Chapter 1

Lambda calculus

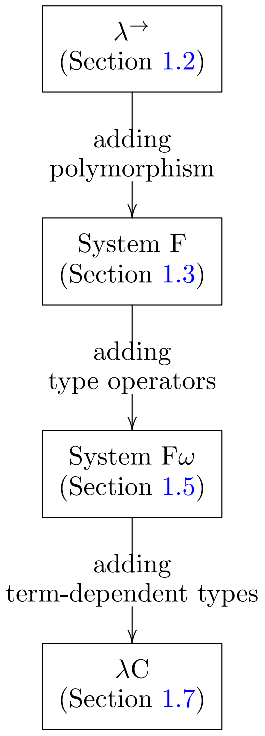
The *lambda calculus* serves as the basis of most functional programming languages. More accurately, we might say that functional programming languages are based on the *lambda calculi* (plural), since there are many variants of lambda calculus. In this chapter we'll introduce four of these variants, starting with the simply typed lambda calculus (Section 1.2), moving on to System F, the polymorphic lambda calculus (Section 1.3) and System Fω, a variant of System F with type operators (Section 1.5) before concluding with the expressive calculus of constructions (λC).

None of the calculi in this chapter are particularly suitable for programming. The simpler systems are insufficiently expressive — for example, they have no facilities for defining data types. In contrast, System Fω and λC, the last systems that we'll look at, have excellent abstraction facilities but are rather too verbose for writing programs. In OCaml, the main language we'll use in the rest of the course, we might define the function that composes two functions as follows:

```
fun f g x→f (g x)
```

We can define the same function in System Fω, but the simple logic is rather lost in a mass of annotations:

```
Λα::*.
   Λβ::*.
      Λγ::*.
         λf:α → β.
            λg:γ → α.
               λx:γ.f (g x)
```



$\lambda^{\rightarrow}$ (Section 1.2)
adding polymorphism
System F (Section 1.3)
adding type operators
System Fω (Section 1.5)
adding term-dependent types
λC (Section 1.7)

Figure 1.1: Chapter plan

If these systems are unsuitable for programming, why should we study them at all? One reason is that many — perhaps most — of the features available in modern functional programming languages have straightforward translations into System Fω or other powerful variants of the lambda calculus[1]. The calculi in this chapter will give us a simple and uniform framework for understanding many language features and programming patterns in the rest of the course that might otherwise seem complex and arbitrary.

The foundational nature of the more expressive calculi means that we'll see them in a variety of roles during the course, including:

- as an elaboration language for type inference (lecture 2).
- as proof systems for reasoning with propositional and predicate logic (lecture 3).
- as the background for parametricity properties (lectures 4-5)
- underlying and motivating higher-order polymorphism in languages such as OCaml (lectures 4-5)
- as the elaboration language for modules (lectures 4-5)
- as the core calculus for GADTs and indexed data (lectures 7-8)
- as a foundation for multi-stage programming (lectures 14-16)

1.1 Typing rules

We'll present the various languages in this chapter using *inference rules*, which take the following form:

$$\frac{\text{premise 1} \qquad \text{premise 2} \qquad \ldots \qquad \text{premise N}}{\text{conclusion}}\ \text{rule name}$$

This is read as follows: from proofs of *premise 1*, *premise 2*, … *premise n* we may obtain a proof of *conclusion* using the rule *rule name*. Here is an example rule, representing one of the twenty-four Aristotelian syllogisms:

[1] Some implementations of functional languages, such as the Glasgow Haskell Compiler, use a variant of System Fω as an internal language into which source programs are translated as an intermediate step during compilation to machine code. OCaml doesn't use this strategy, but understanding how we might translate OCaml programs to System Fω still gives us a useful conceptual model.

$$\frac{\text{all } M \text{ are } P \qquad \text{all } S \text{ are } M}{\text{all } S \text{ are } P}\ \text{modus barbara}$$

The upper-case letters M, P, and S in this rule are *meta-variables*: we may replace them with valid terms to obtain an *instance* of the rule. For example, we might instantiate the rule to

$$\frac{\text{all programs are buggy} \qquad \text{all functional programs are programs}}{\text{all functional programs are buggy}}\ \text{modus barbara}$$

The rules that we will see in this chapter have some additional structure: each statement, whether a premise or a conclusion, typically involves a *context*, a *term* and a *classification*. For example, here is the rule *→-elim*[2]:

$$\frac{\begin{array}{c}\Gamma \vdash M : A \to B \\ \Gamma \vdash N : A\end{array}}{\Gamma \vdash M\ N : B}\ \to\text{-elim}$$

Both the premises and the conclusion take the form $\Gamma \vdash M : A$, which we can read "In the context Γ, M has type A." It is important to note that each occurrence of Γ refers to the same context (and similarly for M, N, A, and B): the *→-elim* rule says that if the term M has type $A \to B$ in a context Γ, and the term N has type A in the same context Γ, then the term $M\ N$ has type B in the same context Γ. As before, Γ, M, etc. are metavariables which we can instantiate with particular contexts and terms to obtain facts about particular programs.

