Title: Introduction to Functional Programming
Lecturer: Mike Gordon
(http://www.cl.cam.ac.uk/users/mjcg/)
Class: Computer Science Tripos, Part II(General) \& Diploma
Term: Lent term 1996
First Lecture: Friday January 191996 at 12am
Location: Heycock Lecture Room
Duration: Twelve lectures (M. W. F. 12)

## Preface

This course aims to teach both the theory and practice of functional programming. The theory consists of the $\lambda$-calculus and the practice will be illustrated using the programming language Standard ML.
The field of Functional Programming splits into those who prefer 'lazy' languages like Haskell and those who prefer 'strict' languages like ML. The practical parts of this course almost exclusively emphasise the latter, but the material on the $\lambda$ calculus underlies both approaches.
The chapters on the $\lambda$-calculus have been largely condensed from Part II of the book:
M.J.C. Gordon, Programming Language Theory and its Implementation, Prentice Hall International Series in Computer Science, 1988 (currently out of print).

The introduction to ML in Chapter 4 started life as part of:
Gordon, M.J.C., Milner, A.J.R.G. and Wadsworth, C.P., Edinburgh LCF: a mechanized logic of computation, Springer Lecture Notes in Computer Science, Springer-Verlag, 1979.

The ML parts of this were updated substantially in the technical report:
G. Cousineau, M. Gordon, G. Huet, R. Milner, L. Paulson, and C. Wadsworth, The ML handbook, INRIA (1986).

I translated the introduction of this report into Standard ML and added some new material to get Chapter 4. The case studies were written by me at great speed, and so are bound to contain numerous mistakes! They aim to show how ML-based functional programming can be used in practice.
The following people have contributed in various ways to the material cited above or to these notes: Graham Birtwistle, Shiu Kai Chin, Avra Cohn, Jan van Eijck, Mike Fourman, Elsa Gunter, Peter Hancock, Martin Hyland, Tom Melham, Allan C. Milne, Nicholas Ourusoff, David Shepherd and Roger Stokes.

## Contents

Preface ..... i
1 Introduction to the $\lambda$-calculus ..... 1
1.1 Syntax and semantics of the $\lambda$-calculus ..... 1
1.2 Notational conventions ..... 3
1.3 Free and bound variables ..... 3
1.4 Conversion rules ..... 4
1.4.1 $\alpha$-conversion ..... 5
1.4.2 $\beta$-conversion ..... 5
1.4.3 $\quad \eta$-conversion ..... 6
1.4.4 Generalized conversions ..... 6
1.5 Equality of $\lambda$-expressions ..... 7
1.6 The $\longrightarrow$ relation ..... 9
1.7 Extensionality ..... 10
1.8 Substitution ..... 10
2 Representing Things in the $\lambda$-calculus ..... 13
2.1 Truth-values and the conditional ..... 13
2.2 Pairs and tuples ..... 15
2.3 Numbers ..... 16
2.4 Definition by recursion ..... 20
2.5 Functions with several arguments ..... 22
2.6 Mutual recursion ..... 25
2.7 Representing the recursive functions ..... 26
2.7.1 The primitive recursive functions ..... 26
2.7.2 The recursive functions ..... 27
2.7.3 The partial recursive functions ..... 29
2.8 Extending the $\lambda$-calculus ..... 29
2.9 Theorems about the $\lambda$-calculus ..... 30
2.10 Call-by-value and $\mathbf{Y}$ ..... 33
3 Combinators ..... 35
3.1 Combinator reduction ..... 36
3.2 Functional completeness ..... 36
3.3 Reduction machines ..... 39
3.4 Improved translation to combinators ..... 41
3.5 More combinators ..... 42
3.6 Curry's algorithm ..... 43
3.7 Turner's algorithm ..... 44
4 A Quick Overview of ML ..... 47
4.1 Interacting with ML ..... 47
4.2 Expressions ..... 47
4.3 Declarations ..... 48
4.4 Comments ..... 49
4.5 Functions ..... 49
4.6 Type abbreviations ..... 51
4.7 Operators ..... 52
4.8 Lists ..... 53
4.9 Strings ..... 53
4.10 Records ..... 54
4.11 Polymorphism ..... 54
4.12 fn-expressions ..... 55
4.13 Conditionals ..... 56
4.14 Recursion ..... 56
4.15 Equality types ..... 57
4.16 Pattern matching ..... 58
4.17 The case construct ..... 60
4.18 Exceptions ..... 61
4.19 Datatype declarations ..... 63
4.20 Abstract types ..... 65
4.21 Type constructors ..... 66
4.22 References and assignment ..... 67
4.23 Iteration ..... 67
4.24 Programming in the large ..... 68
5 Case study 1: parsing ..... 69
5.1 Lexical analysis ..... 69
5.2 Simple special cases of parsing ..... 73
5.2.1 Applicative expressions ..... 73
5.2.2 Precedence parsing of infixes ..... 77
5.3 A general top-down precedence parser ..... 82
6 Case study 2: the $\lambda$-calculus ..... 89
6.1 A $\lambda$-calculus parser ..... 90
6.2 Implementing substitution ..... 92
6.3 The SECD machine ..... 94
Bibliography ..... 97

## Introduction to the $\lambda$-calculus

The $\lambda$-calculus (or lambda-calculus) is a theory of functions that was originally developed by the logician Alonzo Church as a foundation for mathematics. This work was done in the 1930s, several years before digital computers were invented. A little earlier (in the 1920s) Moses Schönfinkel developed another theory of functions based on what are now called 'combinators'. In the 1930s, Haskell Curry rediscovered and extended Schönfinkel's theory and showed that it was equivalent to the $\lambda$-calculus. About this time Kleene showed that the $\lambda$-calculus was a universal computing system; it was one of the first such systems to be rigorously analysed. In the 1950 s John McCarthy was inspired by the $\lambda$-calculus to invent the programming language LISP. In the early 1960s Peter Landin showed how the meaning of imperative programming languages could be specified by translating them into the $\lambda$-calculus. He also invented an influential prototype programming language called ISWIM [24]. This introduced the main notations of functional programming and influenced the design of both functional and imperative languages. Building on this work, Christopher Strachey laid the foundations for the important area of denotational semantics [13, 33]. Technical questions concerning Strachey's work inspired the mathematical logician Dana Scott to invent the theory of domains, which is now one of the most important parts of theoretical computer science. During the 1970s Peter Henderson and Jim Morris took up Landin's work and wrote a number of influential papers arguing that functional programming had important advantages for software engineering $[17,16]$. At about the same time David Turner proposed that Schönfinkel and Curry's combinators could be used as the machine code of computers for executing functional programming languages. Such computers could exploit mathematical properties of the $\lambda$-calculus for the parallel evaluation of programs. During the 1980s several research groups took up Henderson's and Turner's ideas and started working on making functional programming practical by designing special architectures to support it, some of them with many processors.
We thus see that an obscure branch of mathematical logic underlies important developments in programming language theory, such as:
(i) The study of fundamental questions of computation.
(ii) The design of programming languages.
(iii) The semantics of programming languages.
(iv) The architecture of computers.

### 1.1 Syntax and semantics of the $\lambda$-calculus

The $\lambda$-calculus is a notation for defining functions. The expressions of the notation are called $\lambda$-expressions and each such expression denotes a function. It will be seen later how functions can be used to represent a wide variety of data and datastructures including numbers, pairs, lists etc. For example, it will be demonstrated
how an arbitrary pair of numbers $(x, y)$ can be represented as a $\lambda$-expression. As a notational convention, mnemonic names are assigned in bold or underlined to particular $\lambda$-expressions; for example $\underline{1}$ is the $\lambda$-expression (defined in Section 2.3) which is used to represent the number one.
There are just three kinds of $\lambda$-expressions:
(i) Variables: $x, y, z$ etc. The functions denoted by variables are determined by what the variables are bound to in the environment. Binding is done by abstractions (see 3 below). We use $V, V_{1}, V_{2}$ etc. for arbitrary variables.
(ii) Function applications or combinations: if $E_{1}$ and $E_{2}$ are $\lambda$-expressions, then so is $\left(E_{1} E_{2}\right)$; it denotes the result of applying the function denoted by $E_{1}$ to the function denoted by $E_{2} . E_{1}$ is called the rator (from 'operator') and $E_{2}$ is called the rand (from 'operand'). For example, if ( $\underline{m}, \underline{n}$ ) denotes a function representing the pair of numbers $m$ and $n$ (see Section 2.2) and sum denotes the addition function ${ }^{1} \lambda$-calculus (see Section 2.5), then the application $(\operatorname{sum}(\underline{m}, \underline{n}))$ denotes $\underline{m+n}$.
(iii) Abstractions: if $V$ is a variable and $E$ is a $\lambda$-expression, then $\lambda V . E$ is an abstraction with bound variable $V$ and body $E$. Such an abstraction denotes the function that takes an argument $a$ and returns as result the function denoted by $E$ in an environment in which the bound variable $V$ denotes $a$. More specifically, the abstraction $\lambda V . E$ denotes a function which takes an argument $E^{\prime}$ and transforms it into the thing denoted by $E\left[E^{\prime} / V\right]$ (the result of substituting $E^{\prime}$ for $V$ in $E$, see Section 1.8). For example, $\lambda x$. $\operatorname{sum}(x, \underline{1})$ denotes the add-one function.

Using BNF, the syntax of $\lambda$-expressions is just:

$$
\begin{aligned}
<\lambda \text {-expression }>:= & <\text { variable }> \\
& (<\lambda \text {-expression }><\lambda \text {-expression }>) \\
& (\lambda<\text { variable }>.<\lambda \text {-expression }>)
\end{aligned}
$$

If $V$ ranges over the syntax class $<$ variable $>$ and $E, E_{1}, E_{2}, \ldots$ etc. range over the syntax class $<\lambda$-expression $>$, then the BNF simplifies to:


The description of the meaning of $\lambda$-expressions just given above is vague and intuitive. It took about 40 years for logicians (Dana Scott, in fact [32]) to make it rigorous in a useful way. We shall not be going into details of this.

Example: $(\lambda x . x)$ denotes the 'identity function': $((\lambda x . x) E)=E$.
Example: $(\lambda x .(\lambda f .(f x)))$ denotes the function which when applied to $E$ yields $(\lambda f .(f x))[E / x]$, i.e. $(\lambda f .(f E))$. This is the function which when applied to $E^{\prime}$ yields $(f E)\left[E^{\prime} / f\right]$ i.e. $\left(E^{\prime} E\right)$. Thus

$$
((\lambda x .(\lambda f .(f x))) E)=(\lambda f .(f E))
$$

and

$$
\left((\lambda f .(f E)) E^{\prime}\right)=\left(E^{\prime} E\right)
$$

[^0]
## Exercise 1

Describe the function denoted by $(\lambda x .(\lambda y . y))$.
Example: Section 2.3 describes how numbers can be represented by $\lambda$-expressions. Assume that this has been done and that $\underline{0}, \underline{1}, \underline{2}, \ldots$ are $\lambda$-expressions which represent $0,1,2, \ldots$, respectively. Assume also that add is a $\lambda$-expression denoting a function satisfying:

$$
((\mathbf{\operatorname { a d d }} \underline{m}) \underline{n})=\underline{m+n} .
$$

Then $(\lambda x .(($ add 1$) x))$ is a $\lambda$-expression denoting the function that transforms $\underline{n}$ to $\underline{1+n}$, and $(\lambda x$. $(\lambda y .((\operatorname{add} x) y)))$ is a $\lambda$-expression denoting the function that transforms $\underline{m}$ to the function which when applied to $\underline{n}$ yields $\underline{m+n}$, namely $\lambda y$. $((\operatorname{add} \underline{m}) y))$.

The relationship between the function sum in (ii) at the beginning of this section (page 2) and the function add in the previous example is explained in Section 2.5.

### 1.2 Notational conventions

The following conventions help minimize the number of brackets one has to write.

1. Function application associates to the left, i.e. $E_{1} \quad E_{2} \cdots E_{n}$ means $\left(\left(\cdots\left(E_{1} E_{2}\right) \cdots\right) E_{n}\right)$. For example:

$$
\begin{array}{llll}
E_{1} E_{2} & \text { means } & \left(E_{1} E_{2}\right) \\
E_{1} E_{2} E_{3} & \text { means } & \left(\left(E_{1} E_{2}\right) E_{3}\right) \\
E_{1} E_{2} E_{3} E_{4} & \text { means } & \left(\left(\left(E_{1} E_{2}\right) E_{3}\right) E_{4}\right)
\end{array}
$$

2. $\lambda V . E_{1} E_{2} \ldots E_{n} \operatorname{means}\left(\lambda V .\left(\begin{array}{llll}E_{1} & E_{2} & \ldots & E_{n}\end{array}\right)\right.$ ). Thus the scope of ' $\lambda V$ ' extends as far to the right as possible.
3. $\lambda V_{1} \cdots V_{n}$. $E$ means $\left(\lambda V_{1} .\left(\cdots .\left(\lambda V_{n} . E\right) \cdots\right)\right)$. For example:
$\lambda x y . E \quad$ means $\quad(\lambda x .(\lambda y . E))$
$\lambda x y z . E \quad$ means $\quad(\lambda x .(\lambda y .(\lambda z . E)))$
$\lambda x y z w . E \quad$ means $\quad(\lambda x .(\lambda y \cdot(\lambda z .(\lambda w . E))))$

Example: $\lambda x y$. add $y x$ means $(\lambda x .(\lambda y .((\operatorname{add} y) x)))$.

### 1.3 Free and bound variables

An occurrence of a variable $V$ in a $\lambda$-expression is free if it is not within the scope of a ' $\lambda V$ ', otherwise it is bound. For example


### 1.4 Conversion rules

In Chapter 2 it is explained how $\lambda$-expressions can be used to represent data objects like numbers, strings etc. For example, an arithmetic expression like $(2+3) \times 5$ can be represented as a $\lambda$-expression and its 'value' 25 can also be represented as a $\lambda$-expression. The process of 'simplifying' $(2+3) \times 5$ to 25 will be represented by a process called conversion (or reduction). The rules of $\lambda$-conversion described below are very general, yet when they are applied to $\lambda$-expressions representing arithmetic expressions they simulate arithmetical evaluation.
There are three kinds of $\lambda$-conversion called $\alpha$-conversion, $\beta$-conversion and $\eta$ conversion (the original motivation for these names is not clear). In stating the conversion rules the notation $E\left[E^{\prime} / V\right]$ is used to mean the result of substituting $E^{\prime}$ for each free occurrence of $V$ in $E$. The substitution is called valid if and only if no free variable in $E^{\prime}$ becomes bound in $E\left[E^{\prime} / V\right]$. Substitution is described in more detail in Section 1.8.

## The rules of $\lambda$-conversion

## - $\alpha$-conversion.

Any abstraction of the form $\lambda V . E$ can be converted to $\lambda V^{\prime} . E\left[V^{\prime} / V\right]$ provided the substitution of $V^{\prime}$ for $V$ in $E$ is valid.

- $\beta$-conversion.

Any application of the form $\left(\lambda V . E_{1}\right) E_{2}$ can be converted to $E_{1}\left[E_{2} / V\right]$, provided the substitution of $E_{2}$ for $V$ in $E_{1}$ is valid.

- $\eta$-conversion.

Any abstraction of the form $\lambda V .(E V)$ in which $V$ has no free occurrence in $E$ can be reduced to $E$.

The following notation will be used:

- $E_{1} \underset{\alpha}{\longrightarrow} E_{2}$ means $E_{1} \alpha$-converts to $E_{2}$.
- $E_{1} \underset{\beta}{\longrightarrow} E_{2}$ means $E_{1} \beta$-converts to $E_{2}$.
- $E_{1} \underset{\eta}{\longrightarrow} E_{2}$ means $E_{1} \eta$-converts to $E_{2}$.

In Section 1.4.4 below this notation is extended.
The most important kind of conversion is $\beta$-conversion; it is the one that can be used to simulate arbitrary evaluation mechanisms. $\alpha$-conversion is to do with the technical manipulation of bound variables and $\eta$-conversion expresses the fact that two functions that always give the same results on the same arguments are equal (see Section 1.7). The next three subsections give further explanation and examples of the three kinds of conversion (note that 'conversion' and 'reduction' are used below as synonyms).

### 1.4.1 $\alpha$-conversion

A $\lambda$-expression (necessarily an abstraction) to which $\alpha$-reduction can be applied is called an $\alpha$-redex. The term 'redex' abbreviates 'reducible expression'. The rule of $\alpha$-conversion just says that bound variables can be renamed provided no 'nameclashes' occur.

## Examples

$$
\begin{gathered}
\lambda x . x \underset{\alpha}{\longrightarrow} \lambda y . y \\
\lambda x . f x \underset{\alpha}{\longrightarrow} \lambda y . f y
\end{gathered}
$$

It is not the case that

$$
\lambda x . \lambda y . \text { add } x y \underset{\alpha}{\longrightarrow} \lambda y . \lambda y . \text { add } y y
$$

because the substitution $(\lambda y$. add $x y)[y / x]$ is not valid since the $y$ that replaces $x$ becomes bound.

### 1.4.2 $\beta$-conversion

A $\lambda$-expression (necessarily an application) to which $\beta$-reduction can be applied is called a $\beta$-redex. The rule of $\beta$-conversion is like the evaluation of a function call in a programming language: the body $E_{1}$ of the function $\lambda V . E_{1}$ is evaluated in an environment in which the 'formal parameter' $V$ is bound to the 'actual parameter' $E_{2}$.

## Examples

$$
\begin{gathered}
(\lambda x . f x) E \underset{\beta}{\longrightarrow} f E \\
(\lambda x \cdot(\lambda y \cdot \text { add } x y)) \underline{3} \underset{\beta}{\longrightarrow} \lambda y \cdot \text { add } \underline{3} y \\
(\lambda y \cdot \text { add } \underline{3} y) \underline{4} \underset{\beta}{\longrightarrow} \text { add } \underline{3} \underline{4}
\end{gathered}
$$

It is not the case that

$$
(\lambda x .(\lambda y . \text { add } x y))(\text { square } y) \underset{\beta}{\longrightarrow} \lambda y . \text { add (square } y) y
$$

because the substition $(\lambda y$. add $x y)$ [(square $y) / x]$ is not valid, since $y$ is free in (square $y$ ) but becomes bound after substitution for $x$ in ( $\lambda y$. add $x y$ ).
It takes some practice to parse $\lambda$-expressions according to the conventions of Section 1.2 so as to identify the $\beta$-redexes. For example, consider the application:

$$
(\lambda x . \lambda y . \operatorname{add} x y) \underline{3} \underline{4}
$$

Putting in brackets according to the conventions expands this to:

$$
(((\lambda x \cdot(\lambda y \cdot((\operatorname{add} x) y))) \underline{3}) \underline{4})
$$

which has the form:

$$
((\lambda x . E) \underline{3}) \underline{4}
$$

where

$$
E=(\lambda y \cdot \mathbf{a d d} x y)
$$

$(\lambda x . E) \underline{3}$ is a $\beta$-redex and could be reduced to $E[\underline{3} / x]$.

### 1.4.3 $\eta$-conversion

A $\lambda$-expression (necessarily an abstraction) to which $\eta$-reduction can be applied is called an $\eta$-redex. The rule of $\eta$-conversion expresses the property that two functions are equal if they give the same results when applied to the same arguments. This property is called extensionality and is discussed further in Section 1.7. For example, $\eta$-conversion ensures that $\lambda x$. $(\sin x)$ and $\sin$ denote the same function. More generally, $\lambda V$. ( $E V$ ) denotes the function which when applied to an argument $E^{\prime}$ returns $(E V)\left[E^{\prime} / V\right]$. If $V$ does not occur free in $E$ then $(E V)\left[E^{\prime} / V\right]=\left(E E^{\prime}\right)$. Thus $\lambda V . E V$ and $E$ both yield the same result, namely $E E^{\prime}$, when applied to the same arguments and hence they denote the same function.

## Examples

$$
\begin{gathered}
\lambda x . \text { add } x \underset{\eta}{\longrightarrow} \text { add } \\
\lambda y . \text { add } x y \underset{\eta}{\longrightarrow} \text { add } x
\end{gathered}
$$

It is not the case that

$$
\lambda x . \text { add } x x \underset{\eta}{\longrightarrow} \text { add } x
$$

because $x$ is free in add $x$.

### 1.4.4 Generalized conversions

The definitions of $\underset{\alpha}{\longrightarrow}, \underset{\beta}{\longrightarrow}$ and $\underset{\eta}{\longrightarrow}$ can be generalized as follows:

- $E_{1} \underset{\alpha}{\longrightarrow} E_{2}$ if $E_{2}$ can be got from $E_{1}$ by $\alpha$-converting any subterm.
- $E_{1} \xrightarrow[\beta]{\longrightarrow} E_{2}$ if $E_{2}$ can be got from $E_{1}$ by $\beta$-converting any subterm.
- $E_{1} \xrightarrow[\eta]{\longrightarrow} E_{2}$ if $E_{2}$ can be got from $E_{1}$ by $\eta$-converting any subterm.


## Examples

$$
\begin{gathered}
((\lambda x . \lambda y \cdot \operatorname{add} x y) \underline{3}) \underline{4} \underset{\beta}{\longrightarrow}(\lambda y \cdot \operatorname{add} \underline{3} y) \underline{4} \\
(\lambda y \cdot \operatorname{add} \underline{3} y) \underline{4} \underset{\beta}{\longrightarrow} \operatorname{add} \underline{3} \underline{4}
\end{gathered}
$$

The first of these is a $\beta$-conversion in the generalized sense because ( $\lambda y$. add $\underline{3} y$ ) $\underline{4}$ is obtained from $((\lambda x$. $\lambda y$. add $x y) \underline{3}) \underline{4}$ (which is not itself a $\beta$-redex) by reducing the subexpression $(\lambda x$. $\lambda y$. add $x y) \underline{3}$. We will sometimes write a sequence of conversions like the two above as:

$$
((\lambda x . \lambda y . \operatorname{add} x y) \underline{3}) \underline{4} \underset{\beta}{\longrightarrow}(\lambda y . \text { add } \underline{3} y) \underline{4} \underset{\beta}{\longrightarrow} \text { add } \underline{3} \underline{4}
$$

## Exercise 2

Which of the three $\beta$-reductions below are generalized conversions (i.e. reductions of subexpressions) and which are conversions in the sense defined on page 4 ?
(i) $(\lambda x \cdot x) \underset{\beta}{\longrightarrow} \underline{1}$
(ii) $(\lambda y \cdot y)((\lambda x \cdot x) \underline{1}) \underset{\beta}{\longrightarrow}(\lambda y \cdot y) \underline{1} \underset{\beta}{\longrightarrow} \underline{1}$
(iii) $(\lambda y . y)((\lambda x . x) \underline{1}) \underset{\beta}{\longrightarrow}(\lambda x . x) \underset{\beta}{\underset{\beta}{1}}$

In reductions (ii) and (iii) in the exercise above one starts with the same $\lambda$ expression, but reduce redexes in different orders.
An important property of $\beta$-reductions is that no matter in which order one does them, one always ends up with equivalent results. If there are several disjoint redexes in an expression, one can reduce them in parallel. Note, however, that some reduction sequences may never terminate. This is discussed further in connection with the normalization theorem of Chapter 2.9. It is a current hot research topic in 'fifth-generation computing' to design processors which exploit parallel evaluation to speed up the execution of functional programs.

### 1.5 Equality of $\lambda$-expressions

The three conversion rules preserve the meaning of $\lambda$-expressions, i.e. if $E_{1}$ can be converted to $E_{2}$ then $E_{1}$ and $E_{2}$ denote the same function. This property of conversion should be intuitively clear. It is possible to give a mathematical definition of the function denoted by a $\lambda$-expression and then to prove that this function is unchanged by $\alpha$-, $\beta$ - or $\eta$-conversion. Doing this is surprisingly difficult [33] and is beyond the scope of this book.
We will simply define two $\lambda$-expressions to be equal if they can be transformed into each other by a sequence of (forwards or backwards) $\lambda$-conversions. It is important to be clear about the difference between equality and identity. Two $\lambda$-expressions are identical if they consist of exactly the same sequence of characters; they are equal if one can be converted to the other. For example, $\lambda x . x$ is equal to $\lambda y . y$, but not identical to it. The following notation is used:

- $E_{1} \equiv E_{2}$ means $E_{1}$ and $E_{2}$ are identical.
- $E_{1}=E_{2}$ means $E_{1}$ and $E_{2}$ are equal.

Equality $(=)$ is defined in terms of identity $(\equiv)$ and conversion $(\underset{\alpha}{\longrightarrow}, \underset{\beta}{\longrightarrow}$ and $\underset{\eta}{\longrightarrow}$ as follows.

## Equality of $\lambda$-expressions

If $E$ and $E^{\prime}$ are $\lambda$-expressions then $E=E^{\prime}$ if $E \equiv E^{\prime}$ or there exist expressions $E_{1}, E_{2}, \ldots, E_{n}$ such that:

1. $E \equiv E_{1}$
2. $E^{\prime} \equiv E_{n}$
3. For each $i$ either
(a) $E_{i} \underset{\alpha}{\longrightarrow} E_{i+1}$ or $E_{i} \underset{\beta}{\longrightarrow} E_{i+1}$ or $E_{i} \underset{\eta}{\longrightarrow} E_{i+1}$ or
(b) $E_{i+1} \underset{\alpha}{\longrightarrow} E_{i}$ or $E_{i+1} \underset{\beta}{\longrightarrow} E_{i}$ or $E_{i+1} \underset{\eta}{\longrightarrow} E_{i}$.

## Examples

$$
\begin{gathered}
(\lambda x \cdot x) \underline{1}=\underline{1} \\
(\lambda x \cdot x)((\lambda y \cdot y) \underline{1})=\underline{1} \\
(\lambda x . \lambda y \cdot \operatorname{add} x y) \underline{3} \underline{4}=\text { add } \underline{3} \underline{4}
\end{gathered}
$$

From the definition of $=$ it follows that:
(i) For any $E$ it is the case that $E=E$ (equality is reflexive).
(ii) If $E=E^{\prime}$, then $E^{\prime}=E$ (equality is symmetric).
(iii) If $E=E^{\prime}$ and $E^{\prime}=E^{\prime \prime}$, then $E=E^{\prime \prime}$ (equality is transitive).

If a relation is reflexive, symmetric and transitive then it is called an equivalence relation. Thus $=$ is an equivalence relation.
Another important property of $=$ is that if $E_{1}=E_{2}$ and if $E_{1}^{\prime}$ and $E_{2}^{\prime}$ are two $\lambda$-expressions that only differ in that where one contains $E_{1}$ the other contains $E_{2}$, then $E_{1}^{\prime}=E_{2}^{\prime}$. This property is called Leibnitz's law. It holds because the same sequence of reduction for getting from $E_{1}$ to $E_{2}$ can be used for getting from $E_{1}^{\prime}$ to $E_{2}^{\prime}$. For example, if $E_{1}=E_{2}$, then by Leibnitz's law $\lambda V . E_{1}=\lambda V . E_{2}$.
It is essential for the substitutions in the $\alpha$ - and $\beta$-reductions to be valid. The validity requirement disallows, for example, $\lambda x .(\lambda y . x)$ being $\alpha$-reduced to $\lambda y .(\lambda y . y)$ (since $y$ becomes bound after substitution for $x$ in $\lambda y . x$ ). If this invalid substitution were permitted, then it would follow by the definition of $=$ that:

$$
\lambda x \cdot \lambda y \cdot x=\lambda y \cdot \lambda y \cdot y
$$

But then since:

$$
(\lambda x .(\lambda y \cdot x)) \underline{1} \underset{\beta}{\longrightarrow}(\lambda y . \underline{1}) \underset{\beta}{\longrightarrow} \underset{\sim}{1}
$$

and

$$
(\lambda y \cdot(\lambda y \cdot y)) \underline{1} \underline{2} \underset{\beta}{\longrightarrow}(\lambda y \cdot y) \underline{2} \underset{\beta}{\longrightarrow} \underline{2}
$$

one would be forced to conclude that $\underline{1}=\underline{2}$. More generally by replacing $\underline{1}$ and $\underline{2}$ by any two expressions, it could be shown that any two expressions are equal!

## Exercise 3

Find an example which shows that if substitutions in $\beta$-reductions are allowed to be invalid, then it follows that any two $\lambda$-expressions are equal.

Example: If $V_{1}, V_{2}, \ldots, V_{n}$ are all distinct and none of them occur free in any of $E_{1}, E_{2}, \ldots, E_{n}$, then

$$
\begin{array}{cc}
\left(\lambda V_{1}\right. & \left.V_{2} \cdots V_{n} . E\right) E_{1} E_{2} \cdots E_{n} \\
= & \left(\left(\lambda V_{1} \cdot\left(\lambda V_{2} \cdots V_{n} \cdot E\right)\right) E_{1}\right) E_{2} \cdots E_{n} \\
\xrightarrow[\beta]{\longrightarrow} & \left(\left(\lambda V_{2} \cdots V_{n} \cdot E\right)\left[E_{1} / V_{1}\right]\right) E_{2} \cdots E_{n} \\
= & \left(\lambda V_{2} \ldots V_{n} \cdot E\left[E_{1} / V_{1}\right]\right) E_{2} \cdots E_{n} \\
\vdots & \vdots \\
= & E\left[E_{1} / V_{1}\right]\left[E_{2} / V_{2}\right] \cdots\left[E_{n} / V_{n}\right]
\end{array}
$$

## Exercise 4

In the last example, where was the assumption used that $V_{1}, V_{2}, \ldots, V_{n}$ are all distinct and that none of them occur free in any of $E_{1}, E_{2}, \ldots, E_{n}$ ?

