Nominal Wyvern: 
Employing Semantic Separation for Usability 

Yu Xiang Zhu
CMU-CS-19-105
April 2019

School of Computer Science 
Carnegie Mellon University 
Pittsburgh, PA 15213

Thesis Committee: 
Jonathan Aldrich, Chair 
Heather Miller 
Alex Potanin, Victoria University of Wellington, NZ

Submitted in partial fulfillment of the requirements for the degree of Master of Science.

Copyright © 2019 Yu Xiang Zhu
Keywords: Nominality, Wyvern, Dependent Object Types, Subtype Decidability
Abstract

This thesis presents Nominal Wyvern, a nominal type system that emphasizes semantic separation for better usability. Nominal Wyvern is based on the dependent object types (DOT) calculus, which provides greater expressivity than traditional object-oriented languages by incorporating concepts from functional languages. Although DOT is generally perceived to be nominal due to its path-dependent types, it is still a mostly structural system and relies on the free construction of types. This can present usability issues in a subtyping-based system where the semantics of a type are as important as its syntactic structure. Nominal Wyvern overcomes this problem by semantically separating structural type/subtype definitions from ad hoc type refinements and type bound declarations. In doing so, Nominal Wyvern is also able to overcome the subtype undecidability problem of DOT by adopting a semantics-based separation between types responsible for recursive subtype definitions and types that represent concrete data. The result is a more intuitive type system that achieves nominality and decidability while maintaining the expressiveness of f-bounded polymorphism that is used in practice.
Acknowledgments

Foo Bar
# Contents

1 Introduction 1

2 Background and Motivation 5
   2.1 DOT and Path-Dependent Types ................................. 5
   2.2 Subtyping and Undecidability .................................. 7
      2.2.1 Getting Back Decidability .................................. 10
      2.2.2 Material-Shape Separation .................................. 11
      2.2.3 Material-Shape Separation for DOT .......................... 12
   2.3 Nominality .......................................................... 13
      2.3.1 Typing Nominality ............................................. 14
      2.3.2 Subtyping Nominality ........................................ 14

3 Nominal Wyvern Design 17
   3.1 A Store of Named Structures .................................... 17
   3.2 Adding Generics .................................................. 19
   3.3 A Binary Typing Approach ....................................... 21
      3.3.1 Top-Level Well-Formedness .................................. 24
      3.3.2 Nominal Subtyping Graph .................................... 24
   3.4 Material-Shape Separation ...................................... 26
   3.5 Term Typing ........................................................ 29
      3.5.1 Typing Decidability .......................................... 31

4 Subtyping Decidability 37
   4.1 Subtyping Judgments .............................................. 37
   4.2 Decidability ........................................................ 41
### List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Type parameter vs type member</td>
<td>8</td>
</tr>
<tr>
<td>2.2</td>
<td>Variance of type parameter (in Scala’s type parameter syntax)</td>
<td>8</td>
</tr>
<tr>
<td>2.3</td>
<td>F-bounded polymorphism (in Scala)</td>
<td>9</td>
</tr>
<tr>
<td>2.4</td>
<td>Infinite derivation example (from Greenman et al. [2014])</td>
<td>9</td>
</tr>
<tr>
<td>2.5</td>
<td>Diverging context example</td>
<td>10</td>
</tr>
<tr>
<td>2.6</td>
<td>Cloneable list causing expansive inheritance</td>
<td>11</td>
</tr>
<tr>
<td>2.7</td>
<td>Example translation from type parameter to type member</td>
<td>12</td>
</tr>
<tr>
<td>3.1</td>
<td>Nominal Wyvern Grammar</td>
<td>22</td>
</tr>
<tr>
<td>3.2</td>
<td>Nominal Wyvern Top-Level Decl Well-Formedness</td>
<td>25</td>
</tr>
<tr>
<td>3.3</td>
<td>Nominal Wyvern Term Typing</td>
<td>30</td>
</tr>
<tr>
<td>4.1</td>
<td>Nominal Wyvern Subtyping</td>
<td>38</td>
</tr>
<tr>
<td>4.2</td>
<td>Nominal Wyvern Subtyping (continued)</td>
<td>39</td>
</tr>
<tr>
<td>4.3</td>
<td>Ways to get recursive subtyping judgments</td>
<td>41</td>
</tr>
</tbody>
</table>
Listings

2.1 Path-dependent type ................................................. 5
2.2 Bounds on type members .......................................... 6
2.3 Type refinement ...................................................... 7
3.1 A nominal typing system on structural interfaces (Part 1) ........ 18
3.2 A nominal typing system on structural interfaces (Part 2) ....... 18
3.3 Adding generics to Nominal Wyvern ................................. 20
4.1 3 ways nominal subtyping can fail ................................. 37
5.1 Basic path-dependent types in Nominal Wyvern .................. 53
5.2 Expressing SchoolLibrary with a refinement ...................... 55
5.3 F-Movable example in Nominal Wyvern ............................ 56
5.4 More movable types ................................................. 57
5.5 Family polymorphism: general nodes and edges .................. 58
5.6 Family polymorphism with OnOffGraph ........................... 59
5.7 Existential types in Nominal Wyvern ............................... 62
5.8 NatSet in Nominal Wyvern ......................................... 62
5.9 Representing functors as functions ................................. 63
5.10 Pure OOP in Nominal Wyvern ..................................... 65
5.11 Mixing OOP and FP in Nominal Wyvern ........................... 67
Chapter 1

Introduction

The power of abstraction is recognized as one of the greatest ideas in computer science [Vleck 2008]. It allows humans to safely segregate pieces of large systems so they can be reasoned about individually. For programming languages, abstraction provides flexibility by allowing for code reuse and modularization. However, while functions are the agreed upon way for abstracting away specific values to focus on the procedure itself, the exact way to achieve polymorphism, the abstraction of types (aka. Generic programming), is less agreed upon. On the one end is pure object-oriented (OO) languages, where polymorphism exists in the form of subtyping. Types are generally monomorphic except for the ability for specific types to act like general ones. On the other end is functional languages, where parametric polymorphism allows type variables to stand in for real types, and new types can be constructed from abstract type parameters. The differences also correspond to the encouraged ways of achieving data abstraction in each paradigm. OO languages naturally use objects to encapsulate state and procedures for interaction (aka. accepts messages). Functional languages utilize modules to encapsulate abstract type members and operations on them (i.e. abstract data types). Each paradigm has its own benefits: objects enjoy the flexibility of being dynamically generated and treated as first-class values, while modules allow for more efficient implementations [Cook 2009] and more flexible type abstraction via parameterization.

Programming language researchers have long wanted to get the best of both worlds. Scala is one of the languages that sit in between the two paradigms by supporting type members in objects and bounded parametric polymorphism (or “bounded quantification”). In contrast to plain parametric polymorphism, the bounded version allows for an OO-style
restriction that restricts the insantiating type to be the subtype of some type. The benefit is one can now additionally require that the instantiating type to have certain features by giving a bound on how general it can be. For example, a hash table type can now easily require its key type to be hashable by specifying the instantiating type to be a subtype of the general Hashable type that has a `hash` method.

However, the merging of the two paradigms is also the merging of nominal and structural type systems. Traditionally, OO languages are nominal in that the names of the types are significant themselves since they are what the types are identified with in the subtype hierarchy, the backbone of OO abstractions. Two types can have the same structure internally, but having different names means they are different. In contrast, functional languages lean on the structural side, where the structure of a type is what defines the type, and the name is a mere convenience in referring to it. This difference is closely related to how abstraction is achieved in each paradigm, so a careful merging of nominality and structurality is also required to get the best of both worlds in terms of usability. The foundational type system for Scala, dependent object types, merges the two paradigms by preserving nominality only for the subtype topology. Path-dependent types are referred to by name since they, combined with type bound declarations, make up the subtype topology. However, the rest of the type system is structural since it allows the construction of new types freely. This meant new structural types can be created without ever being given any names. For large systems where subtype relations play an important role, this may lead to accidental subtyping, or implicit subtyping relations that are unclear to code readers. Heavily relying on structural types may also lead to types whose purposes are less clear and whose problems cannot be easily communicated to the user by compiler and programming tools.

Additionally, while the merging of the paradigms seem to have provided additional expressiveness, it comes at the cost of decidability. While subtyping and parametric polymorphism by themselves are well-studied and easily decidable, the combination of them is not. It is proved very early on that bounded quantification is fundamentally an undecidable problem [1992]. This is due to the ability to define types that subtype a type parameterized by themselves. The languages that conservatively build on top of it also suffer from the same problem. Scala, for instance, is known to have an undecidable subtyping problem [2014]. Scala also is not the only language that has tried to merge the two concepts. Java and C++ are both object-oriented languages that have parametric polymorphism added (Java Generics, C++ templates), and these systems are also shown to
be problematic as well [Grigore, 2017]. Having an undecidable component in a language can pose real problems to the user. Programmers unfamiliar with the intricacies of their language’s type system may end up unable to compile their “well-written” program. The compiler’s failure also meant no guidance on what the programmer should do to fix the problem.

This thesis presents Nominal Wyvern, a new core type system for the Wyvern programming language [Nistor et al., 2013] based on the dependent object types calculus of Scala [Odersky et al., 2003], that aims to solve the presented problems that arise from the merging of the two paradigms with a clearer semantic separation. In particular, Nominal Wyvern’s nominal typing system separates the definition of a structural type from the declaration of a type member’s bound so that both typing and subtyping are explicit and nominal. It also separates the structural types that are responsible for recursive definitions from the types that represent concrete data (following an adaptation of Greenman et al. [2014]’s material-shape separation) to curb subtype undecidability. The goal is to produce a type system that avoids the usability pitfalls mentioned above in hope of making it easier to write safe and correct code.

The main contribution of this thesis is the design of a more usable core type system for Wyvern that achieves its usability goals via nominality and decidability. More specifically, this thesis presents 1) the design of a more thoroughly nominal system based on DOT, and 2) an adaptation of material-shape separation to this DOT-based type system.

Chapter discusses in detail the earlier research in DOT, subtype decidability, and nominality that motivated this thesis. Chapter presents the grammar and typing rules of Nominal Wyvern with an example, and explains how this design facilitates usability and decidability. Chapter delves into the subtyping rules and gives a proof of why subtyping is decidable when material-shape separation is observed. Chapter shows the expressiveness of Nominal Wyvern by presenting several examples of common programming patterns in Nominal Wyvern syntax. Chapter concludes the thesis and talks about possible directions for future work.
Chapter 2

Background and Motivation

2.1 DOT and Path-Dependent Types

The dependent object types (DOT) calculus [Amin et al., 2012] was developed as a type-theoretic foundation for Scala. The key distinguishing feature of the DOT calculus is objects with type members. Traditionally, objects included only fields and methods. Modules in ML systems supported type members, but modules do not enjoy the benefits of being first-class values like objects do in object-oriented programming languages. Unifying concepts from objects and modules allows DOT to model types that are dependent on objects. Listing 2.1 presents such an example.

```scala
class Library { l =>
  type Book
  def borrow(title: String) : l.Book
  def renew(item: l.Book) : Unit = {}
}

val hunt : Library = ...
val sorrells : Library = ...
val csapp = sorrells.borrow("CSAPP")
sorrells.renew(csapp) // OK
hunt.renew(csapp)    // type mismatch
    // found     : sorrells.Book
    // required  : hunt.Book
```

Listing 2.1: Path-dependent type
The `Library` class defines an abstract type member `Book`. This `Book` type is then used in the definition of the trait by the `borrow` and `renew` methods (the variable `l` is the self variable with which other members of the trait can be accessed). This means if you have a value, `sorrells`, of type `Library`, calling `borrow` on it would return a value of type `sorrells.Book`. This is a path-dependent type because the type is not self-contained. It depends on another variable in the environment. Since the exact type of `sorrells.Book` is unknown (abstracted away in `Library`), renewing it from any other `Library` would not typecheck, even if the underlying `Book` type for the two Libraries are the same. The type system thus allows the code to model the real world restrictions of libraries (cannot renew a book borrowed elsewhere) without restricting the number of possible libraries that can be dynamically created.

Unlike modules, type members in objects do not have to be either completely opaque or completely transparent. The exposed type member can be specified with a bound on its subtyping relation: A type member `t` can be either upper-bounded (`t` must be a subtype of another type), lower-bounded (`t` must be a supertype of another type), or exact-bounded (`t` is exactly another type, i.e. completely transparent).\footnote{In this system, a completely opaque type is usually defined with an upper bound of the top type ($\top$), which is defined as the supertype of all types. On the contrary, a bottom type ($\bot$) typically exists and is defined as the subtype of all types.} This provides the language with not only the ability to represent and typecheck traditional object/record or module types, but also more expressive types that are related to each other.

```plaintext
1 class GenericBook {}
2 class Textbook extends GenericBook {}
3 class Fiction extends GenericBook {}
4
5 class SchoolLibrary extends Library { l =>
6    type Book ≤ Textbook
7        ...
8 }
9
def study(item: Textbook): Unit = ...
10
val hunt : Library = ...
11 val sorrells : SchoolLibrary = ...
12 study(sorrells.borrow("CSAPP"))  // OK
13 study(hunt.borrow("Harry Potter")) // type mismatch
```

Listing 2.2: Bounds on type members
In Listing 2.2, `SchoolLibrary` is defined to be the kind of library whose books are all of an abstract type that is a subtype of `Textbook`. This means books borrowed from a `SchoolLibrary` can be passed to `study()`, while a book borrowed from a generic `Library` whose type member `Book` is completely opaque will not typecheck when studied.