1.2 Simply typed lambda calculus

We'll start by looking at a minimal language. The simply typed lambda calculus lies at the core of typed functional languages such as OCaml. Every typed lambda calculus program is (after a few straightforward syntactic changes) a valid program in OCaml, and every non-trivial OCaml program is built from the constructs of the typed lambda calculus along with some "extra stuff" — polymorphism, data types, modules, and so on — which we will cover in later chapters.

The name "simply typed lambda calculus" is rather unwieldy, so we'll use the traditional and shorter name $\lambda^{\to}$. The arrow $\to$ indicates the centrality of function types $A \to B$.

Let's start by looking at some simple $\lambda^{\to}$ programs.

- The simplest complete program is the **identity function**, which simply returns its argument. Since $\lambda^{\to}$ is not polymorphic we need a separate identity function for each type. At a given type A the identity function is written as follows:

[2]We will often arrange premises vertically rather than horizontally.

λ`x:A.x`

In OCaml we write

fun `x` $\rightarrow$ `x`

or, if we'd like to be explicit about the type of the argument,

fun `(x:a)` $\rightarrow$ `x`

- The **compose function** corresponding to the mathematical composition of two functions is written

  λ`f:`$B \rightarrow C$`.`λ`g:`$A \rightarrow B$`.`λ`x:`A`.f (g x)`

  for types A, B and C. In OCaml we write

  fun `f g x` $\rightarrow$ `f (g x)`

Although simple, these examples illustrate all the elements of $\lambda^{\rightarrow}$: types, variables, abstractions, applications and parenthesization. We now turn to a more formal definition of the calculus. For uniformity we will present both the grammar of the language and the type rules[3] as inference rules.

The elements of the lambda calculi described here are divided into three "sorts":

- **terms**, such as the function application `f p`. We use the letters L, M and N (sometimes subscripted: L_1, L_2, etc.) as metavariables that range over terms, so that we can write statements about arbitrary terms: "For any terms M and N ...".

- **types**, such as the function type $\mathcal{B} \rightarrow \mathcal{B}$. The metavariables A and B range over expressions that form types. We write $M : A$ to say that the term M has the type A.

- **kinds**, which you can think of as the types of type expressions. The metavariable K ranges over kinds. We write $T :: K$ to say that the type expression T has the kind K.

[3]Here and throughout this chapter we will focus almost exclusively on the grammar and typing rules of the language — the so-called static semantics — and treat the dynamic semantics (evaluation rules) as "obvious". There are lots of texts that cover the details of evaluation, if you're interested; Pierce's book is a good place to start.

Kinds in $\lambda^{\to}$ Rules that introduce kinds take the following form:

$$K \text{ is a kind}$$

Kinds play little part in $\lambda^{\to}$, so their structure is trivial. The later calculi will enrich the kind structure. For now there is a single kind, called $*$.

$$\frac{}{* \text{ is a kind}}\ *\text{-kind}$$

Kinding rules in $\lambda^{\to}$ The set of types is defined inductively using rules of the form

$$\Gamma \vdash A :: K$$

which you can read as "type A has kind K in environment Γ. We will have more to say about environments shortly. In $\lambda^{\to}$ there are two kinding rules, which describe how to form types:

$$\frac{}{\Gamma \vdash \mathcal{B} :: *}\ \text{kind-}\mathcal{B} \qquad\qquad \frac{\Gamma \vdash A :: * \qquad \Gamma \vdash B :: *}{\Gamma \vdash A \to B :: *}\ \text{kind-}\to$$

The kind-$\mathcal{B}$ rule introduces a base type $\mathcal{B}$ of kind $*$. Base types correspond to primitive types available in most programming languages, such as the types `int`, `float`, etc. in OCaml. They will not play a very significant part in the development, but without them we would have no way of constructing types in $\lambda^{\to}$.

The kind-$\to$ rule gives us a way of forming function types $A \to B$ from types A and B. The arrow $\to$ associates rightwards: $A \to B \to C$ means $A \to (B \to C)$, not $(A \to B) \to C$. We'll use parentheses in the obvious way when we need something other than the default associativity, but we won't bother to include the (entirely straightforward) rules showing where parentheses can occur.