## Exercise 5

Find an example to show that if $V_{1}=V_{2}$, then even if $V_{2}$ is not free in $E_{1}$, it is not necessarily the case that:

$$
\left(\lambda V_{1} V_{2} . E\right) E_{1} E_{2}=E\left[E_{1} / V_{1}\right]\left[E_{2} / V_{2}\right]
$$

## Exercise 6

Find an example to show that if $V_{1} \neq V_{2}$, but $V_{2}$ occurs free in $E_{1}$, then it is not necessarily the case that:

$$
\left(\lambda V_{1} V_{2} . E\right) E_{1} E_{2}=E\left[E_{1} / V_{1}\right]\left[E_{2} / V_{2}\right]
$$

### 1.6 The $\longrightarrow$ relation

In the previous section $E_{1}=E_{2}$ was defined to mean that $E_{2}$ could be obtained from $E_{1}$ by a sequence of forwards or backwards conversions. A special case of this is when $E_{2}$ is got from $E_{1}$ using only forwards conversions. This is written $E_{1} \longrightarrow E_{2}$.

$$
\text { Definition of } \longrightarrow
$$

If $E$ and $E^{\prime}$ are $\lambda$-expressions, then $E \longrightarrow E^{\prime}$ if $E \equiv E^{\prime}$ or there exist expressions
$E_{1}, E_{2}, \ldots, E_{n}$ such that:
$\quad$ 1. $E \equiv E_{1}$
2. $E^{\prime} \equiv E_{n}$
3. For each $i$ either $E_{i} \underset{\alpha}{\longrightarrow} E_{i+1}$ or $E_{i} \underset{\beta}{\longrightarrow} E_{i+1}$ or $E_{i} \underset{\eta}{\longrightarrow} E_{i+1}$.

Notice that the definition of $\longrightarrow$ is just like the definition of $=$ on page 7 except that part (b) of 3 is missing.

## Exercise 7

Find $E, E^{\prime}$ such that $E=E^{\prime}$ but it is not the case that $E \longrightarrow E^{\prime}$.

## Exercise 8

[very hard!] Show that if $E_{1}=E_{2}$, then there exists $E$ such that $E_{1} \longrightarrow E$ and $E_{2} \longrightarrow E$. (This property is called the Church-Rosser theorem. Some of its consequences are discussed in Chapter 2.9.)

### 1.7 Extensionality

Suppose $V$ does not occur free in $E_{1}$ or $E_{2}$ and

$$
E_{1} V=E_{2} V
$$

Then by Leibnitz's law (see page 8 )

$$
\lambda V . E_{1} V=\lambda V . E_{2} V
$$

so by $\eta$-reduction applied to both sides

$$
E_{1}=E_{2}
$$

It is often convenient to prove that two $\lambda$-expressions are equal using this property, i.e. to prove $E_{1}=E_{2}$ by proving $E_{1} V=E_{2} V$ for some $V$ not occuring free in $E_{1}$ or $E_{2}$. We will refer to such proofs as being by extensionality.

## Exercise 9

Show that

$$
(\lambda f g x . f x(g x))(\lambda x y . x)(\lambda x y . x)=\lambda x \cdot x
$$

### 1.8 Substitution

At the beginning of Section $1.4 E\left[E^{\prime} / V\right]$ was defined to mean the result of substituting $E^{\prime}$ for each free occurrence of $V$ in $E$. The substitution was said to be valid if no free variable in $E^{\prime}$ became bound in $E\left[E^{\prime} / V\right]$. In the definitions of $\alpha$ - and $\beta$-conversion, it was stipulated that the substitutions involved must be valid. Thus, for example, it was only the case that

$$
\left(\lambda V . E_{1}\right) E_{2} \underset{\beta}{\longrightarrow} E_{1}\left[E_{2} / V\right]
$$

as long as the substitution $E_{1}\left[E_{2} / V\right]$ was valid.
It is very convenient to extend the meaning of $E\left[E^{\prime} / V\right]$ so that we don't have to worry about validity. This is achieved by the definition below which has the property that for all expressions $E, E_{1}$ and $E_{2}$ and all variables $V$ and $V^{\prime}$ :

$$
\left(\lambda V . E_{1}\right) E_{2} \longrightarrow E_{1}\left[E_{2} / V\right] \quad \text { and } \quad \lambda V . E \longrightarrow \lambda V^{\prime} . E\left[V^{\prime} / V\right]
$$

To ensure this property holds, $E\left[E^{\prime} / V\right]$ is defined recursively on the structure of $E$ as follows:

| $E$ | $E\left[E^{\prime} / V\right]$ |
| :---: | :---: |
| $V$ | $E^{\prime}$ |
| $V^{\prime} \quad$ (where $V \neq V^{\prime}$ ) | $V^{\prime}$ |
| $E_{1} E_{2}$ | $E_{1}\left[E^{\prime} / V\right] E_{2}\left[E^{\prime} / V\right]$ |
| $\lambda V . E_{1}$ | $\lambda V . E_{1}$ |
| $\lambda V^{\prime} . E_{1}$ (where $V \neq V^{\prime}$ and $V^{\prime}$ is not free in $E^{\prime}$ ) | $\lambda V^{\prime} . E_{1}\left[E^{\prime} / V\right]$ |
| $\lambda V^{\prime} . E_{1}$ (where $V \neq V^{\prime}$ and $V^{\prime}$ is free in $E^{\prime}$ ) | $\lambda V^{\prime \prime} . E_{1}\left[V^{\prime \prime} / V^{\prime}\right]\left[E^{\prime} / V\right]$ <br> where $V^{\prime \prime}$ is a variable not free in $E^{\prime}$ or $E_{1}$ |

This particular definition of $E\left[E^{\prime} / V\right]$ is based on (but not identical to) the one in Appendix C of [2].
To illustrate how this works consider $(\lambda y . y x)[y / x]$. Since $y$ is free in $y$, the last case of the table above applies. Since $z$ does not occur in $y x$ or $y$,

$$
(\lambda y . y x)[y / x] \equiv \lambda z .(y x)[z / y][y / x] \equiv \lambda z .(z x)[y / x] \equiv \lambda z . z y
$$

In the last line of the table above, the particular choice of $V^{\prime \prime}$ is not specified. Any variable not occurring in $E^{\prime}$ or $E_{1}$ will do.
A good discussion of substitution can be found in the book by Hindley and Seldin [19] where various technical properties are stated and proved. The following exercise is taken from that book.

## Exercise 10

Use the table above to work out
(i) $(\lambda y, x(\lambda x, x))[(\lambda y, y x) / x]$.
(ii) $(y(\lambda z \cdot x z))[(\lambda y \cdot z y) / x]$.

It is straightforward, but rather tedious, to prove from the definition of $E\left[E^{\prime} / V\right]$ just given that indeed

$$
\left(\lambda V . E_{1}\right) E_{2} \longrightarrow E_{1}\left[E_{2} / V\right] \quad \text { and } \quad \lambda V . E \longrightarrow \lambda V^{\prime} . E\left[V^{\prime} / V\right]
$$

for all expressions $E, E_{1}$ and $E_{2}$ and all variables $V$ and $V^{\prime}$.
In Chapter 3 it will be shown how the theory of combinators can be used to decompose the complexities of substitution into simpler operations. Instead of combinators it is possible to use the so-called nameless terms of De Bruijn [6]. De Bruijn's idea is that variables can be thought of as 'pointers' to the $\lambda \mathrm{s}$ that bind them. Instead of 'labelling' $\lambda \mathrm{s}$ with names (i.e. bound variables) and then pointing to them via these names, one can point to the appropriate $\lambda$ by giving the number of levels 'upwards' needed to reach it. For example, $\lambda x$. $\lambda y . x y$ would be represented by $\lambda \lambda 21$. As a
more complicated example, consider the expression below in which we indicate the number of levels separating a variable from the $\lambda$ that binds it.


In De Bruijn's notation this is $\lambda \lambda 21 \lambda 311$.
A free variable in an expression is represented by a number bigger than the depth of $\lambda s$ above it; different free variables being assigned different numbers. For example,

$$
\lambda x \cdot(\lambda y \cdot y x z) x y w
$$

would be represented by

$$
\lambda(\lambda 123) 124
$$

Since there are only two $\lambda \mathrm{s}$ above the occurrence of 3 , this number must denote a free variable; similarly there is only one $\lambda$ above the second occurrence of 2 and the occurrence of 4 , so these too must be free variables. Note that 2 could not be used to represent $w$ since this had already been used to represent the free $y$; we thus chose the first available number bigger than 2 ( 3 was already in use representing $z$ ). Care must be taken to assign big enough numbers to free variables. For example, the first occurrence of $z$ in $\lambda x . z(\lambda y . z)$ could be represented by 2 , but the second occurrence requires 3 ; since they are the same variable we must use 3 .

Example: With De Bruijn's scheme $\lambda x . x(\lambda y . x y y)$ would be represented by $\lambda 1\left(\begin{array}{ll}\lambda & 1\end{array}\right)$.

## Exercise 11

What $\lambda$-expression is represented by $\lambda 2(\lambda 2)$ ?

## Exercise 12

Describe an algorithm for computating the De Bruijn representation of the expression $E\left[E^{\prime} / V\right]$ from the representations of $E$ and $E^{\prime}$.

## Representing Things in the $\lambda$-calculus

The $\lambda$-calculus appears at first sight to be a very primitive language. However, it can be used to represent most of the objects and structures needed for modern programming. The idea is to code these objects and structures in such a way that they have the required properties. For example, to represent the truth values true and false and the Boolean function $\neg$ ('not'), $\lambda$-expressions true, false and not are devised with the properties that:

$$
\begin{aligned}
& \text { not true }=\text { false } \\
& \text { not false }=\text { true }
\end{aligned}
$$

To represent the Boolean function $\wedge$ ('and') a $\lambda$-expression and is devised such that:

> and true true $=$ true
> and true false $=$ false
> and false true $=$ false
> and false false $=$ false
and to represent $\vee$ ('or') an expression or such that:

> or true true $=$ true
> or true false $=$ true
> or false true $=$ true
> or false false $=$ false

The $\lambda$-expressions used to represent things may appear completely unmotivated at first. However, the definitions are chosen so that they work together in unison.
We will write

$$
\mathrm{LET} \sim=\lambda \text {-expression }
$$

to introduce $\sim$ as a new notation. Usually $\sim$ will just be a name such as true or and. Such names are written in bold face, or underlined, to distinguish them from variables. Thus, for example, true is a variable but true is the $\lambda$-expression $\lambda x . \lambda y . x$ (see Section 2.1 below) and 2 is a number but $\underline{2}$ is the $\lambda$-expression $\lambda f x . f(f x)$ (see Section 2.3).
Sometimes $\sim$ will be a more complicated form like the conditional notation $(E \rightarrow$ $\left.E_{1} \mid E_{2}\right)$.

### 2.1 Truth-values and the conditional

This section defines $\lambda$-expressions true, false, not and $\left(E \rightarrow E_{1} \mid E_{2}\right)$ with the following properties:

$$
\begin{aligned}
& \text { not true }=\text { false } \\
& \text { not false }=\text { true } \\
& \left(\text { true } \rightarrow E_{1} \mid E_{2}\right)=E_{1} \\
& \left(\text { false } \rightarrow E_{1} \mid E_{2}\right)=E_{2}
\end{aligned}
$$

The $\lambda$-expressions true and false represent the truth-values true and false, not represents the negation function $\neg$ and $\left(E \rightarrow E_{1} \mid E_{2}\right)$ represents the conditional 'if $E$ then $E_{1}$ else $E_{2}$ '.
There are infinitely many different ways of representing the truth-values and negation that work; the ones used here are traditional and have been developed over the years by logicians.

$$
\begin{aligned}
& \text { LET true }=\lambda x \cdot \lambda y \cdot x \\
& \text { LET false }=\lambda x \cdot \lambda y \cdot y \\
& \text { LET not }=\lambda t \cdot t \text { false true }
\end{aligned}
$$

It is easy to use the rules of $\lambda$-conversion to show that these definitions have the desired properties. For example:

$$
\begin{array}{rlr}
\text { not true } & =(\lambda t . t \text { false true }) \text { true } & (\text { definition of not }) \\
& =\text { true false true } & (\beta \text {-conversion }) \\
& =(\lambda x . \lambda y . x) \text { false true } & (\text { definition of true }) \\
& =(\lambda y . \text { false }) \text { true } & (\beta \text {-conversion }) \\
& =\text { false } & (\beta \text {-conversion })
\end{array}
$$

Similarly not false $=$ true.
Conditional expressions $\left(E \rightarrow E_{1} \mid E_{2}\right)$ can be defined as follows:

$$
\operatorname{LET}\left(E \rightarrow E_{1} \mid E_{2}\right)=\left(\begin{array}{ll}
E & E_{1} \\
E_{2}
\end{array}\right)
$$

This means that for any $\lambda$-expressions $E, E_{1}$ and $E_{2},\left(E \rightarrow E_{1} \mid E_{2}\right)$ stands for $\left(E E_{1} E_{2}\right)$.
The conditional notation behaves as it should:

$$
\begin{aligned}
\left(\text { true } \rightarrow E_{1} \mid E_{2}\right) & =\text { true } E_{1} E_{2} \\
& =(\lambda x y . x) E_{1} E_{2} \\
& =E_{1}
\end{aligned}
$$

and

$$
\begin{aligned}
\left(\text { false } \rightarrow E_{1} \mid E_{2}\right) & =\text { false } E_{1} E_{2} \\
& =(\lambda x y \cdot y) E_{1} E_{2} \\
& =E_{2}
\end{aligned}
$$

## Exercise 13

Let and be the $\lambda$-expression $\lambda x y .(x \rightarrow y \mid$ false $)$. Show that:

$$
\begin{aligned}
& \text { and true true }=\text { true } \\
& \text { and true false }=\text { false } \\
& \text { and false true }=\text { false } \\
& \text { and false false }=\text { false }
\end{aligned}
$$

## Exercise 14

Devise a $\lambda$-expression or such that:

$$
\begin{aligned}
& \text { or true true }=\text { true } \\
& \text { or true false }=\text { true } \\
& \text { or false true }=\text { true } \\
& \text { or false false }=\text { false }
\end{aligned}
$$

### 2.2 Pairs and tuples

The following abbreviations represent pairs and $n$-tuples in the $\lambda$-calculus.

$$
\begin{aligned}
& \text { LET } \mathbf{f s t}=\lambda p . p \text { true } \\
& \text { LET snd }=\lambda p . p \text { false } \\
& \text { LET }\left(E_{1}, E_{2}\right)=\lambda f . f E_{1} E_{2}
\end{aligned}
$$

$\left(E_{1}, E_{2}\right)$ is a $\lambda$-expression representing an ordered pair whose first component (i.e. $E_{1}$ ) is accessed with the function fst and whose second component (i.e. $E_{2}$ ) is accessed with snd. The following calculation shows how the various definitions co-operate together to give the right results.

$$
\text { fst } \begin{aligned}
\left(E_{1}, E_{2}\right) & =(\lambda p . p \text { true })\left(E_{1}, E_{2}\right) \\
& =\left(E_{1}, E_{2}\right) \text { true } \\
& =\left(\lambda f \cdot f E_{1} E_{2}\right) \text { true } \\
& =\text { true } E_{1} E_{2} \\
& =(\lambda x y \cdot x) E_{1} E_{2} \\
& =E_{1}
\end{aligned}
$$

## Exercise 15

Show that $\operatorname{snd}\left(E_{1}, E_{2}\right)=E_{2}$.

A pair is a data-structure with two components. The generalization to $n$ components is called an $n$-tuple and is easily defined in terms of pairs.

$$
\operatorname{LET}\left(E_{1}, E_{2}, \ldots, E_{n}\right)=\left(E_{1},\left(E_{2},\left(\cdots\left(E_{n-1}, E_{n}\right) \cdots\right)\right)\right)
$$

$\left(E_{1}, \ldots, E_{n}\right)$ is an $n$-tuple with components $E_{1}, \ldots, E_{n}$ and length $n$. Pairs are 2tuples. The abbreviations defined next provide a way of extracting the components of $n$-tuples.

$$
\begin{aligned}
& \text { LET } E \stackrel{n}{\downarrow} 1=\mathbf{f s t} E \\
& \text { LET } E \stackrel{n}{\downarrow} 2=\mathbf{f s t}(\text { snd } E) \\
& \quad \vdots \\
& \text { LET } E \stackrel{n}{\downarrow} i=\mathbf{f s t}(\underbrace{\operatorname{snd}(\operatorname{snd}(\cdots(\text { snd }}_{i-1 \text { snds }} E) \cdots))) \quad(\text { if } i<n) \\
& \quad \vdots \\
& \text { LET } E \stackrel{n}{\downarrow}=\underbrace{\operatorname{snd}(\mathbf{s n d}(\ldots(\mathbf{s n d}}_{n-1 \text { snds }} E) \ldots)))
\end{aligned}
$$

It is easy to see that these definitions work, for example:

$$
\begin{aligned}
& \left(E_{1}, E_{2}, \ldots, E_{n}\right) \stackrel{n}{\downarrow} 1=\left(E_{1},\left(E_{2},(\ldots)\right)\right) \stackrel{n}{\downarrow} 1 \\
& =\mathbf{f s t}\left(E_{1},\left(E_{2},(\ldots)\right)\right) \\
& =E_{1} \\
& \left(E_{1}, E_{2}, \ldots, E_{n}\right) \stackrel{n}{\downarrow} 2=\left(E_{1},\left(E_{2},(\ldots)\right)\right) \stackrel{n}{\downarrow} 2 \\
& =\mathbf{f s t}\left(\boldsymbol{s n d}\left(E_{1},\left(E_{2},(\ldots)\right)\right)\right) \\
& =\mathbf{f s t}\left(E_{2},(\ldots)\right) \\
& =E_{2}
\end{aligned}
$$

In general $\left(E_{1}, E_{2}, \ldots, E_{n}\right) \stackrel{n}{\downarrow} i=E_{i}$ for all $i$ such that $1 \leq i \leq n$.

## Convention

We will usually just write $E \downarrow i$ instead of $E \stackrel{n}{\downarrow} i$ when it is clear from the context what $n$ should be. For example,

$$
\left(E_{1}, \ldots, E_{n}\right) \downarrow i=E_{i} \quad(\text { where } 1 \leq i \leq n)
$$

### 2.3 Numbers

There are many ways to represent numbers by $\lambda$-expressions, each with their own advantages and disadvantages [38, 22]. The goal is to define for each number $n$ a $\lambda$ expression $\underline{n}$ that represents it. We also want to define $\lambda$-expressions to represent the primitive arithmetical operations. For example, we will need $\lambda$-expressions suc, pre, add and iszero representing the successor function $(n \mapsto n+1)$, the predecessor function ( $n \mapsto n-1$ ), addition and a test for zero, respectively. These $\lambda$-expressions will represent the numbers correctly if they have the following properties:

$$
\begin{aligned}
& \text { suc } \underline{n}=\underline{n+1} \quad(\text { for all numbers } n) \\
& \text { pre } \underline{n}=\underline{n-1} \quad(\text { for all numbers } n) \\
& \text { add } \underline{m} \underline{n}=\underline{m+n} \quad(\text { for all numbers } m \text { and } n) \\
& \text { iszero } \underline{0}=\text { true } \\
& \text { iszero }(\operatorname{suc} \underline{n})=\text { false }
\end{aligned}
$$

The representation of numbers described here is the original one due to Church. In order to explain this it is convenient to define $f^{n} x$ to mean $n$ applications of $f$ to $x$. For example,

$$
f^{5} x=f(f(f(f(f x))))
$$

By convention $f^{0} x$ is defined to mean $x$. More generally:

$$
\begin{aligned}
& \operatorname{LET} E^{0} E^{\prime}=E^{\prime} \\
& \operatorname{LET} E^{n} E^{\prime}=\underbrace{E(E(\cdots(E}_{n E \mathrm{~s}} E^{\prime}) \cdots))
\end{aligned}
$$

Note that $E^{n}\left(E E^{\prime}\right)=E^{n+1} E^{\prime}=E\left(E^{n} E^{\prime}\right)$; we will use the fact later.

## Example:

$$
f^{4} x=f(f(f(f x)))=f\left(f^{3} x\right)=f^{3}(f x)
$$

Using the notation just introduced we can now define Church's numerals. Notice how the definition of the $\lambda$-expression $\underline{n}$ below encodes a unary representation of $n$.

| LET $\underline{0}=\lambda f x . x$ |
| :---: |
| LET $\underline{1}=\lambda f x . f x$ |
| LET $\underline{2}=\lambda f x . f(f x)$ |
| $\vdots$ |
| LET $\underline{n}=\lambda f x . f^{n} x$ |
| $\vdots$ |

The representations of suc, add and iszero are now magically pulled out of a hat. The best way to see how they work is to think of them as operating on unary representations of numbers. The exercises that follow should help.

$$
\begin{aligned}
& \text { LET suc }=\lambda n f x . n f(f x) \\
& \text { LET add }=\lambda m n f x . m f(n f x) \\
& \text { LET iszero }=\lambda n . n(\lambda x . \text { false }) \text { true }
\end{aligned}
$$

## Exercise 16

Show:
(i) $\operatorname{suc} \underline{0}=\underline{1}$
(ii) $\operatorname{suc} \underline{5}=\underline{6}$
(iii) iszero $\underline{0}=$ true
(iv) iszero $\underline{5}=$ false
(v) $\boldsymbol{\operatorname { a d d }} \underline{0} \underline{1}=\underline{1}$
(vi) $\operatorname{add} \underline{2} \underline{3}=\underline{5}$

## Exercise 17

Show for all numbers $m$ and $n$ :
(i) $\operatorname{suc} \underline{n}=\underline{n+1}$
(ii) iszero (suc $\underline{n}$ ) $=$ false
(iii) add $\underline{0} \underline{n}=\underline{n}$
(iv) add $\underline{m} \underline{0}=\underline{m}$
(v) add $\underline{m} \underline{n}=\underline{m+n}$

The predecesor function is harder to define than the other primitive functions. The idea is that the predecessor of $\underline{n}$ is defined by using $\lambda f x . f^{n} x$ (i.e. $\underline{n}$ ) to obtain a function that applies $f$ only $n-1$ times. The trick is to 'throw away' the first application of $f$ in $f^{n}$. To achieve this, we first define a function prefn that operates on pairs and has the property that:
(i) prefn $f($ true,$x)=($ false,$x)$
(ii) prefn $f($ false, $x)=($ false, $f x)$

From this it follows that:
(iii) $(\text { prefn } f)^{n}($ false, $x)=\left(\right.$ false, $\left.f^{n} x\right)$
(iv) $(\text { prefn } f)^{n}($ true,$x)=\left(\right.$ false, $\left.f^{n-1} x\right) \quad($ if $n>0)$

Thus $n$ applications of prefn to (true, $x$ ) result in $n-1$ applications of $f$ to $x$. With this idea, the definition of the predecessor function pre is straightforward. Before giving it, here is the definition of prefn:

$$
\text { LET prefn }=\lambda f p .(\text { false },(\text { fst } p \rightarrow \text { snd } p \mid(f(\text { snd } p))))
$$

## Exercise 18

Show prefn $f(b, x)=($ false,$(b \rightarrow x \mid f x))$ and hence:
(i) prefn $f($ true,$x)=($ false,$x)$
(ii) prefn $f$ (false, $x)=($ false, $f x)$
(iii) $(\text { prefn } f)^{n}($ false,$x)=\left(\right.$ false, $\left.f^{n} x\right)$
(iv) $(\mathbf{p r e f n} f)^{n}($ true,$x)=\left(\right.$ false, $\left.f^{n-1} x\right) \quad($ if $n>0)$

The predecessor function pre can now be defined.

$$
\text { LET pre }=\lambda n f x . \text { snd }(n(\text { prefn } f)(\text { true }, x))
$$

It follows that if $n>0$

$$
\begin{array}{rlr}
\text { pre } \underline{n} f x & =\operatorname{snd}(\underline{n}(\operatorname{prefn} f)(\text { true }, x)) & \text { (definition of pre) } \\
& =\operatorname{snd}\left((\text { prefn } f)^{n}(\text { true }, x)\right) & \text { (definition of } \underline{n}) \\
& =\operatorname{snd}\left(\text { false, } f^{n-1} x\right) & \text { (by (v) above) } \\
& =f^{n-1} x &
\end{array}
$$

hence by extensionality (Section 1.7 on page 10)

$$
\begin{aligned}
\text { pre } \underline{n} & \left.=\lambda f x \cdot f^{n-1} x \quad \text { (definition of } \underline{n-1}\right) \\
& =\underline{n-1}
\end{aligned}
$$

## Exercise 19

Using the results of the previous exercise (or otherwise) show that
(i) $\operatorname{pre}(\operatorname{suc} \underline{n})=\underline{n}$
(ii) pre $\underline{0}=\underline{0}$
$\square$

The numeral system in the next exercise is the one used in [2] and has some advantages over Church's (e.g. the predecessor function is easier to define).