Finally, type refinements are also supported as a flexible way of specifying more specific types. In Listing 2.3, the `borrowAndStudy` function typechecks because it requires its argument to not just be any `Library`, but specifically a `Library` whose `Book` type is “at most” (i.e. no more general than) `TextBook`.

```python
1  def borrowAndStudy(lib: Library{type Book ≤ Textbook}): Unit =
2       study(lib.borrow("CSAPP"))
3
4  borrowAndStudy(sorrells)   // OK
5  borrowAndStudy(hunt)      // type mismatch
```

Listing 2.3: Type refinement

### 2.2 Subtyping and Undecidability

The widespread adoption of object-oriented programming languages has made popular the concept of subtyping\(^2\), a form of declaration-site inclusion polymorphism \cite{cardelli1985type} that allows one type to masquerade as another. At its core, subtyping is characterized by the subsumption principle: if S is a subtype of T (written S <: T), then values of type S can act like values of type T. Clearly, this provides additional expressive power to programmers by enabling a limited form of bounded parametric polymorphism for functions even without traditional parametric polymorphism support. For example, one could write a traversal method that takes in any type of Graph, and not care what particular subtype of Graph (e.g. DAG, Tree) they actually get since the interface would be as expected from Graph.

Subtype checking is the procedure for checking if one type subtypes another. In the basic scenario above, subtype checking is easily decidable. Since all the types are self-contained names, the predefined subtyping relations (usually defined at the declaration site of the type with keywords such as `extends`) define a partial order on all the type names.

\(^2\)Or “subclassing”, since types and classes are often conflated in this context.
Subtype checking is thus checking if the two types are correctly related with respect to the partial order.

However, many modern object-oriented languages also support parametric polymorphism, either in the form of type parameters (e.g. Java generics, C++ templates) or type members (e.g. Scala type members\(^3\)) as shown in Figure 2.1. As a result, not all types are predefined like before. The possibility of constructing new types (by filling in type parameters or refining type members) means subtype checking in these systems must evolve to be structural and recursive. However, \(A \triangleleft B\) does not necessarily mean each type parameter/member of \(A\) is a subtype of the corresponding parameter/member of \(B\). Figure 2.2 illustrates that when \(A \triangleleft B\), a type parameter/member can be covariant (i.e. it preserves this subtyping relation) or contravariant (i.e. it reverses this subtyping relation).

The difficulty of subtyping arises when a type \(S\) is defined as a subtype of some type parameterized with \(S\) itself. Such a recursive definition may seem unfamiliar, but recursive subtype definition is heavily used by F-bounded polymorphism [Canning et al., 1989], a generalization of bounded polymorphism where the bounded type can appear in its bound. One common usage, shown in Figure 2.3, is using recursive bounds to specify features

\(^3\)Scala also supports type parameters.
trait Cloneable[T] {
  def clone(): T
}
class String extends Cloneable[String] {
  def clone(): String = ...
}
def makeClone[T <: Cloneable[T]](x: T) = x.clone()

Figure 2.3: F-bounded polymorphism (in Scala)

Given types:

Eq⟨−T⟩, List⟨+T⟩ <: Eq⟨List⟨Eq⟨T⟩⟩⟩, Tree <: List⟨Tree⟩

Query: Tree <: Eq⟨Tree⟩

Tree <: Eq⟨Tree⟩
List⟨Tree⟩ <: Eq⟨Tree⟩
Eq⟨List⟨Eq⟨Tree⟩⟩⟩ <: Eq⟨Tree⟩
Tree <: List⟨Eq⟨Tree⟩⟩
List⟨Tree⟩ <: List⟨Eq⟨Tree⟩⟩
Tree <: Eq⟨Tree⟩
...

Figure 2.4: Infinite derivation example (from Greenman et al. [2014])

of the bounded type. In this case, String is defined recursively so that functions expecting Cloneable objects can make more specific inferences about the return type of their clone() method.

Subtype checking on these constructed types already involve recursively looking into the structure of both types to make sure all members/parameters satisfy the subtyping relation. Thus, recursive bounds are a potential cause for concern since subtype checking can now possibly loop back to a type it has seen already. Prior research shows this is indeed a challenge. For Java, allowing wildcards with contravariant bounds (i.e. <: ? super T>) is shown to be able to lead to an undecidable subtyping problem [Grigore, 2017]. For Scala, it is the ability to encode system $\text{F}_<\text{c}$, a language already shown to have an undecidable subtyping problem [Pierce, 1992], that makes it undecidable [Amin et al., 2014].

Having an undecidable component in a language can pose real problems to the user
Given types:

\[ N = \{ n \Rightarrow L \geq \bot \}, \quad T = \left\{ z \Rightarrow \begin{array}{l} A \leq T \\ B \leq N \{ n \Rightarrow L \geq \{ z_1 \Rightarrow \begin{array}{l} A \leq z_0.A \\ B \leq z_1.A \end{array} \} \} \end{array} \right\}, \]

\[ T_0 = \left\{ z_0 \Rightarrow \begin{array}{l} A \leq N \{ n \Rightarrow L \geq T \} \\ B \leq z_0.A \end{array} \right\} \]

Query: \( T_0 <: T \)?

\[
\begin{array}{c}
\emptyset \vdash T_0 \quad : \quad T \\
z_0 : T_0 \vdash z_0.A \quad : \quad N \{ n \Rightarrow L \geq \{ z_1 \Rightarrow \begin{array}{l} A \leq z_0.A \\ B \leq z_1.A \end{array} \} \} \\
z_0 : T_0 \vdash N \{ n \Rightarrow L \geq T \} \quad : \quad N \{ n \Rightarrow L \geq \{ z_1 \Rightarrow \begin{array}{l} A \leq z_0.A \\ B \leq z_1.A \end{array} \} \} \\
z_0 : T_0 \vdash \{ z_1 \Rightarrow \begin{array}{l} A \leq z_0.A \\ B \leq z_1.A \end{array} \} \quad : \quad T \\
z_0 : T_0 \vdash z_0.A \quad : \quad N \{ n \Rightarrow L \geq \{ z_2 \Rightarrow \begin{array}{l} A \leq z_1.A \\ B \leq z_2.A \end{array} \} \} \\
z_1 : \{ z_1 \Rightarrow \begin{array}{l} A \leq z_0.A \\ B \leq z_1.A \end{array} \} \vdash z_1.A \quad : \quad N \{ n \Rightarrow L \geq \{ z_2 \Rightarrow \begin{array}{l} A \leq z_1.A \\ B \leq z_2.A \end{array} \} \} 
\end{array}
\]

To check if a type subtypes another type, we look inside the type structure and compare the bounds on each member. Only the comparison on the type bounds of \( B \) is shown above for brevity.

Figure 2.5: Diverging context example

(e.g. programmer, compiler & tools implementer). Programmers unfamiliar with the intricacies of the type system of their language may end up unable to compile their “well-written” program. This problem is worsened in that in these situations the compiler is unable to provide any helpful hints as to why it timed-out/crashed\(^4\), not to mention any guidance on what can be done to the source code to fix the problem (which is a feature expected of modern compilers). By having a decidable system with clearly defined constraints, compilers and other programming tools will be able to much better assist the programmer in expressing what they want.

2.2.1 Getting Back Decidability

Unfortunately, subtype checking in systems like F\(_<\) is not a simple case of cycle detection. Figure 2.5 presents a classic example first discovered by Ghelli [1995] encoded in

\(^4\)Running the example in Figure 2.4, the javac compiler loops until it runs out of stack space, and the Scala compiler complains the class graph is not finitary
DOT (based on the translation from Mackay et al. [2019]). As the derivation progresses, we constantly loop back to checking the same structural types but with different variable names, and the context grows larger with these new types. In the general case, it is not trivial to identify a looping derivation. In fact, it is not even ideal to implement only a simple looping detector since, similar to just adding a time-out in the compiler, it does not help the programmer in fixing the problem.

Many have since proposed enforcing some sort of subtype dependency restrictions so that infinitely looping derivations never occur. The most notable is the restriction on “expansive inheritance” by Kennedy and Pierce [2006] as it is used by the Scala compiler. Among their three proposed solutions, the most reasonable restriction was disallowing dependency cycles with “expansive edges”: where one type parameter of type S may appear at a deeper nested level in the supertype of S. However, as the authors themselves acknowledged, this solution is not immediately applicable to Java wildcards. In addition, Greenman et al. [2014] pointed out that this restriction prevents a common pattern for expressing certain “features” of types: Recall in Figure 2.3 the Cloneable type is used to by its subtypes to signal they have a clone() method. But if we want a generic list to be cloneable we would get the definition in Figure 2.6 which now includes an expansive edge from E to E.

```
class List[E] extends Cloneable[List[E]] {z =>
  def clone(): List[E] = ...
}
```

Figure 2.6: Cloneable list causing expansive inheritance

### 2.2.2 Material-Shape Separation

The solution of Nominal Wyvern is adapted from the “material-shape separation” idea proposed by Greenman et al. [2014] for Java-like languages (instead of DOT, which has type members). Material-shape separation is a conservative way of separating all types in a program into two camps: materials and shapes: A material type represents concrete types that actually represent data, and are passed around in a program. A shape type, on the other hand, are only used to bound other types, typically parameterized with these other types as well (thus creating loops).
The restriction the authors enforced on top of this dichotomy is that all type cycles must go through at least one shape (shapes enable loops), and that shapes cannot be used as type arguments for inherited types. The reasoning behind the feasibility of such a split is that the problematic dependencies in real world programs are not arbitrary, and are usually not representative of the theoretical types that cause the subtyping to loop forever. Indeed, after studying a large corpus of existing code (13.5 million lines of Java), the authors found that current coding practices already mostly follow this separation, and that the rare cases that do not conform can be easily made so.

The benefit of this solution is twofold: 1) The restriction is already compatible with industry programming standards, meaning it would not require any major shift in programming practices for its adoption. 2) The restriction is easy to understand and identify due to a limited number of intuitive uses of shapes. In fact, Greenman et al. identified the two ways in which shapes are used by programmers that are corroborated by their study of existing code: A shape is either used as a bound for the “self” type, as is the case in the aforementioned Equatable example, or used as a bound for the “self” type as part of a type family. This clear and intuitive separation means that instead of arbitrarily restricting what programmers can write (and thus forcing them to adopt an esoteric rule), it can serve as a useful tool in helping them structure their code to be more modular, and, according to the codebase review, in a way that most programmers already do.

2.2.3 Material-Shape Separation for DOT

Getting a similar separation for DOT-based languages requires adapting the original solution from relying on type parameters to rely instead on type members. The most straight-

```java
interface Pair<T> {
    T getLeft()
    T getRight()
}

class Point implements Pair<Int> {
    Int getLeft() { ... }
    Int getRight() { ... }
}
```

```scala
trait Pair { z =>
    type E ≲ T
    def getLeft(): z.E
    def getRight(): z.E
}

class Point extends Pair[Int] { z =>
    type E = Int
    def getLeft(): Int
    def getRight(): Int
}
```

Figure 2.7: Example translation from type parameter to type member
forward translation from a type parameter-based system like Java to a type member-based system like DOT is using type members to represent type parameters (e.g. Figure 2.7). However, if we directly translate Greenman et al.’s restriction that shapes cannot be type parameters to this system, we end up disallowing the use of shapes when defining type member bounds. This is a notably wider restriction than it was originally meant for in Java since type members have more uses than specifying type parameters. For example, given a shape type `Equatable`, if we wanted to write a type storing a pair of equatable objects, we would have to use equatable to define the type members, even if the parent type `Pair` is not part of any subtype chain.

Nominal Wyvern’s adaptation of material-shape separation is inspired by work by Mackay et al. [2019] on Decidable Wyvern, in which the authors proposed an adaptation of material-shape separation to DOT-based systems. Decidable Wyvern’s solution is a combination of semantic and syntactic restrictions. Shapes are still defined as the enabler of cycles in the subtyping dependency graph, but additionally all cycles must pass through structural types (to avoid meaningless cross inter-member dependencies). On the syntactic side, shapes can only serve as upper bounds for materials; shapes can only be refined with purely materials; shapes can only be defined with purely materials; and additionally shapes can only be upper-bounded by purely material refinements on the top type. This ensures subtype derivations always sink towards purely material types, from where no cycles will ever occur and termination of subtyping is guaranteed.

Nominal Wyvern differs from Decidable Wyvern mainly in having a nominal typing and subtyping system. This allows for a slightly simpler set of material-shape separation restrictions. Chapter 3 details our restrictions, and Chapter 4 proves the termination of the subtyping algorithm under the restrictions.

### 2.3 Nominality

The typing of the DOT calculus is already considered nominal as it relies on the names of objects to get path-dependent types. The nominality we present here goes one step further by mandating all structural types be named, and that the subtyping relations between these named structural types be entirely nominal as well. This contributes to a simpler and more usable system.
2.3.1 Typing Nominality

In contrast to DOT where a new structural type can be defined anywhere anonymously by simply writing out its members, Nominal Wyvern requires all structural types to be predeclared and named in the scope. This is due to two reasons:

1. Usability: Explicitly written out structures makes the code easier to understand since the names would be representative of what the structure is used for. This is especially important as the object gets larger and contains more kinds of members (i.e. type members, value members, function members). Equally important is that having easily identifiable names would make any information the compiler or other programming assisting tools generate be more readable.