Using the rules kind-$\mathcal{B}$ and kind-$\to$ we can form a variety of function types. For example,

- $\mathcal{B}$, the base type.
- $\mathcal{B} \to \mathcal{B}$, the type of functions from $\mathcal{B}$ to $\mathcal{B}$.
- $\mathcal{B} \to (\mathcal{B} \to \mathcal{B})$, the type of functions from $\mathcal{B}$ to functions-from-$\mathcal{B}$-to-$\mathcal{B}$. The expression $\mathcal{B} \to \mathcal{B} \to \mathcal{B}$ has the same meaning.
- $(\mathcal{B} \to \mathcal{B}) \to \mathcal{B}$, the type of functions from functions-from-$\mathcal{B}$-to-$\mathcal{B}$ to $\mathcal{B}$.

Since the syntax of types is described using logical rules we can give formal derivations for particular types. For example, the type $(\mathcal{B} \to \mathcal{B}) \to \mathcal{B}$ can be derived as follows:

$$\dfrac{\dfrac{\dfrac{}{\Gamma \vdash \mathcal{B} :: *}\ \text{kind-}\mathcal{B} \qquad \dfrac{}{\Gamma \vdash \mathcal{B} :: *}\ \text{kind-}\mathcal{B}}{\Gamma \vdash \mathcal{B} \to \mathcal{B} :: *}\ \text{kind-}\to \qquad \dfrac{}{\Gamma \vdash \mathcal{B} :: *}\ \text{kind-}\mathcal{B}}{\Gamma \vdash (\mathcal{B} \to \mathcal{B}) \to \mathcal{B} :: *}\ \text{kind-}\to$$

In practice we will assume that all the types that appear in programs and derivations are well-formed, and will sometimes omit explicit statements of well-formedness in the interest of succinctness.

Environment rules in $\lambda^{\rightarrow}$ Environments associate variables with their classifiers. In $\lambda^{\rightarrow}$ there is only one sort of variables, for terms, and so environments associate term variables with types. Later sections extend the type language with variables, too; environments then additionally map type variables to kinds.

Rules for forming environments have the form

$$\Gamma \text{ is an environment}$$

In $\lambda^{\rightarrow}$ there are two rules for forming environments:

$$\frac{}{\cdot \text{ is an environment}}\ \Gamma\text{-}\cdot \qquad \frac{\Gamma \text{ is an environment} \qquad \Gamma \vdash A :: *}{\Gamma, x{:}A \text{ is an environment}}\ \Gamma\text{-:}$$

The rule Γ-$\cdot$ introduces an empty environment. The rule Γ-: extends an existing environment Γ with a binding $x{:}A$ – that is, it associates the variable x with the type A. (We use the letters x, y, z for variables.) We will be a little informal in our treatment of environments, sometimes viewing them as sequences of bindings and sometimes as sets of bindings. We'll also make various simplifying assumptions; for example, we'll assume that each variable can only occur once in a given environment. With more care it's possible to formalise these assumptions, but the details are unnecessary for our purposes here.

As with types, we can use Γ-$\cdot$ and Γ-: to form a variety of environments. For example,

- The empty environment $\cdot$
- An environment with two variable bindings $\cdot, x{:}\mathcal{B}, f{:}(\mathcal{B} \rightarrow \mathcal{B})$

Typing rules in $\lambda^{\rightarrow}$ The rule Γ-: shows how to add variables to an environment. We'll also need a way to look up variables in an environment. The following rule is the first of the three $\lambda^{\rightarrow}$ typing rules, which have the form $\Gamma \vdash M : A$ (read "the term M has the type A in environment Γ") and describes how to form terms:

$$\frac{x{:}A \in \Gamma}{\Gamma \vdash x : A}\ \text{tvar}$$

The tvar rule makes it possible to type open terms (i.e. terms with free variables). If the environment Γ contains the binding $x{:}A$ then the term x has the type A in Γ.

The remaining two typing rules for $\lambda^{\rightarrow}$ show how to introduce and eliminate terms of function type — that is, how to define and apply functions.

$$\frac{\Gamma, x{:}A \vdash M : B}{\Gamma \vdash \lambda x{:}A.M : A \to B}\ \to\text{-intro} \qquad\qquad \frac{\begin{array}{c}\Gamma \vdash M : A \to B \\ \Gamma \vdash N : A\end{array}}{\Gamma \vdash M\ N : B}\ \to\text{-elim}$$

The introduction rule $\to$-intro shows how to form a term $\lambda x{:}A.M$ of type $A \to B$. You can read the rule as follows: "the term $\lambda x{:}A.M$ has type $A \to B$ in Γ if its body M has type B in Γ extended with $x{:}A$". Since we are *introducing* the $\to$ operator, the rule has the function arrow $\to$ below the line but not above it.

The elimination rule $\to$-elim shows how to apply terms of function type. The environment Γ is the same throughout the rule, reflecting the fact that no variables are bound by the terms involved. Since we are *eliminating* the $\to$ operator, the rule has the function arrow $\to$ above the line but not below it.

The $\to$-intro and $\to$-elim form the first *introduction-elimination pair*[4]. We'll see many more such pairs as we introduce further type and kind constructors.