## Exercise 20

$$
\begin{aligned}
& \text { LET } \underline{\hat{0}}=\lambda x . x \\
& \text { LET } \underline{\hat{1}}=(\text { false }, \underline{\widehat{0}}) \\
& \text { LET } \underline{\widehat{2}}=(\text { false }, \underline{\underline{1}}) \\
& \vdots \\
& \text { LET } \underline{n+1}=(\text { false }, \underline{\underline{n}})
\end{aligned}
$$

Devise $\lambda$-expressions $\widehat{\text { suc }}, \widehat{\text { iszero }}, \widehat{\text { pre }}$ such that for all $n$ :
(i) $\widehat{\operatorname{suc}} \underline{\widehat{n}}=\widehat{n+1}$
(ii) iszero $\widehat{\widehat{0}}=$ true
(iii) iszero $(\widehat{\operatorname{suc}} \underline{\widehat{n}})=$ false
(iv) $\widehat{\text { pre }}(\widehat{\operatorname{suc}} \underline{\widehat{n}})=\underline{\widehat{n}}$

### 2.4 Definition by recursion

To represent the multiplication function in the $\lambda$-calculus we would like to define a $\lambda$-expression, mult say, such that:

$$
\text { mult } m n=\underbrace{\operatorname{add} n(\mathbf{a d d} n(\cdots(\mathbf{a d d} n \underline{0}) \cdots))}_{m \mathbf{a d d}_{s}}
$$

This would be achieved if mult could be defined to satisfy the equation:

$$
\text { mult } m n=(\text { iszero } m \rightarrow \underline{0} \mid \text { add } n(\mathbf{m u l t}(\text { pre } m) n))
$$

If this held then, for example,

```
mult \underline{2}\underline{3}=(\mathrm{ iszero }\underline{2}->\underline{0}|\mathrm{ add }\underline{3}(\mathrm{ mult (pre 2)}\underline{3}))
```

$=\operatorname{add} \underline{3}(\mathbf{m u l t} \underline{1} \underline{3})$
(by properties of iszero, the conditional and pre)
$=\operatorname{add} \underline{3}($ iszero $\underline{1} \rightarrow \underline{0} \mid$ add $\underline{3}($ mult $($ pre $\underline{1}) \underline{3}))$
(by the equation)
$=\operatorname{add} \underline{3}(\operatorname{add} \underline{3}(\operatorname{mult} \underline{0} \underline{3}))$
(by properties of iszero, the conditional and pre)
$=\operatorname{add} \underline{3}(\operatorname{add} \underline{3}($ iszero $\underline{0} \rightarrow \underline{0} \mid \operatorname{add} \underline{3}(\operatorname{mult}($ pre $\underline{0}) \underline{3})))$
(by the equation)
$=\mathbf{a d d} \underline{3}(\boldsymbol{\operatorname { a d d }} \underline{3} \underline{0})$
(by properties of iszero and the conditional)
The equation above suggests that mult be defined by:

$$
\mathbf{m u l t}=\lambda m n .(\text { iszero } m \rightarrow \underline{0} \mid \text { add } n(\mathbf{m u l t}(\text { pre } m) n))
$$

Unfortunately, this cannot be used to define mult because, as indicated by the arrow, mult must already be defined for the $\lambda$-expression to the right of the equals to make sense.
Fortunately, there is a technique for constructing $\lambda$-expressions that satisfy arbitrary equations. When this technique is applied to the equation above it gives the desired definition of mult. First define a $\lambda$-expression $\mathbf{Y}$ that, for any expression $E$, has the following odd property:

$$
\mathbf{Y} E=E(\mathbf{Y} E)
$$

This says that $\mathbf{Y} E$ is unchanged when the function $E$ is applied to it. In general, if $E E^{\prime}=E^{\prime}$ then $E^{\prime}$ is called a fixed point of $E$. A $\lambda$-expression Fix with the property that Fix $E=E(\mathbf{F i x} E)$ for any $E$ is called a fixed-point operator. There are known to be infinitely many different fixed-point operators [28]; $\mathbf{Y}$ is the most famous one, and its definition is:

$$
\operatorname{LET} \mathbf{Y}=\lambda f .(\lambda x \cdot f(x x))(\lambda x \cdot f(x x))
$$

It is straightforward to show that $\mathbf{Y}$ is indeed a fixed-point operator:

$$
\begin{array}{rlr}
\mathbf{Y} E & =(\lambda f .(\lambda x \cdot f(x x))(\lambda x \cdot f(x x))) E & \text { (definition of } \mathbf{Y}) \\
& =(\lambda x \cdot E(x x))(\lambda x \cdot E(x x)) & (\beta \text {-conversion) } \\
& =E((\lambda x \cdot E(x x))(\lambda x \cdot E(x x))) & (\beta \text {-conversion) } \\
& =E(\mathbf{Y} E) & \text { (the line before last) }
\end{array}
$$

This calculation shows that every $\lambda$-expression $E$ has a fixed point.
Armed with $\mathbf{Y}$, we can now return to the problem of solving the equation for mult.
Suppose multfn is defined by
LET multfn $=\lambda f m n .($ iszero $m \rightarrow \underline{0} \mid$ add $n(f($ pre $m) n))$
and then define mult by:

$$
\text { LET mult }=\mathbf{Y} \text { multfn }
$$

Then:

$$
\begin{array}{rlr}
\text { mult } m n & =(\mathbf{Y} \text { multfn } m n & \text { (definition of mult) } \\
& =\text { multfn }(\mathbf{Y} \text { multfn }) m n & \text { (fixed-point property of } \mathbf{Y}) \\
& =\text { multfn mult } m n & (\text { definition of mult) } \\
& =(\lambda f m n .(\mathbf{i s z e r o} m \rightarrow \underline{0} \mid \text { add } n(f(\text { pre } m) n))) \text { mult } m n \\
& =(\text { iszero } m \rightarrow \underline{0} \mid \text { add } n(\text { mult }(\text { pre } m) n)) & (\beta \text {-conversion) })
\end{array}
$$

An equation of the form $f x_{1} \cdots x_{n}=E$ is called recursive if $f$ occurs free in $E$. Y provides a general way of solving such equations. Start with an equation of the form:

$$
\mathbf{f} x_{1} \ldots x_{n}=\sim \mathbf{f} \sim
$$

where $\sim \mathbf{f} \sim$ is some $\lambda$-expression containing $\mathbf{f}$. To obtain an $\mathbf{f}$ so that this equation holds define:

$$
\operatorname{LET} \mathbf{f}=\mathbf{Y}\left(\lambda f x_{1} \ldots x_{n} \sim f \sim\right)
$$

The fact that the equation is satisfied can be shown as follows:

$$
\begin{array}{rlr}
\mathbf{f} x_{1} \ldots x_{n} & =\mathbf{Y}\left(\lambda f x_{1} \ldots x_{n} \sim f \sim\right) x_{1} \ldots x_{n} & \text { (definition of } \mathbf{f}) \\
& =\left(\lambda f x_{1} \ldots x_{n} \sim f \sim\right)\left(\mathbf{Y}\left(\lambda f x_{1} \ldots x_{n} \sim f \sim\right)\right) x_{1} \ldots x_{n}
\end{array}
$$

(fixed-point property)

$$
=\left(\lambda f x_{1} \ldots x_{n} \sim f \sim\right) \mathbf{f} x_{1} \ldots x_{n} \quad \quad(\text { definition of } \mathbf{f})
$$

$$
=\sim \mathbf{f}_{\sim}
$$

## Exercise 21

Construct a $\lambda$-expression eq such that

$$
\begin{aligned}
\text { eq } m n= & (\text { iszero } m \rightarrow \text { iszero } n \mid \\
& (\text { iszero } n \rightarrow \text { false } \mid \text { eq }(\text { pre } m)(\text { pre } n)))
\end{aligned}
$$

## Exercise 22

Show that if $\mathbf{Y}_{1}$ is defined by:

$$
\operatorname{LET} \mathbf{Y}_{1}=\mathbf{Y}(\lambda y f . f(y f))
$$

then $\mathbf{Y}_{1}$ is a fixed-point operator, i.e. for any $E$ :

$$
\mathbf{Y}_{1} E=E\left(\mathbf{Y}_{1} E\right)
$$

The fixed-point operator in the next exercise is due to Turing (Barendregt [2], page 132).

## Exercise 23

Show that $(\lambda x y . y(x x y))(\lambda x y . y(x x y))$ is a fixed-point operator.
The next exercise also comes from Barendregt's book, where it is attributed to Klop.

## Exercise 24

Show that $\mathbf{Y}_{2}$ is a fixed-point operator, where:

$$
\begin{aligned}
\text { LET } £= & \lambda a b c d e f g h i j k l m n o p q s t u v w x y z r . \\
& r(\text { thisisafixedpointcombinator }) \\
\text { LET } \mathbf{Y}_{2}= & £ £ £ £ £ £ £ £ £ £ £ £ £ £ £ £ £ £ £ £ £ £ £ £ £ £
\end{aligned}
$$

## Exercise 25

Is it the case that $\mathbf{Y} f \longrightarrow f(\mathbf{Y} f)$ ? If so prove it; if not find a $\lambda$-expression $\widehat{\mathbf{Y}}$ such that $\widehat{\mathbf{Y}} f \longrightarrow f(\widehat{\mathbf{Y}} f)$.

In the pure $\lambda$-calculus as defined on page $1, \lambda$-expressions could only be applied to a single argument; however, this argument could be a tuple (see page 16). Thus one can write:

$$
E\left(E_{1}, \ldots, E_{n}\right)
$$

which actually abbreviates:

$$
E\left(E_{1},\left(E_{2},\left(\cdots\left(E_{n-1}, E_{n}\right) \cdots\right)\right)\right)
$$

For example, $E\left(E_{1}, E_{2}\right)$ abbreviates $E\left(\lambda f . f E_{1} E_{2}\right)$.

### 2.5 Functions with several arguments

In conventional mathematical usage, the application of an $n$-argument function $f$ to arguments $x_{1}, \ldots, x_{n}$ would be written as $f\left(x_{1}, \ldots, x_{n}\right)$. There are two ways of representing such applications in the $\lambda$-calculus:
(i) as $\left(\begin{array}{llll}f & x_{1} & \ldots & x_{n}\end{array}\right)$, or
(ii) as the application of $f$ to an $n$-tuple $\left(x_{1}, \ldots, x_{n}\right)$.

In case (i), $f$ expects its arguments 'one at a time' and is said to be curried after a logician called Curry (the idea of currying was actually invented by Schönfinkel [31]). The functions and, or and add defined earlier were all curried. One advantage of curried functions is that they can be 'partially applied'; for example, add 1 is the result of partially applying add to $\underline{1}$ and denotes the function $n \mapsto n+1$.
Although it is often convenient to represent $n$-argument functions as curried, it is also useful to be able to represent them, as in case (ii) above, by $\lambda$-expressions expecting a single $n$-tuple argument. For example, instead of representing + and $\times$ by $\lambda$-expressions add and mult such that

$$
\begin{aligned}
\text { add } \underline{m} \underline{n} & =\underline{m+n} \\
\text { mult } \underline{m} \underline{n} & =\underline{m \times n}
\end{aligned}
$$

it might be more convenient to represent them by functions, sum and prod say, such that

$$
\begin{aligned}
\operatorname{sum}(\underline{m}, \underline{n}) & =\underline{m+n} \\
\operatorname{prod}(\underline{m}, \underline{n}) & =\underline{m \times n}
\end{aligned}
$$

This is nearer to conventional mathematical usage and has applications that will be encountered later. One might say that sum and prod are uncurried versions of add and mult respectively.
Define:

```
LET curry \(=\lambda f x_{1} x_{2} . f\left(x_{1}, x_{2}\right)\)
LET uncurry \(=\lambda f p . f(\mathbf{f s t} p)(\operatorname{snd} p)\)
```

then defining

$$
\begin{array}{ll}
\operatorname{sum}= & \text { uncurry add } \\
\text { prod }= & \text { uncurry mult }
\end{array}
$$

results in sum and prod having the desired properties; for example:

$$
\begin{aligned}
\operatorname{sum}(\underline{m}, \underline{n}) & =\operatorname{uncurry} \operatorname{add}(\underline{m}, \underline{n}) \\
& =(\lambda f p . f(\mathbf{f s t} p)(\operatorname{snd} p)) \mathbf{a d d}(\underline{m}, \underline{n}) \\
& =\operatorname{add}(\mathbf{f s t}(\underline{m}, \underline{n}))(\operatorname{snd}(\underline{m}, \underline{n})) \\
& =\operatorname{add} \underline{m} \underline{n} \\
& =\underline{m+n}
\end{aligned}
$$

## Exercise 26

Show that for any $E$ :

$$
\begin{aligned}
& \operatorname{curry}(\text { uncurry } E)=E \\
& \text { uncurry }(\operatorname{curry} E)=E
\end{aligned}
$$

hence show that:

$$
\begin{aligned}
& \text { add }=\text { curry sum } \\
& \text { mult }=\text { curry prod }
\end{aligned}
$$

We can define $n$-ary functions for currying and uncurrying. For $n>0$ define:

$$
\begin{aligned}
& {\text { LET } \operatorname{curry}_{n}=\lambda f x_{1} \cdots x_{n} . f\left(x_{1}, \ldots, x_{n}\right)}^{\text {LET uncurry }_{n}=\lambda f p . f(p \stackrel{n}{\downarrow} 1) \cdots(p \stackrel{n}{\downarrow} n)}
\end{aligned}
$$

If $E$ represents a function expecting an $n$-tuple argument, then curry ${ }_{n} E$ represents the curried function which takes its arguments one at a time. If $E$ represents a curried function of $n$ arguments, then uncurry ${ }_{n} E$ represents the uncurried version which expects a single $n$-tuple as argument.

## Exercise 27

Show that:
(i) $\operatorname{curry}_{n}\left(\operatorname{uncurry}_{n} E\right)=E$
(ii) $\operatorname{uncurry}_{n}\left(\operatorname{curry}_{n} E\right)=E$

## Exercise 28

Devise $\lambda$-expressions $E_{1}^{n}$ and $E_{2}^{n}$ built out of curry and uncurry such that $\operatorname{curry}_{n}=E_{1}^{n}$ and uncurry ${ }_{n}=E_{2}^{n}$.

The following notation provides a convenient way to write $\lambda$-expressions which expect tuples as arguments.

## Generalized $\lambda$-abstractions

$$
\operatorname{LET} \lambda\left(V_{1}, \ldots, V_{n}\right) . E=\operatorname{uncurry}_{n}\left(\lambda V_{1} \ldots V_{n} . E\right)
$$

Example: $\lambda(x, y)$. mult $x y$ abbreviates:

$$
\begin{aligned}
\text { uncurry }_{2}(\lambda x y . \text { mult } x y) & =(\lambda f p . f(p \stackrel{2}{\downarrow} 1)(p \stackrel{2}{\downarrow} 2))(\lambda x y . \text { mult } x y) \\
& =(\lambda f p . f(\mathbf{f s t} p)(\text { snd } p))(\lambda x y . \text { mult } x y) \\
& =\lambda p . \text { mult }(\mathbf{f s t} p)(\text { snd } p)
\end{aligned}
$$

Thus:

$$
\begin{aligned}
(\lambda(x, y) . \text { mult } x y)\left(E_{1}, E_{2}\right) & =(\lambda p . \operatorname{mult}(\mathbf{f s t} p)(\operatorname{snd} p))\left(E_{1}, E_{2}\right) \\
& =\operatorname{mult}\left(\mathbf{f s t}\left(E_{1}, E_{2}\right)\right)\left(\operatorname{snd}\left(E_{1}, E_{2}\right)\right) \\
& =\operatorname{mult} E_{1} E_{2}
\end{aligned}
$$

This example illustrates the rule of generalized $\beta$-conversion in the box below. This rule can be derived from ordinary $\beta$-conversion and the definitions of tuples and generalized $\lambda$-abstractions. The idea is that a tuple of arguments is passed to each argument position in the body of the generalized abstraction; then each individual argument can be extracted from the tuple without affecting the others.

$$
\begin{aligned}
& \qquad \text { Generalized } \beta \text {-conversion } \\
& \qquad\left(\lambda\left(V_{1}, \ldots, V_{n}\right) . E\right)\left(E_{1}, \ldots, E_{n}\right)=E\left[E_{1}, \ldots, E_{n} / V_{1}, \ldots, V_{n}\right] \\
& \text { where } E\left[E_{1}, \ldots, E_{n} / V_{1}, \ldots, V_{n}\right] \text { is the simultaneous substitution of } E_{1}, \ldots, E_{n} \\
& \text { for } V_{1}, \ldots, V_{n} \text { respectively and none of these variables occur free in any of } \\
& E_{1}, \ldots, E_{n} .
\end{aligned}
$$

It is convenient to extend the notation $\lambda V_{1} V_{2} \ldots V_{n} . E$ described on page 3 so that each $V_{i}$ can either be an identifier or a tuple of identifiers. The meaning of $\lambda V_{1} V_{2} \ldots V_{n} . E$ is still $\lambda V_{1} \cdot\left(\lambda V_{2} \cdot\left(\cdots\left(\lambda V_{n} . E\right) \cdots\right)\right)$, but now if $V_{i}$ is a tuple of identifiers then the expression is a generalized abstraction.

Example: $\lambda f(x, y) . f x y$ means $\lambda f .(\lambda(x, y) . f x y)$ which in turn means $\lambda f$. uncurry $(\lambda x y . f x y)$ which equals $\lambda f .(\lambda p . f($ fst $p)($ snd $p))$.

## Exercise 29

Show that if the only free variables in $E$ are $x_{1}, \ldots, x_{n}$ and $f$, then if:

$$
\mathbf{f}=\mathbf{Y}\left(\lambda f\left(x_{1}, \ldots, x_{n}\right) . E\right)
$$

then

$$
\mathbf{f}\left(x_{1}, \ldots, x_{n}\right)=E[\mathbf{f} / f]
$$

## Exercise 30

Define a $\lambda$-expression div with the property that:

$$
\operatorname{div}(\underline{m}, \underline{n})=(\underline{q}, \underline{r})
$$

where $q$ and $r$ are the quotient and remainder of dividing $n$ into $m$.

### 2.6 Mutual recursion

To solve a set of mutually recursive equations like:

$$
\begin{aligned}
\mathbf{f}_{1} & =F_{1} \mathbf{f}_{1} \cdots \mathbf{f}_{n} \\
\mathbf{f}_{2} & =F_{2} \mathbf{f}_{1} \cdots \mathbf{f}_{n} \\
\vdots & \\
\mathbf{f}_{n} & =F_{n} \mathbf{f}_{1} \cdots \mathbf{f}_{n}
\end{aligned}
$$

we simply define for $1 \leq i \leq n$

$$
\mathbf{f}_{i}=\mathbf{Y}\left(\lambda\left(f_{1}, \ldots f_{n}\right) .\left(F_{1} f_{1} \cdots f_{n}, \ldots, F_{n} f_{1} \cdots f_{n}\right)\right) \downarrow i
$$

This works because if

$$
\overrightarrow{\mathbf{f}}=\mathbf{Y}\left(\lambda\left(f_{1}, \ldots f_{n}\right) \cdot\left(F_{1} f_{1} \cdots f_{n}, \ldots, F_{n} f_{1} \cdots f_{n}\right)\right)
$$

then $\mathbf{f}_{i}=\overrightarrow{\mathbf{f}} \downarrow i$ and hence:

$$
\begin{aligned}
\overrightarrow{\mathbf{f}} & =\left(\lambda\left(f_{1}, \ldots, f_{n}\right) .\left(F_{1} f_{1} \cdots f_{n}, \ldots, F_{n} f_{1} \cdots f_{n}\right)\right) \overrightarrow{\mathbf{f}} \\
& =\left(F_{1}(\overrightarrow{\mathbf{f}} \downarrow 1) \cdots(\overrightarrow{\mathbf{f}} \downarrow n), \ldots, F_{n}(\overrightarrow{\mathbf{f}} \downarrow 1) \cdots(\overrightarrow{\mathbf{f}} \downarrow n)\right) \\
& \left.=\left(F_{1} \mathbf{f}_{1} \cdots \mathbf{f}_{n}, \ldots, F_{n} \mathbf{f}_{1} \cdots \mathbf{f}_{n}\right) \quad \text { (since } \overrightarrow{\mathbf{f}} \downarrow i=\mathbf{f}_{i}\right) .
\end{aligned}
$$

Hence:

$$
\mathbf{f}_{i}=F_{i} \mathbf{f}_{1} \cdots \mathbf{f}_{n}
$$

### 2.7 Representing the recursive functions

The recursive functions form an important class of numerical functions. Shortly after Church invented the $\lambda$-calculus, Kleene proved that every recursive function could be represented in it. This provided evidence for Church's thesis, the hypothesis that any intuitively computable function could be represented in the $\lambda$-calculus. It has been shown that many other models of compution define the same class of functions that can be defined in the $\lambda$-calculus.
In this section it is described what it means for an arithmetic function to be represented in the $\lambda$-calculus. Two classes of functions, the primitive recursive functions and the recursive functions, are defined and it is shown that all the functions in these classes can be represented in the $\lambda$-calculus.
In Section 2.3 it was explained how a number $n$ is represented by the $\lambda$-expression $\underline{n}$. A $\lambda$-expression $f$ is said to represent a mathematical function $f$ if for all numbers $x_{1}, \ldots, x_{n}$ :

$$
\underline{f}\left(\underline{x_{1}}, \ldots, \underline{x_{n}}\right)=\underline{y} \quad \text { if } \quad f\left(x_{1}, \ldots, x_{n}\right)=y
$$

### 2.7.1 The primitive recursive functions

A function is called primitive recursive if it can be constructed from 0 and the functions $S$ and $U_{n}^{i}$ (defined below) by a finite sequence of applications of the operations of substitution and primitive recursion (also defined below).
The successor function $S$ and projection functions $U_{n}^{i}$ (where $n$ and $i$ are numbers) are defined by:
(i) $S(x)=x+1$
(ii) $U_{n}^{i}\left(x_{1}, x_{2}, \ldots, x_{n}\right)=x_{i}$

## Substitution

Suppose $g$ is a function of $r$ arguments and $h_{1}, \ldots, h_{r}$ are $r$ functions each of $n$ arguments. We say $f$ is defined from $g$ and $h_{1}, \ldots, h_{r}$ by substitution if:

$$
f\left(x_{1}, \ldots, x_{n}\right)=g\left(h_{1}\left(x_{1}, \ldots, x_{n}\right), \ldots, h_{r}\left(x_{1}, \ldots, x_{n}\right)\right)
$$

## Primitive recursion

Suppose $g$ is a function of $n-1$ arguments and $h$ is a function of $n+1$ arguments. We say $f$ is defined from $g$ and $h$ by primitive recursion if:

$$
\begin{aligned}
f\left(0, x_{2}, \ldots, x_{n}\right) & =g\left(x_{2}, \ldots, x_{n}\right) \\
f\left(S\left(x_{1}\right), x_{2}, \ldots, x_{n}\right) & =h\left(f\left(x_{1}, x_{2}, \ldots, x_{n}\right), x_{1}, x_{2}, \ldots, x_{n}\right)
\end{aligned}
$$

$g$ is called the base function and $h$ is called the step function. It can proved that for any base and step function there always exists a unique function defined from them by primitive recursion. This result is called the primitive recursion theorem; proofs of it can be found in textbooks on mathematical logic.

Example: The addition function sum is primitive recursive because:

$$
\begin{aligned}
\operatorname{sum}\left(0, x_{2}\right) & =x_{2} \\
\operatorname{sum}\left(S\left(x_{1}\right), x_{2}\right) & =S\left(\operatorname{sum}\left(x_{1}, x_{2}\right)\right)
\end{aligned}
$$

It is now shown that every primitive recursive function can be represented by $\lambda$ expressions.
It is obvious that the $\lambda$-expressions $\underline{0}$, suc, $\lambda p . p \stackrel{n}{\downarrow} i$ represent the initial functions $0, S$ and $U_{n}^{i}$ respectively.
Suppose function $g$ of $r$ variables is represented by $\mathbf{g}$ and functions $h_{i}(1 \leq i \leq r)$ of $n$ variables are represented by $\mathbf{h}_{i}$. Then if a function $f$ of $n$ variables is defined by substitution by:

$$
f\left(x_{1}, \ldots, x_{n}\right)=g\left(h_{1}\left(x_{1}, \ldots, x_{n}\right), \ldots, h_{r}\left(x_{1}, \ldots, x_{n}\right)\right)
$$

then $f$ is represented by $\mathbf{f}$ where:

$$
\mathbf{f}=\lambda\left(x_{1}, \ldots, x_{n}\right) \cdot \mathbf{g}\left(\mathbf{h}_{1}\left(x_{1}, \ldots, x_{n}\right), \ldots, \mathbf{h}_{r}\left(x_{1}, \ldots, x_{n}\right)\right)
$$

Suppose function $f$ of $n$ variables is defined inductively from a base function $g$ of $n-1$ variables and an inductive step function $h$ of $n+1$ variables. Then

$$
\begin{aligned}
f\left(0, x_{2}, \ldots, x_{n}\right) & =g\left(x_{2}, \ldots, x_{n}\right) \\
f\left(S\left(x_{1}\right), x_{2}, \ldots, x_{n}\right) & =h\left(f\left(x_{1}, x_{2}, \ldots, x_{n}\right), x_{1}, x_{2}, \ldots, x_{n}\right)
\end{aligned}
$$

Thus if $\mathbf{g}$ represents $g$ and $\mathbf{h}$ represents $h$ then $\mathbf{f}$ will represent $f$ if

$$
\begin{aligned}
\mathbf{f}\left(x_{1}, x_{2}, \ldots,\right. & \left.x_{n}\right)= \\
\quad\left(\text { iszero } x_{1}\right. & \rightarrow \mathbf{g}\left(x_{2}, \ldots, x_{n}\right) \mid \\
& \left.\mathbf{h}\left(\mathbf{f}\left(\text { pre } x_{1}, x_{2}, \ldots, x_{n}\right), \text { pre } x_{1}, x_{2}, \ldots, x_{n}\right)\right)
\end{aligned}
$$

Using the fixed-point trick, an $\mathbf{f}$ can be constructed to satisfy this equation by defining $\mathbf{f}$ to be:

$$
\begin{aligned}
& \mathbf{Y}\left(\lambda f . \lambda\left(x_{1}, x_{2}, \ldots, x_{n}\right)\right. \\
& \quad\left(\text { iszero } x_{1}\right. \rightarrow \mathbf{g}\left(x_{2}, \ldots, x_{n}\right) \mid \\
&\left.\left.\mathbf{h}\left(f\left(\text { pre } x_{1}, x_{2}, \ldots, x_{n}\right), \text { pre } x_{1}, x_{2}, \ldots, x_{n}\right)\right)\right)
\end{aligned}
$$

Thus any primitive recursive function can be represented by a $\lambda$-expression.

### 2.7.2 The recursive functions

A function is called recursive if it can be constructed from 0 , the successor function and the projection functions (see page 26) by a sequence of substitutions, primitive recursions and minimizations.

## Minimization

Suppose $g$ is a function of $n$ arguments. We say $f$ is defined from $g$ by minimization if:

$$
f\left(x_{1}, x_{2}, \ldots, x_{n}\right)=\text { 'the smallest } y \text { such that } g\left(y, x_{2}, \ldots, x_{n}\right)=x_{1} \text { ' }
$$

The notation $\operatorname{MIN}(f)$ is used to denote the minimization of $f$. Functions defined by minimization may be undefined for some arguments. For example, if one is the function that always returns 1, i.e. one $(x)=1$ for every $x$, then $\operatorname{MIN}$ (one) is only defined for arguments with value 1. This is obvious because if $f(x)=\operatorname{MIN}($ one $)(x)$, then:

$$
f(x)=\text { 'the smallest } y \text { such that one }(y)=x \text { ' }
$$

and clearly this is only defined if $x=1$. Thus

$$
\operatorname{MIN}(\text { one })(x)= \begin{cases}\underline{0} & \text { if } x=1 \\ \text { undefined } & \text { otherwise }\end{cases}
$$

To show that any recursive function can be represented in the $\lambda$-calculus it is necessary to show how to represent the minimization of an arbitrary function. Suppose $\mathbf{g}$ represents a function $g$ of $n$ variables and $f$ is defined by:

$$
f=\operatorname{MIN}(g)
$$

Then if a $\lambda$-expression min can be devised such that $\min \underline{x} \mathbf{f}\left(\underline{x}_{1}, \ldots, \underline{x}_{n}\right)$ represents the least number $y$ greater than $x$ such that

$$
f\left(y, x_{2}, \ldots, x_{n}\right)=x_{1}
$$

then $\mathbf{g}$ will represent $g$ where:

$$
\mathbf{g}=\lambda\left(x_{1}, x_{2}, \ldots, x_{n}\right) \cdot \min \underline{0} \mathbf{f}\left(x_{1}, x_{2}, \ldots, x_{n}\right)
$$

min will clearly have the desired property if:

$$
\begin{aligned}
& \min x f\left(x_{1}, x_{2}, \ldots, x_{n}\right)= \\
& \left.\quad\left(\mathbf{e q}\left(f\left(x, x_{2}, \ldots, x_{n}\right)\right) x_{1}\right) \rightarrow x \mid \min (\text { suc } x) f\left(x_{1}, x_{2}, \ldots, x_{n}\right)\right)
\end{aligned}
$$

where eq $\underline{m} \underline{n}$ is equal to true if $m=n$ and false otherwise (a suitable definition of eq occurs on page 21). Thus min can simply be defined to be:

$$
\begin{aligned}
& \mathbf{Y}(\lambda m . \\
& \quad \lambda x f\left(x_{1}, x_{2}, \ldots, x_{n}\right) \\
& \left.\quad\left(\mathbf{e q}\left(f\left(x, x_{2}, \ldots, x_{n}\right)\right) x_{1} \rightarrow x \mid m(\text { suc } x) f\left(x_{1}, x_{2}, \ldots, x_{n}\right)\right)\right)
\end{aligned}
$$

Thus any recursive function can be represented by a $\lambda$-expression.

## Higher-order primitive recursion

There are functions which are recursive but not primitive recursive. Here is a version of Ackermann's function, $\psi$, defined by:

$$
\begin{aligned}
& \psi(0, n)=n+1 \\
& \psi(m+1,0)=\psi(m, 1) \\
& \psi(m+1, n+1)=\psi(m, \psi(m+1, n))
\end{aligned}
$$

However, if one allows functions as arguments, then many more recursive functions can be defined by a primitive recursion. For example, if the higher-order function rec is defined by primitive recursion as follows:

$$
\begin{aligned}
& \operatorname{rec}\left(0, x_{2}, x_{3}\right)=x_{2} \\
& \operatorname{rec}\left(S\left(x_{1}\right), x_{2}, x_{3}\right)=x_{3}\left(\operatorname{rec}\left(x_{1}, x_{2}, x_{3}\right)\right)
\end{aligned}
$$

then $\psi$ can be defined by:

$$
\psi(m, n)=\operatorname{rec}(m, S, f \mapsto(x \mapsto \operatorname{rec}(x, f(1), f)))(n)
$$

where $x \mapsto \theta(x)$ denotes the function ${ }^{1}$ that maps $x$ to $\theta(x)$. Notice that the third argument of rec, viz. $x_{3}$, must be a function. In the definition of $\psi$ we also took $x_{2}$ to be a function, viz. $S$.

[^1]
## Exercise 31

Show that the definition of $\psi$ in terms of rec works, i.e. that with $\psi$ defined as above:

$$
\begin{aligned}
& \psi(0, n)=n+1 \\
& \psi(m+1,0)=\psi(m, 1) \\
& \psi(m+1, n+1)=\psi(m, \psi(m+1, n))
\end{aligned}
$$

A function which takes another function as an argument, or returns another function as a result, is called higher-order. The example $\psi$ shows that higher-order primitive recursion is more powerful than ordinary primitive recursion ${ }^{2}$. The use of operators like $r e c$ is one of the things that makes functional programming very powerful.