2. Performance: Having named structures allows the typechecker to easily store pre-checked subtyping relations so that later subtyping queries (or ones with the same name but with slight refinements) can avoid repeating work. The next subsection also explains how this benefit is augmented with a nominal subtyping system.

2.3.2 Subtyping Nominality

Traditionally, structural subtyping is done by comparing each member between two types to ensure each satisfied the subtyping relation defined on structures. Since Nominal Wyvern disallows anonymous structural types, structural subtyping is no longer required. Following a similar reasoning as typing nominality, Nominal Wyvern requires all subtyping relations to be explicitly declared. Type S subtypes type T only if the programmer writes that they want S to subtype T. The two benefits are similar:

1. Usability: Explicitly naming subtyping relations avoids accidental subtyping by making sure that types whose signatures match are not automatically considered related. Since all structural types are named, types that are structurally compatible should not be related if the meaning associated with the names do not match. This makes the type system better match up with what a programmer reading the code expects, and prevents accidental passing of the wrong argument to a function even if naively the structures match up. The added benefit is a more understandable subtyping relation to the code reader (many subtyping checks are quite involved as the rules will soon show).
2. Performance: Having all subtyping relations defined explicitly means more type-checks can be done on the type signatures (and subtyping assertions) alone. The result of these checks can also be saved and reused throughout the typechecking of the dynamic expression, saving repeated checks that may be long and recursive.
Chapter 3

Nominal Wyvern Design

This chapter discusses how the nominal design of Nominal Wyvern facilitates subtype decidability and usability. Sections N.N to N.N builds up a motivating example; Sections N.N to N.N explains the design while referencing the earlier example.

3.1 A Store of Named Structures

Suppose we want to write a system for keeping track of the stock at a store. Each piece of fruit is labelled with an ID number, and they’re weighed when entered into the system. Similar to DOT, this warrants a record (aka “structural”) type with the two member values. As evident from the following code, the interface for both apple and orange are the same. If not careful, a completely structural type system would fail miserably as any function that is supposed to operate on Apples would “physically” work just as well when given an Orange. Having a nominal type system means the type system will stop the programmer from mixing apples and oranges since structural types are given names in addition to its members. Similarly, a nominal subtyping system means that while both McIntosh and Macintosh are structurally compatible with Apple (in this case, a form of width subtyping\(^1\)), only the one that is declared to be a subtype explicitly by the programmer can be used as an Apple. In Nominal Wyvern, this is declared with a special subtype declaration as seen on line \(26\) of Listing \[3.1\].

\(^1\)Width subtyping refers to allowing subtypes to have more members (hence ‘width’) than its supertype. This is often supplemented with depth subtyping, which allows members to have more specific type bounds in the subtype than in the supertype.
Listing 3.1: A nominal typing system on structural interfaces (Part 1)

With all the basic setup completed, the following code uses the declared interfaces to keep track of the stock with two simple counters encapsulated in a `StockCounter` interface. `StockTracker` operates on the counter, and the type system makes sure that subtyping is completely nominal by disallowing the mixing of apples and oranges.
Let's add generics to our stock system. We want to make sure we do not overestimate our stock if we accidentally scan the same piece of fruit twice. We want to make the set generic so it can work for any type of fruit, but in order for the set to properly operate, the element type must have an `equals` method defined on it so we can remove duplicates. In Listing 3.3, we update the code so Apple and Orange both subtype Fruit, which in turn subtypes a new `Equatable` type that has an equals method.

`Equatable` is defined with F-bounded polymorphism. Similar to the Cloneable example in Figure 2.6, the type member `T`'s purpose is for future subtypes to refine with themselves. In order for `equals` to refer to it, we introduce a self variable to the named interface.
like with Scala. This self variable (named \( z \) in the example code) represents the object that the member is being accessed on. For example, for any object that instantiates Fruit, its equals method must take in an object of its own \( T \) type.

The Set type consists of an element type \( \text{elemT} \) that must subtype Equatable. A refinement serves as a qualification of the Equatable type that additionally requires the \( T \) type of Equatable to be lower bounded by \( \text{elemT} \) itself. This is key in enabling any client method to pass objects of \( \text{elemT} \) type to the equals method of another object of \( \text{elemT} \) type.

```haskell
name Equatable {z =>
  type T >= BOT
  fun equals (x : z.T) : Bool
}

name Fruit {z =>
  type T >= BOT
  val ID : Int
  val weight : Float
  fun equals (x : z.T) : Bool
}

subtype Fruit <: Equatable

name Apple {z =>
  type T >= BOT
...
}

name Orange {z => ...
}

subtype Apple {type T >= Apple} <: Fruit

subtype Orange {type T >= Orange} <: Fruit

name Set {z =>
  type elemT <= Equatable {type T >= z.elemT}
...
}

name SetTracker {z =>
  type S <= Set
  fun empty () : z.S
  fun insert (stock : z.S, item : stock.elemT) : z.S
...
}

let apple_tracker = new SetTracker {z =>
```
### 3.3 A Binary Typing Approach

The main difference of Nominal Wyvern compared to DOT is its heterogenous typing system. In DOT, all types are type members of some other types, whereas in Nominal Wyvern, there are two sorts of types: The concrete structural types (aka “named types”) and the abstract member types. At first glance Nominal Wyvern may seem like an extraction of all structural types (along with any width-expanding refinement) to a global object type in DOT, but there is a more fundamental difference between named types and type members that contributes to Nominal Wyvern’s theme of usability and decidability.

The key difference between named types and type members is the way they are specified. Type members are declared with a bound: either with only a lower bound, only an upper bound, or both bounds that are the same in the case of an exact bound. In contrast, named types are defined as a named record for representing entities with the given prop-
Figure 3.1: Nominal Wyvern Grammar
erties. Semantically, this more closely resembles the usual definition of a “definition”: If our store thinks of each Apple as a record with an integer ID and a floating point weight, then that is exactly what the named type “Apple” is. Contrast this with the declaration of a type member “Apple” with a bound on the subtype relation, which more closely resembles a guideline and guarantee on how generic/specific this type may be for future instantiators and users of the parent type (Recall in Listing 3.3, the type member \texttt{elemT} of \texttt{Set} is defined with an upper bound to make sure all instantiators use an equatable type in sets). This dichotomy separates responsibility by making named types the definer of types, and type members merely users of the pre-defined types. Consequently, all types are named types behind abstraction boundaries.

The separation of named types from member types also requires a separate way of defining the subtype relation between named types. Traditionally in DOT, the subtype relation between types (which are all type members) is defined solely by the bounds declared on the type members. Syntactically, this no longer works as named types are separately defined. Semantically, bounds are also no longer suitable since the subtype relation between named types should be a semantic definition instead of a guide/guarantee. In Nominal Wyvern, bounds are complemented by the explicit subtype declaration statements. The purpose of this 	exttt{subtype} declaration is for programmers to declare semantically (and the type system to verify structurally) suitable base cases for the subsumption principle: If we think of McIntosh as a special kind of Apple, then a McIntosh should be able to be used anywhere a generic Apple is needed. The declarations are only base cases because depth subtyping in the form of type refinement is automatic (e.g. in Listing 3.3, \texttt{Apple \{type \texttt{T = Apple} \} <: \texttt{Apple}}).

In addition to the benefit of semantic separation, explicit subtype declarations also allow for more flexibility than traditional bounds in two ways:

1. 	extit{Multiple Subtyping}: Type \texttt{T} can be a subtype of multiple types. This is often used when one type wants to have the features of many other types. For example, a resource type (such as Apple) can declare itself a subtype of both Equatable and Hashable to signal it has both an \texttt{equals} method and a \texttt{hash} method, so types that require either one (e.g. Set, Hashtable) can use it as the key. In the absence of arbitrary intersection types, this circumvents the problem by essentially requiring each intersection type to have a unique name.
2. **Conditional Subtyping**: Type \( T \) can be a subtype of another type only if \( T \) is under certain refinements. It may be the case that named type \( T \) is not inherently a subtype of named typed \( S \) (this could be due to either structural compatibility or semantic compatibility). The syntax allows a refinement on the LHS of the subtype symbol \(<:\) to make a subtype relation hold conditionally. Section P.Q.R presents an example where conditional subtyping is used.

This difference carries over to the subtyping algorithm as well. While the type bound on a type member \( S \) represents the authoritative “next” type to check after \( S \), the subtype declarations with \( S \) as the base type on the LHS present us with multiple conditional options for what this type could also be seen as by the subsumption principle. Section N details the binary subtyping algorithm and why it is a decidable problem after material-shape separation is applied.

### 3.3.1 Top-Level Well-Formedness

The static contexts \( \Delta \Sigma \) are derived solely from the top-level declarations \( D \), and are used in the typing and subtyping rules detailed in the following sections. Figure 3.2 presents the judgment rules for top-level declaration well-formedness. Note that all named types are considered to be declared at the same time (in the same scope and can reference each other). One additional constraint is that the bounds on type members cannot reference any of its sibling fields. This is to prevent infinite typing derivations, which will be explained in detail in section 3.5.

### 3.3.2 Nominal Subtyping Graph

To aid in nominal subtyping, we define a “nominal subtyping graph” to capture the multiple conditional subtyping relation between named types.

**Definition 1** (Nominal subtyping graph). For a set of top level declarations \( \overline{D} \), the nominal subtyping graph is a graph \( (V, E) \). The vertices, \( V \), consist of all the named types in \( \overline{D} \). The edges, \( E \), each represent an explicit subtype declaration in \( \overline{D} \), with the refinement
\[\text{\(\overline{D}\) wf}\]

\[
\begin{align*}
\text{names}(\overline{D}) &= \Delta \\
\text{subs}(\overline{D}) &= \Sigma
\end{align*}
\]

\[
\begin{align*}
\forall n : \{x \mapsto \sigma\} \in \Delta. \Delta \vdash n : \{x \mapsto \sigma\} \text{wf} \\
\forall n_1 r_1 <: n_2 \in \Sigma. \Delta\Sigma \vdash n_1 r_1 <: n_2
\end{align*}
\]

\[\text{\(\Delta\) \(\vdash\) \(n : \{x \mapsto \sigma\}\) \(\text{wf}\)}\]

\[
\begin{align*}
\forall \text{type} t \text{ } B \tau_t \in \overline{\sigma}. \forall \text{val} v : \tau_v \in \overline{\sigma}. v \not\in \tau_t \\
\Delta \vdash n : \{x \mapsto \sigma\} \text{wf}
\end{align*}
\]

\[\Delta\Sigma \vdash n_1 r_1 <: n_2\]

\[
\begin{align*}
\Delta(n_1) &= \{x_1 \mapsto \overline{\sigma_1}\} \\
\Delta(n_2) &= \{x_2 \mapsto \overline{\sigma_2}\} \\
\Delta\Sigma \vdash n_1 r_1 S +_\sigma \overline{\delta_1} <: [x_1/x_2]\overline{\sigma_2}
\end{align*}
\]

Figure 3.2: Nominal Wyvern Top-Level Decl Well-Formedness
Example 1. The nominal subtyping graph for the code in Listing 3.3 is:

```
Equatable

Fruit

{type T >= Apple}  {type T >= Orange}

Apple  Orange
```

3.4 Material-Shape Separation

The definition of materials and shapes is based on the discovery by Greenman et al. [2014] that shapes should be the only types that enable cycles during subtype derivation. Practically, shapes are used either to define features of the self-type, or to define features of a type family that the self-type is a part of. In the fruit shop example, Equatable is an obvious case of the former use of shape since it prescribes that whatever subtypes it must have an equals method that takes in a value of their own T type.

For Java, Greenman et al. defines a subtype dependency graph (called “inheritance usage graph” in their paper) that maps each type S to the types mentioned in its inheriting type so that when S becomes the LHS type, we know what types will be visited by the subtype derivation afterwards. By preventing the cycle-inducing shapes from appearing as type arguments, no new usages of shapes are created and eventually all shapes are reduced into materials.

Nominal Wyvern follows a similar adaptation of the subtype dependency graph as Mackay et al.’s adaptation in Decidable Wyvern. The difference is due to the binary typing approach that Nominal Wyvern employs. Before defining this graph, we first define the
refinement tree for types.

**Definition 2 (Refinement tree).** For a type \( \tau = \beta r \), the refinement tree is a tree \( \langle V, r, E \rangle \). The vertices, \( V \), consist of all the base types referred to in \( \tau \). The root, \( r \), is \( \beta \). The edges, \( E \), each go from a base type, \( \beta' \), to the root of the refinement tree generated by the type used as the bound in a refinement to \( \beta' \). It is formally generated by \( \text{GenTree} \):

\[
\text{GenTree}(\tau) = \langle V, r, E \rangle. \quad \text{(Each edge is represented by } v_1 \to v_2: \text{An edge from node } n_1 \text{ to } n_2 \text{).}
\]

\[
r = \text{type } t_1 B_1 \tau_1 \ldots \text{type } t_n B_n \tau_n
\]

\[
\text{GenTree}(\tau_1) = \langle V_1, r_1, E_1 \rangle \quad \ldots \quad \text{GenTree}(\tau_n) = \langle V_n, r_n, E_n \rangle
\]

\[
\text{GenTree}(\beta r) = \langle V_1 \ldots V_n \beta, \beta, E_1 \ldots E_n (\beta \to r_1) \ldots (\beta \to r_n) \rangle
\]

**Example 2.** The refinement tree for type

```plaintext
SetTracker {
    type S = Set {
        type elemT = Apple {T = Apple}
    }
}
```

is:

```
SetTracker

Set

Apple

Apple
```

The subtype dependency graph is defined partially using refinement trees.