We illustrate the typing rules by giving derivations for the identity and compose functions (Figures 1.2 and 1.3)[5]

1.2.1 Adding products

Interesting programs typically involve more than functions. Useful programming languages typically provide ways of aggregating data together into larger structures. For example, OCaml offers various forms of variants, tuples and records, besides more elaborate constructs such as modules and objects. The $\lambda^{\to}$ calculus doesn't support any of these: there are no built-in types beyond functions, and its abstraction facilities are too weak to define interesting new constructs. In order to make the language a little more realistic we therefore introduce a built-in type of binary pairs (also called "products").

As before, we'll start by giving some programs that we can write using products:

- An `apply` function, that applies the first component of a pair to the second:

  ```
  λp:(A→B)×A.fst p (snd p)
  ```

  In OCaml we write

  ```
  fun (f,p) → f p
  ```

- A `dup` function that builds a pair with equal components:

[4]It is possible to consider the environment and variable-lookup rules as an introduction and elimination pair for environments, but we won't take this point any further here.

[5]While the derivations may appear complex at first glance, they are constructed mechanically from straightforward applications of the three typing rules for $\lambda^{\to}$. If typing derivations featured extensively in the course we might adopt various conventions to make them simpler to write down, since much of the apparent complexity is just notational overhead.

$$\dfrac{\cdot, x{:}A \vdash x : A}{\cdot \vdash \lambda x{:}A.x : A \to A}\ \to\text{-intro}$$

Figure 1.2: The derivation for the identity function.

$$\dfrac{\dfrac{\dfrac{\dfrac{\cdot, f{:}B \to C, g{:}A \to B, x{:}A \vdash f : B \to C \qquad \dfrac{\begin{array}{c}\cdot, f{:}B \to C, g{:}A \to B, x{:}A \vdash g : A \to B \\ \cdot, f{:}B \to C, g{:}A \to B, x{:}A \vdash x : A\end{array}}{\cdot, f{:}B \to C, g{:}A \to B, x{:}A \vdash g\ x : B}\ \to\text{-elim}}{\cdot, f{:}B \to C, g{:}A \to B, x{:}A \vdash f\ (g\ x) : C}\ \to\text{-elim}}{\cdot, f{:}B \to C, g{:}A \to B \vdash \lambda x{:}A.f\ (g\ x) : A \to C}\ \to\text{-intro}}{\cdot, f{:}B \to C \vdash \lambda g{:}A \to B.\lambda x{:}A.f\ (g\ x) : (A \to B) \to A \to C}\ \to\text{-intro}}{\cdot \vdash \lambda f{:}B \to C.\lambda g{:}A \to B.\lambda x{:}A.f\ (g\ x) : (B \to C) \to (A \to B) \to A \to C}\ \to\text{-intro}$$

Figure 1.3: The derivation for `compose`

λ`x:A.`$\langle$`x,x`$\rangle$

In OCaml we write

fun `x` $\rightarrow$ `(x, x)`

- The (bi)map function for pairs:

  λ`f:A`$\rightarrow$`C.`λ`g.B`$\rightarrow$`C.`λ`p.A`$\times$`B.`$\langle$`f` **fst** `p,g` **snd** `p`$\rangle$

  In OCaml we write

  fun `f g (x,y)` $\rightarrow$ `(f x, g y)`

- The function which swaps the elements of a pair:

  λ`p:A`$\times$`B.`$\langle$**snd** `p,` **fst** `p`$\rangle$

  In OCaml we write

  fun `(x,y)` $\rightarrow$ `(y,x)`

Kinding rules for $\times$ There is a new way of forming types $A \times B$, and so we need a new kinding rule.

$$\frac{\Gamma \vdash A :: * \qquad \Gamma \vdash B :: *}{\Gamma \vdash A \times B :: *}\ \text{kind-}\times$$

The kind-$\times$ rule is entirely straightforward: if A and B have kind $*$, then so does $A \times B$.

Typing rules for $\times$ There are three new typing rules:

$$\frac{\begin{array}{c}\Gamma \vdash M : A \\ \Gamma \vdash N : B\end{array}}{\Gamma \vdash \langle M, N\rangle : A \times B}\ \times\text{-intro}$$

$$\frac{\Gamma \vdash M : A \times B}{\Gamma \vdash \ \textbf{fst}\ M : A}\ \times\text{-elim-1}$$

$$\frac{\Gamma \vdash M : A \times B}{\Gamma \vdash \ \textbf{snd}\ M : B}\ \times\text{-elim-2}$$

The $\times$-intro rule shows how to build pairs: a pair $\langle M, N\rangle$ of type $A \times B$ is built from terms M and N of types A and B.