### 2.7.3 The partial recursive functions

A partial function is one that is not defined for all arguments. For example, the function $\operatorname{MIN}$ (one) described above is partial. Another example is the division function, since division by 0 is not defined. Functions that are defined for all arguments are called total.
A partial function is called partial recursive if it can be constructed from 0 , the successor function and the projection functions by a sequence of substitutions, primitive recursions and minimizations. Thus the recursive functions are just the partial recursive functions which happen to be total. It can be shown that every partial recursive function $f$ can be represented by a $\lambda$-expression $\underline{f}$ in the sense that
(i) $\underline{f}\left(\underline{x_{1}}, \ldots, \underline{x_{n}}\right)=\underline{y}$ if $f\left(x_{1}, \ldots, x_{n}\right)=y$
(ii) If $f\left(x_{1}, \ldots, x_{n}\right)$ is undefined then $\underline{f}\left(\underline{x_{1}}, \ldots, \underline{x_{n}}\right)$ has no normal form.

Note that despite (ii) above, it is not in general correct to regard expressions with no normal form as being 'undefined'.

## Exercise 32

Write down the $\lambda$-expression that represents $\operatorname{MIN}(f)$, where $f(x)=0$ for all $x$.

### 2.8 Extending the $\lambda$-calculus

Although it is possible to represent data-objects and data-structures with $\lambda$ expressions, it is often inefficient to do so. For example, most computers have hardware for arithmetic and it is reasonable to use this, rather than $\lambda$-conversion, to compute with numbers. A mathematically clean way of 'interfacing' computation rules to the $\lambda$-calculus is via so called $\delta$-rules.
The idea is to add a set of new constants and then to specify rules, called a $\delta$-rules, for reducing applications involving these constants. For example, one might add numerals and + as new constants, together with the $\delta$-rule:

$$
+m n \underset{\delta}{\longrightarrow} m+n
$$

$\left(E_{1} \underset{\delta}{\longrightarrow} E_{2}\right.$ means $E_{2}$ results by applying a $\delta$-rule to some subexpression of $\left.E_{1}\right)$.
When adding such constants and rules to the $\lambda$-calculus one must be careful not to destroy its nice properties, e.g. the Church-Rosser theorem (see page 31).

[^2]It can be shown that $\delta$-rules are safe if they have the form:

$$
c_{1} c_{2} \cdots c_{n} \underset{\delta}{\longrightarrow} e
$$

where $c_{1}, \ldots, c_{n}$ are constants and $e$ is either a constant or a closed abstraction (such $\lambda$-expressions are sometimes called values).
For example, one might add as constants Suc, Pre, IsZero, $\Delta_{0}, \Delta_{1}, \Delta_{2}, \cdots$ with the $\delta$-rules:

$$
\begin{aligned}
& \text { Suc } \Delta_{n} \xrightarrow[\delta]{\longrightarrow} \Delta_{n+1} \\
& \text { Pre } \Delta_{n+1} \xrightarrow[\delta]{\longrightarrow} \Delta_{n} \\
& \text { IsZero } \Delta_{0} \underset{\delta}{\longrightarrow} \text { true } \\
& \text { IsZero } \Delta_{n+1} \xrightarrow[\delta]{\longrightarrow} \text { false }
\end{aligned}
$$

Here $\Delta_{n}$ represents the number $n$, Suc, Pre, IsZero are new constants (not defined $\lambda$-expressions like suc, pre, iszero), and true and false are the previously defined expressions (which are both closed abstractions).

### 2.9 Theorems about the $\lambda$-calculus

If $E_{1} \longrightarrow E_{2}$ then $E_{2}$ can be thought of as having been got from $E_{1}$ by 'evaluation'. If there are no ( $\beta$ - or $\eta$-) redexes in $E_{2}$ then it can be thought of as 'fully evaluated'. A $\lambda$-expression is said to be in normal form if it contains no $\beta$ - or $\eta$-redexes (i.e. if the only conversion rule that can be applied is $\alpha$-conversion). Thus a $\lambda$-expression in normal form is 'fully evaluated'.

## Examples

(i) The representations of numbers are all in normal form.
(ii) $(\lambda x . x) \underline{0}$ is not in normal form.

Suppose an expression $E$ is 'evaluated' in two different ways by applying two different sequences of reductions until two normal forms $E_{1}$ and $E_{2}$ are obtained. The Church-Rosser theorem stated below shows that $E_{1}$ and $E_{2}$ will be the same except for having possibly different names of bound variables.
Because the results of reductions do not depend on the order in which they are done, separate redexes can be evaluated in parallel. Various research projects are currently trying to exploit this fact by designing multiprocessor architectures for evaluating $\lambda$-expressions. It is too early to tell how successful this work will be. There is a possibility that the communication overhead of distributing redexes to different processors and then collecting together the results will cancel out the theoretical advantages of the approach. Let us hope this pessimistic possibility can be avoided. It is a remarkable fact that the Church-Rosser theorem, an obscure mathematical result dating from before computers were invented, may underpin the design of the next generation of computing systems.
Here is the statement of the Church-Rosser theorem. It is an example of something that is intuitively obvious, but very hard to prove. Many properties of the $\lambda$-calculus share this property.

## The Church-Rosser theorem

If $E_{1}=E_{2}$ then there exists an $E$ such that $E_{1} \longrightarrow E$ and $E_{2} \longrightarrow E$.

It is now possible to see why the Chuch-Rosser theorem shows that $\lambda$-expressions can be evaluated in any order. Suppose an expression $E$ is 'evaluated' in two different ways by applying two different sequences of reductions until two normal forms $E_{1}$ and $E_{2}$ are obtained. Since $E_{1}$ and $E_{2}$ are obtained from $E$ by sequences of conversions, it follows by the definition of $=$ that $E=E_{1}$ and $E=E_{2}$ and hence $E_{1}=E_{2}$. By the Church-Rosser theorem there exists an expression, $E^{\prime}$ say, such that $E_{1} \longrightarrow E^{\prime}$ and $E_{2} \longrightarrow E^{\prime}$. Now if $E_{1}$ and $E_{2}$ are in normal form, then the only redexes they can contain are $\alpha$-redexes and so the only way that $E_{1}$ and $E_{2}$ can be reduced to $E^{\prime}$ is by changing the names of bound variables. Thus $E_{1}$ and $E_{2}$ must be the same up to renaming of bound variables (i.e. $\alpha$-conversion).
Another application of the Church-Rosser theorem is to show that if $m \neq n$ then the $\lambda$-expressions representing $m$ and $n$ are not equal, i.e. $\underline{m} \neq \underline{n}$. Suppose $m \neq n$ but $\underline{m}=\underline{n}$; by the Church-Rosser theorem $\underline{m} \longrightarrow E$ and $\underline{n} \longrightarrow E$ for some $E$. But it is obvious from the definitions of $\underline{m}$ and $\underline{n}$, namely

$$
\begin{aligned}
\underline{m} & =\lambda f x . f^{m} x \\
\underline{n} & =\lambda f x . f^{n} x
\end{aligned}
$$

that no such $E$ can exist. The only conversions that are applicable to $\underline{m}$ and $\underline{n}$ are $\alpha$-conversions and these cannot change the number of function applications in an expression ( $\underline{m}$ contains $m$ applications and $\underline{n}$ contains $n$ applications).
A $\lambda$-expression $E$ has a normal form if $E=E^{\prime}$ for some $E^{\prime}$ in normal form. The following corollary relates expressions in normal form to those that have a normal form; it summarizes some of the statements made above.

## Corollary to the Church-Rosser theorem

(i) If $E$ has a normal form then $E \longrightarrow E^{\prime}$ for some $E^{\prime}$ in normal form.
(ii) If $E$ has a normal form and $E=E^{\prime}$ then $E^{\prime}$ has a normal form.
(iii) If $E=E^{\prime}$ and $E$ and $E^{\prime}$ are both in normal form, then $E$ and $E^{\prime}$ are identical up to $\alpha$-conversion.

## Proof

(i) If $E$ has a normal form then $E=E^{\prime}$ for some $E^{\prime}$ in normal form. By the Church-Rosser theorem there exists $E^{\prime \prime}$ such that $E \longrightarrow E^{\prime \prime}$ and $E^{\prime} \longrightarrow E^{\prime \prime}$. As $E^{\prime}$ is in normal form the only redexes it can have are $\alpha$-redexes, so the reduction $E^{\prime} \longrightarrow E^{\prime \prime}$ must consist of a sequence of $\alpha$-conversions. Thus $E^{\prime \prime}$ must be identical to $E^{\prime}$ except for some renaming of bound variables; it must thus be in normal form as $E^{\prime}$ is.
(ii) Suppose $E$ has a normal form and $E=E^{\prime}$. As $E$ has a normal form, $E=E^{\prime \prime}$ where $E^{\prime \prime}$ is in normal form. Hence $E^{\prime}=E^{\prime \prime}$ by the transitivity of $=$ (see page 8 ) and so $E^{\prime}$ has a normal form.
(iii) This was proved above.

## Exercise 33

For each of the following $\lambda$-expressions either find its normal form or show that it has no normal form:
(i) add $\underline{3}$
(ii) add $\underline{3} \underline{5}$
(iii) $(\lambda x . x x)(\lambda x . x)$
(iv) $(\lambda x . x x)(\lambda x . x x)$
(v) $\mathbf{Y}$
(vi) $\mathbf{Y}(\lambda y . y)$
(vii) $\mathbf{Y}(\lambda f x$. (iszero $x \rightarrow \underline{0} \mid f($ pre $x))) \underline{7}$

Notice that a $\lambda$-expression $E$ might have a normal form even if there exists an infinite sequence $E \longrightarrow E_{1} \longrightarrow E_{2} \cdots$. For example $(\lambda x . \underline{1})$ (Y $f$ ) has a normal form 1 even though:

$$
(\lambda x . \underline{1})(\mathbf{Y} f) \longrightarrow(\lambda x . \underline{1})(f(\mathbf{Y} f)) \longrightarrow \cdots(\lambda x . \underline{1})\left(f^{n}(\mathbf{Y} f)\right) \longrightarrow \cdots
$$

The normalization theorem stated below tells us that such blind alleys can always be avoided by reducing the leftmost $\beta$ - or $\eta$-redex, where by 'leftmost' is meant the redex whose beginning $\lambda$ is as far to the left as possible.
Another important point to note is that $E_{1}$ may not have a normal form even though $E_{1} E_{2}$ does have one. For example, $\mathbf{Y}$ has no normal form, but $\mathbf{Y}(\lambda x . \underline{1}) \longrightarrow 1$. It is a common mistake to think of $\lambda$-expressions without a normal form as denoting 'undefined' functions; $\mathbf{Y}$ has no normal form but it denotes a perfectly well defined function ${ }^{3}$. Analysis beyond the scope of this book (see Wadsworth's paper [37]) shows that a $\lambda$-expression denotes an undefined function if and only if it cannot be converted to an expression in head normal form, where $E$ is in head normal form if it has the form

$$
\lambda V_{1} \cdots V_{m} . V E_{1} \cdots E_{n}
$$

where $V_{1}, \ldots, V_{m}$ and $V$ are variables and $E_{1}, \ldots, E_{n}$ are $\lambda$-expressions ( $V$ can either be equal to $V_{i}$, for some $i$, or it can be distinct from all of them). It follows that the fixed-point operator $\mathbf{Y}$ is not undefined because it can be converted to

$$
\lambda f . f((\lambda x . f(x x))(\lambda x . f(x x)))
$$

which is in head normal form.
It can be shown that an expression $E$ has a head normal form if and only if there exist expressions $E_{1}, \ldots, E_{n}$ such that $E E_{1} \ldots E_{n}$ has a normal form. This supports the interpretation of expressions without head normal forms as denoting undefined functions: $E$ being undefined means that $E E_{1} \ldots E_{n}$ never terminates for any $E_{1}, \ldots, E_{n}$. Full details on head normal forms and their relation to definedness can be found in Barendregt's book [2].

[^3]
## The normalization theorem

If $E$ has a normal form, then repeatedly reducing the leftmost $\beta$ - or $\eta$-redex (possibly after an $\alpha$-conversion to avoid invalid substitutions) will terminate with an expression in normal form.

The remark about $\alpha$-conversion in the statement of the theorem is to cover cases like:

$$
(\lambda x .(\lambda y . x y)) y \longrightarrow \lambda y^{\prime} . y y^{\prime}
$$

where $\lambda y . x y \longrightarrow \lambda y^{\prime} . x y^{\prime}$ has been $\alpha$-converted so as to avoid the invalid substitution $(\lambda y . x y)[y / x]=\lambda y . y y$.
A sequence of reductions in which the leftmost redex is always reduced is called a normal order reduction sequence.
The normalization theorem says that if $E$ has a normal form (i.e. for some $E^{\prime}$ in normal form $E=E^{\prime}$ ) then it can be found by normal order reduction. This, however, is not usually the 'most efficient' way to find it. For example, normal order reduction requires

$$
(\lambda x \sim x \sim x \sim) E
$$

to be reduced to

$$
\sim E \sim E \sim
$$

If $E$ is not in normal form then it would be more efficient to first reduce $E$ to $E^{\prime}$ say (where $E^{\prime}$ is in normal form) and then to reduce

$$
(\lambda x . \sim x \sim x \sim) E^{\prime}
$$

to

$$
\sim E^{\prime} \sim E^{\prime} \sim
$$

thereby avoiding having to reduce $E$ twice.
Note, however, that this 'call-by-value' scheme is disastrous in cases like

$$
(\lambda x . \underline{1})((\lambda x . x x)(\lambda x . x x))
$$

It is a difficult problem to find an optimal algorithm for choosing the next redex to reduce. For recent work in this area see Levy's paper [25].
Because normal order reduction appears so inefficient, some programming languages based on the $\lambda$-calculus, e.g. LISP, have used call by value even though it doesn't always terminate. Actually, call by value has other advantages besides efficiency, especially when the language is 'impure', i.e. has constructs with side effects (e.g. assignments). On the other hand, recent research suggests that maybe normal order evaluation is not as inefficient as was originally thought if one uses cunning implementation tricks like graph reduction (see page 40). Whether functional programming languages should use normal order or call by value is still a controversial issue.

### 2.10 Call-by-value and $Y$

Recall Y:

$$
\operatorname{LET} \mathbf{Y}=\lambda f .(\lambda x . f(x x))(\lambda x . f(x x))
$$

Unfortunately Y doesn't work with call-by-value, because applicative order causes it to go into a loop.

$$
\begin{aligned}
\mathbf{Y f} & \longrightarrow f(\mathbf{Y} f) \\
& \longrightarrow f(f(\mathbf{Y} f)) \\
& \longrightarrow f(f(f(\mathbf{Y} f))) \\
& \vdots
\end{aligned}
$$

To get around this, define:

$$
\operatorname{LET} \hat{\mathbf{Y}}=\lambda f .(\lambda x . f(\lambda y . x x y))(\lambda x . f(\lambda y . x x y))
$$

Note that $\hat{\mathbf{Y}}$ is $\mathbf{Y}$ with " $x x$ " $\eta$-converted to " $\lambda y . x x y$ ". $\hat{\mathbf{Y}}$ doesn't goes into a loop with call-by-value:

$$
\hat{\mathbf{Y}} f \quad \longrightarrow f(\lambda y . \hat{\mathbf{Y}} f y)
$$

Call-by-value doesn't evaluate $\lambda \mathrm{s}$, hence the looping is avoided.

## Chapter 3

## Combinators

Combinators provide an alternative theory of functions to the $\lambda$-calculus. They were originally introduced by logicians as a way of studying the process of substitution. More recently, Turner has argued that combinators provide a good 'machine code' into which functional programs can be compiled [34]. Several experimental computers have been built based on Turner's ideas (see e.g. [8]) and the results are promising. How these machines work is explained in Section 3.3. Combinators also provide a good intermediate code for conventional machines; several of the best compilers for functional languages are based on them (e.g. [11, 1]).
There are two equivalent ways of formulating the theory of combinators:
(i) within the $\lambda$-calculus, or
(ii) as a completely separate theory.

The approach here is to adopt (i) as it is slightly simpler, but (ii) was how it was done originally ${ }^{1}$. It will be shown that any $\lambda$-expression is equal to an expression built from variables and two particular expressions, $\mathbf{K}$ and $\mathbf{S}$, using only function application. This is done by mimicking $\lambda$-abstractions using combinations of $\mathbf{K}$ and S. It will be demonstrated how $\beta$-reductions can be simulated by simpler operations involving $\mathbf{K}$ and $\mathbf{S}$. It is these simpler operations that combinator machines implement directly in hardware. The definitions of $\mathbf{K}$ and $\mathbf{S}$ are

$$
\begin{aligned}
& \text { LET } \mathbf{K}=\lambda x y \cdot x \\
& \text { LET } \mathbf{S}=\lambda f g x .(f x)(g x)
\end{aligned}
$$

From these definitions it is clear by $\beta$-reduction that for all $E_{1}, E_{2}$ and $E_{3}$ :

$$
\begin{aligned}
& \mathbf{K} E_{1} E_{2}=E_{1} \\
& \mathbf{S} E_{1} E_{2} E_{3}=\left(E_{1} E_{3}\right)\left(E_{2} E_{3}\right)
\end{aligned}
$$

Any expression built by application (i.e. combination) from $\mathbf{K}$ and $\mathbf{S}$ is called a combinator; $\mathbf{K}$ and $\mathbf{S}$ are the primitive combinators.
In BNF, combinators have the following syntax:

$$
<\text { combinator }>::=\mathbf{K}|\mathbf{S}|(<\text { combinator }><\text { combinator }>)
$$

A combinatory expression is an expression built from $\mathbf{K}, \mathbf{S}$ and zero or more variables. Thus a combinator is a combinatory expression not containing variables. In

[^4]BNF, the syntax of combinatory expressions is:

```
<combinatory expression>
    ::= K | S
        | <variable>
        | (<combinatory expression> <combinatory expression>)
```

Exercise 34
Define I by:

$$
\text { LET } \mathbf{I}=\lambda x \cdot x
$$

Show that $\mathbf{I}=\mathbf{S} \mathbf{K} \mathbf{K}$.
The identity function $\mathbf{I}$ defined in the last exercise is often taken as a primitive combinator, but as the exercise shows this is not necessary as it can be defined from $\mathbf{K}$ and $\mathbf{S}$.

### 3.1 Combinator reduction

If $E$ and $E^{\prime}$ are combinatory expressions then the notation $E \xrightarrow[\mathrm{c}]{\longrightarrow} E^{\prime}$ is used if $E \equiv E^{\prime}$ or if $E^{\prime}$ can be got from $E$ by a sequence of rewritings of the form:
(i) $\mathbf{K} E_{1} E_{2} \underset{\mathrm{c}}{\longrightarrow} E_{1}$
(ii) $\mathbf{S} E_{1} E_{2} E_{3} \underset{\mathrm{c}}{\longrightarrow}\left(E_{1} E_{3}\right)\left(E_{2} E_{3}\right)$
(iii) $\mathbf{I} E \underset{\mathrm{c}}{\longrightarrow} E$

Note that the reduction $\mathbf{I} E \underset{\mathrm{c}}{\longrightarrow} E$ is derivable from (i) and (ii).

## Example

$$
\begin{array}{rlr}
\mathbf{S} \mathbf{K} \mathbf{K} x & \underset{\mathrm{c}}{\longrightarrow} \mathbf{K} x(\mathbf{K} x) & \text { by (ii) } \\
\underset{\mathrm{c}}{\longrightarrow} x & \text { by }(\mathrm{i})
\end{array}
$$

This example shows that for any $E: \mathbf{I} E \underset{\mathrm{c}}{\longrightarrow} E$.
Any sequence of combinatory reductions, i.e. reductions via $\underset{c}{ }$, can be expanded into a sequence of $\beta$-conversions. This is clear because $\mathbf{K} E_{1} E_{2}$ and $\mathbf{S} E_{1} E_{2} E_{3}$ reduce to $E_{1}$ and $\left(E_{1} E_{3}\right)\left(E_{2} E_{3}\right)$, respectively, by sequences of $\beta$-conversions.

### 3.2 Functional completeness

A surprising fact is that any $\lambda$-expression can be translated to an equivalent combinatory expression. This result is called the functional completeness of combinators and is the basis for compilers for functional languages to the machine code of combinator machines.
The first step is to define, for an arbitrary variable $V$ and combinatory expression $E$, another combinatory expression $\lambda^{*} V . E$ that simulates $\lambda V . E$ in the sense that $\lambda^{*} V . E=\lambda V . E$. This provides a way of using $\mathbf{K}$ and $\mathbf{S}$ to simulate adding ' $\lambda V$ ' to an expression.

If $V$ is a variable and $E$ is a combinatory expression, then the combinatory expression $\lambda^{*} V . E$ is defined inductively on the structure of $E$ as follows:
(i) $\lambda^{*} V \cdot V=\mathbf{I}$
(ii) $\lambda^{*} V \cdot V^{\prime}=\mathbf{K} V^{\prime} \quad\left(\right.$ if $\left.V \neq V^{\prime}\right)$
(iii) $\lambda^{*} V \cdot C=\mathbf{K} C \quad$ (if $C$ is a combinator)
(iv) $\lambda^{*} V .\left(E_{1} E_{2}\right)=\mathbf{S}\left(\lambda^{*} V . E_{1}\right)\left(\lambda^{*} V . E_{2}\right)$

Note that $\lambda^{*} V . E$ is a combinatory expression not containing $V$.
Example: If $f$ and $x$ are variables and $f \neq x$, then:

$$
\begin{aligned}
\lambda^{*} x . f x & =\mathbf{S}\left(\lambda^{*} x . f\right)\left(\lambda^{*} x . x\right) \\
& =\mathbf{S}(\mathbf{K} f) \mathbf{I}
\end{aligned}
$$

The following theorem shows that $\lambda^{*} V . E$ simulates $\lambda$-abstraction.
Theorem $\quad\left(\lambda^{*} V . E\right)=\lambda V . E$

## Proof

We show that $\left(\lambda^{*} V . E\right) V=E$. It then follows immediately that $\lambda V \cdot\left(\lambda^{*} V . E\right) V=$ $\lambda V . E$ and hence by $\eta$-reduction that $\lambda^{*} V . E=\lambda V . E$.
The proof that $\left(\lambda^{*} V . E\right) V=E$ is by mathematical induction on the 'size' of $E$. The argument goes as follows:
(i) If $E=V$ then:

$$
\left(\lambda^{*} V . E\right) V=\mathbf{I} V=(\lambda x \cdot x) V=V=E
$$

(ii) If $E=V^{\prime}$ where $V^{\prime} \neq V$ then:

$$
\left(\lambda^{*} V . E\right) V=\mathbf{K} V^{\prime} V=(\lambda x y . x) V^{\prime} V=V^{\prime}=E
$$

(iii) If $E=C$ where $C$ is a combinator, then:

$$
\left(\lambda^{*} V . E\right) V=\mathbf{K} C=(\lambda x y . x) C V=C=E
$$

(iv) If $E=\left(E_{1} E_{2}\right)$ then we can assume by induction that:

$$
\begin{aligned}
& \left(\lambda^{*} V . E_{1}\right) V=E_{1} \\
& \left(\lambda^{*} V . E_{2}\right) V=E_{2}
\end{aligned}
$$

and hence

$$
\begin{aligned}
\left(\lambda^{*} V . E\right) V & =\left(\lambda^{*} V .\left(E_{1} E_{2}\right)\right) V \\
& =\left(\mathbf{S}\left(\lambda^{*} V . E_{1}\right)\left(\lambda^{*} V . E_{2}\right)\right) V \\
& =(\lambda f g x . f x(g x))\left(\lambda^{*} V . E_{1}\right)\left(\lambda^{*} V . E_{2}\right) V \\
& =\left(\lambda^{*} V . E_{1}\right) V\left(\left(\lambda^{*} V . E_{2}\right) V\right) \quad \text { (by induction assumption) } \\
& =E_{1} E_{2} \\
& =E
\end{aligned}
$$

The notation

$$
\lambda^{*} V_{1} V_{2} \cdots V_{n} . E
$$

is used to mean

$$
\lambda^{*} V_{1} . \lambda^{*} V_{2} . \quad \cdots \lambda^{*} V_{n} . E
$$

Now define the translation of an arbitrary $\lambda$-expression $E$ to a combinatory expres$\operatorname{sion}(E)_{\mathrm{C}}$ :
(i) $(V)_{\mathrm{C}}=V$
(ii) $\left(E_{1} E_{2}\right)_{\mathrm{C}}=\left(E_{1}\right)_{\mathrm{C}}\left(E_{2}\right)_{\mathrm{C}}$
(iii) $(\lambda V \cdot E)_{\mathrm{C}}=\lambda^{*} V \cdot(E)_{\mathrm{C}}$

Theorem For every $\lambda$-expression $E$ we have: $E=(E)_{\mathrm{C}}$

Proof
The proof is by induction on the size of $E$.
(i) If $E=V$ then $(E)_{\mathrm{C}}=(V)_{\mathrm{C}}=V$
(ii) If $E=\left(\begin{array}{ll}E_{1} & E_{2}\end{array}\right)$ we can assume by induction that

$$
\begin{aligned}
& E_{1}=\left(E_{1}\right)_{\mathrm{c}} \\
& E_{2}=\left(E_{2}\right)_{\mathrm{c}}
\end{aligned}
$$

hence

$$
(E)_{\mathrm{C}}=\left(E_{1} E_{2}\right)_{\mathrm{C}}=\left(E_{1}\right)_{\mathrm{C}}\left(E_{2}\right)_{\mathrm{C}}=E_{1} E_{2}=E
$$

(iii) If $E=\lambda V . E^{\prime}$ then we can assume by induction that

$$
\left(E^{\prime}\right)_{\mathrm{C}}=E^{\prime}
$$

hence

$$
\begin{array}{rlr}
(E)_{\mathrm{C}} & =\left(\lambda V \cdot E^{\prime}\right)_{\mathrm{C}} & \\
& =\lambda^{*} V \cdot\left(E^{\prime}\right)_{\mathrm{C}} & \text { (by translation rules) } \\
& =\lambda^{*} V \cdot E^{\prime} & \text { (by induction assumption) } \\
& =\lambda V \cdot E^{\prime} & \text { (by previous theorem) } \\
& =E &
\end{array}
$$

This theorem shows that any $\lambda$-expression is equal to a $\lambda$-expression built up from $\mathbf{K}$ and $\mathbf{S}$ and variables by application, i.e. the class of $\lambda$-expressions $E$ defined by the BNF:

$$
E::=V|\mathbf{K}| \mathbf{S} \mid E_{1} E_{2}
$$

is equivalent to the full $\lambda$-calculus.
A collection of $n$ combinators $C_{1}, \ldots, C_{n}$ is called an $n$-element basis (Barendregt [2], Chapter 8) if every $\lambda$-expression $E$ is equal to an expression built from $C_{i} \mathrm{~s}$ and variables by function applications. The theorem above shows that $\mathbf{K}$ and $\mathbf{S}$ form a 2-element basis. The exercise below (from Section 8.1.5. of Barendregt) shows that there exists a 1-element basis.