**Definition 3 (Subtype dependency graph (SDG)).** For a set of top level declarations \( \overline{D} \), the subtype dependency graph is a graph \( \langle V, E \rangle \). The vertices, \( V \), consist of all available
base types, which are: \(\top, \bot\), all declared named types in \(\mathcal{D}\) and their type members. To disambiguate type members with the same name from different named types, each type member \(t\) of name \(n\) is denoted \(n::t\) (will be referred to as a “pseudotype”). The edges, \(E\), are generated as follows:

For each type member declaration, \(\text{type} \ t \ B \ \beta, r_t\), in named type \(n\), generate edges:

\[
\begin{align*}
\text{DIRECT} & : n::t \rightarrow \beta \\
\text{INDIRECT} & : n::t \xrightarrow{rta(\beta, \beta_r)} \beta_r
\end{align*}
\]

And for each subtype declaration \(\text{subtype} \ n_1 \ r_1 <: n_2\) in \(\mathcal{D}\), generate edges:

\[
\begin{align*}
\text{BACK} & : n_2 \rightarrow n_1 \\
\text{BACK-REF-ROOT} & : n_2 \xrightarrow{\text{rta}(\beta, \beta)} n_r
\end{align*}
\]

where \(\delta \in^* r\) is true if the refinement member \(\delta\) occurs anywhere syntactically within \(r\), and \(\delta \in r\) is true only if \(\delta\) is immediately within \(r\), and where \(\text{rta}(\tau, \beta)\) returns all the ancestors of \(\beta\) in the refinement tree of \(\tau\) in the order from root of tree to \(\beta\) (not including \(\beta\) itself), and

where \(v_1 \rightarrow v_2\) represents an unlabeled edge, and \(v_1 \xrightarrow{\beta_1, \beta_2, \ldots} v_2\) represents an edge labeled with the base types \(\beta_1, \beta_2, \ldots\) in that order.

The idea is if \(\beta\) has an edge to (“depends on”) \(\beta'\), then when \(\beta\) is encountered as a base type during subtype derivation, \(\beta'\) may appear as the base type later in the derivation. The graph is thus a conservative guarantee of what types will not be visited again in later derivations. This serves as the key idea for proving decidability in Chapter 4.

Example 3. The subtype dependency graph for the declarations in Listing 3.3 is:
Shapes are still defined to be the types that enable cycles in the subtype dependency graph:

**Definition 4** (Shape type). *Shapes are the set of types such that if all edges labeled with at least one shape are removed from the subtype dependency graph, then the graph is acyclic. All other types are material types.*

Clearly, the solution to which set of types are the shapes in a program is not unique: one can simply label all types as shapes. However, not only does that not follow the semantic meaning of shapes and materials detailed earlier (the two uses of shapes), the material-shape separation requirements will further limit which types are allowed to be shapes. Observe that as long as the separation requirements are fulfilled, the number of possible solutions does not matter for decidability as long as there is one valid set of shapes.

**Definition 5** (Material-shape separation). *A program is material-shape separated if there exists a set of shape types such that:*

- A shape is never used as part of a lower bound syntactically (i.e. after ≥ or =)
- The upper bound of a shape is always a shape, and named shapes can only subtype named shapes
- A named shape cannot be refined in subtype declarations

**Example 4.** *For the code in Listing 3.3, the only shape type is Equatable because removing the self loop that is labeled with it will make the subtype dependency graph in Example 3 be acyclic. It is a valid use of shape since it follows the restrictions in Definition 5.*

### 3.5 Term Typing

All expressions in Nominal Wyvern correspond to objects. All expressions are given names, either assigned with a `let` expression, or as a `val` member in another object. Objects are then used by referring to the path that refers to their names, either directly as the
\[ \Delta \Sigma G S \vdash e : \tau \]

\[
\begin{array}{ll}
\Gamma(x) = \tau & \Delta \Sigma G S \vdash x : \tau \\
S(l) = \tau & \text{T-VAR} \\
\Delta \Sigma G S \vdash l : \tau & \text{T-LOC} \\
\Delta \Sigma G S \vdash p : \tau & \Delta \Sigma G S \vdash \tau < n r \\
\Delta \Sigma G S \vdash n r \ni \text{fun } f : \tau_a x_a \rightarrow \tau_r \\
\Delta \Sigma G S \vdash p' : \tau' & \Delta \Sigma G S \vdash \tau' < : [p/x]r_{a r} \\
\Delta \Sigma G S \vdash p.f(p') : [p, p_a/x, x_a]r_r & \text{T-APP} \\
\Delta \Sigma G S \vdash \tau \ni \text{val } v : \tau_v & \text{T-NEW} \\
\Delta \Sigma G S \vdash e : \tau & \Delta \Sigma G S \vdash e_1 : \tau_1 \quad \Delta \Sigma G S \vdash \tau_1 < : \tau \\
\Delta \Sigma G S \vdash e_2 : \tau_2 & \Delta \Sigma G S \vdash \tau_2 < : \tau \\
\Delta \Sigma G S \vdash \text{let } x = e_x \text{ in } e : \tau & \text{T-LET} \\
\Delta \Sigma G S \vdash \text{if } \tau \ni p_1 = p_2 \text{ then } e_1 \text{ else } e_2 : \tau & \text{T-IF} \\
\end{array}
\]

\[ \Delta \Sigma G S \vdash \tau \ni \{ x \Rightarrow d \} \text{ wf} \]

\[ \begin{array}{ll}
\Delta \Sigma G S \vdash \tau < n r & \Delta(n) = \{ x_n \Rightarrow \overline{\sigma_n} \} \\
\tau_x = n \{ r_n \ni \text{ref}(\text{sig}(\overline{d})) \} & \Gamma' = \Gamma, x : \tau_x \\
\Delta \Sigma G S' \vdash \text{sig}(\overline{d}) < : [x/x_n]r_{\overline{\sigma}} & \forall \text{val } v : \tau_v \ni e \Rightarrow d, \Delta \Sigma G S' \vdash e : \tau_v \\
\forall \text{fun } f : \tau_a x_a \rightarrow \tau_r & \ni e \Rightarrow d, \Delta \Sigma G S' \vdash x_a : \tau_a S' \vdash e : \tau_r \\
\Delta \Sigma G S \vdash \tau \ni \{ x \Rightarrow d \} \text{ wf} \\
\end{array} \]

\[ \Delta \Sigma G S \vdash \tau \text{ wf} \]

\[ \begin{array}{ll}
\Delta \Sigma G S \vdash \top \text{ wf} & \Delta \Sigma G S \vdash \bot \text{ wf} \\
\Delta \Sigma G S \vdash \beta r_{\beta} < n r_n & \Delta(n) = \{ x_n \Rightarrow \overline{\sigma_n} \} \\
\forall \text{type } t B \ni \delta, \Delta \Sigma G S \vdash n r_n \ni \text{type } t B' \tau' & \text{and} \\
\Delta \Sigma G, x_n : \beta r_{\beta} S \vdash \text{type } t B \ni \tau < : \text{type } t B' \tau' \\
\Delta \Sigma G S \vdash \beta r_{\beta} \text{ wf} \\
\end{array} \]

where \( +_\sigma \) and \( +_\tau \) are merge operations on \( \overline{\sigma} \) and refinements, respectively (On conflict, RHS is preserved and LHS is discarded).

\( \text{sig} : \overline{d} \rightarrow \overline{\sigma} \) transforms object member definitions into member declarations by removing the dynamic expression part of vals and funs.

\( \text{ref} : \overline{\sigma} \rightarrow v \) filters member declarations by preserving only type member declarations.

\( \Delta \Sigma G S \vdash \tau < \tau_u \) is true if following the upper bound of \( \tau \) leads to \( \tau_u \), whose upper bound is itself. Judgments not formally defined here are defined with the subtyping rules.

Figure 3.3: Nominal Wyvern Term Typing
assigned variable in a let expression, or by selecting a val member from another object. This simplification makes objects slightly easier to work with in the type system while not hindering expressiveness at all, since to use “anonymous” objects, one only needs to wrap the object in a local let expression with a fresh variable name and immediately use it.

All objects are created with the new expression and are given a type explicitly. This is in line with the goals of a nominal type system. Without a name provided by the programmer, a structure can potentially be mapped to many un-related (w.r.t. subtyping) named types. However, the explicit type given to new need not be exactly the same type as the following structure. It is only required that the structure is proper structural subtype to the exposed type. This differing view of the same object provides an easy way to abstract the types and other members of the new’ed object. Section N will show why this separate view of the same type is type safe.

Method application typing is done by first checking the argument type is a subtype of the required argument type specified by the type of the parent object of the method. Note that the argument type of the function comes from the exposed type of the object rather than the actual internal type. The resulting type of the application is the exposed type of the return type with the argument variable and self variable replaced with paths from the current context.

### 3.5.1 Typing Decidability

Term typing ($\Delta \Sigma \Gamma S \vdash e : \tau$) is heavily reliant on path typing, which judges the type that a path represents. In order to find out the type of a path, we need to look at the type of the object whose val field is being accessed. To understand the type of the inner object, which is also a path, we need to recursively apply this procedure until we get down to a single variable. However, since the type of a variable can be a path-dependent type, we need to find out the underlying named structure it is based on in order to know the type of its fields. This is reliant on continually following the upper bound of a path-dependent type until a name is reached (This is possible because well-formed programs can only access members of objects whose type is upper bounded, since otherwise we are unable to know if it even has the desired member). The type expansion rules ($\Delta \Sigma \Gamma S \vdash \tau \prec \tau$) capture this process. Therefore, as we back out of the recursive path typing judgments, we repeatedly apply type expansion to find out the type of each successive field access.
To understand why this process always terminates, we first look at how variables depend on each other in the dynamic context.

**Theorem 1.** Each record $x : \tau$ in $\Gamma$ can only mention the variables that were added to $\Gamma$ before it.

**Proof.** There are only two rules that add to $\Gamma$ during term typing: $T\text{-LET}$ and $T\text{-NEW}$. 

- Case $T\text{-LET}$: For each “let” expression $\text{let } x = e_x \text{ in } e$, the rule adds to the context $x : \tau_e$ when typing $e$, where $\tau_e$ is the type of the enclosed expression $e_x$. Since $e_x$ is typed under $\Gamma$, its type $\tau_e$ can only refer to variables in $\Gamma$, which does not yet contain $x$.

- Case $T\text{-NEW}$: For each “new” expression $\text{new } \tau \{x \Rightarrow d\}$, for each method declaration $\text{fun } f : \tau_x x \rightarrow \tau_r = e$, the rule adds to the context $x : \tau_x$ when typing $e$. $\tau_x$ cannot refer to $x$ because $x$ is not defined to be in scope for $\tau_x$.

This means every successive member of $\Gamma$ can only refer to the variables that were added before it. \qed

To make referring to positions easier, define the position of a variable $x$ in $\Gamma$ as its rank (The first variable has rank 1). Intuitively, this means the first variable can only refer to static types (names, top, or bottom), and higher ranked variables can refer to path-dependent types whose paths are rooted at the lower ranked variables (the variable that begins a path is denoted the root of the path, and the rank of a path is defined as the rank of its root). This hints at the possibility that as path typing progresses, its range of reference decreases, and derivation eventually stops. For the following theorems, we define the length of a path $p$ (written $|p|$) as the number of field accesses it has. A single variable has length 0, and $x.v_1.v_2.\cdots.v_m$ has length $m$.

**Theorem 2.** $\Delta\Sigma\Gamma S \vdash p : \tau_p$ is decidable for all well-formed path $p$.

**Theorem 3.** $\Delta\Sigma\Gamma S \vdash p.t.r < \tau$ is decidable for all well-formed path-dependent type $p.t.r$.

Since the path-typing and type expansion are mutually recursive, the two theorems are proved at the same time with a nested induction proof.
Proof. The two theorems are combined and strengthened to get the following statement.

$P_{\Delta \Sigma \Gamma S}(p.tr)$: let $x_p$ be the root of path $p$. $\Delta \Sigma \Gamma S \vdash p.tr \prec \tau$ and $\Delta \Sigma \Gamma S \vdash p : \tau_p$, and $\tau_p$ can only contain either static types, path-dependent types whose path has a lower rank than $p$, or path-dependent types rooted at $x_p$ whose length is less than $|p|$.


Base Case: The rank of $p$ is 1.

Prove by induction on the length of the longest $x_p$-rooted path in $p.tr$.

Base Case: The length of the longest $x_p$-rooted path is 0.

According to rule T-VAR, $\tau_p = \Gamma(x_p)$, which is easily decidable by looking up $\Gamma$.

Theorem 1 indicates $\tau_p$ cannot contain any path-dependent types. Therefore, it does not contain any $x_p$-rooted path at all.