The $\times$-elim-1 and $\times$-elim-2 rules show how to deconstruct pairs. Given a pair M of type $A \times B$, **fst** `M` and **snd** `M` are respectively the first and second elements of the pair. Unlike in OCaml, **fst** and **snd** are "keywords" rather than first-class functions. For particular types A and B we can define abstractions λ`p` :$A \times B$**.fst** `p` and λ`p`:$A \times B$**.snd** `p`, but we do not yet have the polymorphism required to give definitions corresponding to the polymorphic OCaml functions:

```
val fst : 'a * 'b → 'a          val snd : 'a * 'b → 'b
let fst (a, _) = a              let snd (_, b) = b
```

1.2.2 Adding sums

Product types correspond to a simple version of OCaml's records and tuples. We next extend $\lambda^{\to}$ with sum types, which correspond to a simple version of variants.

Here are some programs that we can write with $\lambda^{\to}$ extended with sums:

- The (bi-)map function over sums:

  ```
  λf:A→C.λg:B→C.λs:A+B.case s of x.f x | y.g y
  ```

  In OCaml we write

  ```
  fun f g s → match s with Inl x → f x | Inr y → g y
  ```

- The function of type `A+B→B+A` which swaps the **inl** and **inr** constructors:

  ```
  λs:A+B.case s of x.inr [B] x | y. inl [A] y
  ```

  In OCaml we write

  ```
  function Inl x → Inr x | Inr y → Inl y
  ```

- The function of type `A+A→A` that projects from either side of a sum:

  ```
  λs:A+A.case s of x.x | y.y
  ```

  In OCaml we write

  ```
  function Inl x → x | Inr y → y
  ```

Kinding rules for + The kinding rule for sum types follows the familiar pattern:

$$\frac{\Gamma \vdash A :: * \qquad \Gamma \vdash B :: *}{\Gamma \vdash A + B :: *}\ \text{kind-+}$$

Typing rules for + There are three new typing rules:

$$\frac{\Gamma \vdash M : A}{\Gamma \vdash \mathbf{inl}\ [B]\ M : A + B}\ \text{+-intro-1} \qquad \frac{\Gamma \vdash N : B}{\Gamma \vdash \mathbf{inr}\ [A]\ N : A + B}\ \text{+-intro-2}$$

$$\frac{\begin{array}{c}\Gamma \vdash L : A + B \\ \Gamma, x{:}A \vdash M : C \\ \Gamma, y{:}B \vdash N : C\end{array}}{\Gamma \vdash \textbf{case}\ L\ \textbf{of}\ x.M \mid y.N : \mathrm{C}}\ \text{+-elim}$$

The +-intro-1 and +-intro-2 rules show how to build values of sum type by *injecting* with **inl** or **inr**. In order to maintain the property that each term has a unique type we also require a type argument to **inl** and **inr**[6]

The +-elim rule shows how to deconstruct sums. We can deconstruct sum values L of type $A+B$ if we can deconstruct both the left and the right summand. Given an OCaml variant definition

```
type plus = Inl of a | Inr of b
```

the **case** expression

case L **of** x.M | y.N

corresponds to the OCaml match statement

```
match l with Inl x → m | Inr y → n
```

There is an appealing symmetry between the definitions of products and sums. Products have one introduction rule and two elimination rules; sums have two introduction rules and one elimination rule. We shall consider this point in more detail in lecture 6 (*The Curry-Howard correspondence*).

1.3 System F

The simply typed lambda calculus $\lambda^{\rightarrow}$ captures the essence of programming with functions as first-class values, an essential feature of functional programming languages. Our next calculus, **System F** (also known as the *polymorphic lambda calculus*) captures another fundamental feature of typed functional programming languages like OCaml and Haskell: parametric polymorphism.

We have already seen an example of the problems that arise in languages that lack support for parametric polymorphism. In Section 1.2.1 the **fst** and **snd** operations which project the elements of a pair were introduced as built-in operators with special typing rules. It would be preferable to be able to define **fst** and **snd** using the other features of the language, but it is clear that they cannot be so defined, since $\lambda^{\rightarrow}$ lacks even the facilities necessary to express their types. A similar situation often arises during the development of a programming language: the language is insufficiently expressive to support a feature that is useful or even essential for writing programs. The most pragmatic solution is often to add new built-in operations for common cases, as we have done with products[7]. In this chapter we take the alternative approach of systematically

[6] These kinds of annotations are not needed in OCaml; can you see why?

[7] Some examples from real languages: OCaml has a polymorphic equality operation that works for all first-order types, but that cannot be defined within the language; Haskell's

enriching the core language until it is sufficiently powerful to define new data types directly.

The difficulties caused by the lack of polymorphism go beyond pairs. We introduced $\lambda^{\rightarrow}$ by showing how to write the identity and compose functions and comparing the implementations with the corresponding OCaml code. In fact, the comparison is a little misleading: in OCaml it is possible to define a single identity function and a single compose function that work at all types, whereas in $\lambda^{\rightarrow}$ we must introduce a separate definition for each type. As the following examples show, System F allows us to define the functions in a way that works for all types.