## Exercise 35

Find a combinator, $\mathbf{X}$ say, such that any $\lambda$-expression is equal to an expression built from $\mathbf{X}$ and variables by application. Hint: Let $\left\langle E_{1}, E_{2}, E_{3}\right\rangle=\lambda p . p E_{1} E_{2} E_{3}$ and consider $\langle\mathbf{K}, \mathbf{S}, \mathbf{K}\rangle\langle\mathbf{K}, \mathbf{S}, \mathbf{K}\rangle\langle\mathbf{K}, \mathbf{S}, \mathbf{K}\rangle$ and $\langle\mathbf{K}, \mathbf{S}, \mathbf{K}\rangle\langle\langle\mathbf{K}, \mathbf{S}, \mathbf{K}\rangle\langle\mathbf{K}, \mathbf{S}, \mathbf{K}\rangle\rangle$

## Examples:

$$
\begin{aligned}
\lambda^{*} f . \lambda^{*} x . f(x x) & =\lambda^{*} f .\left(\lambda^{*} x . f(x x)\right) \\
& =\lambda^{*} f .\left(\mathbf{S}\left(\lambda^{*} x . f\right)\left(\lambda^{*} x . x x\right)\right) \\
& =\lambda^{*} f .\left(\mathbf{S}(\mathbf{K} f)\left(\mathbf{S}\left(\lambda^{*} x . x\right)\left(\lambda^{*} x . x\right)\right)\right) \\
& =\lambda^{*} f .(\mathbf{S}(\mathbf{K} f)(\mathbf{S} \mathbf{I} \mathbf{~})) \\
& =\mathbf{S}\left(\lambda^{*} f . \mathbf{S}(\mathbf{K} f)\right)\left(\lambda^{*} f . \mathbf{S} \mathbf{I} \mathbf{~}\right) \\
& =\mathbf{S}\left(\mathbf{S}\left(\lambda^{*} f . \mathbf{S}\right)\left(\lambda^{*} f . \mathbf{K} f\right)\right)(\mathbf{K}(\mathbf{S} \mathbf{I} \mathbf{~})) \\
& =\mathbf{S}\left(\mathbf{S}(\mathbf{K ~ S})\left(\mathbf{S}\left(\lambda^{*} f . \mathbf{K}\right)\left(\lambda^{*} f . f\right)\right)\right)(\mathbf{K}(\mathbf{S} \mathbf{I} \mathbf{I})) \\
& =\mathbf{S}(\mathbf{S}(\mathbf{K} \mathbf{S})(\mathbf{S}(\mathbf{K} \text { K }))(\mathbf{K}(\mathbf{S ~ I ~ I}))
\end{aligned}
$$

$$
\begin{aligned}
(\mathbf{Y})_{\mathrm{c}} & =(\lambda f .(\lambda x . f(x x))(\lambda x \cdot f(x x)))_{\mathrm{C}} \\
& =\lambda^{*} f \cdot\left((\lambda x \cdot f(x x))(\lambda x \cdot f(x x))_{\mathrm{C}}\right. \\
& =\lambda^{*} f \cdot\left((\lambda x \cdot f(x x))_{\mathrm{C}}(\lambda x \cdot f(x x))_{\mathrm{C}}\right) \\
& =\lambda^{*} f \cdot\left(\lambda^{*} x \cdot(f(x x))_{\mathrm{c}}\right)\left(\lambda^{*} x \cdot(f(x x))_{\mathrm{C}}\right) \\
& =\lambda^{*} f \cdot\left(\lambda^{*} x \cdot f(x x)\right)\left(\lambda^{*} x \cdot f(x x)\right) \\
& =\mathbf{S}\left(\lambda^{*} f \cdot \lambda^{*} x \cdot f(x x)\right)\left(\lambda^{*} f \cdot \lambda^{*} x \cdot f(x x)\right) \\
& =\mathbf{S}(\mathbf{S}(\mathbf{S}(\mathbf{K S})(\mathbf{S}(\mathbf{K K}) \mathbf{I}))(\mathbf{K}(\mathbf{S I I})))(\mathbf{S}(\mathbf{S}(\mathbf{K S})(\mathbf{S}(\mathbf{K K}) \mathbf{I}))(\mathbf{K}(\mathbf{S I I})))
\end{aligned}
$$

### 3.3 Reduction machines

Until David Turner published his paper [34], combinators were regarded as a mathematical curiosity. In his paper Turner argued that translating functional languages, i.e. languages based on the $\lambda$-calculus, to combinators and then reducing the resulting expressions using the rewrites given on page 36 is a practical way of implementing these languages.
Turner's idea is to represent combinatory expressions by trees. For example, $\mathbf{S}(f x)(\mathbf{K} y) z$ would be represented by:


Such trees are represented as pointer structures in memory. Special hardware or firmware can then be implemented to transform such trees according to the rules of combinator reduction defining $\underset{\mathrm{c}}{\longrightarrow}$.
For example, the tree above could be transformed to:

using the transformation

which corresponds to the reduction $\mathbf{S} E_{1} E_{2} E_{3} \underset{\mathrm{c}}{\longrightarrow}\left(E_{1} E_{3}\right)\left(E_{2} E_{3}\right)$.

## Exercise 36

What tree transformation corresponds to $\mathbf{K} E_{1} E_{2} \longrightarrow{ }_{\mathrm{c}} E_{1}$ ? How would this transformation change the tree above?

Notice that the tree transformation for $\mathbf{S}$ just given duplicates a subtree. This wastes space; a better transformation would be to generate one subtree with two pointers to it, i.e.


This generates a graph rather than a tree. For further details of such graph reductions see Turner's paper [34].
It is clear from the theorem above that a valid way of reducing $\lambda$-expressions is:
(i) Translating to combinators (i.e. $\left.E \mapsto(E)_{\mathrm{C}}\right)$.
(ii) Applying the rewrites

$$
\begin{aligned}
& \mathbf{K} E_{1} E_{2} \longrightarrow E_{1} \\
& \mathbf{S} E_{1} E_{2} \quad E_{3} \xrightarrow[\mathrm{c}]{\longrightarrow}\left(E_{1} E_{3}\right)\left(E_{2} E_{3}\right)
\end{aligned}
$$

until no more rewriting is possible.
An interesting question is whether this process will 'fully evaluate' expressions. If some expression $E$ is translated to combinators, then reduced using $\underset{\mathrm{c}}{\longrightarrow}$, is the resulting expression as 'fully evaluated' as the result of $\lambda$-reducing $E$ directly, or is it only partially evaluated? Surprisingly, there doesn't seem to be anything in the literature on this important question ${ }^{2}$. However, combinator machines have been built and they appear to work [8]!
It is well known that if $E_{1} \longrightarrow E_{2}$ in the $\lambda$-calculus, then it is not necessarily the case that $\left(E_{1}\right)_{\mathrm{c}} \underset{\mathrm{c}}{\longrightarrow}\left(E_{2}\right)_{\mathrm{c}}$. For example, take

$$
\begin{aligned}
& E_{1}=\lambda y \cdot(\lambda z \cdot y)(x y) \\
& E_{2}=\lambda y \cdot y
\end{aligned}
$$

## Exercise 37

With $E_{1}$ and $E_{2}$ as above show that $E_{1} \longrightarrow E_{2}$ in the $\lambda$-calculus, but it is not the case that $\left(E_{1}\right)_{\mathrm{C}} \underset{\mathrm{c}}{\longrightarrow}\left(E_{2}\right)_{\mathrm{C}}$.

A combinatory expression is defined to be in combinatory normal form if it contains no subexpressions of the form $\mathbf{K} E_{1} E_{2}$ or $\mathbf{S} E_{1} E_{2} E_{3}$. Then the normalization theorem holds for combinatory expressions, i.e. always reducing the leftmost combinatory redex will find a combinatory normal form if it exists.
Note that if $E$ is in combinatory normal form, then it does not necessarily follow that it is a $\lambda$-expression in normal form.

Example: $\mathbf{S} \mathbf{K}$ is in combinatory normal form, but it contains a $\beta$-redex, namely:

$$
(\lambda f .(\lambda g x .(f x(g x)))(\lambda x y . x)
$$

## Exercise 38

Construct a combinatory expression $E$ which is in combinatory normal form, but has no normal form.

### 3.4 Improved translation to combinators

The examples on page 39 show that simple $\lambda$-expressions can translate to quite complex combinatory expressions via the rules on page 38 .
To make the 'code' executed by reduction machines more compact, various optimizations have been devised.

## Examples

[^5](i) Let $E$ be a combinatory expression and $x$ a variable not occurring in $E$. Then:
$$
\mathbf{S}(\mathbf{K} E) \mathbf{I} x \underset{\mathrm{c}}{\longrightarrow}(\mathbf{K} E x)(\mathbf{I} x) \underset{\mathrm{c}}{\longrightarrow} E x
$$
hence $\mathbf{S}(\mathbf{K} E) \mathbf{I} x=E x$ (because $E_{1} \xrightarrow[\mathrm{c}]{ } E_{2}$ implies $E_{1} \longrightarrow E_{2}$ ), so by extensionality (Section 1.7 , see on page 10 ):
$$
\mathbf{S}(\mathbf{K} E) \mathbf{I}=E
$$
(ii) Let $E_{1}, E_{2}$ be combinatory expressions and $x$ a variable not occurring in either of them. Then:
$$
\mathbf{S}\left(\mathbf{K} E_{1}\right)\left(\mathbf{K} E_{2}\right) x \underset{\mathrm{c}}{\longrightarrow} \mathbf{K} E_{1} x\left(\mathbf{K} E_{2}\right) x \underset{\mathrm{c}}{\longrightarrow} E_{1} E_{2}
$$

Thus

$$
\mathbf{S}\left(\mathbf{K} E_{1}\right)\left(\mathbf{K} E_{2}\right) x=E_{1} E_{2}
$$

Now

$$
\mathbf{K}\left(E_{1} E_{2}\right) x \underset{\mathrm{c}}{\longrightarrow} E_{1} E_{2}
$$

hence $\mathbf{K}\left(E_{1} E_{2}\right) x=E_{1} E_{2}$. Thus

$$
\mathbf{S}\left(\mathbf{K} E_{1}\right)\left(\mathbf{K} E_{2}\right) x=E_{1} E_{2}=\mathbf{K}\left(E_{1} E_{2}\right) x
$$

It follows by extensionality that:

$$
\mathbf{S}\left(\mathbf{K} E_{1}\right)\left(\mathbf{K} E_{2}\right)=\mathbf{K}\left(E_{1} E_{2}\right)
$$

Since $\mathbf{S}(\mathbf{K} E) \mathbf{I}=E$ for any $E$, whenever a combinatory expression of the form $\mathbf{S}(\mathbf{K} E) \mathbf{I}$ is generated, it can be 'peephole optimized' to just $E$. Similarly, whenever an expression of the form $\mathbf{S}\left(\mathbf{K} E_{1}\right)\left(\mathbf{K} E_{2}\right)$ is generated, it can be optimized to $\mathbf{K}\left(E_{1} E_{2}\right)$.

Example: On page 39 it was shown that:

$$
\lambda^{*} f . \lambda^{*} x . f(x x)=\mathbf{S}(\mathbf{S}(\mathbf{K} \mathbf{S})(\mathbf{S}(\mathbf{K} \mathbf{K}) \mathbf{I}))(\mathbf{K}(\mathbf{S} \mathbf{I} \mathbf{~ I}))
$$

Using the optimization $\mathbf{S}(\mathbf{K} E) \mathbf{I}=E$ this simplifies to:

$$
\lambda^{*} f . \lambda^{*} x . f(x x)=\mathbf{S}(\mathbf{S}(\mathbf{K} \mathbf{S}) \mathbf{K})(\mathbf{K}(\mathbf{S} \mathbf{I} \mathbf{I}))
$$

### 3.5 More combinators

It is easier to recognize the applicability of the optimization $\mathbf{S}(\mathbf{K} E) \mathbf{I}=E$ if $\mathbf{I}$ has not been expanded to $\mathbf{S} \mathbf{K} \mathbf{K}$, i.e. if $\mathbf{I}$ is taken as a primitive combinator. Various other combinators are also useful in the same way; for example, $\mathbf{B}$ and $\mathbf{C}$ defined by:

$$
\begin{aligned}
\text { LET } \mathbf{B} & =\lambda f g x . f(g x) \\
\text { LET } \mathbf{C} & =\lambda f g x . f x g
\end{aligned}
$$

These have the following reduction rules:

$$
\begin{aligned}
& \mathbf{B} E_{1} E_{2} E_{3} \xrightarrow[\mathrm{c}]{\longrightarrow} E_{1}\left(E_{2} E_{3}\right) \\
& \mathbf{C} E_{1} E_{2} E_{3} \xrightarrow[\mathrm{c}]{\longrightarrow} E_{1} E_{3} E_{2}
\end{aligned}
$$

## Exercise 39

Show that with B, C defined as above:

$$
\begin{aligned}
& \mathbf{S}\left(\mathbf{K} E_{1}\right) E_{2}=\mathbf{B} E_{1} E_{2} \\
& \mathbf{S} E_{1}\left(\mathbf{K} E_{2}\right)=\mathbf{C} E_{1} E_{2}
\end{aligned}
$$

(where $E_{1}, E_{2}$ are any two combinatory expressions).

Using $\mathbf{B}$ and $\mathbf{C}$, one can further optimize the translation of $\lambda$-expressions to combinators by replacing expressions of the form $\mathbf{S}\left(\mathbf{K} E_{1}\right) E_{2}$ and $\mathbf{S} E_{1}\left(\mathbf{K} E_{2}\right)$ by B $E_{1} E_{2}$ and $\mathbf{C} E_{1} E_{2}$.

### 3.6 Curry's algorithm

Combining the various optimizations described in the previous section leads to Curry's algorithm for translating $\lambda$-expressions to combinatory expressions. This algorithm consists in using the definition of $(E)_{\mathrm{C}}$ given on page 38 , but whenever an expression of the form $\mathbf{S} E_{1} E_{2}$ is generated one tries to apply the following rewrite rules:

1. $\mathbf{S}\left(\mathbf{K} E_{1}\right)\left(\mathbf{K} E_{2}\right) \longrightarrow \mathbf{K}\left(E_{1} E_{2}\right)$
2. $\mathbf{S}(\mathbf{K} E) \mathbf{I} \longrightarrow E$
3. $\mathbf{S}\left(\mathbf{K} E_{1}\right) E_{2} \longrightarrow \mathbf{B} E_{1} E_{2}$
4. $\mathbf{S} E_{1}\left(\mathbf{K} E_{2}\right) \longrightarrow \mathbf{C} E_{1} E_{2}$

If more than one rule is applicable, the earlier one is used. For example, $\mathbf{S}\left(\mathbf{K} E_{1}\right)\left(\mathbf{K} E_{2}\right)$ is translated to $\mathbf{K}\left(E_{1} E_{2}\right)$, not to $\mathbf{B} E_{1}\left(\mathbf{K} E_{2}\right)$.

## Exercise 40

Show that using Curry's algorithm, $\mathbf{Y}$ is translated to the combinator:

$$
\text { S }\left(\mathbf{C B}\left(\text { S I I }^{\prime}\right)\right)(\text { C B }(\text { S I I }))
$$

## Exercise 41

Show that:

### 3.7 Turner's algorithm

In a second paper, Turner proposed that Curry's algorithm be extended to use another new primitive combinator called $\mathbf{S}^{\prime}$ [35]. This is defined by:

$$
\operatorname{LET} \mathbf{S}^{\prime}=\lambda c f g x . c(f x)(g x)
$$

and has the reduction rule:

$$
\mathbf{S}^{\prime} C E_{1} E_{2} E_{3} \underset{\mathrm{c}}{\longrightarrow} C\left(E_{1} E_{3}\right)\left(E_{2} E_{3}\right)
$$

where $C, E_{1}, E_{2}, E_{3}$ are arbitrary combinatory expressions. The reason why ' $C$ ' is used is that $\mathbf{S}^{\prime}$ has the property that if $C$ is a combinator (i.e. contains no variables), then for any $E_{1}$ and $E_{2}$ :

$$
\lambda^{*} x . C E_{1} E_{2}=\mathbf{S}^{\prime} C\left(\lambda^{*} x . E_{1}\right)\left(\lambda^{*} x . E_{2}\right)
$$

This can be shown using extensionality. Clearly $x$ is a variable not occurring in $\lambda^{*} x . C E_{1} E_{2}$ or $\mathbf{S}^{\prime} C\left(\lambda^{*} x . E_{1}\right)\left(\lambda^{*} x . E_{2}\right)$ (exercise: why?), so it is sufficient to show:

$$
\left(\lambda^{*} x . C E_{1} E_{2}\right) x=\left(\mathbf{S}^{\prime} C\left(\lambda^{*} x . E_{1}\right)\left(\lambda^{*} x . E_{2}\right)\right) x
$$

From the definition of $\lambda^{*} x$ it easily follows that:

$$
\lambda^{*} x . C E_{1} E_{2}=\mathbf{S}\left(\mathbf{S}(\mathbf{K} C)\left(\lambda^{*} x . E_{1}\right)\right)\left(\lambda^{*} x . E_{2}\right)
$$

hence

$$
\begin{aligned}
\left(\lambda^{*} x . C E_{1} E_{2}\right) x & =\left(\mathbf{S}\left(\mathbf{S}(\mathbf{K} C)\left(\lambda^{*} x . E_{1}\right)\right)\left(\lambda^{*} x . E_{2}\right) x\right. \\
& \left.=\mathbf{S}(\mathbf{K} C)\left(\lambda^{*} x . E_{1}\right) x\left(\left(\lambda^{*} x . E_{2}\right)\right) x\right) \\
& =\mathbf{K} C x\left(\left(\lambda^{*} x . E_{1}\right) x\right)\left(\left(\lambda^{*} x . E_{2}\right) x\right) \\
& \left.=C\left(\left(\lambda^{*} x . E_{1}\right) x\right)\left(\left(\lambda^{*} x . E_{2}\right)\right) x\right)
\end{aligned}
$$

But $\left(\mathbf{S}^{\prime} C\left(\lambda^{*} x . E_{1}\right)\left(\lambda^{*} x . E_{2}\right) x=C\left(\left(\lambda^{*} x . E_{1}\right) x\right)\left(\left(\lambda^{*} x . E_{2}\right)\right) x\right)$ also, and so:

$$
\left(\lambda^{*} x . C E_{1} E_{2}\right) x=\left(\mathbf{S}^{\prime} C\left(\lambda^{*} x . E_{1}\right)\left(\lambda^{*} x . E_{2}\right)\right) x
$$

## Exercise 42

Where in the argument above did we use the assumption that $C$ is a combinator?

Turner's combinator $\mathbf{S}^{\prime}$ is useful when translating $\lambda$-expressions of the form $\lambda V_{n} \cdots V_{2} V_{1} . E_{1} E_{2}$ (it will be seen shortly why it is convenient to number the bound variables in descending order). To see this, following Turner [35], temporarily define

| $E^{\prime}$ | to mean | $\lambda^{*} V_{1} \cdot E$ |
| :--- | :--- | :--- |
| $E^{\prime \prime}$ | to mean | $\lambda^{*} V_{2} \cdot\left(\lambda^{*} V_{1} \cdot E\right)$ |
| $E^{\prime \prime \prime}$ | to mean | $\lambda^{*} V_{3} \cdot\left(\lambda^{*} V_{2} \cdot\left(\lambda^{*} V_{1} \cdot E\right)\right)$ |
|  | $\vdots$ |  |

Recall that:

$$
\left.\left(\lambda V_{n} \cdots V_{2} V_{1} . E_{1} E_{2}\right)_{\mathrm{c}}=\lambda^{*} V_{n} \cdot\left(\cdots\left(\lambda^{*} V_{2} \cdot\left(\lambda^{*} V_{1} \cdot\left(E_{1} E_{2}\right)_{\mathrm{C}}\right)\right)\right) \cdots\right)
$$

The next exercise shows that:

$$
\lambda^{*} V_{n} . \ldots \lambda^{*} V_{2} \cdot \lambda^{*} V_{1} .\left(E_{1} E_{2}\right)
$$

gets very complicated as $n$ increases.

## Exercise 43

Show that:
(i) $\lambda^{*} x_{1} \cdot E_{1} E_{2}=\mathbf{S} E_{1}^{\prime} E_{2}^{\prime}$
(ii) $\lambda^{*} x_{2} \cdot\left(\lambda^{*} x_{1} \cdot E_{1} E_{2}\right)=\mathbf{S}\left(\mathbf{B} \mathbf{S} E_{1}^{\prime \prime}\right) E_{2}^{\prime \prime}$
(iii) $\lambda^{*} x_{3} \cdot\left(\lambda^{*} x_{2} \cdot\left(\lambda^{*} x_{1} \cdot E_{1} E_{2}\right)\right)=\mathbf{S}\left(\mathbf{B} \mathbf{S}\left(\mathbf{B}(\mathbf{B} \mathbf{S}) E_{1}^{\prime \prime \prime}\right)\right) E_{2}^{\prime \prime \prime}$
(iv) $\lambda^{*} x_{4} \cdot\left(\lambda^{*} x_{3} \cdot\left(\lambda^{*} x_{2} \cdot\left(\lambda^{*} x_{1} \cdot E_{1} E_{2}\right)\right)\right)=$
$\mathbf{S}\left(\mathbf{B} \mathbf{S}\left(\mathbf{B}(\mathbf{B} \mathbf{S})(\mathbf{B}(\mathbf{B}(\mathbf{B} \mathbf{S}))) E_{1}^{\prime \prime \prime \prime}\right)\right) E_{2}^{\prime \prime \prime \prime}$

The size of $\lambda^{*} V_{n} . \ldots \lambda^{*} V_{2} . \lambda^{*} V_{1} .\left(E_{1} E_{2}\right)$ is proportional to the square of $n$. Using $\mathbf{S}^{\prime}$, the size can be made to grow linearly with $n$ :

$$
\begin{aligned}
\lambda^{*} x_{2} \cdot\left(\lambda^{*} x_{1} \cdot E_{1} E_{2}\right) & =\lambda^{*} x_{2} \cdot \mathbf{S} E_{1}^{\prime} E_{2}^{\prime} \\
& =\mathbf{S}^{\prime} \mathbf{S}\left(\lambda^{*} x_{2} \cdot E_{1}^{\prime}\right)\left(\lambda^{*} x_{2} \cdot E_{2}^{\prime}\right) \\
& =\mathbf{S}^{\prime} \mathbf{S} E_{1}^{\prime \prime} E_{2}^{\prime \prime} \\
\lambda^{*} x_{3} \cdot\left(\lambda^{*} x_{2} \cdot\left(\lambda^{*} x_{1} \cdot E_{1} E_{2}\right)\right) & =\lambda^{*} x_{3} \cdot \mathbf{S}^{\prime} \mathbf{S} E_{1}^{\prime \prime} E_{2}^{\prime \prime} \\
& =\mathbf{S}^{\prime}\left(\mathbf{S}^{\prime} \mathbf{S}\right)\left(\lambda^{*} x_{3} . E_{1}^{\prime \prime}\right)\left(\lambda^{*} x_{3} . E_{2}^{\prime \prime}\right) \\
& =\mathbf{S}^{\prime}\left(\mathbf{S}^{\prime} \mathbf{S}\right) E_{1}^{\prime \prime \prime} E_{2}^{\prime \prime \prime} \\
& \\
\lambda^{*} x_{4} \cdot\left(\lambda^{*} x_{3} \cdot\left(\lambda^{*} x_{2} \cdot\left(\lambda^{*} x_{1} \cdot E_{1} E_{2}\right)\right)\right) & =\lambda^{*} x_{4} \cdot \mathbf{S}^{\prime}\left(\mathbf{S}^{\prime} \mathbf{S}\right) E_{1}^{\prime \prime \prime} E_{2}^{\prime \prime \prime} \\
& =\mathbf{S}^{\prime}\left(\mathbf{S}^{\prime}\left(\mathbf{S}^{\prime} \mathbf{S}\right)\right)\left(\lambda^{*} x_{4} \cdot E_{1 \prime \prime}^{\prime \prime \prime}\right)\left(\lambda^{*} x_{4} \cdot E_{2}^{\prime \prime \prime \prime}\right) \\
& =\mathbf{S}^{\prime}\left(\mathbf{S}^{\prime}\left(\mathbf{S}^{\prime} \mathbf{S}\right)\right) E_{1}^{\prime \prime \prime} E_{2}^{\prime \prime \prime \prime}
\end{aligned}
$$

Just as $\mathbf{B}$ and $\mathbf{C}$ were introduced to simplify combinatory expressions of the form $\mathbf{S}\left(\mathbf{K} E_{1}\right) E_{2}$ and $\mathbf{S} E_{1}\left(\mathbf{K} E_{2}\right)$ respectively, Turner also devised $\mathbf{B}^{\prime}$ and $\mathbf{C}^{\prime}$ with an analogous role for $\mathbf{S}^{\prime}$. The properties required are:

$$
\begin{aligned}
& \mathbf{S}^{\prime} C\left(\mathbf{K} E_{1}\right) E_{2}=\mathbf{B}^{\prime} C E_{1} E_{2} \\
& \mathbf{S}^{\prime} C E_{1}\left(\mathbf{K} E_{2}\right)=\mathbf{C}^{\prime} C E_{1} E_{2}
\end{aligned}
$$

(where $C$ is any combinator, and $E_{1}, E_{2}$ are arbitrary combinatory expressions). This is achieved if $\mathbf{B}^{\prime}$ and $\mathbf{C}^{\prime}$ are defined by:

$$
\begin{aligned}
\text { LET } \mathbf{B}^{\prime} & =\lambda c f g x \cdot c f(g x) \\
\text { LET } \mathbf{C}^{\prime} & =\lambda c f g x \cdot c(f x) g
\end{aligned}
$$

Clearly $\mathbf{B}^{\prime}$ and $\mathbf{C}^{\prime}$ will have the property that for arbitrary $\lambda$-expressions $C, E_{1}$, $E_{2}$ and $E_{3}$ :

$$
\begin{aligned}
& \mathbf{B}^{\prime} C E_{1} E_{2} E_{3} \xrightarrow[\mathrm{c}]{\longrightarrow} C E_{1}\left(E_{2} E_{3}\right) \\
& \mathbf{C}^{\prime} C E_{1} E_{2} E_{3} \underset{\mathrm{c}}{\longrightarrow} C\left(E_{1} E_{3}\right) E_{2}
\end{aligned}
$$

## Exercise 44

Show that for arbitrary $\lambda$-expressions $E_{1}, E_{2}$ and $E_{3}$ :
(i) $\mathbf{S}^{\prime} E_{1}\left(\mathbf{K} E_{2}\right) E_{3}=\mathbf{B}^{\prime} E_{1} E_{2} E_{3}$
(ii) $\mathbf{S}^{\prime} E_{1} E_{2}\left(\mathbf{K} E_{3}\right)=\mathbf{C}^{\prime} E_{1} E_{2} E_{3}$
(iii) $\mathbf{S}\left(\mathbf{B} E_{1} E_{2}\right) E_{3}=\mathbf{S}^{\prime} E_{1} E_{2} E_{3}$
(iv) $\mathbf{B}\left(E_{1} E_{2}\right) E_{3}=\mathbf{B}^{\prime} E_{1} E_{2} E_{3}$
(v) $\mathbf{C}\left(\mathbf{B} E_{1} E_{2}\right) E_{3}=\mathbf{C}^{\prime} E_{1} E_{2} E_{3}$

Turner's algorithm for translating $\lambda$-expressions to combinatory expressions is described by him [35] as follows:

Use the algorithm of Curry but whenever a term beginning in $\mathbf{S}, \mathbf{B}$ or $\mathbf{C}$ is formed use one of the following transformations if it is possible to do so

$$
\begin{array}{r}
\mathbf{S}(\mathbf{B} K A) B \longrightarrow \mathbf{S}^{\prime} K A B \\
\mathbf{B}(K A) B \longrightarrow \mathbf{B}^{\prime} K A B \\
\mathbf{C}(\mathbf{B} K A) B \longrightarrow \mathbf{C}^{\prime} K A B
\end{array}
$$

Here $A$ and $B$ stand for arbitrary terms as usual and $K$ is any term composed entirely of constants. The correctness of the new algorithm can be inferred from the correctness of the Curry algorithm by demonstrating that in each of the above transformations the left- and right-hand sides are extensionally equal. In each case this follows directly from the definitions of the combinators involved.

Since Turner's pioneering papers appeared, many people have worked on improving the basic idea. For example, John Hughes has devised a scheme for dynamically generating an 'optimal' set of primitive combinators (called supercombinators) for each program [20]. The idea is that the compiler will generate combinatory expressions built out of the supercombinators for the program being compiled. It will also dynamically produce 'microcode' to implement the reduction rules for these supercombinators. The result is that each program runs on a reduction machine tailored specially for it. Most current high-performance implementations of functional languages use supercombinators [1, 11]. Another avenue of research is to use combinators based on the De Bruijn notation briefly described on page 11. The 'Categorical Abstract Machine' [26] uses this approach.

## A Quick Overview of ML

There are two widely use descendents of the original ML: Standard ML and Caml ${ }^{1}$. These notes ${ }^{2}$ describe the former. Several implementations of Standard ML exist. These all support the same core language, but differ in extensions, error message details etc. AT\&T's public domain "Standard ML of New Jersey" (SML/NJ) and the commercial system PolyML ${ }^{3}$ are used for research applications in the Cambridge Computer Laboratory. The ML implementation on Thor for teaching is "Edinburgh ML" from the University of Edinburgh (with enhancements due to Arthur Norman of Cambridge). The different outputs produced by SML/NJ and Edinburgh ML will be sometimes shown, but the examples that follow are presented in the system neutral style of Paulson's book (which is closer to Edinburgh ML than SML/NJ).

### 4.1 Interacting with ML

ML is an interactive language. A common way to run it is inside a shell window from emacs. The programs are then tested by 'cutting and pasting' from the text window to the shell window.
The two main things one does in ML are evaluate expressions and perform declarations.
What follows is a session in which simple uses of various ML constructs are illustrated. To make the session easier to follow, it is split into a sequence of boxed sub-sessions.

### 4.2 Expressions

The top-level ML prompt is "-". As ML reads a phrase it prompts with "=" until a complete expression or declaration is found. Neither the initial prompt - nor the intermediate prompt = will normally be shown here, except in sessions which are included to illustrate the behaviour of particular ML implementations (e.g. the next two boxes).
SML/NJ is called "sml" on Computer Lab machines. The following seesion shows it being run and the expression $2+3$ being evaluated.