To expand a $p.tr$, expansion rules first expands $\tau_p$. Since $\tau_p$ only contains static types, its expansion is $\tau_p$ itself (rule TE-NAME). When accessing type member $t$ from $\tau_p$, if the bound is a lower bound, the judgment ends, so we only consider when $t$ is not just lower bounded. There are two cases (let $x$ denote the self-variable from the membership judgment ($\Delta \Sigma \Gamma S \vdash \tau_p \ni x$ type $t \leq \beta t r$)):

If $t$ is in the refinement of $\tau_p$: Then $\beta t r$ can only contain static types or 0-length $x_p$-rooted paths. Therefore, the replacement $[p/x]$ has no effect. Combining two refinements does not create any new types, so the resulting type $\beta t r +_r \tau_1$ has the same property (can only contain static types or 0-length $x_p$-rooted paths). If $\beta t$ is a name, the second expansion also terminates. If $\beta t$ is $x_p.t'$ (for some $t'$), then expansion recurses on this new path-dependent type. However, material-shape separation rules require there be no cyclic dependencies between type members unless it goes through a refinement (i.e. a labeled edge). Therefore, this limited recursion either ends when it accesses a type member that was refined, or when it naturally ends before going through all type members of $p$.

If $t$ is not in the refinement (i.e. the bound of $t$ comes from the name type definitions): The bound of $t$ can only contain static types or 0-length $x$-rooted paths (Due to restriction on length of paths for type member bounds). The replacement $[p/x]$ replaces $x$ with $x_p$. This means we get to the exact same situation as in the previous case when $t$ is in the refinement.

Inductive Case: The length of the longest $x_p$-rooted path in $p.tr$ is $l$ ($l > 1$). Let length of $p$ be $m$ ($m \leq l$). Denote $p$ as $x.v_1.\cdots.v_m$. 
**Inductive Hypothesis:** The statement $P_{\Delta \Sigma \Gamma S}(\cdot)$ holds for all $p.t\,r$ whose longest $x_p$-rooted path (where $x_p$ is root of $p$) is less than $|p|$, and $p$ has rank 1 in $\Gamma$.

Let $p' = x.v_1\cdots.v_{m-1}$.

Rule $T\text{-}SEL$ depends on the typing of $p'$. The IH determines that $\Delta \Sigma \Gamma S \vdash p' : \tau_{p-1}$. Since $\tau_{p-1}$ does not contain $x_p$-rooted paths with length $\geq |p| - 1$, type expansion is decidable for $\tau_{p-1}$ according to IH.

Now consider when the field declaration of $v_m$ is accessed in $\tau_{p-1}$. Again, let $x$ denote the self-variable. If the type of $v_m$ in $\tau_{p-1}$ (denote as $\tau_v$) does not refer to $x$, it will be static only. If it does refer to $x$, the any $x$-rooted paths will have 0 length. After performing $[p'/x]$, the longest $x_p$-rooted path in $\tau_v$ is still no longer than $m - 1$.

Next consider the expansion of $p.t\,r$. Expansion rules first expand the type of $p$, which was just showed to not contain any $x_p$-rooted paths longer than $m - 1$. According to IH, this expansion is decidable. The proof for accessing the type member $t$ is very similar to the proof in the base case: The bound of $t$ (denote as $\beta_t\,r_t$) will either have a name type as base, in which case the second expansion terminates, or it is a $x_p$-rooted path with length equal to $m$, and $r$ can only contain static types or $x_p$-rooted paths with length no greater than $l$. In the latter case, the recursion is with another type member of $p$, but this cannot happen infinitely due to material-shape separation.

**Inductive Case:** The rank of $p$ is $r$ ($r > 1$).

**Inductive Hypothesis:** The statement $P_{\Delta \Sigma \Gamma S}(\cdot)$ holds for all $p.t\,r$ where $p$ has rank less than $r$ in $\Gamma$.

Prove by induction on the length of the longest $x_p$-rooted path in $p.t\,r$.

This part is omitted for brevity: Similar to the base case on rank, when the longest length is 0, the type of $p$ can only refer to lower-ranked paths. By IH, they are all typeable and expandable. When the longest length is $l$, it can refer to both lower-ranked and shorter paths. The key is always that the material-shape separation rules and the restriction on path-length of type member bounds prevent infinite cyclic expansion for any one particular path.

One additional lemma we are able gain from this strengthened proof is that typing a path will either decrease the rank of the path or the length of the path. We can thus define...
Definition 6 (rank-length (RL)). The rank-length measure of a path, \( p \), given context \( \Gamma \) (written \( RL_\Gamma(p) \), or just \( RL(p) \)) is the pair of natural numbers:

\[
(rank(p), length(p))
\]

, where \( rank(p) \) and \( length(p) \) are the rank and length of \( p \), respectively.

Ordering of RL follows dictionary ordering. Two RL measures \((i_r, i_l) < (i'_r, i'_l)\) iff \( i_r < i'_r \), or \( i_r = i'_r \wedge i_l < i'_l \). They are equal if both components are equal.

Lemma 1. If \( \Delta \Sigma \Gamma S \vdash p : \tau_p \), then for any path-dependent type \( p'.tr \) in \( \tau_p \), \( RL_\Gamma(p') < RL_\Gamma(p) \).

In addition, as can be seen from the proof, given any type \( p.tr \), if \( \Delta \Sigma \Gamma S \vdash p.tr \prec \tau \), then the largest RL of any path in \( \tau \) cannot be larger than the largest RL of any path in \( p.tr \). This is because at no point during expansion can a path type get longer. The only time a new path type is created is when the type bound came from a name type definition, in which case any self variable \( x \) (length 0) is replaced with \( p \).

Lemma 2. If \( \Delta \Sigma \Gamma S \vdash p.t r \prec \tau \), then the largest RL of any path in \( \tau \) cannot be larger than the largest RL of any path in \( p.t r \).
Chapter 4

Subtyping Decidability

4.1 Subtyping Judgments

Figures 4.1 and 4.2 present the subtyping judgments for Nominal Wyvern. They consist of three main parts:

1. Nominal type subtyping: subtyping between two named types.

2. Member type subtyping: subtyping between two type members.

3. Structural subtyping: subtyping between two structural type.

Nominal type subtyping follows the nominal subtyping graph (Definition 1). To check $n_1 r_1 <: n_2 r_2$, we first check if the two are related by the nominal subtype relation by finding a path from $n_1$ to $n_2$ in the nominal subtyping graph. If there are no such path, we can immediately conclude ‘false’. Due to conditional subtyping, even if there are paths from $n_1$ to $n_2$, we must check if there is a path such that the refinement labeled on every edge of the path (the “conditions”) are each satisfied by $r_1$. Finally, if we have a conditional path, we still have to check if the refinements in $r_2$ are still supertypes of the corresponding types in $n_1 r_1$. It is possible that $r_2$ makes some type member of $n_2$ too specific for even $n_1 r_1$. Listing 4.1 contains one example for each of these three cases.

```plaintext
1   name A {z => ...}
2   name B {z => ...}
3   name C {z => ...}
```
\[ \Delta \Sigma \Gamma S \vdash \tau <: \tau \]

\[ \Delta \Sigma \Gamma S \vdash \tau : \top \quad \Delta \Sigma \Gamma S \vdash \bot <: \tau \]

\[ \Delta \Sigma \Gamma S \vdash p : \tau \quad \Delta \Sigma \Gamma S \vdash \tau_p < nr \quad \Delta \Sigma \Gamma S \vdash nr \ni \text{type } t \geq \beta r_t \]

\[ \Delta \Sigma \Gamma S \vdash [p/x](\beta_t r_t + r r_1) <: \tau_2 \]

\[ \Delta \Sigma \Gamma S \vdash \tau : \bot \quad \Delta \Sigma \Gamma S \vdash \bot \ni \text{type } t \leq \beta r_t \]

\[ \Delta \Sigma \Gamma S \vdash p.tr_1 <: \tau_2 \quad \Delta \Sigma \Gamma S \vdash \tau_1 <: p.tr_2 \]

\[ \Delta \Sigma \Gamma S \vdash r_1 <: r_2 \quad \Delta \Sigma \Gamma S \vdash p.tr_1 <: p.tr_2 \]

\[ \Delta \Sigma \Gamma S \vdash n_1 r_1 <: n_2 \quad \Delta \Sigma \Gamma S \vdash n_1 r_1 <: n_2 r_2 \]

where type \( B_1 \tau \) matches a type member declaration with either bounds \( B_1 \) or \( B_2 \).

Figure 4.1: Nominal Wyvern Subtyping
$\Delta \Sigma \Gamma S \vdash r <: r$

$\Delta \Sigma \Gamma S \vdash r <: \{\}$

$\Delta \Sigma \Gamma S \vdash {\delta} <: {\delta}$

$\Delta \Sigma \Gamma S \vdash \tau_1 <: \tau_2$

$\Delta \Sigma \Gamma S \vdash \text{type } t \leq \tau_1 <: \text{type } t \leq \tau_2$

$\Delta \Sigma \Gamma S \vdash \tau_2 <: \tau_1$

$\Delta \Sigma \Gamma S \vdash \text{type } t \geq \tau_1 <: \text{type } t \geq \tau_2$

$\Delta \Sigma \Gamma S \vdash \tau_1 <: \tau_2$

$\Delta \Sigma \Gamma S \vdash \text{type } t = \tau_1 <: \text{type } t = \tau_2$

$\Sigma \ni n_1 r_1 <: n_2$

$\Delta \Sigma \Gamma S \vdash r'_1 <: r_1$

$\Delta \Sigma \Gamma S \vdash n_2 \stackrel{r_1}{\rightarrow} n_3$

$\Delta \Sigma \Gamma S \vdash n r \ni x \sigma$

$\sigma \in \delta$ $x \notin \Gamma$ M-REF

$\Delta \Sigma \Gamma S \vdash n \{\delta\} \ni x \sigma$

$\sigma \notin \delta$ $\Delta(n) = \{x \mapsto \sigma\}$ $\sigma \in \sigma_n$ M-NAME

$\sigma \in \delta$ is true if $\sigma$ is a type member declaration (i.e. $\delta$), and is part of $\delta$.

$x \notin \Gamma$ is true when $x$ is a fresh variable under the current variable typing context.

Figure 4.2: Nominal Wyvern Subtyping (continued)
subtype A <: B
subtype B <: C

name N1 {z => type t <= ⊤}
name N2 {z => type t <= C}
name N3 {z => type t <= C}

subtype N1 {type t <= B} <: N2

/* Query 1: N1 {type t <= B} <: N3
* --> false, no path in graph
* Query 2: N1 {type t <= C} <: N2
* --> false, condition not met
* Query 3: N1 {type t <= B} <: N2 {type t <= A}
* --> false, r_2 too specific
*/

Listing 4.1: 3 ways nominal subtyping can fail

Member type subtyping is similar to DOT. Follow the upper bound for LHS base types, and follow the lower bound for RHS base types. Existing refinements are merged with new refinements by discarding the new ones on conflict. Reflexivity applies for when base types on both sides are exactly the same path, in which case structural subtyping applies to the refinements. If any one side’s base type becomes a name type, it waits for the other side to also reduce into a name type, at which point nominal type subtyping applies. If both sides get stuck without reflexivity or nominal subtyping applying, the query concludes to ‘false’.

Structural subtyping follows standard subtyping on record types. Width subtyping allows LHS to contain more members than the RHS. Depth subtyping allows the bounds on the LHS to be more specific than the corresponding bounds on the RHS.

- For type members, the LHS type bound must be in the same direction (or be an exact bound) as the RHS, and be no less specific than the RHS

- For field members, the LHS type of the val must be a subtype of the RHS type

- For method members, first replace the argument variable so that both sides use the same name. The argument type of LHS must be a supertype of the RHS, and the result type must be a subtype of RHS.
For all subsequent subtype queries, the self variable (and the argument variable in the case of method) are added to the single context with the LHS parent type.

In all three parts, whenever a bottom type appears on the LHS, or a top type appears on the RHS, derivation immediately returns ‘true’.

### 4.2 Decidability

Observe from the subtyping rules the only ways nested subtype judgments can occur in a derivation tree. Figure 4.3 shows the ways one subtype judgment (any $S-*$ rule) can call back to a subtype judgment. Each node represents one family of subtype judgment rules (based on the prefix of the rule names in Figure 4.1): $SN-*$ rules judge subtype relations between named types; $SR-*$ rules judge structural subtyping relations between refinements; $SS-*$ rules judge subtyping relations between individual refinement member definitions. Each edge (black arrow) represents a possibility of a rule from one rule family calling into another family, labeled with the inducing rule names. Note that there are two outgoing edges from $S-*$ labeled with $S-NAME$. This is because the $S-NAME$ rule both directly calls into $SR-*$ and calls $SR-*$ as well inside $SN-TRANS$. The three colored paths cover the three general ways recursion can occur:
• [RED] S-UPPER or S-LOWER: called directly.

• [GREEN] S-STRUCT or S-NAME directly: called directly via structural subtyping rules, specifically SS-LOWER or SS-UPPER.

• [BLUE] S-NAME indirectly: called while checking conditional subtyping between names in the nominal subtyping graph.

All the potential ways of getting nested subtype judgments can be partitioned into the following cases (the name of each case is given in parentheses):

1. [RED] via S-UPPER (RSU): The new LHS is the upper bound of the old LHS type. RHS stays the same.