- The polymorphic identity function of type $\forall\alpha{::}*.\alpha \rightarrow \alpha$:

  $\Lambda\alpha{::}*.\lambda\mathtt{x}{:}\alpha.\mathtt{x}$

- The polymorphic compose function of type $\forall\alpha{::}*.\forall\beta{::}*.\forall\gamma{::}*.(\beta \rightarrow \gamma) \rightarrow (\alpha \rightarrow \beta) \rightarrow \alpha \rightarrow \gamma$:

  $\Lambda\alpha{::}*.\Lambda\beta{::}*.\Lambda\gamma{::}*.\lambda\mathtt{f}{:}\beta\rightarrow\gamma.\lambda\mathtt{g}{:}\alpha\rightarrow\beta.\lambda\mathtt{x}{:}\alpha.\mathtt{f}\ (\mathtt{g}\ \mathtt{x})$

- The polymorphic apply function of type $\forall\alpha{::}*.\forall\beta{::}*.(\alpha \rightarrow \beta) \times \alpha \rightarrow \beta$:

  $\Lambda\alpha{::}*.\Lambda\beta{::}*.\lambda\mathtt{p}{:}(\alpha\rightarrow\beta)\times\alpha.\,\mathbf{fst}\ \mathtt{p}\ (\mathbf{snd}\ \mathtt{p})$

To put it another way, we can now use abstraction *within* the calculus ($\forall\alpha ::*.A$) where we previously had to use abstraction *about* the calculus (For all types A ...).

Kinding rules for $\forall$ As the examples show, System F extends $\lambda^{\rightarrow}$ with a type-level operator $\forall\alpha{::}*.-$ that binds a variable α within a particular scope. There are two new kinding rules:

$$\frac{\Gamma, \alpha{::}K \vdash A :: *}{\Gamma \vdash \forall\alpha{::}K.A :: *}\ \text{kind-}\forall \qquad\qquad \frac{\alpha{::}K \in \Gamma}{\Gamma \vdash \alpha :: K}\ \text{tyvar}$$

The kind-$\forall$ rule builds *universal types* $\forall\alpha{::}K.A$. The type $\forall\alpha{::}K.A$ has kind $*$ under an environment Γ if the type A has kind $*$ under Γ extended with an entry for α.

The tyvar rule is a type-level analogue of the tvar rule: it allows type variables to appear within type expressions, making it possible to build open types (i.e. types with free variables). If the environment Γ contains the binding $\alpha{::}K$ then the type α has the kind K in Γ

These rules involve a new kind of variable into the language. *Type variables* are bound by $\forall$ and (as we shall see) by Λ, and can be used in place

`deriving` keyword supports automatically creating instances of a fixed number of built-in type classes; C99 provides a number of type-generic macros which work across numeric types, but does not offer facilities that allow the user to define such macros.

of concrete types in expressions. It is important to distinguish between type variables (for which we write α, β, γ) and metavariables (written A, B, C). We use metavariables when we wish to talk about types without specifying any particular type, but type variables are part of System F itself, and can appear in concrete programs.

Environment rules for ∀ The tyvar rule requires that we extend the definition of environments to support type variable bindings (associations of type variables with their kinds):

$$\frac{\Gamma \text{ is an environment} \qquad K \text{ is a kind}}{\Gamma, \alpha{::}K \text{ is an environment}}\ \Gamma\text{-::}$$

Typing rules for ∀ Since we have a new type constructor ∀, we need a new pair of introduction and elimination rules:

$$\frac{\Gamma, \alpha{::}K \vdash M : A}{\Gamma \vdash \Lambda\alpha{::}K.M : \forall\alpha{::}K.A}\ \forall\text{-intro} \qquad \frac{\Gamma \vdash M : \forall\alpha{::}K.A \qquad \Gamma \vdash B :: K}{\Gamma \vdash M\ [B] : A[\alpha := B]}\ \forall\text{-elim}$$

The ∀-intro rule shows how to build values of type $\forall\alpha{::}K.A$ — that is, polymorphic values. The term $\Lambda\alpha{::}K.M$ has type $\forall\alpha{::}K.A$ in Γ if the body M has the type A in Γ extended with a binding for α.

The ∀-elim rule shows how to use values of polymorphic type via a second form of application: applying a (suitably-typed) term to a *type*. If M has a polymorphic type $\forall\alpha{::}*.A$ then we can apply it to the type B (also written "*instantiate* it at type B") to obtain a term of type $A[\alpha := B]$ — that is, the type A with all free occurrences of α replaced by the type B. (Once again, substitution needs to be carefully defined to avoid inadvertently capturing variables, and once again we omit the details.)