[^6]| woodcock\% sml |  |
| :--- | :--- |
| Standard ML of New Jersey, Version 0.93, February 15, 1993 |  |
| val it $=$ () : unit | 1 |
| $-2+3$; |  |
| val it $=5$ : int |  |
| -it; |  |
| val it $=5$ : int |  |

After SML/NJ starts up it prints a message followed by val it = () : unit, (this will be explained later). It then prompts for user input with -, the user then input $2+3$; followed by a carriage return; ML then responded with val it = 5 : int, a new line, and then prompted again. This output shows that $2+3$ evaluates to the value 5 of type int.
The user then input it; followed by a carriage return, and the system responded with val it = 5 : int again. In general, to evaluate an expression $e$ one inputs $e$ followed by a semi-colon and then a carriage return; the system then prints $e$ 's value and type in the format shown. The value of the last expression evaluated at top level is remembered in the identifier it. This is shown explicitly in the output from SML/NJ, but not in the output from Edinburgh ML shown in the following box (which, after Edinburgh ML has been run, has the same input as the preceding one).

```
hammer.thor.cam.ac.uk% /group/clteach/acn/ml/unix/cml
                            2
FAM /group/clteach/acn/ml/unix/fam started on 02-Jan-1996 16:03:07
    (version 4.2.01 of Jan 25 1995)
Image file /group/clteach/acn/ml/unix/cml.exp
    (written on 25-Jan-1995 15:42:47 by FAM version 4.2.01)
[Loading Generic Heap...resexing...relocating by efff1ff8 (bytes)]
Edinburgh ML for DOS/Win32s/Unix (C) Edinburgh University & A C Norman
- 2+3;
> 5 : int
- it;
> 5 : int
```

Unless explicity indicated otherwise, the boxed sessions that follow use the format illustrated by:

```
2+3;
3
> val it = 5 : int
it;
> val it = 5 : int
```

Prompts (-) are not shown, system output is indicated by $>$ and the values of expressions are shown explicitly bound to it. Sometimes part of the output will be omitted (e.g. the type).

### 4.3 Declarations

The declaration val $x=e$ evaluates $e$ and binds the resulting value to $x$.

```
val x=2*3; 4
> val x = 6 : int
it=x;
> val it = false : bool
```

Notice that declarations do not affect it.
Inputting $e$; at top level is actually treated as inputting the declaration let it $=e ;$. The ML system (both SML/NJ and Edinburgh ML) initially binds it to a special value (), which is the only value of the one-element type unit.
To bind the variables $x_{1}, \ldots, x_{n}$ simultaneously to the values of the expressions $e_{1}, \ldots, e_{n}$ one can perform:

- either the declaration val $x_{1}=e_{1}$ and $x_{2}=e_{2} \ldots$ and $x_{n}=e_{n}$
- or val $\left(x_{1}, x_{2}, \ldots, x_{n}\right)=\left(e_{1}, e_{2}, \ldots, e_{n}\right)$.

These two declarations are equivalent.

```
val \(y=10\) and \(z=x\);
5
\(>\) val \(y=10\) : int
\(>\) val \(z=6\) : int
val ( \(\mathrm{x}, \mathrm{y}\) ) \(=(\mathrm{y}, \mathrm{x})\);
> val \(\mathrm{x}=10\) : int
\(>\) val \(\mathrm{y}=6\) : int
```

A declaration $d$ can be made local to the evaluation of an expression $e$ by evaluating the expression let $d$ in $e$ end.

```
let val x=2 in x*y end
> val it = 12 : int
x;
> val it = 10 : int
```

6

### 4.4 Comments

Comments start with (* and end with *). They nest like parentheses, can extend over many lines and can be inserted wherever spaces are allowed.

```
tr(* comments can't go in the middle of names *)ue;
> Error: unbound variable or constructor: tr
> Error: unbound variable or constructor: ue
1 (* this comment is ignored *) < 2;
> val it = true : bool
(* Inside this comment (* another one is nested *) ! *)
```


### 4.5 Functions

To define a function $f$ with formal parameter $x$ and body $e$ one performs the declaration: fun $f x=e$. To apply the function $f$ to an actual parameter $e$ one evaluates the expression: $f e$.
fun $f x=2 * x ;$
$>$ val $f=\mathrm{fn}:$ int $\rightarrow$ int
$f 4 ;$
$>$ val it $=8:$ int

Functions are printed as fn in SML/NJ and Fn in Edinburgh ML, since a function as such is not printable. After fn or Fn is printed, the type of the function is also printed. Functions are printed as fn in these notes.
Applying a function to an argument of the wrong type results in a typechecking error. The particular error message depends on the ML system used. In SML/NJ:

```
- f true; 10, 9
std_in:12.1-12.6 Error: operator and operand don't agree (tycon mismatch)
    operator domain: int
    operand: bool
    in expression:
        f true
```

In Edinburgh ML:

| -f true; |  | 10 |
| :--- | :--- | :--- |
| Type clash in: (f true) |  |  |
| Looking for a: int |  |  |
| I have found a: bool |  |  |

Application binds more tightly than anything else in the language; thus, for example, $f 3+4$ means $(f 3)+4$ not $f(3+4)$. Functions of several arguments can be defined:

```
fun add (x:int) (y:int) \(=x+y\);
11
> val add \(=\) fn : int \(->\) int \(->\) int
add 3 4;
> val it = 7 : int
val \(f=\) add 3 ;
> val \(f=\mathrm{fn}\) : int \(->\) int
f 4;
> val it \(=7\) : int
```

Application associates to the left, so add 34 means (add 3)4. In the expression add 3 , the function add is applied to 3 ; the resulting value is the function of type int -> int which adds 3 to its argument. Thus add takes its arguments 'one at a time'.
Without the explicit typing of the formal parameters, ML cannot tell whether the + is addition of integers or reals. The symbol + is overloaded. If the extra type information is omitted, an error results. In SML/NJ:

| - fun add $x$ y $=x+y$; | 12 |
| :--- | :---: |
| std_in:5.16 Error: overloaded variable "+" cannot be resolved |  |

In Edinburgh ML:

```
- fun add x y = x+y; 13
Type checking error in: (syntactic context unknown)
Unresolvable overloaded identifier: +
Definition cannot be found for the type: ('a * 'a) -> 'a
```

This kind of typechecking error is relatively rare. Much more common are errors resulting from applying functions to arguments of the wrong type.
The function add could alternatively have been defined to take a single argument of the product type int * int:

```
fun add(x,y):int = x+y; 14
> val add = fn : int * int -> int
add(3,4);
> val it = 7 : int
let val z = (3,4) in add z end;
> val it = 7 : int
add 3;
> std_in:2.1-2.5 Error: operator and operand don't agree (tycon mismatch)
    operator domain: int * int
    operand: int
    in expression:
        add 3
```

The error message shown here is the one generated by SML/NJ. Notice that this time the result of the function has had its type given explicitly. In general, it is sufficient to explicitly type any subexpression as long as this disambiguates all overloaded operators.
As well as taking structured arguments (e.g. $(3,4)$ ) functions may also return structured results.

```
fun sumdiff(x:int,y:int) \(=(x+y, x-y)\);
15
> val sumdiff \(=\) fn \(:\) int \(*\) int -> int \(*\) int
sumdiff (3,4);
> val it \(=(7, \sim 1)\) : int \(*\) int
```


### 4.6 Type abbreviations

Types can be given names:

```
type intpair = int * int; 
> type intpair defined
fun addpair ((x,y) :intpair) = x+y;
> val addpair = fn : intpair -> int
(3,5);
> val it = (3,5) : int * int
(3,5) :intpair;
> val it = (3,5) : intpair
addpair(3,5);
> val it = 8 : int
```

The new name is simply an abbreviation; intpair and int*int are completely equivalent.

### 4.7 Operators

+ (addition) and * are built-in infix operators. Users can define their own infixes using infix (for left associative operators) and infixr for right associative ones.

```
\begin{tabular}{l|l|}
\hline infix op1; & 17 \\
infixr op2; & \\
\(>\) infix op1 \\
\(>\) infixr op2 & \\
\hline
\end{tabular}
```

This merely tells the parser to parse $e_{1}$ op1 $e_{2}$ as op1 $\left(e_{1}, e_{2}\right)$ and $e_{1}$ op2 $e_{2}$ as op2 ( $e_{1}, e_{2}$ ).

```
fun (x:int) op1 (y:int) = x + y;
> val op1 = fn : int * int -> int
1 op1 2;
> val it = 3 : int
fun (x:int) op2 (y:int) = x * y;
> val op2 = fn : int * int -> int
2 op2 3;
> val it = 6 : int
```18

An infix of precedence \(n\) can be created by using infix \(n\) instead of just infix (and infixr \(n\) instead of just infixr). If the \(n\) is omitted a default precedence of 0 is assumed.

The ML parser can be told to ignore the infix status of an occurrence of an identifier by preceding the occurrence with op.
\begin{tabular}{|l|l|}
\hline op1; \\
\(>\) Error \(: ~ n o n f i x ~ i d e n t i f i e r ~ r e q u i r e d ~\) \\
op op1; \\
\(>\) val it \(=\) fn \(:\) int \(*\) int \(\rightarrow\) int
\end{tabular}

The infix status of an operator can be permanently removed using the directive nonfix.
```

1 + 2; 20
> val it = 3 : int
nonfix +;
> nonfix +
1 + 2;
Error: operator is not a function
operator: int
in expression:
1 + : overloaded

```

Removing the infix status of built-in operators is not recommended. Let's restore it before chaos results: + is left-associative with precedence 6 .

\subsection*{4.8 Lists}

If \(e_{1}, \ldots, e_{n}\) all have type \(t y\) then the ML expression \(\left[e_{1}, \ldots, e_{n}\right]\) has type ( \(t y\) list). The standard functions on lists are hd (head), tl (tail), null (which tests whether a list is empty-i.e. is equal to []), and the infixed operators : : (cons) and © (append, or concatenation).
```

val m = [1,2,(2+1), 4];
> val m = [1,2,3,4] : int list
(hd m , tl m);
> val it $=(1,[2,3,4])$ : int $*$ int list
(null m , null []);
> val it = (false,true) : bool * bool
0: :m;
> val it $=[0,1,2,3,4]$ : int list
$[1,2]$ © $[3,4,5,6] ;$
> val it = [1,2,3,4,5,6] : int list
[1, true, 2];
> std_in:3.1-3.10 Error: operator and operand don't agree (tycon mismatch)
operator domain: bool * bool list
operand: bool * int list
in expression:
true :: 2 :: nil

```

All the members of a list must have the same type (the error message shown is from SML/NJ).

\subsection*{4.9 Strings}

A sequence of characters enclosed between quotes (") is a string.
```

"this is a string";
> val it = "this is a string" : string
"";
> val it = "" : string

```

The empty string is "". A string can be 'exploded' into a list of single-character strings with the function explode. The inverse of this is implode, which concatenates a list of single-character strings into a single string.
```

explode;
24
> val it = fn : string -> string list
explode "this is a string";
> val it =
> ["t","h","i","s"," ","i","s"," ","a"," ","s","t","r","i","n","g"]
> : string list
implode it;
> val it = "this is a string" : string

```

\subsection*{4.10 Records}

Records are data-structures with named components. They can be contrasted with tuples whose components are determined by position.
A record with fields \(x_{1}, \ldots, x_{n}\) whose values are \(v_{1}, \ldots, v_{n}\) is created by evaluating the expression: \(\left\{x_{1}=v_{1}, \ldots, x_{n}=v_{n}\right\}\).
```

val MikeData =
{userid = "mjcg", sex = "male", married = true, children = 2};
> val MikeData = {children=2,married=true,sex="male",userid="mjcg"}
> : {children:int, married:bool, sex:string, userid:string}

```

The type of \(\left\{x_{1}=v_{1}, \ldots, x_{n}=v_{n}\right\}\) is \(\left\{x_{1}: \sigma_{1}, \ldots, x_{n}: \sigma_{n}\right\}\), where \(\sigma_{i}\) is the type of \(v_{i}\). The order in which record components are named does not matter:
```

val MikeData' =
{sex = "male", userid = "mjcg", children = 2, married = true};
> val MikeData' = {children=2,married=true,sex="male",userid="mjcg"}
: {children:int, married:bool, sex:string, userid:string}
MikeData = MikeData';
> val it = true : bool

```

The component named \(x\) of a record can be extracted using the special operator \#x.
\begin{tabular}{|l|l|}
\hline \#children MikeData; & 27 \\
\(>\) val it \(=2\) : int
\end{tabular}

Functions which access record components need to be explicitly told the type of the record they are accessing, since there may be several types of records around with the same field names.
```

fun Sex p = \#sex p;
> Error: unresolved flex record in let pattern
type persondata = {userid:string, children:int, married:bool, sex:string};
> type persondata = {children:int, married:bool, sex:string, userid:string}
fun Sex(p:persondata) = \#sex p;
> val Sex = fn : persondata -> string

```

A tuple \(\left(v_{1}, \ldots, v_{n}\right)\) is equivalent to the record \(\left\{1=v_{1}, \ldots, n=v_{n}\right\}\) (i.e. tuples in ML are special cases of records).
```

{1 = "Hello", 2 = true, 3 = 0};
> val it = ("Hello",true,0) : string * bool * int
\#2 it;
> val it = true : bool

```

\subsection*{4.11 Polymorphism}

The list processing functions hd, tl etc. can be used on all types of lists.
```

hd [1,2,3]; 草 30
> val it = 1 : int
hd [true,false,true];
> val it = true : bool
hd [(1,2),(3,4)];
> val it = (1,2) : int * int

```

Thus hd has several types: above it is used with types (int list) -> int, (bool list) -> bool and (int * int) list -> (int * int). In fact if ty is any type then hd has the type ( \(t y\) list) -> ty. Functions, like hd, with many types are called polymorphic, and ML uses type variables ' a , ' b , ' c etc. to represent their types.
\begin{tabular}{|l|l|}
\hline hd ; \\
\(>\) val it \(=f n: ~ ' a ~ l i s t ~\)
\end{tabular}\(>\) 'a \(\quad 31\)

The ML function map takes a function \(f\) (with argument type 'a and result type \({ }^{\prime} \mathrm{b}\) ), and a list \(l\) (of elements of type 'a), and returns the list obtained by applying \(f\) to each element of \(l\) (which is a list of elements of type 'b).

map can be used at any instance of its type: above, both 'a and 'b were instantiated to int; below, 'a is instantiated to (int list) and 'b to bool. Notice that the instance need not be specified; it is determined by the type checker.
\begin{tabular}{|l|l|}
\hline \begin{tabular}{l} 
map null \([[1,2], ~[], ~[3], ~[]] ; ~\) \\
\(>\) \\
val it \(=[f a l s e\), true,false,true]
\end{tabular} & bool list
\end{tabular}

A useful built-in operator is function composition o
```

op o; 34
> val it = fn : ('a -> 'b) * ('c -> 'a) -> 'c -> 'b
fun add1 n = n+1
and add2 n = n+2;
> val add1 = fn : int -> int
> val add2 = fn : int -> int
(add1 o add2) 5;
> val it = 8 : int

```

\subsection*{4.12 fn-expressions}

The expression \(\mathrm{fn} x \Rightarrow e\) evaluates to a function with formal parameter \(x\) and with body \(e\). Thus the declaration fun \(f x=e\) is equivalent to val \(f=\mathrm{fn} x \Rightarrow e\). Similarly fun \(f(x, y) z=e\) is equivalent to val \(f=\mathrm{fn}(x, y) \Rightarrow \mathrm{fn} z \Rightarrow e\). In the theory of functions, the symbol \(\lambda\) is used instead of fn ; expressions like \(\mathrm{fn} x \Rightarrow e\) are sometimes called \(\lambda\)-expressions, because they correspond to \(\lambda\)-calulus abstractions \(\lambda x . e(\) see Chapter 1).
\begin{tabular}{|l|l|}
\hline fn \(x=>x+1 ;\) \\
\(>\) val it \(=\) fn \(:\) int \(\rightarrow\) int & 35 \\
it \(3 ;\) \\
\(>\) val it \(=4:\) int & \\
\hline
\end{tabular}

The higher order function map applies a function to each element of a list in turn and returns the list of results.
```

map (fn x => x*x) [1,2,3,4]; 36
$>$ val it $=[1,4,9,16]$ : int list
val doubleup $=\operatorname{map}(f n x=x @ x)$;
> val doubleup $=$ fn : 'a list list -> 'a list list
doubleup $[[1,2]$, $[3,4,5]]$;
> val it $=[[1,2,1,2],[3,4,5,3,4,5]]$ : int list list
doubleup [];
> val it $=[]$ : 'a list list

```

\subsection*{4.13 Conditionals}

ML has conditionals with syntax if \(e\) then \(e_{1}\) else \(e_{2}\) with the expected meaning. The truthvalues are true and false, both of type bool.
```

if true then 1 else 2;
> val it = 1 : int
if 2<1 then 1 else 2;
> val it = 2 : int

```
\(e_{1}\) orelse \(e_{2}\) abbreviates if \(e_{1}\) then true else \(e_{2}\) and \(e_{1}\) andalso \(e_{2}\) abbreviates if \(e_{1}\) then \(e_{2}\) else false.

\subsection*{4.14 Recursion}

The following defines the factorial function:
```

fun fact n = if n=0 then 1 else n*fact(n-1);
> val fact = fn : int -> int
fact 5;
> val it = 120 : int

```

Notice that the compiler automatically detects recursive calls. In earlier versions of ML, recursion had to be explicitly indicated.
Consider:
```

fun f n : int = n+1;
39
> val f = fn : int -> int
fun f n = if n=0 then 1 else n*f(n-1);
> val f = fn : int -> int
f 3;
> val it = 6 : int

```

Here f 3 results in the evaluation of \(3 * f(2)\). In earlier versions of ML, the first \(f\) would have been used, so that \(f(2)\) would have evaluated to \(2+1=3\), hence the expression f 3 would have evaluated to \(3 * 3=9\).
An alternative style of defining functions in Standard ML that avoids enforced recursion uses val and fn.
```

fun f n : int = n+1;
> val f = fn : int -> int
val f = fn n => if n=0 then 1 else n*f(n-1);
> val f = fn : int -> int
f 3;
> val it = 9 : int

```

Here, the occurrence of \(f\) in \(n * f(n-1)\) is interpreted as the previous version of \(f\). The keyword rec after val can be used to force a recursive interpretation:
```

fun f n : int = n + 1;
> val f = fn : int -> int
val rec f = fn n => if n=0 then 1 else n*f(n-1);
> val f = fn : int -> int
f 3;
> val it = 6 : int

```

With val rec the occurrence of \(f\) in \(n * f(n-1)\) is interpreted recursively.

\subsection*{4.15 Equality types}

Simple 'concrete' values like integers, booleans and strings are easy to test for equality. Values of simple datatypes, like pairs and records, whose components have concrete types are also easy to test for equality. For example, ( \(v_{1}, v_{2}\) ) is equal to ( \(v_{1}^{\prime}, v_{2}^{\prime}\) ) if and only if \(v_{1}=v_{1}^{\prime}\) and \(v_{2}=v_{2}^{\prime}\). There is thus a large class of types whoses values can be tested for equality. However, in general it is undecidable to test the equality of functions. It is thus not possible to overload = to work properly on all types. In old versions of ML, = was interpreted on functions by testing the equality of the addresses in memory of the data-structure representing the functions. If such a test yielded true then the functions were certainly equal, but many mathematically (i.e. extensionally) equal functions were different using this interpretation of \(=\).
In Standard ML, those types whose values can be tested for equality are called "equality types" and are treated specially. Special type variables that are constrained only to range over equality types are provided. These have the form ' ' \(\alpha\), whereas ordinary type variables have the form ' \(\alpha\). The built-in function \(=\) has type ''a * ''a -> bool. Starting from this, the ML typechecker can infer types containing equality type variables.
\begin{tabular}{|c|c|}
\hline \multicolumn{2}{|l|}{fun Eq \(x\) y \(=(x=y)\);} \\
\hline > val Eq = fn : ''a -> ''a -> bool & \\
\hline ```
fun EqualHd l1 12 = (hd l1 = hd 12);
> val EqualHd = fn : ''a list -> ''a list -> bool
``` & \\
\hline
\end{tabular}

Trying to instantiate an equality type variable to a functional type results in an error. In SML/NJ:
```

hd = hd;
> Error: operator and operand don't agree (equality type required)
operator domain: ''Z * ''Z
operand: ('Y list -> 'Y) * ('X list -> 'X)
in expression:
= (hd,hd)
EqualHd [hd] [hd];
Error: operator and operand don't agree (tycon mismatch)
operator domain: 'Z * 'Z
operand: ,'Y list -> ',Y list -> bool
in expression:
- : overloaded EqualHd

```

The use of equality types in Standard ML is considered controversial: some people think they are too messy for the benefit they provide. It is possible that future versions of ML will drop equality types.

\subsection*{4.16 Pattern matching}

Functions can be defined by pattern matching. For example here is another definition of the factorial function.
```

fun fact 0 = 1
| fact n = n * (fact(n-1));
> val fact = fn : int -> int

```

Here is the Fibonacci function:
```

fun fib 0 = 0 45
fib 1 = 1
| fib n = fib(n-1) + fib(n-2);
> val fib = fn : int -> int

```

Suppose function \(f\) is defined by:
\begin{tabular}{|c|c|}
\hline fun \(f p_{1}=e_{1}\) & 46 \\
\hline | \(f p_{2}=e_{2}\) & \\
\hline & \\
\hline \(f p_{n}=e_{n}\) & \\
\hline
\end{tabular}

An expression \(f e\) is evaluated by successively matching the value of \(e\) with the patterns \(p_{1}, p_{2}, \ldots, p_{n}\) (in that order) until a match is found, say with \(p_{i}\). Then the value of \(f e\) is the value of \(e_{i}\). During the match variables in the patterns may be bound to components of \(e\) 's value and then the variables have these values during the evaluation of \(e_{i}\). For example, evaluation fib 8 causes 8 to be matched with 0 then 1 , both of which fail, and then with \(n\) which succeeds, binding \(n\) to 8 . The result is then the value of \(f i b(8-1)+f i b(8-2)\) which (after some recursive calls), evaluates to 21 .
\begin{tabular}{|l|c|}
\hline fib \(8 ;\) \\
\(>\) val it \(=21:\) int & 47 \\
\hline
\end{tabular}

Patterns can be quite elaborate and are typically composed with 'constructors' (see Section 4.19 below).
The patterns in a function definition need not be exhaustive. In SML/NJ;
```

- fun foo 0 = 0;
std_in:33.1-33.13 Warning: match nonexhaustive
0 => ...
val foo = fn : int -> int

```
48

In Edinburgh ML:
- fun foo \(0=0 ;\)
\(* *\) Warning: Patterns in Match not exhaustive: \(0=>0\)
\(>\) val foo \(=\) Fn \(:\) int \(\rightarrow\) int

If a function is defined with a non-exhaustive match, and then applied to an argument whose value doesn't match any pattern a special kind of run-time error called an exception results (see Section 4.18).
In SML/NJ:
\begin{tabular}{|l|l|}
\hline- foo \(0 ;\) \\
val it \(=0\) : int & 50 \\
- foo \(1 ;\) & \\
uncaught Match exception std_in:33.1-33.13 & \\
\hline
\end{tabular}

In Edinburgh ML:
```

- foo 1; 51
Exception raised at top level
Warning: optimisations enabled -
some functions may be missing from the trace
Exception: Match raised

```

Messages warning that a match is non-exhaustive will sometimes be omitted from the output shown here.
The built-in list-processing functions hd and tl can be defined by:
```

fun hd(x::l) = x; 52
> Warning: match nonexhaustive
> val hd = fn : 'a list -> 'a
fun tl(x::l) = l;
> Warning: match nonexhaustive
> val tl = fn : 'a list -> 'a list

```

These definitions give exactly the same results as the built-in functions except on the empty list [], where they differ in the exceptions raised - exceptions are described in Section 4.18.
if \(x\) is a variable and \(p\) a pattern, then the pattern \(x\) as \(p\) is a pattern that matches the same things as \(p\), but has the additional effect that when a match succeeds the value matched is bound to \(x\). Consider the function RemoveDuplicates:
The wildcard "_" matches anything:
```

fun null [] = true
| null _ = false;
> val null = fn : 'a list -> bool

```
\begin{tabular}{|c|c|}
\hline fun RemoveDuplicates [] = [] & 54 \\
\hline | RemoveDuplicates[x] = [x] & \\
\hline | RemoveDuplicates(x1::x2::1) = & \\
\hline if \(\mathrm{x} 1=\mathrm{x} 2\) then RemoveDuplicates(x2: 1 ) & \\
\hline else x1::RemoveDuplicates(x2::1) ; & \\
\hline >val RemoveDuplicates \(=\) fn : ', a list -> ', a list & \\
\hline RemoveDuplicates[1, 1, 1, 2, 3, 4, 5, 5, 5, 5, 5, 6, 7, 8, 8, 8] ; & \\
\hline > val it \(=[1,2,3,4,5,6,7,8]\) : int list & \\
\hline
\end{tabular}

The repetition (and extra list conses) of \(x 2:: 1\) can be avoided as follows:


Incidently, note that (alas) duplicate variables are not allowed in patterns:


Anonymous functions (fn-expressions) can be defined by pattern matching using the syntax: fn \(p_{1} \Rightarrow e_{1}|\ldots| p_{n} \Rightarrow e_{n}\)
```

fn [] => "none" | [_] => "one" | [_,_] => "two" | _ => "many";
> val it = fn : 'a list -> string
(it [], it[true], it[1,2], it[1,2,3]);
> val it = ("none","one","two","many") : string * string * string * string

```

Patterns can be constructed out of records, with "..." as a wildcard.
```

fun IsMale({sex="male",...}:persondata) = true
58
| IsMale _ = false;
> val IsMale = fn : persondata -> bool
IsMale MikeData;
> val it = true : bool

```

An alternative definition is:
\begin{tabular}{|c|c|}
\hline fun IsMale (\{sex=x, ...\}:persondata) \(=(x=\) "male"); & 59 \\
\hline
\end{tabular}

A more compact form of this is allowed.
\begin{tabular}{|l|l|}
\hline fun IsMale(\{sex,...\}:persondata) \(=(\) sex \(=\) "male"); & 60 \\
\hline
\end{tabular}

The field name sex doubles as a variable. Think of a pattern \(\{\cdots, v, \cdots\}\) as abbreviating \(\{\cdots, \mathrm{v}=\mathrm{v}, \cdots\}\).

\subsection*{4.17 The case construct}

The case construct permits one to compute by cases on an expression of a datatype. The expression case \(e\) of \(p_{1} \Rightarrow e_{1}|\ldots| p_{n} \Rightarrow e_{n}\), is an equivalent form for the application (fn \(p_{1} \Rightarrow e_{1}|\ldots| p_{n} \Rightarrow e_{n}\) ) \(e\).

\subsection*{4.18 Exceptions}

Some standard functions raise exceptions at run-time on certain arguments. When this happens a special kind of value (called an exception packet) is propagated which identifies the cause of the exception. These packets have names which usually reflect the function that raised the exception; they may also contain values.
\begin{tabular}{|l|c|}
\hline hd (tl[2]); & 61 \\
\(>\) uncaught exception Hd & \\
1 div 0; & \\
\(>\) uncaught exception Div & \\
(1 div 0\()+1000 ;\) \\
\(>\) uncaught exception Div & \\
\hline
\end{tabular}

Exceptions must be declared using the keyword exception; they have type exn. Exceptions can be explicitly raised by evaluating an expression of the form raise \(e\) where \(e\) evaluates to an exception value. Exceptions are printed slightly differently in SML/NJ and Edinburgh ML. In SML/NJ:
```

- exception Ex1;exception Ex2;
62
exception Ex1
exception Ex2
- [Ex1,Ex2];
val it = [Ex1(-),Ex2(-)] : exn list
- raise hd it;
uncaught exception Ex1

```

In Edinburgh ML:
```

exception Ex1; exception Ex2;
$>$ type exn
con Ex1 = - : exn
$>$ type exn
con Ex2 $=$ - : exn
$-\quad[E x 1, E x 2] ;$
$>[-,-]$ : exn list