2. [RED] via S-LOWER (RSL): The new RHS is the lower bound of the old RHS type. LHS stays the same.

3. [GREEN] via S-STRUCT (GSS): The new type on each side comes from inside the refinement of each side. (Left and right may be swapped if path also went through SS-LOWER or SS-EXACT).

4. [GREEN] via S-NAME (GSN): The new RHS comes from inside the old RHS refinement. The new LHS comes from either the old LHS refinement or a type bound in the definition of the LHS named type. (Left and right may be swapped if path also went through SS-LOWER or SS-EXACT).

5. [BLUE] (BSN): The new LHS comes from inside the old LHS refinement. The new RHS comes from the refinement labeled on an edge between the old LHS named type and old RHS named type in the nominal subtyping graph. (Left and right may be swapped if path also went through SS-LOWER or SS-EXACT).

To prove decidability, we define the notion of a “lineage” in the context of a subtype derivation. The idea is that given a subtype query \( \Delta \Sigma \Gamma S \vdash \tau <: \tau \), a lineage \( \mathcal{L} \) captures the trace a type goes through during the subtype derivation that starts with that initial query. Concretely, a lineage is a tree with types as nodes. The shape of the tree corresponds exactly to the derivation tree of the initial subtype query. Each subtyping derivation creates two lineages: an initial left-lineage rooted at the initial LHS type, and an initial right-lineage rooted at the initial RHS type. For each nested subtyping judgment
in the derivation tree, the inner judgment’s (deeper in tree) LHS and RHS types are linked to the outer judgment’s types depending on which rules were used between the inner and outer subtype judgments. In most cases, the inner type is added as a child of the outer type of the same side. However, if the judgments involve a \texttt{SS-LOWER} or \texttt{SS-EXACT} (i.e. the bound on the type in the structure involved a lower bound), the inner LHS type is added as a child of the outer RHS type (“the left lineage swings to the right”), and the inner RHS links to the outer LHS. This is the only case when the two lineages swap sides.

**Definition 7 (Lineage).** Given a subtype derivation tree rooted at \( \Delta \Sigma S \vdash \tau_{\text{initl}} \prec < \tau_{\text{initr}} \), the two lineages of the derivation tree are each a tree. Each subtype judgment in the derivation tree is given a label, and for each subtype judgment \( J: \Delta \Sigma S \vdash \tau \prec < \tau_r \), two vertices are created: \( J \# \tau_l \) and \( J \# \tau_r \). For each pair of judgments \( J_1: \Delta \Sigma S \vdash \tau_{l1} \prec < \tau_{r1} \) and \( J_2: \Delta \Sigma S \vdash \tau_{l2} \prec < \tau_{r2} \) such that \( J_1 \) is the closest ancestor of \( J_2 \) that is a subtype judgment \( S-\ast \), denote new sets of edges:

- **Covariant edges:** \( J_1 \# \tau_{l1} \rightarrow J_2 \# \tau_{l2} \), \( J_1 \# \tau_{r1} \rightarrow J_2 \# \tau_{r2} \)
- **Contravariant edges:** \( J_1 \# \tau_{l1} \rightarrow J_2 \# \tau_{r2} \), \( J_1 \# \tau_{r1} \rightarrow J_2 \# \tau_{l2} \)

, with each edge labeled with how the recursive call was made (one of \texttt{RSU}, \texttt{RSL}, \texttt{GSS}, \texttt{GSN}, \texttt{BSN}). Then generate edges depending on the path from \( J_1 \) to \( J_2 \):

- **If** \( J_1 \) **calls to** \( J_2 \) **via a** \texttt{SS-EXACT}, **add both contravariant and covariant edges**
- **If** \( J_1 \) **calls to** \( J_2 \) **via a** \texttt{SS-LOWER}, **add contravariant edges**
- **Otherwise, add covariant edges**

*The initial left-lineage and right-lineage are the trees rooted at \( \tau_{\text{initl}} \) and \( \tau_{\text{initr}} \), respectively.*

A lineage captures the relation between types in recursively dependent subtyping judgments. As long as all paths in a lineage (from the root downwards) are finite, the entire corresponding subtype derivation is finite. We can consider any path starting from the root of a lineage as made up of many segments, divided by edges labeled with \texttt{S-NAME} recursions (i.e. \texttt{GSN} or \texttt{BSN}). Below, we first study the behavior within a segment, and then extend to across segments.
To prove decidability, we define a measure $E$ on types that will decrease during derivation. To define $E$, we first define two measures, $M$ and $A$, on type members of a given name type $n$. $M(n::t)$ captures the number of other type members of $n$ reachable from $t$. $A(n::t)$ captures the other dependencies of $t$.

**Definition 8** ($M$ and $A$ measures of pseudotypes). Given a name type definition $n : \{x \mapsto \sigma\} \in \Delta$ and the subtype dependency graph derived from $\Delta \Sigma$, the $M_{\Delta \Sigma}$ and $A_{\Delta \Sigma}$ measures of a pseudotype $n::t$ is defined as:

\[
\text{type } t \; B \; \tau \in \sigma \; \Rightarrow \; T = \{n::t' | x.t' \in \tau \land n::t \rightarrow_B n::t'\}
\]

\[
M_{\Delta \Sigma}(n::t) = \sum_{x.t' \in T} M_{\Delta \Sigma}(n::t')
\]

\[
A_{\Delta \Sigma}(n::t) = 1 + \sum_{x.t' \in T} A_{\Delta \Sigma}(n::t') + \sum_{n'' \in \tau} E_{\Delta \Sigma \Gamma S}(n'')
\]

where $n::t \rightarrow_B n::t'$ is true if there is a path in the subtype dependency graph from $n::t$ to $n::t'$ that only consists of edges whose variance is $B$ and are not labeled with shapes, and the path does not consist of nodes that are not pseudotypes of $n$.

$E_{\Delta \Sigma \Gamma S}()$ is the measure on types to be defined below.

The $M$ and $A$ measures are computable thanks to the shape-material separation rules. Since there are no cycles without going through labeled edges, a partial order based on dependence exists on all type members of a given name type. Both measurements can be computed “bottom-up” from this partial order.

We now wish to define the measure $E$ on types. It can be thought of as the “potential energy” of a type, and that continued subtype derivation requires spending energy.

**Definition 9** (Energy measure of a type). Given the contexts $\Delta \Sigma \Gamma S$, first define the energy
The energy $\mathcal{E}_{\Delta \Sigma \Gamma S}$ of a type $\tau$ is the sum of the energies of the base types that occur in $\tau$ (i.e. the nodes of the refinement tree of $\tau$).

$$\mathcal{E}_{\Delta \Sigma \Gamma S}(\tau) = \sum_{\beta \in \tau} \mathcal{E}_{\Delta \Sigma \Gamma S}(\beta)$$

The material-shape separation requirements ensure that the energy measurement is defined on name types: The rules dictate a partial order over all name types, which make sure that it is always possible for the recursive energy calculation to terminate. For the path-dependent types, lemmas $[\text{1}]$ and $[\text{2}]$ make sure that every recursive $\mathcal{E}$ call can only contain paths strictly smaller in terms of RL. This means recursion eventually all stops on static types.

The intuition behind the energy equation of path-dependent types is to make sure that even if one type member $t$ depends on many other type members, the energy of $p.t$ for any $p$ should be greater than the combined energy of all that $t$ depends on. Due to the existence of cyclic dependencies through shapes, this is not always possible to compute. However, the definition of $M$ and $A$ takes this into account by ignoring any dependencies that are inside the refinement of a shape. This means during subtype derivation, when following the type bound of $t$, as long as the bound does not include any shapes, the energy before will be bigger than after. If there is a shape, we’ll show that it is still fine.

First we wish to show that, within any segment of a lineage, the energy of nodes decrease when the lineage is on the right-hand side.

**Theorem 4.** Given a subtype derivation under the contexts $\Delta \Sigma \Gamma S$ and one of its lineages
L, if \( J_1 \# \tau_1 \) is the parent of \( J_2 \# \tau_2 \) in the same segment in L and \( \tau_2 \) is on the RHS of \( J_2 \), then \( E_{\Delta \Sigma \Gamma S}(\tau_2) < E_{\Delta \Sigma \Gamma S}(\tau_1) \) (unless the edge between \( J_1 \# \tau_1 \) and \( J_2 \# \tau_2 \) is RSU, in which case the energy stays the same).

**Proof.** Within a segment we only need to case on the following three recursion pathways:

- **RSU**: \( \tau_2 = \tau_1 \), and the energy stays the same.

- **RSL**: Let \( \tau_1 = p.t \). S-_LOWER shows that \( \Delta \Sigma \Gamma S \vdash p : \tau_p, \Delta \Sigma \Gamma S \vdash \tau_p < n.r \), and \( \Delta \Sigma \Gamma S \vdash n.r \exists_x \text{type } t \geq \beta_t r_t \). There are three places the bound \( \beta_t r_t \) could come from:
  - Case - \( t \) is from \( r \): The energy of \( p.t \) already includes the energy from \( n.r \), which includes the energy of \( \beta_t r_t \).
  - Case - \( t \) is from \( r_1 \): The energy of \( r_1 \) already includes the energy from \( \beta_t r_t \).
  - Case - \( t \) is not refined in either \( r \) of \( r_1 \): \( \beta_t r_t \) came from \( \Delta \). The energy measure \( E_{\Delta \Sigma \Gamma S}(\lfloor p/x \rfloor \beta_t r_t) \) is the sum of the energies of all types in \( \lfloor p/x \rfloor \beta_t r_t \). Observe that since \( n::t \) is lower bounded, any pseudotype that bound it are reachable from \( n::t \) without going through shapes. Therefore, \( M_{\Delta \Sigma}(n::t) = \sum_{x.t' \in \beta_t r_t} M_{\Delta \Sigma}(n::t'), \) and \( A_{\Delta \Sigma}(n::t) = 1+\sum_{x.t' \in \beta_t r_t} A_{\Delta \Sigma}(n::t') + \sum_{n' \in \tau} E_{\Delta \Sigma \Gamma S}(n') \). This means the energy-calculating equation of \( p.t \) already subsumed all the energy from its bound.

In all cases, \( E_{\Delta \Sigma \Gamma S}(\lfloor p/x \rfloor \beta_t r_t) \) is no greater than the energy contributed by one part of \( p.t r_1 \). Therefore, total energy decreases for the RHS after RSL.

- **GSS**: \( \tau_2 \) is a bound in the refinement part of \( \text{tau}_1 \), which means \( E_{\Delta \Sigma \Gamma S}(\tau_2) \) is strictly smaller since it contains strictly less types.

\[ \square \]

The same is true for the left-hand side, but since we allow certain cyclic dependencies to exist through shapes, the LHS energy can increase when introducing a shape.

**Definition 10.** A path-dependent base type \( p.t \) is considered a shape if \( \Delta \Sigma \Gamma S \vdash p : \tau_p, \Delta \Sigma \Gamma S \vdash \tau_p < n.r, \) and \( n::t \) is marked as a shape in the subtype dependency graph.
Theorem 5. Given a subtype derivation under the contexts $\Delta \Sigma \Gamma S$ and one of its lineages $\mathcal{L}$, if $J_1 \# \tau_1$ is the parent of $J_2 \# \tau_2$ in the same segment in $\mathcal{L}$ and $\tau_2$ is on the LHS of $J_2$, then $E_{\Delta \Sigma \Gamma S}(\tau_2) < E_{\Delta \Sigma \Gamma S}(\tau_1)$ unless 1) $\tau_2$ contains a shape, or 2) the edge from $J_1 \# \tau_1$ to $J_2 \# \tau_2$ is RSL, in which case the energy stays the same.

Proof. Within a segment we only need to case on the following three recursion pathways:

- **RSL**: $\tau_2 = \tau_1$, and the energy stays the same.

- **RSU**: Let $\tau_1 = p.t r_1$. $S$-$\text{UPPER}$ shows that $\Delta \Sigma \Gamma S \vdash p : \tau_p$, $\Delta \Sigma \Gamma S \vdash \tau_p < n r$, and $\Delta \Sigma \Gamma S \vdash n r \ni \exists_x \text{type } t \subseteq \beta_t r_t$. There are three places the bound $\beta_t r_t$ could come from:
  - Case - $t$ is from $r$: The energy of $p.t$ already includes the energy from $n r$, which includes the energy of $\beta_t r_t$.
  - Case - $t$ is from $r_1$: The energy of $r_1$ already includes the energy from $\beta_t r_t$.
  - Case - $t$ is not refined in either $r$ of $r_1$: $\beta_t r_t$ came from $\Delta$. The energy measure $E_{\Delta \Sigma \Gamma S}([p/x]\beta_t r_t)$ is the sum of the energies of all types in $[p/x]\beta_t r_t$.
    * If $[p/x]\beta_t r_t$ does not contain a shape: Similar to the RHS proof, if the bound does not contain any shapes, then all type members and names that are in this bound are already included in the energy calculation of $p.t$.
    * If $[p/x]\beta_t r_t$ contains a shape: Then the energy calculation of $p.t$ only included the energies from the types that are not refining the shape. If the shape does not refine anything, the energy of the entire $[p/x]\beta_t r_t$ is still covered by $p.t$, but if the shape is refined, the LHS may incur an increase in energy.