There is nothing in the OCaml language that quite corresponds to the explicit introduction and elimination of polymorphic terms. However, one way to view OCaml's implicit polymorphism is as a syntactic shorthand for (some) System F programs, where constructs corresponding to ∀-intro and ∀-elim are automatically inserted by the type inference algorithm. We will consider this view in more detail in lecture 2 (*Type Inference*) and describe OCaml's alternatives to System F-style polymorphism in lectures 4 & 5 (*Abstraction & Parametricity*).

1.3.1 Adding existentials

For readers of a logical bent the name of the new type operator ∀ is suggestive. Might there be a second family of type operators ∃? It turns out that there is indeed a useful notion of ∃ types: just as ∀ is used for terms which can be instantiated at *any* type, ∃ can be used to form types for which we have *some* implementation, but prefer to leave the details abstract. These *existential types* play several important roles in programming, and we will return to them on various occasions throughout the course. For example,

- There is a close connection between the $\forall$ and $\exists$ operators in logic and the type operators that we introduce here, which we will explore in lecture 6.
- Just as in logic, there is also a close connection between the $\forall$ operator and the $\exists$ operator.
- The types of modules can be viewed as a kind of existential type, as we shall see in lecture 4.
- OCaml's variant types support a form of existential quantification, as we shall see in lecture 4.

It is possible to encode existential types using universal types. For now we will find it more convenient to introduce them directly as an extension to System F.

Kinding rules for $\exists$ Adding existentials involves one new kinding rule, which says that $\exists\alpha{::}K.A$ has kind $*$ if A has kind $*$ in an extended environment:

$$\frac{\Gamma, \alpha{::}K \vdash A :: *}{\Gamma \vdash \exists\alpha{::}K.A :: *}\ \text{kind-}\exists$$

Typing rules for $\exists$ The typing rules for $\exists$ follow the familiar introduction-elimination pattern:

$$\frac{\Gamma \vdash M : A[\alpha := B] \qquad \Gamma \vdash \exists\alpha{::}K.A :: *}{\Gamma \vdash \textbf{pack}\ B, M\ \textbf{as}\ \exists\alpha{::}K.A : \exists\alpha{::}K.A}\ \exists\text{-intro}$$

$$\frac{\begin{array}{c}\Gamma \vdash M : \exists\alpha{::}K.A \qquad \Gamma \vdash C :: * \\ \Gamma, \alpha{::}K, x{:}A \vdash M' : C\end{array}}{\Gamma \vdash \textbf{open}\ M\ \textbf{as}\ \alpha, x\ \textbf{in}\ M' : C}\ \exists\text{-elim}$$

The $\exists$-intro rule shows how to build values of existential type using a new construct, **pack**. A **pack** expression associates two types with a particular term M: if M may be given the type $A[\alpha := B]$ in the environment Γ for suitable A, α and B then we may pack M together with B under the type $\exists\alpha{::}K.A$. It is perhaps easiest to consider the conclusion first: the expression **pack** `B,M` **as** $\exists\alpha{::}K.A$ has the existential type $\exists\alpha{::}K.A$ if replacing every occurrence of α in A with B produces the type of M.

The $\exists$-elim rule shows how to use values of existential type using a new construct **open**. "Opening" a term M with the existential type $\exists\alpha{::}K.A$ involves binding the existential variable α to the type A and a term variable x to the term M within some other expression M'. It is worth paying careful attention to the contexts of the premises and the conclusion. Since α is only in scope for the typing of M' it cannot occur free in the result type of the conclusion C. That is, the existential type is not allowed to "escape" from the body of the **open**.

∃ **examples** Existential types are more complex than the other constructs we have introduced so far, so we will consider several examples. Each of our examples encodes a data type using the constructs available in System F.

1.3.2 Encoding data types in System F

Our first example is a definition of the simplest datatype — that is, the "unit" type with a single constructor and no destructor. OCaml has a built-in unit type, but if it did not we might define its signature as follows:

```
type t
val u : t
```

The following System F expression builds a representation of the unit type using the type of the polymorphic identity function, which also has a single inhabitant:

```
pack (∀α::*.α → α,
      Λα.λa:α.a)
  as ∃u::*.u
```

In the examples that follow we will write `1` to denote the unit type and $\langle\rangle$ to denote its sole inhabitant.

