below the typical syntax for declaring a subclass—the combination of superclass and subclass body can be viewed as just a hybrid type. This is precisely the view that our formalism encodes.

```scala
class EnhancedEmployee extends Employee { def extra() = println("more") }
```

Anonymous classes allow us to simulate several syntactic conveniences without integrating them in the core language. F00 relies on such encodings, omitting explicit support for fields, constructors, or multiple method parameters.

**Example 4 (Encoding of fields).** We see in this example the construction of `Employee` objects with different field values. The simulation of (final) fields with methods is faithful.

```scala
abstract class Employee {
  def id(): Integer
  def name(): String
  def salary(): Integer
};

def newEmployee ( cid : Integer, cname : String, csalary : Integer ) : Employee
= new Employee { def id() = cid
                def name() = cname
                def salary() = csalary }
```

Similarly to the encoding of fields, we can encode multiple function parameters. This is a conventional encoding, listed here for completeness.

**Example 5 (Encoding of multiple formal parameters).** Let us assume that we want to encode a function adding two integers as follows:

```scala
class Add {
  def apply(x : Integer, y : Integer) = x + y
}
(new Add).apply(5, 10)
```

The above snippet can be encoded by capturing the first parameter inside an instance of `Add`, using the technique of the previous example, and by invoking method `apply` which now takes only the second argument.

```scala
abstract class Add {
  def x(): Integer
  def apply(y : Integer) = x() + y
}
(new Add{ def x() = 5 }).apply(10)
```

3. **Language Description**

In this section, we introduce the syntax of the F00 calculus and present its formal semantics.

Our formalism captures the salient features of class-based OO languages with nominal and structural subtyping elements, but eliminates unnecessary complexity: we do not model redundant language features such as fields, constructors, multiple parameters as they can be encoded in our calculus.

3.1 **Syntax**

The syntax of F00 is presented below:

| Member type | $\Psi$ ::= $m : N \rightarrow N$ |
| Hybrid Type | $N ::= C \& \Psi$ |
| Member | $M ::= m(x) e$ |
| Program Value | $v ::= new N \{F\} \mid x$ |
| Expression | $e ::= v \mid v.m(e)$ |
| Top-level classes | $P ::= class C ::= N \{F\}$ |

We adopt many of the notational conventions of well-known formal calculi such as FJ [10]: $C$ denotes constant class names, $N$ denotes object types, $M$ denotes method names and $x$ denotes argument names. Classes and methods are explicitly typed in our calculus.

There are two kinds of type annotations: member types $\Psi$ and hybrid types $N$. A method signature maps method identifier $m$ to a method type $N \rightarrow N'$, indicating methods taking a single argument of type $N$ and returning a value of type $N'$. A class type $N$ consists of two components expressing both nominal and structural types: the first component is the parent class $C$, which is extended by the method signatures defined in $\Psi$. The two components are separated by the ampersand ($\&$) symbol.

Type annotations and definitions of methods are placed separately: a class is defined by listing its hybrid type (i.e., its superclass, as well as a list $\Psi$ of extra method signatures) and then the definitions, $F$, of these methods (i.e., their bodies, without type annotations) are listed. Both type annotations and definitions are associated with a unique method identifier. Thus, annotations and definitions having the same identifier are associated.

Similarly to FJ we use the bar notation to express an ordered sequence of symbols. For sequences, we use $\sigma$ to denote an empty one, $[\sigma]$ for the sequence holding the single element $\sigma$, and the comma operator to add a new element to the front of an existing sequence.

```scala
$\Psi ::= m : N \rightarrow N$ |
$N ::= C \& \Psi$ |
$M ::= m(x) e$ |
$v ::= new N \{F\} \mid x$ |
$e ::= v \mid v.m(e)$ |
$P ::= class C ::= N \{F\}$ |
```
sequence. We also treat ordered sequences as functions when applicable. Therefore, \( \text{dom}(\overline{v}) \) performs a look up on the elements of \( \overline{v} \) and returns the method signature associated with \( m \). Similarly, \( \text{dom}(\overline{v}) \) returns the set of method names in \( \overline{v} \). Furthermore, we implicitly overload all predicates applicable to members of a sequence to apply to the sequence itself, in the usual way.

Class members \( H \) consist only of method definitions of the form \( m(x) \), where \( m \) is a method name and \( x \) is the formal parameter of \( m \), that is bound for the scope of its body \( e \).

A value \( v \) can either be a variable \( x \) or an object instantiation expression \( \text{new } N \{ \overline{e} \} \), where \( N \) is the class to be instantiated and \( \overline{e} \) denotes the set of additional methods extending \( N \). Expressions can either be values or method invocations of the form \( v.m(e) \), where \( v \) is the receiver of method \( m \) and \( e \) an expression. Without loss of generality, we restrict method receivers to values in order to simplify our formal semantics.

Finally, \( P \) is a set of class declarations mapping class names to their definitions. The special class \( \text{Object} \) is considered to be the root of the nominal class hierarchy and contains no members.