In all cases other than when $[p/x]\beta_t r_t$ contains a refined shape, $E_{\Delta \Sigma \Gamma S}([p/x]\beta_t r_t)$ is no greater than the energy contributed by one part of $p.t r_1$.

- **GSS**: $\tau_2$ is a bound in the refinement part of $\text{tau}_1$, which means $E_{\Delta \Sigma \Gamma S}(\tau_2)$ is strictly smaller since it contains strictly less types.
An energy increase on the LHS is fine because the proof shows that to do so, a path-dependent shape was encountered on the LHS, and it requires an S-STRUCT for both lineages in order to progress. However, this means the RHS base type was also a shape. Note that shapes rarely appear as base type on the RHS within a segment:

**Theorem 6.** Given a subtype derivation under the contexts $\Delta \Sigma \Gamma S$ and one of its lineages $L$, if $J_1 \# \tau_1$ is the parent of $J_2 \# \tau_2$ within a segment in $L$ (and $\tau_1 \neq \tau_2$), and $\tau_2$ is on the RHS of $J_2$, and $\tau_2$ is a shape, then the edge from $J_1 \# \tau_1$ to $J_2 \# \tau_2$ must be GSS and $\tau_1$ must also be on the RHS of $J_1$.

**Proof.** Within a segment we only need to case on the following three recursion pathways:

- **RSU:** $\tau_2 = \tau_1$, the RHS type does not change.

- **RSL:** $\tau_2$ is the base type of the lower bound of $\tau_1$. However, since shapes can never be used after a lower bound, $\tau_2$ is not a shape.

- **GSS:**
  - If recursion is via SS-LOWER: $\tau_2$ is the base type of a lower bound in the refinement of $\tau_1$. Therefore, $\tau_2$ cannot be a shape.
  - If recursion is via SS-UPPER: $\tau_2$ is the base type of an upper bound in the refinement of $\tau_1$. It is possible that it is a shape.
  - If recursion is via SS-EXACT: both of the above will happen, though in two separate branches of the lineage (i.e. reduces to the above two cases).

As a result, the only shapes that can appear on the RHS as base types are the ones that were already in the refinement on the RHS. We know every time the LHS increases its energy by going through a shape, the RHS loses a shape. However, even though the LHS can replenish energy, the RHS cannot replenish its number of shapes. When the RHS runs out of shapes, it will no longer be able to aid in the energy increase of the LHS by performing S-STRUCT. Therefore, any infinite derivation cannot be entirely within one segment. Both lineages must eventually go through a S-NAME.
S-NAME can trigger two kinds of recursions: one through the green path (GSN), one through the blue path (BSN). Denote the types on the LHS and RHS of the judgment immediately before S-NAME as \( n_1 r_1 \) and \( n_2 r_2 \), respectively.

First, consider BSN recursions. BSN recursion is when S-NAME calls into SN-TRANS, which in turn recurses by replacing the RHS with types defined on the edges of the nominal subtyping graph. After BSN, the old left-lineage’s energy decreases since we recursed on its refinement members (i.e. strictly smaller refinement tree). The old right-lineage turns into a completely new type, but it contains only static types. In fact, the energy of the old right-lineage also decreased due to the way energy is defined on named types. The old name type \( n_2 \) has a greater energy than the entire new RHS type.

BSN has a great impact on the right-lineage. It contains only static types, and the only shapes it contain are not refined (due to material-shape restrictions). This means the only recursions it and its paired lineage can perform are S-NAME and S-UPPER. Without S-STRUCT, the LHS can no longer increase energy indefinitely. Therefore, any derivation that passes through BSN will monotonically decrease its energy until the RHS type becomes either top, bottom, a path-dependent type that is not lower-bounded, or a named type that is not declared to be the supertype of any other name type, all of which ends the derivation.

Now consider GSN recursions. GSN recursion is when S-NAME recursively checks subtyping between the commonly refined members of both sides. It behaves exactly like an S-STRUCT: There are no replenishing of energies or shapes on either side. Instead, the energy of both sides decrease because the recursion is on a proper subtree of the original type’s refinement tree. This means the depth of the refinement trees of types on both sides strictly decrease every time recursion passes through GSN. Recursion continues as before, which means any infinite derivation will eventually visit S-NAME again. As shown earlier, the branch that goes through BSN is determined to end. The branch that goes through GSN again will decrease in depth. Since no available recursion paths (other than BSN) can increase the depth of any type, recursion eventually stops when there are no refinements on the RHS type and GSN no longer occurs.

Due to the combined result of energy decreasing, number of RHS shapes decreasing, and depth decreasing, no lineage can have infinite length. All subtype derivations must eventually terminate.
Chapter 5

Expressiveness

This chapter discusses the expressiveness of Nominal Wyvern by showing the encoding of some of the important features of both object-oriented languages and functional languages that we identified. Section 5.1 introduces the syntax sugar that will be used as shorthand in this section. Sections 5.2 through 5.4 shows how some of the common patterns expressible in DOT are encoded in Nominal Wyvern. Sections 5.5 and 5.6 show how Nominal Wyvern can be compatible with common patterns in both functional and object-oriented programming languages.

5.1 Syntax Sugar

For brevity, this chapter makes use of the following syntax sugar for expressing common patterns.

- **Omitting self-variables**
  The self-variable that immediately follows the open curly brace of name declarations and new object definitions may be omitted if they are never referred to in the structure. Desugaring would just be adding a fresh identifier as the self-variable wherever it was omitted.

- **Omitting val types during object creation**
  In new expressions, the type annotation on val can be omitted if they are identical to the type of the val defined by the type (or its upper bound) that is being created.
This is always known during `new` because of the need to check well-formedness. Desugaring would just be adding back the required type of the `val`.

- **Inline member declarations**
  Members of named type definitions ($\sigma$), refinements ($\delta$), and new object definitions ($d$) are allowed to be separated with commas to make it clearer to the reader when placed on a single line. This can be easily desugared by removing the commas.

- **Inline expressions**
  The abstract syntax requires the argument to methods be pre-bound to a variable. This is relaxed to allow using any arbitrary expression directly as the argument to a method. The desugaring would be moving the expression into a `let` expression (with a fresh variable) that wraps the method call, and instead invoking the method with the fresh variable.

- **Multiple method arguments**
  The abstract syntax only allows one argument in method declarations. This is obviously not practical, but it is also not as limiting as it appears to be. Since all expressions are objects, multiple arguments are simply considered a new object that encapsulates them. For the examples shown in this chapter, the type of a function argument is allowed to be a structural type with a self-variable ($\{x \Rightarrow \sigma\}$). Desugaring would be declaring the structure as a fresh name in the top level and then using that name in the method signature. Any references to type members of the parent type are swapped for either $\top$ or $\bot$ when moved outside. The use site subsequently refines the fresh name type with the referenced type members.
  This serves as a useful feature for a concrete syntax since it does not break the theme of semantic separation of Nominal Wyvern: The arguments to a function should be considered to be defined by the function itself. It would therefore be perfectly fine if each method also declared a named type `$\langle \text{method\_ID} \rangle$ _arg`. Therefore, to spare the need to separately declare each such type, they are declared together with the method signatures.

- **Passing new objects to methods**
  Since the argument to a method can now be of an anonymous named type, it is impossible to `new` such an object. As a result, we introduce a syntax sugar that allows the creation of a new object by simply defining a structure ($\{x \Rightarrow d\}$) to be
passed to any method whose argument type is an anonymous name. The desugaring for this is paired with the desugaring of the method signature: the anonymous named type is declared with a name, and the object is re-defined with a \texttt{new} expression using this name.

5.2 Basic Path-Dependent Types

Since Nominal Wyvern is based on DOT, the first example re-implements the Library example of Listing 2.1 and Listing 2.2 in Nominal Wyvern to show that it preserves the abstraction boundaries and provides the same guarantees.

```plaintext
// assume a pre-defined String type. String objects are created // directly with string literals.
name Unit {}

name GenericBook {}
name Textbook {}
subtype Textbook <= GenericBook
name Fiction {}
subtype Fiction <= GenericBook

name Library {l =>
type Book
fun borrow : String title -> l.Book
fun renew : l.Book item -> Unit
}

name SchoolLibrary {l =>
type Book <= Textbook
fun borrow : String title -> l.Book
fun renew : l.Book item -> Unit
}
subtype SchoolLibrary <= Library

name Utils {
def study : Textbook item -> Unit
}

let hunt = new Library {l =>
```

53
The types of books are empty structures whose names give them meaning. Two types of libraries are defined. Note that the name declarations are equivalent to interface declarations in other languages, therefore they do not contain implementation. Instead, the implementation is given to the particular object at creation time with `new`.

The guarantees provided by DOT are preserved in Nominal Wyvern:

- A book can only be renewed from the library it was borrowed from: Each library object $x$ carries with it a type $x$.Book that is hidden from the outside. In this case, the object csapp has type sorrells.Book (derived directly from the signature of the borrow() method of SchoolLibrary, the type of sorrells). Therefore, it is incompatible with the input type of hunt.renew(), which requires the argument to be of the entirely opaque type hunt.Book.

1This difference is discussed more in depth in section 5.6
• Only a textbook can be “studied”: Even though the exact type of sorrells.Book is unknown, SchoolLibrary provides the hint that its book type is a subtype of TextBook. As a result, it fits the requirement of util.study(). In contrast, since hunt.Book is completely abstract, it is not safe to pass into util.study().

Since Nominal Wyvern supports refinements, one can avoid the need to declare a specific library type if the difference can be succinctly represented by a refinement (and preferably only when the new type is not used as a common concept in the code). For example, the sorrells library can be written without a SchoolLibrary:

```haskell
... let sorrells = new Library {type Book <= Textbook} {l =>
  type Book = Textbook
  fun borrow : String title -> l.Book = new Textbook {}
  fun renew : l.Book item -> Unit = new Unit {} } in ...
```

Listing 5.2: Expressing SchoolLibrary with a refinement

The underlined type declared in the new expression denotes the type that this new object has. The Book type is refined so that books borrowed here can be studied. Note that it does not need to match the exact type declared inside the definition. This allows a more fine-tuned approach to expressing the degree of abstraction desired.

5.3 F-Bounded Polymorphism

Recall from section 2.2 that one of the benefits of combining subtype polymorphism with parametric polymorphism is the ability to express f-bounded polymorphism.

5.3.1 Positive Recursion

Positive recursion refers to when the recursive type variable is at an “output” position, or covariant. For example, since the clone() method in Figure 2.3 returns the parameterized type, the Cloneable class is considered a positive recursion. When positive recursive usages are encoded in mainstream object-oriented languages that do not support
f-bounded polymorphism, the output type is usually the most general type, and a dynamic cast is performed to get back the original type. With f-bounded polymorphism, the type of the output can be guaranteed statically.

The example in the original paper ([Canning et al.][1989]) was able to express a type \( t \) as "movable" by bounding it with a special constructor \( F\text{-Movable}\[t\] that represents a type with a \texttt{move()} method. This can be expressed in Nominal Wyvern with subtyping.

```plaintext
1 // assume pre-defined Real type with "+" operator
2 name RealPair {
3     val l : Real
4     val r : Real
5 }
6
7 name Movable {m =>
8     type t <= T
9     fun move : RealPair amount -> m.t
10 }

11 name Point {p =>
12     type t <= Point
13     val x : Real
14     val y : Real
15     fun move : RealPair amount -> p.t
16 }
17
18 subtype Point <: Movable
19 // constructor for points
20 name PointCons {pc =>
21     fun create : RealPair pos -> Point
22 }

23 name Utils {
24     // arbitrarily translate any movable object
25     fun translate : {x =>
26         type t <= Movable {type t <= x.t}
27         val obj : x.t
28     } arg -> arg.t
29 }

30 let utils = new Utils {
31     // translate any movable object by 1.0 (:Real) in both directions
32     fun translate : {x =>
33         type t <= Movable {type t <= x.t}
34         val obj : x.t
35     } arg -> arg.t
36 }
```
Listing 5.3: F-Movable example in Nominal Wyvern

Any structure that structurally satisfies the Movable interface and semantically supports such a move operation may subtype Movable. As a result, it will be allowed to be passed to the `utils.translate` method to get a translated version of itself.

Listing 5.4: More movable types
5.3.2 Negative Recursion

In contrast to positive recursion, negative recursion is when the parameterized type is an input to a method, or contravariant. One popular use of this is the built-in equals() methods of Object in Java. Traditionally in Java, any object that wants to override the equals method needs to put a boilerplate at the beginning to make sure the object that is passed in is indeed of the same type as the parent type. This is because overridden methods have to preserve the original signature. Thus, all equals() methods takes in a generic Object. With f-bounded polymorphism, this boilerplate can be checked by the type system. For Nominal Wyvern, this is illustrated as the motivating example of chapter 3 in section 3.2. The property of having an equals() method is expressed with the named type Equatable.