Our next example defines a simple abstract type of booleans. OCaml has a built-in boolean type, but if it did not we might define a signature for bools as follows:

```
type t
val ff : t
val tt : t
val _if_ : t → 'a → 'a → 'a
```

That is: there are two ways of constructing boolean values, `tt` and `ff`, and a branching construct `_if_` which takes a boolean and two alternatives, one of which it returns. We might represent boolean values using a variant type:

```
type boolean =
    False : boolean
  | True : boolean
```

Using `boolean` we can give an implementation of the signature:

```
module Bool :
sig
  type t
  val ff : t
  val tt : t
  val _if_ : t → 'a → 'a → 'a
end =
struct
  type t = boolean
  let ff = False
  let tt = True
  let _if_ cond _then_ _else_ =
    match cond with True → _then_ | False → _else_
end
```

If we ask OCaml to type-check the body of the module, omitting the signature, it produces the following output:

```
sig
  type t = boolean
  val ff : boolean
  val tt : boolean
  val _if_ : boolean → 'a → 'a → 'a
end
```

The relationship between the inferred module type and the signature corresponds closely to the relationship between the supplied existential type and the type of the body in the rule ∃-intro. Substituting the actual representation type `boolean` for the abstract type `t` in the signature gives the module type of the body. We will explore this behaviour more fully in lecture 4.

Just as we can use variants to define booleans in OCaml, we can use sums to define booleans in System F. Here is a definition analogous to the `Bool` module above, using the left injection for **false** and the right injection for **true**:

```
pack (1+1,
       ⟨inr [1] ⟨⟩,
       ⟨inl [1] ⟨⟩,
        λb:Bool.Λα::*.λr:α.λs:α.case b of x.s | y.r⟩⟩)
   as ∃β::*.
        β ×
        β ×
        β → ∀α::*.α → α → α
```

The unit and boolean examples encode data types with small fixed numbers of inhabitants. However System F is also sufficiently powerful to encode datatypes with infinitely many members, as we now proceed to show.

A type representing the natural numbers can be defined in OCaml as a variant with two constructors:

```
type nat =
    Zero : nat
  | Succ : nat → nat
```

Using these constructors we can represent every non-negative integer: for example, the number three is represented as follows:

```
let three = Succ (Succ (Succ Zero))
```

We can encode the natural numbers in System F as follows:

$$
\begin{array}{l}
\textbf{pack}\ \ (\forall\alpha::\star.\alpha \to (\alpha \to \alpha) \to \alpha,\\
\qquad \langle \Lambda\alpha::\star.\lambda z{:}\alpha.\lambda s{:}\alpha \to \alpha.z,\\
\qquad \langle \lambda n{:}\forall\alpha::\star.\alpha \to (\alpha \to \alpha) \to \alpha.\\
\qquad\quad \Lambda\alpha::\star.\lambda z{:}\alpha.\lambda s{:}\alpha \to \alpha.s\ \ (n\ [\alpha]\ z\ s),\\
\qquad \langle \lambda n{:}\forall\alpha::\star.\alpha \to (\alpha \to \alpha) \to \alpha.n\rangle\rangle\rangle)\\
\quad \textbf{as}\ \exists\mathbb{N} :: *.\\
\qquad\quad \mathbb{N}\ \times\\
\qquad\quad (\mathbb{N} \to \mathbb{N})\ \times\\
\qquad\quad (\mathbb{N} \to \forall\alpha.\alpha \to (\alpha \to \alpha) \to \alpha)
\end{array}
$$

As for booleans there are two constructors corresponding to the constructors of the data type and a branching operation (which we will call `foldℕ`) which makes it possible to define functions that discriminate between different numbers. The branching operation accepts a natural number of type $\mathbb{N}$, a type argument α specifying the type of the result, a value of type α to return in case the input is zero and a function of type $\alpha \to \alpha$ to use in case the input is the successor of some other number n. Using `foldℕ` we might define the function which tests whether a number is equal to zero by instantiating the result type with `Bool` and passing suitable arguments for the zero and successor cases:

```
λm:ℕ.λn:ℕ.foldℕ m [Bool] true (λb:Bool.false)
```

Similarly we might define the addition function using `foldℕ` by instantiating the result type with $\mathbb{N}$:

```
λm:ℕ.λn:ℕ.foldℕ m [ℕ] n succ
```

1.4 Exercises

1. [★]: Give a typing derivation for `swap`
2. [★★]: We have seen how to implement booleans in System F using sums and the unit type. Give an equivalent implementation of booleans using polymorphism, using $\mathbb{N}$ encoding as an example.
3. [★★★]: Implement a signed integer type, either using booleans, products and $\mathbb{N}$, or directly in System F. Give an addition function for your signed integers.

These notes aim to be self-contained, but fairly terse. There are many more comprehensive introductions to the typed lambda calculi available. The following two books are highly recommended:

- **Types and Programming Languages**
  Benjamin C. Pierce
  MIT Press (2002)
  `http://www.cis.upenn.edu/~bcpierce/tapl/`
  There are copies in the Computer Laboratory library and many of the college libraries.

- **Lambda Calculi with Types**
  Henk Barendregt
  in Handbook of Logic in Computer Science Volume II, Oxford University Press (1992)
  Available online: `http://ttic.uchicago.edu/~dreyer/course/papers/barendregt.pdf`