### 3.2 Operational Semantics

Figure 1 defines the formal semantics of FOO. Our operational semantics is defined in terms of the reduction relation \( \rightarrow_p \), which transforms expressions from \( e \) to \( e' \), given the entire program \( P \) that is passed as an implicit parameter. The reduction relation is defined by two rules, a congruence rule \( R-C \) and a reduction rule \( R-I \). The former rule is standard and applies the reduction relation recursively until a reducible expression of the form \( v.m(e') \) is reached. The purpose of the latter rule \( R-I \) is to reduce method invocation expressions to the corresponding method bodies. The first premise of \( R-I \) requires that the argument \( v' \) passed to method \( m \) is an object instantiation expression. The second premise employs function \( \text{mbody} \) to lookup the definition of method \( m \) using the receiver object \( N \{ \overline{r} \} \) and the entire program \( P \).

### 3.3 Static Semantics

The typing rules of FOO are presented in Figure 1. There are two environments used in typing judgments, the typing context \( \Gamma \) that maps method variables to hybrid types and the program type schema \( H \), which maps class names to hybrid types for the given program \( P \). In contrast with the context \( \Gamma \), the type schema \( H \) is constant during type checking and is provided as input. The type checker also assumes that the types residing in \( H \) are well-formed, that is, their definitions satisfy the judgment established by rule \( T-C \). The type schema \( H \) is defined as follows:

\[
H \equiv \{ C : N \mid \text{class } C = N \{ \overline{r} \} \}
\]

The typing relation is defined as \( \Gamma \vdash e : N \). The type schema \( H \) is placed as a subscript of the entails symbol to indicate that it is a constant context, as opposed to \( \Gamma \), which may expand. Given the typing environments, the typing relation assigns a unique type \( N \) to expression \( e \). The static semantics rules can be divided into five groups:

**Hierarchy computation.** The hierarchy computation relation is \( \vdash_{\overline{r}} N \Rightarrow \overline{v} \) and is realized by rules \( H-O \) and \( H-C \). The first component of the hierarchy is a sequence of method type signatures; it computes the intermediate “interface” of a hybrid type. It is crucially used in rule \( T-I \) to guarantee type-safe method lookup. The second component of the hierarchy records the chain of superclasses (full hybrid types) towards the \( \text{Object} \) root class. This chain is used in the subtyping rule, \( S-N \), to distinguish between two different hybrid types, even if these happen to have the same method signatures. Given a class type \( N \) and an implicit type schema \( H \), the hierarchy rules compute the full set of method signatures in a class hierarchy \( \overline{v} \), as well as the hierarchy \( \overline{v} \) itself.

This relation guarantees the following invariants: (a) there exist no cyclic definitions in the class hierarchy (no separate cycle checking is required or implied), (b) the returned \( \overline{v} \) contains all methods of the hierarchy, and (c) methods having the same name also have the same type signature. Most importantly, the hierarchy computation rule simplifies subtyping, which involves both structural and nominal types.

**Well-formedness.** Rules \( W-O \) and \( W-C \) are the class type well-formedness rules, which ensure a non-cyclic hierarchy. Rule \( W-M \) lifts well-formed class types to well-formed method types. We use abbreviations for elements of sequences: for instance, \( \vdash_{\overline{r}} \overline{v} \) is an abbreviation for \( \forall m : N \rightarrow N' \in \overline{v} \).
Subtyping. The subtyping relation (rule S-N) is defined as $\bowtie \bowtie$ holds when $\xi$ is a subtype of $\xi'$. Our subtyping relation does not permit depth structural subtyping; we only allow width subtyping for structural types.

Expression typing. The expression typing relation is $\Gamma \vdash_{\bowtie} \xi : \eta$, where $\Gamma$ is the method variable typing context, $\xi$ is the type schema. There are three rules for typing expressions: the method invocation rule, $T-I$, the object instantiation rule, $T-N$, and the variable typing rule, $T-V$. Notice that the ability to define new classes within methods has the following consequences: (a) method signatures $\xi$ and method definitions $\eta$ must be validated at the instantiation point, (b) the outer $\xi$ variable binding is replaced by the new type binding of $\xi$ (i.e., $\Gamma \setminus \{\xi\}$) and (c) methods may capture variables from the enclosing context, since a method can see the formals of all enclosing methods.

Declaration typing. There are two kinds of declarations. Method declarations are of the form $\Gamma \vdash_{\bowtie} \xi \in \mathcal{M}$ and are validated by $\Gamma \vdash_{\bowtie} \xi \in \mathcal{M}$ (rule T-M), while class declarations are validated by rule T-C.

4. Conclusions

We presented a minimal calculus for class-based OO languages that permits hybrid representation of nominal and structural types. There have been several related calculi either combining structural and nominal systems or providing core formal bases for language design work. However, such calculi either lack some key features or add significant complexity with features such as external dispatch, traits, or dependent types. An interesting quick comparison is with the recent Tinygrace calculus [13], which also aims for minimality while capturing different core features. In contrast to Tinygrace, FOO does not model casts and has classes introduce new types, thus does not model casts and has classes introduce new types.

We seek input on avenues for improving FOO or validating its suitability for language modeling. In this direction, we are currently completing a Coq proof of soundness for FOO.

Acknowledgments. We thank the anonymous reviewers for their constructive comments. Our work was funded by the Greek Secretariat for Research and Technology under the “MorphPL” Excellence (Aristeia) award.

References