5.4 Family Polymorphism

Family polymorphism [Ernst, 2001] is useful when subtyping a set of types that are mutually dependent. A classic example is a node type and an edge type. The types are inter-dependent because a node references its incident edges, and an edge references its two endpoint nodes. To create a specific type of graph, one may wish to subtype both node and edge types. Without variances, the subtypes still refer to the general node and edge types, and dynamic checks have to be performed. This can be made statically safe with path-dependent types.

```plaintext
// assume pre-defined type Bool with constructors "true" and "false"
name Unit {}

name Node {n =>
  type e <= Edge
  // checks if edge is incident on this node
  fun touches : n.e edge -> Bool
}

name Edge {e =>
  type n <= Node
  // the two endpoints of this edge
  val l : e.n
  val r : e.n
}
```
Listing 5.5: Family polymorphism: general nodes and edges

The example above sets up the general Node and Edge types for creating a general graph. Now we can subtype the two inter-dependent types to create the OnOffGraph from [Ernst 2001]. (Identical structures are omitted below for conciseness).

```plaintext
// assume pre-defined type Bool with constructors "true" and "false"
name Unit {}
name Node {...}
name Edge {...}
```
// constructors for both types

name Graph {...}

name OnOffNode {n =>
  type e <= OnOffEdge
  fun touches : n.e edge -> Bool
}

name OnOffEdge {e =>
  type n <= OnOffNode
  val enabled : Bool // each edge can be on or off
  val l : e.n
  val r : e.n
}

subtype OnOffNode <: Node
subtype OnOffEdge <: Edge

name Utils {u =>
  fun build : {arg =>
    val g : Graph
    val a : arg.g.n
    val b : arg.g.n
  } arg -> arg.g.e
}

let g = new Graph {...} in
let oog = new Graph a{type n <= OnOffNode, type e <= OnOffEdge} {c =>
type n = OnOffNode
type e = OnOffEdge
fun createNode : Unit x -> OnOffNode =
  new Node {n =>
    type e = OnOffEdge
    fun touches : OnOffEdge edge -> Bool =
      if[Bool] edge.l = n then edge.enabled
      else if[Bool] edge.r = n then edge.enabled
      else false
  }
fun createEdge : {val a : OnOffNode, val b : OnOffNode} arg
-> OnOffEdge =
  new Edge {e =>
    type n = OnOffNode
    val enabled = true // default to enabled
OnOffEdges can be turned on or off, so they are a special type of edge. Family polymorphism guarantees that `utils.build` will only work if the two types are of the same graph family.

### 5.5 Representing ML Modules

Data abstraction in ML is based on abstract data types (ADT). An ADT encapsulates an abstract type along with operations on the type. This serves as an interface that clients of the ADT can use without depending on (or even having any knowledge of) the implementation details, including what the abstract type actually represents.

Formally, ADTs are modeled with existential types: $\exists t.\tau$, where $\tau$ is typically a product of functions that operate on the abstract type $t$. This can simply be represented in DOT-based systems by an object with a type member. For example, a `natlist` type in System FE (modified from Harper [2016]) is

$$\exists (t, \langle \text{emp} \leftrightarrow t, \text{ins} \leftrightarrow \text{nat} \times t \rightarrow t, \text{rem} \leftrightarrow t \rightarrow (\text{nat} \times t) + \text{void} \rangle)$$
where \(emp, \text{ins}, \text{rem}\) are the empty (i.e., create new), insert, and remove operations on the abstract list type.

Note that the implementation of the functions are coupled with the type. Any two expressions both of type \(\text{natlist}\) will use the same implementation (hidden, but fixed nonetheless). In DOT based systems, an interface to a type does not define its implementation. On this front, objects in DOT, and by extension Nominal Wyvern, are more similar to objects than ADTs. ADTs can still be represented, though, with a pre-defined interface and a “standard” implementation.

```
// assume pre-defined Nat type.
// assume the Option type has the following signature
name Option {o =>
  type elem <= T       // type of the enclosed element
  fun isSome : Unit x -> Bool
  fun get : Unit x -> o.elem
}

name Product {p =>
  type Ta <= T
  type Tb <= T
  val a : Ta
  val b : Tb
}

name NatListInterface {nl =>
  type t <= T
  val emp : nl.t
  fun ins : {val elem : Nat, val list : nl.t} arg -> nl.t
  fun rem : nl.t list ->
    Option {type elem = Product {type Ta = Nat, type Tb = nl.t}}
}

let natlist = new NatListInterface {nl => ...} in ...
```

Listing 5.7: Existential types in Nominal Wyvern

ML modules solve the single implementation problem of ADTs by wrapping them in named structures. Signatures define interfaces, and structures ascribe to signatures and define their own implementation. This is very closely modeled by Nominal Wyvern. Below is a classic \(\text{NatSet}\) example translated into Nominal Wyvern.

```
// assume pre-defined Bool type with constructors "true" and "false"
// NAT_SET interface
```
Listing 5.8: NatSet in Nominal Wyvern

The example follows the SML naming convention. Signatures are named in ALL_CAPS and structures are named in CamelCase. Module NatSet ascribing to signature NAT_SET in SML is translated into object NatSet exhibiting type NAT_SET. The benefit of having objects represent modules is the ability to have first-class modules. A direct result is we get generative functors for free. For the previous example, a function could take in a generic list module (represented as an object with type List), and use it to produce a nat_set module (an object with type NAT_SET).

```ml
// a generic List interface
name LIST {l =>
  type elem <= T // element type
type t <= T
val emp : l.t
fun ins : {val x : l.elem, val L : l.t} arg -> l.t
fun rem : l.t list ->
  Option {type elem = Product {type Ta = l.elem, type Tb = l.t}}
}
```
Listing 5.9: Representing functors as functions

Note that signature modifications with `where` are modeled directly with type refinements.

5.6 Object-Oriented Programming

One of the main differences between pure objects in object-oriented programming (OOP) languages and ADTs is how each paradigm relates interfaces to implementations. The interface of an object type is defined separately from its implementation, whereas the implementation of the functions in an ADT is part of its type. While modules allowed the separation of the interface and implementation via signatures and structures, it is not able to overcome the problem of having the implementation tied to the type it provides. Even if two modules both ascribe to the `List` signature, they cannot operate on each other’s list type. This is due to the internal need to unpack abstract types when operating on them, which only the type-providing module can do. Objects, however, do not provide any types. They instead provide implementations for a common type with a common interface. In fact, multiple objects of the same type can have wildly different implementations. Yet they can still interact with each other with no regard to the internal differences since, instead of unpacking the implementation type, they only rely on dynamically dispatched method calls over the common interface. This added interoperability contributes to the success of
The real world’s version of OOP languages paints a different picture than just described. In the aforementioned “pure” OOP system, interoperability is enabled by autognosis [Cook, 2009], or not caring about the implementation of other objects: An object can only be interacted with over its public interface, which is considered its type. In contrast, popular OOP languages such as Java and C++ are heavily based on classes. Instead of knowing only about an object’s type/interface, we can also know about its class, which reveals information about its implementation. This additional information breaches autognosis, making different parts of a system more inter-dependent than in pure OOP languages.

Nominal Wyvern’s semantic separation naturally supports a pure OOP approach: Named types serve as interface definitions, and objects created from named types serve as constructors, or “classes”. This way, the syntax guarantees interfaces are not tied to any implementation, and classes are syntactically different constructs than types. Classes are thus able to serve as pure organizers of implementations. The following listing translates the sets example from Cook [2009] into Nominal Wyvern. In Cook’s paper, ISet defines the interface for sets, while the classes are simply constructor functions. Once created, an object is no longer associated with its constructing class, and can be freely used with objects created from other classes.

```plaintext
// assume pre-defined Int and Bool types.
// Int type has built-in constructors from literals, and an equals() method. Bool type has built-in constructors "true" and "false", and binary operator "||" for logical or.

// interface for sets
name ISet {s =>
  fun isEmpty() : Bool
  fun contains(i: Int) : Bool
  fun insert(i: Int) : ISet
  fun union(s: ISet) : ISet
}

// define classes/constructors
name SET_CONS {c =>
  fun Empty() : ISet
  fun Insert(s: ISet, n: ISet) : ISet
  fun Union(s1: ISet, s2: ISet) : ISet
}
```
let Set = new SET_CONS {c =>
    def Empty() =
        new ISet {z =>
            def isEmpty() = true
            def contains(i: Int) = false
            def insert(i: Int) = c.Insert(z, i)
            def union(s: ISet) = s
        }
    def Insert(s: ISet, n: Int) =
        if [Bool] s.contains(n) = true then s else
        new ISet {z =>
            def isEmpty() = false
            def contains(i: Int) = (i.equals(n)) || (s.contains(i))
            def insert(i: Int) = c.Insert(z, i)
            def union(s: ISet) = c.Union(z, s)
        }
    def Union(s1: ISet, s2: ISet) =
        new ISet {z =>
            def isEmpty() = s1.isEmpty() || s2.isEmpty()
            def contains(i: Int) = (s1.contains(i)) || (s2.contains(i))
            def insert(i: Int) = c.Insert(z, i)
            def union(s: ISet) = c.Union(z, s)
        }
} in

let s1 = Set.Empty() in  // {}
let s2 = Set.Insert(s1, 1) in  // {1}
let s3 = s1.insert(2) in  // {2}
let s4 = Set.union(s2, s3) in  // {1,2}
...

Listing 5.10: Pure OOP in Nominal Wyvern

5.6.1 Mixing functional and OOP

The following listing shows an interesting combined usage of objects and functional modules. The pair type is used to store two objects of the same generic type. Like modules, the PAIR_MOD interface provides a p type that only the providing module can open. However,
the provided type can act like an object in that it is self-contained and thus interoperable. It can be used like any PAIR without regard to who constructed it (although its constructor can be easily accessed via its class member.)

```ocaml
// interface for pairs
name PAIR {p =>
  type t <= T    // type of the elements
  val class : PAIR_MOD   // reference to its class
  fun l : Unit x -> arg.t
  fun r : Unit x -> arg.t
}
// module providing pair types
name PAIR_MOD {c =>
  type p <= PAIR
  fun create :
    {arg => type t
     val left : arg.t
     val right : arg.t} arg
    -> c.impl {type t = arg.t}
  fun l : c.impl arg -> arg.t
  fun r : c.impl arg -> arg.t
}
// define a custom implementation
// ValPair uses two vals to store its info
name ValPair {p =>
  type t
  val class : PAIR_MOD
  fun l : Unit x -> arg.t
  fun r : Unit x -> arg.t
  val a : p.t
  val b : p.t
}
subtype ValPair <: PAIR

let ValPairMod = new PAIR_MOD {c =>
  type p = ValPair
  def create {arg => type t
               val left : arg.t
               val right : arg.t} arg
    -> c.p {type t = arg.t}
```
new ValPair {type t = arg.t} {p =>

type t = arg.t
val class = c
val a = arg.left
val b = arg.right
fun l : Unit x -> arg.t = c.l(p)
fun r : Unit x -> arg.t = c.r(p)
}
def l : c.p arg -> arg.t = arg.a
def r : c.p arg -> arg.t = arg.b
} in
let origin = ValPairMod.create(
{type t = Int, val left = 0, val right = 0}) in
let zero1 = origin.l() in // OO
let zero2 = ValPairMod.l(origin) // functional

Listing 5.11: Mixing OOP and FP in Nominal Wyvern
Chapter 6

Conclusion and Future Work

This thesis presents Nominal Wyvern, a new core type system for Wyvern based on the DOT calculus. Nominal Wyvern achieves a higher degree of nominality in a DOT-based system by semantically separating the definition of structures and their subtype relations from arbitrary type refinements and the declaration of type bounds. This contributes to a system with more explicit meanings and relations, useful for both human readers to reason about and programming tools to refer to. Nominality also helps with achieving subtype decidability. In line with the theme of semantic separation, Nominal Wyvern adapts material-shape separation so that decidability results from an intuitive separation of types with different roles. This contributes to a restriction that is more easily understandable and articulable. The resulting system preserves the ability to express common patterns expressible with DOT, at the same time allowing for patterns that will be familiar to programmers already used to traditional functional or object-oriented programming languages.

Some further areas of study are discussed below.

- More flexible width subtyping. Nominal Wyvern separated out depth subtyping from width subtyping, making the formal automatic and the latter disallowed (have to define new names). One may argue that width subtyping has more to do with what the structure contains, which is more integral to the semantic meaning of the object than the absolute types of each member. If a member needs to be a more specific type, there is likely no need for an altogether semantically different structural name. In fact, the type system additionally constraints how one can refine a type by disal-
lowing additional self-references in refinements (since that would likely change the semantics of the structure). However, there are likely exceptions to both: there may be cases where a width-refinement does not add much to the structure semantically, and there may be cases where a depth-refinement significantly changes the meaning of a structure. There are two potential solutions worth looking into:

– Allow certain kinds of structural width subtyping. The downside is too much freedom risks taking away the benefits of the nominal system.

– Encourage programmers to write a new structure when a certain depth refinement significantly alters the meaning.

Both are subjective and warrant further study.

• Bringing back uniformity. One of the benefits of the DOT system is that the entire program is a first class value. In contrast, Nominal Wyvern breaks this uniformity by introducing a set of second-class declarations at the top level. It would be nice to bring back the uniformity of DOT while maintaining the benefits of nominality presented in this thesis. One potential solution is integrating the named type definitions and explicit subtype declarations as members of object. A separate notation would need to be created to refer to the named types of objects, as the binary typing approach (section 3.3) makes it necessary to keep type members and named types separate.
Bibliography


Nada Amin, Adriaan Moors, and Martin Odersky. Dependent object types - towards a foundation for scala’s type system. In 19th International Workshop on Foundations of Object-Oriented Languages, 2012. 2.1


William R. Cook. On understanding data abstraction, revisited. In Proceedings of