- raise hd it;
Exception raised at top level
Warning: optimisations enabled -
some functions may be missing from the trace
Exception: Ex1 raised

```

An exception packet constructor called name and which constructs packets containing values of type \(t y\) is declared by exception name of \(t y\).
```

exception Ex3 of string;
64
> exception Ex3
Ex3;
> val it = fn : string -> exn
raise Ex3 "foo";
> uncaught exception Ex3

```

The type exn is a datatype (see Section 4.19 below) whose constructors are the exceptions. It is the only datatype that can be dynamically extended. All other datatypes have to have all their constructors declared at the time when the datatype is declared.
Because exn is a datatype, exceptions can be used in patterns like other constructors. This is useful for handling exceptions.
An exception can be trapped (and its contents extracted) using an exception handler. An important special case is unconditional trapping of all exceptions. The value of the expression \(e_{1}\) handle _ \(\Rightarrow e_{2}\) is that of \(e_{1}\), unless \(e_{1}\) raises an exception, in which case it is the value of \(e_{2}\).
```

hd[1,2,3] handle _ => 0; 65
> val it = 1 : int
hd [] handle _ => 0;
> val it $=0$ : int
hd(tl[2]) handle _ => 0;
> val it = 0 : int
1 div 0 handle _ => 1000;
> val it $=1000$ : int

```

The function half, defined below, only succeeds (i.e. doesn't raise an exception) on non-zero even numbers; on 0 it raises Zero, and on odd numbers it raises Odd.
```

exception Zero; exception Odd; }6
> exception Zero
> exception Odd
fun half n =
if n=0 then raise Zero
else let
val m = n div 2
in
if n=2*m then m else raise Odd
end;
> val half = fn : int -> int

```

Some examples of using half:
```

half 4;
> val it = 2 : int
half 0;
> uncaught exception Zero
half 3;
> uncaught exception Odd
half 3 handle _ => 1000;
> val it = 1000 : int

```

Failures may be trapped selectively by matching the exception packet; this is done by replacing the wildcard _ by a pattern. For example, if \(e\) raises \(E x\), then the value of \(e\) handle \(E x_{1} \Rightarrow e_{1}|\ldots| E x_{n} \Rightarrow e_{n}\) is the value of \(e_{i}\) if \(E x\) equals \(E x_{i}\) otherwise the handle-expression raises \(E x\).
```

half(0) handle Zero => 1000; 68
> val it = 1000 : int
half(1) handle Zero => 1000;
> uncaught exception Odd
half(0) handle Zero => 1000 | Odd => 1001;
> val it = 1000 : int
half(3) handle Zero => 1000 | Odd => 1001;
> val it = 1001 : int

```

Instead of having the two exceptions Zero and Odd, one could have a single kind of exception containing a string.
```

exception Half of string; 6 6 %
> exception Half
fun half n =
if n=O then raise Half "Zero"
else let
val m = n div 2
in
if n=2*m then m else raise Half "Odd"
end;
> val half = fn : int -> int

```

A disadvantage of this approach is that the kind of exception is not printed when the exceptions are uncaught.
```

half 0; 年 70
> uncaught exception Half
half 3;
> uncaught exception Half
half(0) handle Half "Zero" => 1000 | Half "Odd" => 1001;
> val it = 1000 : int
half(3) handle Half "Zero" => 1000 | Half "Odd" => 1001;
> val it = 1001 : int

```

Alternatively, one can match the contents of the exception packet to a variable, s say, and then branch on the value matched to s.
```

half(0) handle Half s => (if s="Zero" then 1000 else 1001)
71
> val it = 1000 : int
half(3) handle Half s => (if s="Zero" then 1000 else 1001);
> val it = 1001 : int

```

\subsection*{4.19 Datatype declarations}

New types (rather than mere abbreviations) can also be defined. Datatypes are types defined by a set of constructors which can be used to create objects of that type and also (in patterns) to decompose objects of that type. For example, to define a type card one could use the construct datatype:
```

datatype card = king | queen | jack | other of int;
datatype card
con jack : card
con king : card
con other : int -> card
con queen : card

```

Such a declaration declares king, queen, jack and other as constructors and gives them values. The value of a 0 -ary constructor such as king is the constant value king. The value of a constructor such as other is a constructor function that given an integer value \(n\) produces other ( \(n\) ).
```

king; % 73
> val it = king : card
other(4+5);
> val it = other 9 : card

```

To define functions that take their argument from a concrete type, fn-expressions of the form fn \(p_{1} \Rightarrow e_{1}|\ldots| p_{n} \Rightarrow e_{n}\) can be used. Such an expression denotes a function that given a value \(v\) selects the first pattern that matches \(v\), say \(p_{i}\), binds the variables of \(p_{i}\) to the corresponding components of the value and then evaluates the expression \(e_{i}\). For example, the values of the different cards can be defined in the following way:


Alternatively, and perhaps more lucidly, this could be defined using a fun declaration.
\begin{tabular}{|cl|l|}
\hline fun value king & \(=500\) \\
value queen & \(=200\) \\
\(\mid\) value jack & \(=100\) \\
\(\mid\) value (other n) & \(=5 * n ;\) \\
\(>\) val value \(=\) fn \(:\) card \(->\) int & \\
\hline
\end{tabular}

The notion of datatype is very basic and could enable us to build ML's elementary types from scratch. For example, the booleans could be defined simply by:
\begin{tabular}{|l|l|}
\hline datatype bool = true | false; & 76 \\
\(>\) datatype bool & \\
\(>\) con false : bool & \\
\(>\) con true : bool & \\
\hline
\end{tabular}
and the positive integers by:
\begin{tabular}{|l|c|}
\hline datatype int \(=\) zero \(\mid\) suc of int; & 77 \\
\(>\) datatype int & \\
\(>\) con suc \(:\) int \(\rightarrow\) int & \\
\(>\) con zero \(:\) int & \\
\hline
\end{tabular}
datatype int \(=\) zero | suc of int;
77
con suc : int -> int
    con zero : int

In a similar way, LISP S-expressions could be defined by:
```

datatype sexp = litatom of string }7
| numatom of int
| cons of sexp * sexp;
> datatype sexp
> con cons : sexp * sexp -> sexp
> con litatom : string -> sexp
> con numatom : int -> sexp
fun car (cons(x,y)) = x and cdr (cons(x,y)) = y;
> Warning: match nonexhaustive
> val car = fn : sexp -> sexp
> Warning: match nonexhaustive
> val cdr = fn : sexp -> sexp
val a1 = litatom "Foo" and a2 = numatom 1;
> val a1 = litatom "Foo" : sexp
> val a2 = numatom 1 : sexp
car(cons(a1,a2));
> val it = litatom "Foo" : sexp
cdr(cons(a1,a2));
> val it = numatom 1 : sexp

```

Notice the warning from the compiler that the patterns in the definitions of car and cdr are not exhaustive; these funtions are only partially specified - namely only on lists built with cons (i.e. non-atoms).
```

car (litatom "foo");
> uncaught exception Match

```

\subsection*{4.20 Abstract types}

New types can also be defined by abstraction. For example, a type time could be defined as follows:
```

exception BadTime; }\quad8
> exception BadTime
abstype time = time of int * int
with
fun maketime(hrs,mins) = if hrs<0 orelse 23<hrs orelse
mins<0 orelse 59<mins
then raise BadTime
else time(hrs,mins)
and hours(time(t1,t2)) = t1
and minutes(time(t1,t2)) = t2
end;
> type time
> val maketime = fn : int * int -> time
val hours = fn : time -> int
val minutes = fn : time -> int

```

This defines an abstract type time and three primitive functions: maketime, hours and minutes.
In general, an abstract type declaration has the form abstype \(d\) with \(b\) end where \(d\) is a datatype specification and \(b\) is a binding, i.e. the kind of phrase that can follow val. Such a declaration introduces a new type, \(t y\) say, as specified by the datatype declaration \(d\). However, the constructors declared on \(t y\) by \(d\) are only available within \(b\). The only bindings that result from executing the abstype declaration are those specified in \(b\).

Thus an abstract type declaration simultaneously declares a new type together with primitive functions for the type; the representation datatype is not accessible outside the with-part of the declaration.

```

$>$ val $\mathrm{t}=-$ : time
(hours t , minutes t ) ;
$>$ val it $=(8,30):$ int $*$ int

```

Notice that values of an abstract type are printed as -, since their representation is hidden from the user.

\subsection*{4.21 Type constructors}

Both list and * are examples of type constructors; list has one argument (hence 'a list) whereas * has two (hence 'a * 'b). Type constructors may have various predefined operations associated with them, for example list has null, hd, tl, ... etc. Because of pattern matching, it is not necessary to have any predefined operations for *. One can define, for example, fst and snd by.
```

fun fst(x,y) = x and snd(x,y) = y;
> val fst = fn : 'a * 'b -> 'a
> val snd = fn : 'a * 'b -> 'b
val p = (8,30);
> val p = (8,30) : int * int
fst p;
> val it = 8 : int
snd p;
> val it = 30 : int

```
        82

A type constructor set, that represents sets by lists without repetitions, can be defined in the following way:
```

abstype 'a set = set of 'a list
with
val emptyset = set[]
fun isempty(set s) = null s
fun member(_, set[]) = false
| member(x, set(y::z)) = (x=y) orelse member(x, set z)
fun add(x, set[]) = set[x]
| add(x, set(y::z)) = if x=y
then set(y::z)
else let val set l = add(x, set z) in
set(y::l)
end
end
val emptyset = [] : 'a list
val isempty = fn : 'a set -> bool
val member = fn : ''a * ''a set -> bool
val add = fn :''a *''a set ->''a set

```

Note that the operation add ensures that no repetitions of elements occur in the list representing the set. Here is an example using these sets:
```

val s = add(1,(add(2,(add(3,emptyset)))));
> val s = - : int set
member(3,s);
> val it = true : bool
member(5,s);
> val it = false : bool

```

\subsection*{4.22 References and assignment}

References are 'boxes' that can contain values. The contents of such boxes can be changed using the assignment operator : \(=\). The type \(t y\) ref is possessed by references containing values of type \(t y\).
References are created using the ref operator. This takes a value of type ty to a value of type ty ref. \({ }^{4}\) The expression \(x:=e\) changes the contents of the reference that is the value of \(x\) to be the value of \(e\). The value of this assignment expression is the dummy value (); this is the unique value of the one-element type unit. Assignments are executed for a 'side effect', not for their value.
The contents of a reference can be extracted using the ! operator (error message below from SML/NJ).
```

x:=1; 85
> std_in:7.1-7.4 Error: operator and operand don't agree (tycon mismatch)
operator domain: 'Z ref * 'Z
operand: int * int
in expression:
:= (x,1)
val x = ref 1 and y = ref 2;
> val x = ref 1 : int ref
> val y = ref 2 : int ref
x;
> val it = ref 1 : int ref
x:=6;
> val it = () : unit
x;
> val it = ref 6 : int ref
!x;
> val it = 6 : int

```

References should only be resorted to in exceptional circumstances as experience shows that their use increases the probability of errors.

\subsection*{4.23 Iteration}

Here is an iterative definition of fact using two local references: count and result.

\footnotetext{
\({ }^{4}\) There are some horrible subtleties associated with the types of references, which are ignored here. The treatment of references in ML is currently in a state of flux.
}
```

fun fact n
let val count = ref n and result = ref 1
in while !count > 0
do (result := !count * !result;
count := !count-1);
!result
end;
> val fact = fn : int -> int
fact 6;
>val it = 720 : int

```

The semicolon denotes sequencing. When an expression \(e_{1} ; \ldots ; e_{n}\) is evaluated, each \(e_{i}\) is evaluated in turn and the value of the entire expression is the value of \(e_{n}\).
Evaluating while \(e\) do \(c\) consists in evaluating \(e\) and if the result is true \(c\) is evaluated for its side-effect and then the whole process repeats. If \(e\) evaluates to false, then the evaluation of while \(e\) do \(c\) terminates with value ().

\subsection*{4.24 Programming in the large}

Sophisticated features for structuring collections of declarations ('programming in the large') are provided in Standard ML (but not in earlier versions of ML). These are designed to support the use of ML for large scale system building. They account for much of the complexity of the language.
Standard ML of New Jersey is implemented in itself and makes extensive use of these features. Edinburgh ML does not implement them.
The concepts of structures, signatures and functors, which provide the structuring constructs for programming in the large, are not covered in this course (hence their absence from Edinburgh ML on Thor will not be a problem).

\section*{Case study 1: parsing}

The lexical analysis and parsing programs described here are intended to illustrate functional programming methods and ML, rather than parsing theory. The style of parsing presented is quite reasonable for small lightweight ad hoc parsers, but would be inappropriate for large applications, which should be handled using heavyweight parser generators like YACC. \({ }^{1}\)

\subsection*{5.1 Lexical analysis}

Lexical analysis converts sequences of characters into sequences of tokens (also called "words" or "lexemes").
For us, a token will be either a number (sequence of digits), an identifier (a sequence of letters or digits starting with a letter) or a 'special symbol' such as \(+, *,<,==>\) or ++. Special symbols are specified by a table (see below).
A number is a sequence of digits, The ML infix operator \(<=\) is overloaded and can be applied to strings. If \(x\) and \(y\) are single-character strings, then \(x<=y\) just tests whether the ASCII code of \(x\) is less then or equal to that of \(y\). Thus a singlecharacter string representing a digit can be characterised by the predicate IsDigit:
\begin{tabular}{|l|l|}
\hline fun IsDigit \(x=" 0 "<=x\) andalso \(x<=" 9 " ;\) & 87 \\
\(>\) val IsDigit \(=\) fn \(:\) string \(\rightarrow\) bool & \\
\hline
\end{tabular}

A letter can similarly be characterised by making use of the fact that the ASCII codes of all lower case letters are adjacent and also the codes of all upper case latters are adjacent.
```

fun IsLetter x = 88
("a" <= x andalso x <= "z") orelse ("A" <= x andalso x <= "Z");
> val IsLetter = fn : string -> bool

```

Token are separated by 'separators', which will be taken to be spaces, newlines and tabs, hence:
\begin{tabular}{|l|l|}
\hline fun IsSeparator \(x=(x=" ~ "\) orelse \(x=" \backslash n "\) orelse \(x=" \backslash t ") ;\) & 89 \\
\(>\) val IsSeparator \(=f n:\) string \(\rightarrow>\) bool & \\
\hline
\end{tabular}

Single characters that are not digits, letters or separators will be assumed to be special symbols. Multi-character special symbols (e.g. ==>) are considered later.
The input is assumed to be supplied as a list of single-charater strings. Lexical analysis consists on converting such a list to a list of tokens.
Suppose the input just consists of numbers separated by separators. A function Tokenise that did lexical analysis for just this case would need to repeatedly remove

\footnotetext{
\({ }^{1}\) There is an ML based version of YACC. The parser for SML/NJ uses this.
}
digits until a non-digit (e.g. a separator) was reached, and then implode the removed characters into a string representing a token and add that to the list of tokens.
The function GetNumber takes a list, \(l\) say, of single-character strings and returns a pair consisting of (i) a string representing a number consisting of all the digits in \(l\) up to the first non-digit and (ii) the remainder of \(l\) after these digits have been removed. It is convenient to define GetNum using an auxiliary function GetNumAux that has an extra argument buf for accumulating a (reversed) list of characters making up the number.
```

fun GetNumAux buf [] = (implode(rev buf), [])
| GetNumAux buf (l as (x::l')) =
if IsDigit x then GetNumAux (x::buf) l'
else (implode(rev buf),l);
> val GetNumAux = fn : string list -> string list -> string * string list
GetNumAux ["a","b","c"] ["1","2","3"," ","4","5"];
> val it = ("cba123",[" ","4","5"]) : string * string list

```

Then GetNum is simply defined by:
val GetNum = GetNumAux [];
> val GetNum = fn : string list \(\rightarrow\) string * string list
GetNum ["1","2","3"," ","4","5"];
\(>\) val it \(=(" 123 ",[" 4, " 4 ", " 5 "])\) : string * string list

The definition of GetNumAux could have been localised to GetNum using local...in...end.
Notice that if the list argument of GetNum doesn't start with a number, then the empty token (implode[]) will be returned.
\begin{tabular}{|l|l|}
\hline GetNum ["a","0","1"]; & 92 \\
\(>\) val it \(=(" ",[" a ", " 0 ", " 1 "]):\) string \(*\) string list & \\
\hline
\end{tabular}

This problem will go away when we improve the code later on.
Identifiers can be lexically analysed by similar programs:
```

fun GetIdentAux buf [] = (implode(rev buf), [])
93
| GetIdentAux buf (l as (x::l')) =
if IsLetter x orelse IsDigit x
then GetIdentAux (x::buf) l'
else (implode(rev buf),l);
> val GetIdentAux = fn : string list -> string list -> string * string list
GetIdentAux ["a","b","c"] ["e","f","g","4","5"," ","6","7"];
> val it = ("cbaefg45",[" ","6","7"]) : string * string list

```

An identifier must start with a letter, so GetIdent is defined by:
```

exception GetIdentErr; }9
> exception GetIdentErr
fun GetIdent (x::1) =
if IsLetter x then GetIdentAux [x] l else raise GetIdentErr;
> val GetIdent = fn : string list -> string * string list

```

The lexical analysis of numbers and identifiers can be streamlined and unified by defining a single general function GetTail that takes a predicate as an argument
and then uses this to test whether to keep accumulating characters in buf or to terminate. Then GetNumAux corresponds to GetTail IsDigit and GetIdentAux to GetTail ( \(\mathrm{fn} \mathrm{x}=>\) IsLetter x orelse IsDigit x ).
The definition of GetTail is similar to that of GetNumAux and GetIdentAux.
```

fun GetTail p buf [] = (implode(rev buf),[])
| GetTail p buf (l as x::l') =
if p x then GetTail p (x::buf) l' else (implode(rev buf),l);
> val GetTail = fn
> : (string->bool) -> string list -> string list -> string * string list

```

Using GetTail, a function to get the next token is easy to define:


To lexically analyse a list of characters, GetNextToken is repeatedly called and separators are discarded.
```

fun Tokenise [] = []
| Tokenise (l as x::1') =
if IsSeparator x
then Tokenise l'
else let val (t,l'') = GetNextToken l
in t::(Tokenise l',) end;
> val Tokenise = fn : string list -> string list
Tokenise (explode "123abcde1][ ] 56a");
> val it = ["123","abcde1","]","[","]","56","a"] : string list

```

Tokenise does not handle multi-character special symbols. These will be specified by a table, represented as a list of pairs, that shows which characters can follow each initial segment of each special symbol (such a table represents a state-transition function for an automaton). For example, suppose the special symbols are <=, <<, =>, =, ==> and ->, then the table would be:
\begin{tabular}{|ll|l|}
\hline\([("<"\), & \(["="), "<"])\), & 98 \\
\(("="\), & \([">", "="])\), & \\
\(("-"\), & \([">"])\), \\
\(("=="\), & \([">"])]\)
\end{tabular}

This is not fully general because if \(=\Rightarrow\) is a special symbol, then the representation above forces \(==\) to be also. A fully general treatment of special symbols is left as an exercise.
Some utility functions are needed. Mem tests whether an element occurs in a list.
```

fun Mem x [] = false
99
| Mem x (x'::l) = (x=x') orelse Mem x l;
> val Mem = fn : ''a -> ''a list -> bool
Mem 3 [1,2,3,4,5,6,7];
> val it = true : bool
Mem 9 [1,2,3,4,5,6,7];
> val it = false : bool

```

Notice the equality types.
Get looks up the list of possible successors of a given string in a special-symbol table.
```

fun Get x [] = [] 100
| Get x ((x',l)::rest) = if x=x' then l else Get x rest;
> val Get = fn : ''a -> (''a * 'b list) list -> 'b list
Get "=" [("<",["=","<"]), ("=", [">","="]), ("-", [">"]), ("==", [">"])];
> val it = [">","="] : string list
Get "?" [("<",["=","<"]), ("=",[">","="]), ("-",[">"]), ("==", [">"])];
> val it = [] : string list

```

The function GetSymbol takes a special-symbol table and a token and then extends the token by removing characters from the input until the table says that no further extension is possible.
```

fun GetSymbol spectab tok [] = (tok,[]) 101
| GetSymbol spectab tok (l as x::l') =
if Mem x (Get tok spectab) then GetSymbol spectab (tok`x) l'
else (tok,l);
val GetSymbol = fn
: (string * string list) list
-> string -> string list -> string * string list

```

The function GetNextToken can be enhanced to handle special symbols. It needs to take a special-symbol table as an argument.
```

fun GetNextToken spectab [x] = (x,[]) 102
| GetNextToken spectab (x::(l as x'::l')) =
if IsLetter x
then GetTail (fn x => IsLetter x orelse IsDigit x) [x] l
else if IsDigit x
then GetTail IsDigit [x] l
else if Mem x' (Get x spectab)
then GetSymbol spectab (implode[x,x']) l'
else (x,l);
> val GetNextToken = fn
> : (string * string list) list -> string list -> string * string list

```

Now Tokenise can be enhanced to use the new GetNextToken.
```

fun Tokenise spectab [] = [] 103
| Tokenise spectab (l as x::l') =
if IsSeparator x
then Tokenise spectab l'
else let val (t,l'') = GetNextToken spectab l
in t::(Tokenise spectab l'') end;
> val GetNextToken = fn
> : (string * string list) list -> string list -> string * string list

```

Here is a particular table:
```

val SpecTab = [("=", ["<",">","="]),
("<", ["<",">"]),
(">", ["<",">"]),
("==", [">"])];
val SpecTab =
[("=", ["<",">","="]), ("<", ["<",">"]), (">", ["<", ">"]) , ("==", [">"])]
: (string * string list) list
Tokenise SpecTab (explode "a==>b c5 d5==ff+gg7");
> val it = ["a","==>","b","c5","d5","==","ff","+","gg7"] : string list

```

In the next section the lexical analyset Lex will be used.
```

val Lex = Tokenise SpecTab o explode; }10
> val Lex = fn : string -> string list
Lex "a==>b c5 d5==ff+gg7";
> val it = ["a","==>","b","c5","d5","==","ff","+","gg7"] : string list

```

\subsection*{5.2 Simple special cases of parsing}

Before giving a complete parser, some special cases are considered.

\subsection*{5.2.1 Applicative expressions}

Examples of applicative expressions are \(x, f x,(f x) y, f(f x), f(g x)(h x)\) etc. Parse trees for such expression can be represented by the recursive datatype tree.
```

datatype tree = Atom of string | Comb of tree * tree;
> datatype tree
> con Atom : string -> tree
> con Comb : tree * tree -> tree

```

\section*{Right associative without brackets}

Suppose for the moment: (i) the input is supplied as a list of atoms, (ii) brackets are ignored and (ii) application is taken to be right associative. Then a simple parser is:
```

fun Parse [next] = Atom next
| Parse (next::rest) = Comb(Atom next, Parse rest);
> Warning: match nonexhaustive
> val Parse = fn : string list -> tree
Parse["f", "x", "y", "z"];
> val it = Comb (Atom "f",Comb (Atom "x",Comb (Atom "y",Atom "z"))) : tree

```

\section*{Left associative without brackets}

The usual convention is for application to be left associative. A parser for this is only slightly more complex.
To parse \(f \mathrm{x} y \mathrm{z}\) the following intermediate parsings need to be done:
1. f is parsed to Atom " f "
2. \(f x\) is parsed to Comb(Atom "f", Atom "x")
3. \(f x y\) is parsed to Comb (Comb(Atom "f", Atom "x"), Atom "y")

Intermediate parse trees will be 'passed forward' via a variable \(t\) of an auxiliary function Parser.
```

fun Parser t [] = t 108
| Parser t (next::rest) = Parser (Comb(t, Atom next)) rest;
> val Parser = fn : tree -> string list -> tree
fun Parse [next] = Atom next
| Parse (next::rest) = Parser (Atom next) rest;
> Warning: match nonexhaustive
> val Parse = fn : string list -> tree
Parse["f", "x", "y", "z"];
> val it = Comb (Comb (Comb (Atom "f",Atom "x"),Atom "y"),Atom "z") : tree

```

\section*{Right associative with brackets}

Brackets will now be considered. To parse \(\cdots(e) \cdots\), the parser must be called recursively inside the brackets to parse \(e\), and then the presence of the closing bracket must be checked. Such a recursive call needs to return the parse tree for \(e\) and the rest of the input list. Thus the type of Parse changes to string list \(->\) tree * string list.
If the parser encounters an unexpected closing bracket then it returns the parse tree so far and the rest of the input. For example, \([" x ", " y ", ") ", " z "]\) should parse to (Comb(Atom "x", Atom "y"), [")","z"]).
However, there may be no "parse tree so far" and to handle this case it is convenient to add an empty tree Nil to the type tree.
```

datatype tree = Nil | Atom of string | Comb of tree * tree; }10
> datatype tree
con Atom : string -> tree
con Comb : tree * tree -> tree
con Nil : tree

```

Right associative function application is considered first. A first attempt is:
```

exception MissingClosingBracket; 110
> exception MissingClosingBracket
fun Parse [] = (Nil,[])
| Parse (rest as ")"::_) = (Nil,rest)
| Parse ("("::rest) =
(case Parse rest
of (t, ")"::rest') => let val (t',rest'') = Parse rest'
in (Comb(t,t'), rest'') end
| _ => raise MissingClosingBracket)
| Parse (next::rest) = let val (t,rest') = Parse rest
in (Comb(Atom next,t),rest') end;
> val Parse = fn : string list -> tree * string list

```

This doesn't quite work:
```

Parse ["x"];
111
> val it = (Comb (Atom "x",Nil),[]) : tree * string list
Parse ["x","y","z"];
> val it = (Comb (Atom "x",Comb (Atom "y",Comb (Atom "z",Nil))),[])
Parse ["x","y",")","z"];
> val it =
(Comb (Atom "x",Comb (Atom "y",Nil)),[")","z"]) : tree * string list

```

The empty parse tree Nil returned when Parse exits needs to be removed. This is easily done by replacing Comb by MkComb, where:
\begin{tabular}{|ll|l|}
\hline fun MkComb \(\mathrm{t}, \mathrm{Nil})\) & \(=\) t & 112 \\
| MkComb p & \(=\) Comb p; & \\
\hline
\end{tabular}

Then Parse is redefined:


Parse now works on well-formed expressions.
\begin{tabular}{|l|l|}
\hline \begin{tabular}{l} 
Parse ["x"]; \\
\(>\) \\
val it \(=(A t o m ~ " x ", ~[]) ~: ~ t r e e ~ * ~ s t r i n g ~ l i s t ~\)
\end{tabular} & \\
Parse ["x", "y", "z"]; \\
\(>\) val it \(=\) \\
\(>\) (Comb (Atom "x", Comb (Atom "y",Atom "z")), []) : tree * string list \\
Parse ["x","y",")","z"]; \\
val it = (Comb (Atom "x",Atom "y"), [")","z"]) : tree * string list
\end{tabular}

However, the empty parse tree Nil can still be generated, but only in pathological situations.
```

Parse ["(",")"];
115
> val it = (Nil,[]) : tree * string list
Parse ["(",")","a"];
> val it = (Comb (Nil,Atom "a"),[]) : tree * string list
Parse ["(",")","(",")"];
> val it = (Nil,[]) : tree * string list
Parse [")","x"];
> val it = (Nil,[")","x"]) : tree * string list

```

This might be acceptable, but probably it is better to distinguish the first three examples from the last. In the modified version of Parse that follows, () parses to the 'empty atom' Atom "", where "" is the empty string.
Here is the revised definition of Parse:
\begin{tabular}{|c|c|}
\hline  & 116 \\
\hline
\end{tabular}

The pathological examples now become:
```

Parse [" (",")"]; 117
> val it = (Atom "", []) : tree * string list
Parse ["(",")","a"];
> val it = (Comb (Atom "",Atom "a"), []) : tree * string list
Parse ["(",")","(",")"];
> val it = (Comb (Atom "",Atom ""), []) : tree * string list

```

The second, fourth and fifth clauses of this latest definition of Parse contain some repetition. This can be mitigated by defining an auxiliary function for building a combination. BuildComb parse \(t\) inp builds a combination whose operator is \(t\) and whose operand is got by calling the supplied parser function parse on the supplied input inp. The combination and the remainder of the input are returned.
```

fun BuildComb parse t inp = 118
let val (t', rest) = parse inp
in (MkComb(t,t'), rest) end;
> val BuildComb = fn : ('a -> tree * 'b) -> tree -> 'a -> tree * 'b
fun Parse [] = (Nil, []
Parse ("("::")"::rest) = BuildComb Parse (Atom "") rest
Parse (rest as ")"::_) = (Nil,rest)
| Parse ("("::rest)
(case Parse rest
of (t, ")"::rest') => BuildComb Parse t rest'
| _ $\quad$ 在 raise MissingClosingBracket)
| Parse (next: :rest) = BuildComb Parse (Atom next) rest;

```

Notice how a mutual recursion is set up between Parse and BuildComb via the function argument parse. This technique will be exploited again.
The messy fourth clause of Parse could be simplified with another auxiliary function that called a supplied parser and then checked for a given input symbol.
```

fun CheckSym parse sym exn inp = 119
case Parse inp
of (t, next::rest) => if next=sym then BuildComb parse t rest
else raise exn
| - => raise exn;
val CheckSym = fn
: (string list -> tree * 'a)
-> string -> exn -> string list -> tree * 'a
fun Parse [] = (Nil,[])
Parse ("("::")"::rest) = BuildComb Parse (Atom "") rest
Parse (rest as ")"::_) = (Nil,rest)
Parse ("("::rest) = CheckSym Parse ")" MissingClosingBracket rest
Parse (next::rest) = BuildComb Parse (Atom next) rest;

```

Whether or not this is an improvement is a matter of taste.

\section*{Left associative with brackets}

The technique used above - passing intermediate parse trees as arguments to an auxiliary function - will be used to parse left-associating applicative expressions with brackets. Here is a first attempt.
```

fun Parser t [] = (t, []) 120
| Parser t ("("::")"::rest) = Parser (Comb(Atom "",t)) rest
| Parser t (inp as ")"::_) = (t, inp)
| Parser t [next] = (Comb (t,Atom next), [])
| Parser t ("("::rest) =
(case Parser Nil rest
of (t', ")"::rest') => Parser (Comb(t,t')) rest'
| - => raise MissingClosingBracket)
| Parser t (next::rest) = Parser (Comb(t,Atom next)) rest;
val Parse = Parser Nil;

```

This has a familiar problem with Nil
```

Parse ["x"]; 121
> val it = (Comb (Nil,Atom "x"),[]) : tree * string list
Parse ["x","(","у","z",")","т"];
> val it =
(Comb
(Comb
(Comb (Nil,Atom "x"),
Comb (Comb (Nil,Atom "y"),Atom "z")),Atom "w"),
[]) : tree * string list

```

The solution is to use MkComb instead of Comb, where:


Then:
```

Parse ["x"];
123
> val it = (Atom "x",[]) : tree * string list
Parse ["x","(","y","z",")","w"];
> val it = (Comb (Comb (Atom "x",Comb (Atom "y",Atom "z")),Atom "w"),[])
> : tree * string list

```

\subsection*{5.2.2 Precedence parsing of infixes}

The parsing of expressions like \(x+y \times z\) will now be considered. Parse trees are represented by:
\begin{tabular}{rl|l|}
\hline datatype tree & \(=\) Nil & 124 \\
& \(\mid\) Atom of string \\
& BinOp of string * tree * tree; & \\
\hline
\end{tabular}

Binary operators are assumed to have a precedence given by a table represented as a list of pairs.
\begin{tabular}{|l|l|}
\hline val BinopTable \(=\) & 125 \\
\(\quad[(" * ", 7)\), & \\
\(\left.\left({ }^{\prime \prime}+{ }^{\prime \prime}, 6\right)\right] ;\) & \\
\hline
\end{tabular}

The function Lookup gets the precedence of an operator from such a table.
```

fun Lookup ((s,n)::tab) x = if x=s then n else Lookup tab x; 笽 (26
Warning: match nonexhaustive
val Lookup = fn : (''a * 'b) list -> ''a -> 'b

```

Note that the use of = forces an equality type. A string is an operator if it is assigned a precedence by the precedence table:
```

fun InTable [] x = false 127
| InTable ((s,n)::tab) x = (x=s orelse InTable tab x);
> val InTable = fn : (''a * 'b) list -> ''a -> bool

```

The parser function Parser that follows is quite tricky, but uses many of the principles that have already been illustrated. The main new ingredient are precedences. Assume * (i.e. the ASCII version of \(\times\) ) has higher precedence than \(=\). Consider the parsing of \(x * y+z\) versus the parsing of \(x+y * z\).
For \(\mathrm{x} * \mathrm{y}+\mathrm{z}\) the parser must proceed by:
A1: first building BinOp("*", Atom "x", Atom "y")
A2: then building Atom " z "
A3: finally building Binop("+", BinOp("*", Atom "x", Atom "y"), Atom "z")
For \(x+y * z\) the parser must proceed by:
B1: first building Atom "x"
B2: then building Bin0p("*", Atom "y", Atom "z")
A3: finally building Binop("+", Atom "x", Bin0p("*", Atom "y", Atom "z"))
In both cases, the left hand argument to the operator must be held whilst the right hand argument is parsed. This will be done by giving Parser an extra parameter: the parse-tree of the already-parsed argument (or Nil). This is the same idea already used to parse left associative applicative expressions (in that case, the extra parameter holding the already-parsed rator). The name of the extra parameter will be \(t\).
Precedences are used to decide between A1-A3 and B1-B3. During the parsing there is a current precedence of the parse. For example, if the data in BinopTable above is used, then when parsing the second argument of + the precedence will be 6 and when parsing the second argument of \(*\) the precedence will be 7 .
If the current precedence is \(m\) and a binary operator, op say, is encountered whose precedence is \(n\), then:

A: if \(m>n\) the expression just parsed is the second argument of the operator that proceeds it (case A1-A3)

B: if not \(m>n\) then the expression just parsed is the first argument of the alreadyencountered operator op (case B1-B3).

The expression just parsed will be the parse tree bound to the parse tree parameter ( \(t\) ) of Parser. So in case A above this parameter should be returned immediately. In case B, the parser should be called recursively to get the parse tree, t' say, of the second argument of \(o p\) and then \(\operatorname{Bin} 0 \mathrm{p}(o p, \mathrm{t}, \mathrm{t}\) ') returned.
With this explanation, I hope the ML code to achieve this is comprehensible.
```

fun Parser tab m t [] = (t, [])
| Parser tab m t (inp as next::rest) =
if InTable tab next
then let val n = Lookup tab next
in if (m:int) > n
then (t, inp)
else let val (t',rest') = Parser tab n Nil rest
in Parser tab m (BinOp(next,t, t')) rest' end
end
else Parser tab m (Atom next) rest;
val Parser = fn
: (string * int) list
-> int -> tree -> string list -> tree * string list
val Parse = Parser BinopTable O Nil;
> val Parse = fn : string list -> tree * string list

```

Here are some examples,
```

Parse ["x","*","y","+","z"]; [________________ 129
> val it = (BinOp ("+",BinOp ("*",Atom "x",Atom "y"),Atom "z"),[])
Parse ["x","+","y","*","z"];
> val it = (BinOp ("+",Atom "x",BinOp ("*",Atom "y",Atom "z")),[])
Parse ["x","+","y","+","z"];
> val it = (BinOp ("+",Atom "x",BinOp ("+",Atom "y",Atom "z")),[])
> : tree * string list

```

The last of these examples shows that binary operators are parsed as right associative. This is because > is used to compare the current precedence with that of an encountered operator. Since \(m>m\) is always false, the effect is as though the the operator on the left is not of higher precedence than the one on the right, hence right associativity. If \(>\) is changed to \(>=\) then operators will parse as left associative. The general case where some operators are left associative and some right associative can be handled by giving operators both left and right precedences. This is considered in Section 5.3.
A property of the above parser is that if next is not a known operator (i.e. is not in \(t a b)\), then the last parse tree parsed ( \(\operatorname{viz} t\) ) is thrown away.
\begin{tabular}{|l|l|}
\hline \begin{tabular}{c} 
Parse ["x","y","z"]; \\
\(>\) val it \(=(\) Atom "z", []) : tree * string list
\end{tabular} & 130 \\
\hline
\end{tabular}

Instead of doing this, juxtaposed expressions without any intervening binary operators can be interpreted as function applications. First, parse trees have to be updated to permit this.
\begin{tabular}{|l|l|l|}
\hline datatype tree \begin{tabular}{l} 
Nil \\
\\
\\
\\
\\
| Atom of string \\
\\
| CinOp of tree * tree
\end{tabular} & 131 \\
\hline
\end{tabular}

Then the last line of Parser is modified:
```

fun Parser tab m t [] = (t, []) 132
| Parser tab m t (inp as next::rest) =
if InTable tab next
then let val n = Lookup tab next
in if (m:int) > n
then (t, inp)
else let val (t',rest') = Parser tab n Nil rest
in Parser tab m (Bin0p(next,t, t')) rest' end
end
else Parser tab m (Comb(t, Atom next)) rest;
val Parser = fn
: (string * int) list
-> int -> tree -> string list -> tree * string list
val Parse = Parser BinopTable O Nil;
> val Parse = fn : string list -> tree * string list

```

This almost works:
\begin{tabular}{|l|l|}
\hline \begin{tabular}{l} 
Parse ["x", "y","z"]; \\
\(>\) val it \(=(C o m b ~(C o m b ~(C o m b ~(N i l, A t o m ~ " x "), ~ A t o m ~ " y "), ~ A t o m ~ " z "), ~[]) ~\)
\end{tabular} & 133 \\
\hline
\end{tabular}

The usual trick of using MkComb instead of Comb is needed.
```

fun MkComb(Nil,t2) = t2
| MkComb p = Comb p;
fun Parser tab m t [] = (t, [])
| Parser tab m t (inp as next::rest) =
if InTable tab next
then let val n = Lookup tab next
in if (m:int) > n
then ( }t\mathrm{ , inp)
else let val (t',rest') = Parser tab n Nil rest
in Parser tab m (Bin0p(next,t, t')) rest' end
end
else Parser tab m (MkComb(t, Atom next)) rest;
val Parse = Parser BinopTable O Nil;

```

This works.
```

Parse ["x","y","z"]; 135
> val it = (Comb (Comb (Atom "x",Atom "y"),Atom "z"),[])
Parse ["x","y","+","z"];
> val it = (BinOp ("+",Comb (Atom "x",Atom "y"),Atom "z"),[])
> : tree * string list

```

Notice that function application binds tighter than binary operators, which is normally what is wanted (though achieved here in a rather ad hoc and accidental manner). Two things that Parse doesn't handle are unary operators and brackets. The datatype of parse trees needs to be adjusted to handle unary operators.
\begin{tabular}{|c|c|}
\hline datatype tree = Nil & 136 \\
\hline ```
| Atom of string
| Comb of tree * tree
| Unop of string * tree
| BinOp of string * tree * tree;
``` & \\
\hline \[
\begin{array}{cl}
\text { fun } M k \operatorname{Comb}(\text { Nil }, \mathrm{t} 2) & =\mathrm{t} 2 \\
\text { | MkComb p } & =\text { Comb p; }
\end{array}
\] & \\
\hline
\end{tabular}

Unary operators need a precedence. If ~ has higher precedence than + then \({ }^{\sim} \mathrm{x}+\mathrm{y}\) parses as BinOp("+", Unop("~", Atom "x"), Atom "y"). If ~ has lower precedence than + then \({ }^{\sim} \mathrm{x}+\mathrm{y}\) parses as Unop("~", BinOp("+", Atom "x", Atom "y")).
The precendences of unary operators will be held in a table.
\begin{tabular}{|l|l|}
\hline val UnopTable = & 137 \\
\(\quad[(1 \sim ", 8)\), \\
\(("!", 5)] ;\) & \\
\hline
\end{tabular}

Parsing is now straightforward: both the unary and binory operator tables need to be passed to Parser and an extra clauses is added to test for unary operators.
```

fun Parser (tab as (utab,btab)) m t [] = (t, [])
138
| Parser (tab as (utab,btab)) m t (inp as next::rest) =
if InTable utab next
then let val n = Lookup utab next
in let val (t',rest') = Parser tab n Nil rest
in Parser tab m (MkComb(t, Unop(next,t'))) rest' end
end
else if InTable btab next
then let val n = Lookup btab next
in if (m:int) > n
then (t, inp)
else let val (t',rest') = Parser tab n Nil rest
in Parser tab m (Bin0p(next,t, t')) rest' end
end
else Parser tab m (MkComb(t, Atom next)) rest;
> val Parse = Parser (UnopTable,BinopTable) 0 Nil;

```

Here are some examples:
```

Parse ["x"]; 139
> val it = (Atom "x",[]) : tree * string list
Parse ["~","x"];
val it = (Unop ("~",Atom "x"),[]) : tree * string list
Parse ["~","x","+","y"];
> val it = (BinOp ("+",Unop ("~",Atom "x"),Atom "y"), [])
Parse ["!","x","+","y"];
> val it = (Unop ("!",BinOp ("+",Atom "x",Atom "y")), [])

```

Brackets can be handled in the same way as they were for (left associated) applicative expressions.
```

fun Parser (tab as (utab,btab)) m t [] = (t, []) 140
| Parser tab m t ("("::")"::rest) = Parser tab m (Comb(Atom "",t)) rest
| Parser tab m t (inp as ")"::_) = (t, inp)
| Parser tab m t ("("::rest) =
(case Parser tab O Nil rest
of (t', ")"::rest') => Parser tab m (MkComb(t,t')) rest'
| _ => raise MissingClosingBracket)
| Parser (tab as (utab,btab)) m t (inp as next::rest) =
if InTable utab next
then let val n = Lookup utab next
in let val (t',rest') = Parser tab n Nil rest
in Parser tab m (MkComb(t, Unop(next,t'))) rest' end
end
else if InTable btab next
then let val n = Lookup btab next
in if (m:int) > n
then (t, inp)
else let val (t',rest') = Parser tab n Nil rest
in Parser tab m (Bin0p(next,t, t')) rest' end
end
else Parser tab m (MkComb(t, Atom next)) rest;
val Parse = Parser (UnopTable,BinopTable) O Nil;

```

Here are some more examples.
```

fun P s = Parse(explode s);
> val P = fn : string -> tree * string list
P "~(x+y)";
> val it = (Unop ("~",BinOp ("+",Atom "x",Atom "y")),[])
"(~x)+y";
> val it = (BinOp ("+",Unop ("~",Atom "x"),Atom "y"), [])
P "(x+y)(zw)";
> val it =
(Comb (Bin0p ("+",Atom "x",Atom "y"),Comb (Atom "z",Atom "w")),[])

```

\subsection*{5.3 A general top-down precedence parser}

The parser just given works by looking at the next item being input and then invokes some action, which depends on the item, to parse the rest of the input. A more general scheme is to associate actions with items and then to have a simple parsing loop that consists in repeatedly reading an item and then executing the associated action. \({ }^{2}\)
A rather general datatype of parse trees is the following.
\begin{tabular}{|l|l|l|}
\hline datatype tree & \(=\) Nil & 142 \\
& | Atom of string \\
& Comb of tree * tree \\
& Node of string * tree list; & \\
\hline
\end{tabular}

Unary operator expressions will parse to trees of the form Node (name, [arg]) and binary operator expressions to trees of the form Node (name, \(\left[\arg g_{1}, \arg _{2}\right]\) ).
The familiar MkComb hack will be needed. The empty parse tree Nil should never arise as the right component of a combination (since left associative application will be adopted), so an exception will be raised if it does.

\footnotetext{
\({ }^{2}\) The parser described here is loosely based on Vaughan Pratt's CGOL system (MIT, 1974).
}
```

exception NilRightArg;
fun MkComb(Nil,t2) = t2
MkComb(t1,Nil) = raise NilRightArg
| MkComb(t1, t2) = Comb(t1,t2);

```
143

The action associated with an item may involve recursive calls to the parser. To handle this the techique described earlier of passing a parse function as an argument can be used (see BuildComb and CheckSym described above). The type of parse functions is given by the the following type abbreviation.
\begin{tabular}{|l|l|l|}
\hline type parser \(=\) int \(->\) tree \(->\) string list \(->\) tree * string list; & 144 \\
\hline
\end{tabular}

Selected input items will have precedences and actions associated with them. Precedences are integers. Intuitively, actions are represented by a functions of type parser. However, since an action might need to recursively invoke the whole parser, it should be passed a parsing function. In general, an action must be represented by a function of type parser->parser. A symbol table associates precedences and actions to strings.
```

type symtab = string -> int * (parser -> parser);

```

The main parsing function is now very simple, since all the detail has been hived-off into the symbol table.
```

fun Parser symtab (m:int) t [] = (t,[])
| Parser symtab m t (inp as next::rest) =
let val (n,parsefn) = symtab next
in if m>=n then ( }t,inp\mathrm{ )
else parsefn (Parser symtab) m t inp
end;

```
146

The parse stops on the empty string. If the input isn't empty, then the next item is looked up in the symbol table. Left association will be taken as the default, so if the current precedence equals or exceeds the precedence of the next item, then the parse stops and the last item parsed \((\mathrm{t})\) is returned, with the rest of the input. If the current precedence is less than the precedence of the next item, then the parse action associated with the next item in the symbol table is executed. The parse function Parser symtab is passed to the parse action, so that it can (if necessary) invoke the whole parse recursively.
The definition of Parser intuitively has type symtab->parser. However, the actual type assigned by ML is more general:
```

('a
-> int
* ((int -> 'b -> 'a list -> 'b * 'a list)
-> int -> 'b -> 'a list -> 'b * 'a list))
-> int -> 'b -> 'a list -> 'b * 'a list

```

To constrain the types so that typechecking yields the intuitive type requires some contortions. The following does it.

Standard ML seems rather worse than its predecessors in the flexibility it allows for writing type constraints.
Notice that every input item is supposed to have an entry in the symbol table. The kind of items that might be encountered include atoms, unary operators, binary operators, brackets (both opening and closing) and keywords associated with other kinds of constructs (e.g. if, then, else, while).
Generic functions to construct appropriate symbol table entries for these will now be described.
The parser is initially invoked with a specific symbol table, precedence 0 and t set to Nil.
The action associated with an atom, \(a\) say, is just to return Atom \(a\). Since the atom may be the argument of some preceding function, whose parse tree will be bound to \(t\), the parse tree that is actually returned by the parse action of an atom is MkComb (t,Atom a).
\begin{tabular}{|c|c|}
\hline fun ParseAtom parse p t (next: : rest) \(=\) & 148 \\
parse p (MkComb(t,Atom next)) rest; & \\
\hline
\end{tabular}

The action associated with an opening bracket is to recursively call the parser, check that there is a matching closing bracket, remove it, and then continue the parse. The function ParseBracket below is the action invoked by an opening bracket. It takes as a parameter the closing bracket it should check for. The parse tree \(t\) is combined, using MkComb, with the parse tree t' of the stuff parsed inside of the brackets. If \(t\) is Nil then the definition of MkComb ensures that \(t\) ' becomes the new 'last-thing-parsed' bound to \(t\) in the rest of the parse. However if \(t\) is not Nil, then what is being parsed must have the form \(e_{1}\left(e_{2}\right)\), where t is the parse tree of \(e_{1}\), so a combination is generated (namely, the parse tree of \(e_{1}\) applied to the parse tree of \(e_{2}\) ).
```

exception MissingClosingBracket;
149
fun ParseBracket close parse p t (_::rest) =
let val (t', next'::rest') = parse O Nil rest
in if close=next' then parse p (MkComb(t,t')) rest'
else raise MissingClosingBracket

```

One can ensure that the parsing initiated by an opening bracket will terminate at a closing bracket by giving the closing bracket a sufficiently low precedence in the symbol table (e.g. 0). Closing brackets should always terminate the current parse, so it is an error to try to execute the parse acton associated with them in the symbol table (the type of the symbol table is such that all items have some action - in the case of closing brackets this should never actually be invoked).
\begin{tabular}{|l|c|}
\hline exception TerminatorParseErr; & 150 \\
fun Terminator parse _ = raise TerminatorParseErr; & \\
\hline
\end{tabular}

The next function provides a rather general way of specifying parser actions. The idea is that to parse a given kind of construct the parser is called recursively to get each constituent and then a node containing the resulting constituent parse trees is returned. Each recursive invokation of the parser might require some local checking for keywords etc. For example, to parse if \(e\) then \(e_{1}\) else \(e_{2}\) the parser is called to get the parse tree of the \(e\), then the presence of then is checked (then must be a terminator) and it is removed, then the parser is invoked to get the parse tree for \(e_{1}\), then the presence of else is checked (else must be a terminator) and it it removed, then the parser is invoked again to get the parse tree of \(e_{2}\) and finally a node like Node("COND", [t,t1,t2]) is returned.
The function ParseSeq below takes a constructor function mktree (for building a node), invokes the parser a number of times and then builds a parse tree by applying mktree to the resulting constituent parse trees.
Each invokation of the parser can be 'wrapped around' with some extra checking activity. This is specified by providing a list of functions of type parser->parser: applying such a function to a parser produces a new parser with the checking added on. The simplest case of this is no checking, which is specified by the identity function.
\begin{tabular}{|l|l|}
\hline fun Id \(\mathrm{x}=\mathrm{x}\); & 151 \\
\hline
\end{tabular}

ChkAft is used to modify a parser to check that a given keyword occurs after the parser is invoked. If \(p\) :parser is a parser function, then ChkAft \(s p\) is a parser function that first invokes \(p\), then checks for \(s\) and deletes it if found and raises an exception otherwise.
\begin{tabular}{|l|l|}
\hline exception ChkAftErr; & 152 \\
fun ChkAft s parse m t inp \(=\) \\
case parse m t inp \\
of (t', s'::rest) \(\Rightarrow\) if \(s=s^{\prime}\) then (t',rest) else raise ChkAftErr;
\end{tabular}

The function ParseSeq below takes a parse tree constructor function mktree of type tree * tree list \(\rightarrow\) tree for building a node. The first parse tree is the one passed as a parameter \((\mathrm{t})\) to the parser and the list of parse trees are the constituents that have just been parsed.
Constructors for building parse trees of unary operator expressions and binary operator expressions are MkUnop and MkBinop, respectively.
Suppose \(u\) is a unary operator and consider \(e_{1} u e_{2}\) : this should parse to \(\operatorname{Comb}\left(\hat{e}_{1}, \operatorname{Unop}\left(u, \hat{e}_{2}\right)\right)\), where \(\hat{e}_{1}\) and \(\hat{e}_{2}\) are the parse trees of \(e_{1}\) and \(e_{2}\), respectively. The parse tree constructor for unary operators is thus:
\begin{tabular}{|l|l|}
\hline fun MkUnop unop ( \(t, t l)=M k C o m b(t, N o d e(u n o p, t l)) ;\) & 153 \\
\hline
\end{tabular}

Suppose \(b\) is a binary operator and consider \(e_{1} b e_{2}\) : this should parse to \(\operatorname{Binop}\left(b, \hat{e}_{1}, \hat{e}_{2}\right)\), where \(\hat{e}_{1}\) and \(\hat{e}_{2}\) are the parse trees of \(e_{1}\) and \(e_{2}\), respectively. The parse tree constructor for binary operators is thus:
\begin{tabular}{|l|l|}
\hline fun MkBinop bnop (t,tl) \(=\) Node(bnop,t::tl); & 154 \\
\hline
\end{tabular}

The function ParseSeq also takes as a parameter a list of parser transformations (e.g. Id or ChkAft \(f\) ) and returns a parser that recursively invokes the parser once for each parser transformation and then builds a parse tree using mktree applied
to the resulting constituent parse trees. For example, the parsing of conditionals is specified by:
[ChkAft "then", ChkAft "else", Id]
ParseSeq uses an auxiliary function ParseSeqAux that iterates down the list of supplied parse tree transformers invoking them in turn. The definitions are short, but admittedly cryptic! To try to improve their readability type constraints have been added to constrain excess polymophism. Without the constraints, ParseSeq gets the incomprehensibly general type:
```

(tree * 'a list -> 'b)
-> int
-> ((int -> 'b -> 'c list -> 'd)
-> int -> tree -> 'c list -> 'a * 'c list) list
-> (int -> 'b -> 'c list -> 'd) -> int -> tree -> 'c list -> 'd

```
with the constraints the type is:
```

(tree * tree list -> tree)
-> int -> (parser -> parser) list -> parser -> parser

```

Unfortunately, as with Parser, it is necesary to go to some contortions to achieve this type constraint. Instead of writing:
```

ParseSeq mktree m fl parse n t (_::rest) = ...

```

It is necessary to write
```

fun ParseSeq mktree m fl parse =
fn n => fn t => fn (_::rest) => ...

```
and then add the type constraints shown below.
```

fun ParseSeqAux m [f:parser->parser] (parse:parser) n inp =
let val (t, rest1) = f parse m Nil inp
in ([t], rest1) end
| ParseSeqAux m (f::fl : (parser->parser)list) parse n inp =
let val (t, rest1) = f parse O Nil inp
in let val (l,rest2) = ParseSeqAux m fl parse 0 rest1
in (t::l, rest2) end
end;
fun ParseSeq mktree m (fl:(parser->parser)list) (parse:parser) : parser =
fn n => fn t => fn (_::rest) =>
let val (l,rest1) = ParseSeqAux m fl parse n rest
in parse n (mktree(t,l)) rest1 end;

```

A symbol table is a function of type string -> int * (parser -> parser). Here is an example:
\begin{tabular}{|c|c|c|c|c|c|}
\hline fun & SymTab "*" & \(=(7\), & ParseSeq (MkBinop & "MULT") 8 [Id]) & 156 \\
\hline | & SymTab "+" & \(=\) ( 6 , & ParseSeq (MkBinop & "ADD") 5 [Id]) & \\
\hline I & SymTab "~" & \(=(10\), & ParseSeq (MkUnop ' & "MINUS") 9 [Id]) & \\
\hline I & SymTab "if" & \(=(10\), & ParseSeq (MkUnop & \[
\begin{gathered}
\text { "COND") } 0 \text { [ChkAft "then", } \\
\text { ChkAft "else", } \\
\text { Id]) }
\end{gathered}
\] & \\
\hline I & SymTab "(" & \(=(10\), & ParseBracket ")") & & \\
\hline 1 & SymTab ")" & \(=(0\), & Terminator) & & \\
\hline 1 & SymTab "then" & \(=(0\), & Terminator) & & \\
\hline I & SymTab "else" & \(=(0\), & Terminator) & & \\
\hline | & SymTab x & \(=(10\), & ParseAtom) ; & & \\
\hline
\end{tabular}

Notice that the left precedence of \(*\) is 7 which is the same as its right precedence. However, the left precedence of + is 6 which is greater than its right precedence 5. The effect of this is to make \(*\) left associative and + right associative. In general, if the left precedence is less than or equal to the right precedence, then left associativity results, otherwise right associativity results.
The complete parser P defined below uses the lexical analyser Lex and the symbol table above. Assume the code for Lex, as described in Section 5.1, is in the file Lex.ml.
```

use "Lex.ml";
val P = Parser SymTab O Nil o Lex;
> val P = fn : string -> tree * string list
"f if x then y + z else y * z";
val it =
(Comb
(Atom "f",
Node
("COND",
[Atom "x",Node ("ADD",[Atom "y",Atom "z"]),
Node ("MULT",[Atom "y",Atom "z"])])),[])

```
157

\section*{Chapter 6}

\section*{Case study 2: the \(\lambda\)-calculus}

It is assumed that integers and (unary and binary) operations over integers are primitive. The type atom packages these up into a single datatype. Both unary operator atoms (0p1) and binary operator atoms (0p2) have a name and a semantics.
\begin{tabular}{|l|l|l|}
\hline datatype atom \(=\) Num of int & 158 \\
& \(\mid\) Op1 of string * (int->int) \\
& Op2 of string * (int*int->int); & \\
\hline
\end{tabular}

The application of an atomic operation to a value is defined by the function ConApply (see below). The application of a binary operator b to m results in a unary operator named mb expecting the other argument.
To convert the argument \(m\) to a string that can be concatenated with the name of the operator, a function to convert a number to a string giving its decimal representation is defined.
```

fun StringOfNum 0 = "0"
| StringOfNum 1 = "1"
| StringOfNum 2 = "2"
| StringOfNum 3 = "3"
| StringOfNum 4 = "4"
| StringOfNum 5 = "5"
| StringOfNum 6 = "6"
| StringOfNum 7 = "7"
| StringOfNum 8 = "8"
| StringOfNum 9 = "9"
| StringOfNum n =
(StringOfNum(n div 10)) - (StringOfNum(n mod 10));
StringOfNum 1574;
> val it = "1574" : string

```
                                    159

Now ConApply can be defined.
```

fun ConApply(Op1(_,f1), Num m) = Num(f1 m) 160
| ConApply(Op2(x,f2), Num m) = Op1((StringOfNum m^x), fn n => f2(m,n));
>val ConApply = fn : atom * atom -> atom
ConApply(Op2("+",op +), Num 2);
> val it = Op1 ("2+",fn) : atom
ConApply(it, Num 3);
> val it = Num 5 : atom

```
\(\lambda\)-expressions are represented by the datatype lam.
\begin{tabular}{|l|l|l|}
\hline datatype lam \(=\) Var of string & 161 \\
& \(\mid\) Con of atom \\
& \(\mid\) App of (lam \(*\) lam \()\) \\
& \(\mid\) Abs of (string \(*\) lam); & \\
\hline
\end{tabular}

\subsection*{6.1 A \(\lambda\)-calculus parser}

It is convenient to have a \(\lambda\)-calculus parser. Assume the code of the parser described in Section 5.3 is in the file Parser.ml.
\begin{tabular}{|l|l|}
\hline use "Parser.ml"; & 162 \\
\hline
\end{tabular}

A sutable symbol table for the \(\lambda\)-calculus is:


Note that " \(\backslash \backslash\) " is our ASCII representation of \(\lambda\). This is actually just a single backslash; the first one is the escape character needed to include the second one in the string!
The following function lexically analyses and then parses a string (recall that Parser returns a parse tree and the remaining input).
```

fun ParseLam s = 164
let val (t,[]) = Parser LamSymTab 0 Nil (Lex s)
in t end;
> Warning: binding not exhaustive
val ParseLam = fn : string -> tree
ParseLam "(<br>x.x+1) 200";
val it =
Comb (Node ("Abs",[Atom "x",Node ("ADD",[Atom "x",Atom "1"])]),
Atom "200")
: tree

```

The output of ParseLam is an element of the general purse parse tree type tree defined on page 82 . This is easily converted to type lam. A function for testing whether a string represents a number (i.e. is a string of digits) is needed.
```

fun IsNumber s = 165
let fun TestDigList [] = true
| TestDigList (x::l) = IsDigit x andalso TestDigList l
in TestDigList(explode s)
end;
val IsNumber = fn : string -> bool

```

If a string represents a number then the following provides a way of converting it to a number (i.e. value of type int).
```

fun DigitVal "O" = 0 106
| DigitVal "1" = 1
| DigitVal "2" = 2
| DigitVal "3" = 3
| DigitVal "4" = 4
| DigitVal "5" = 5
| DigitVal "6" = 6
| DigitVal "7" = 7
| DigitVal "8" = 8
| DigitVal "9" = 9;
> Warning: match nonexhaustive
> val DigitVal = fn : string -> int
fun NumOfString s =
let fun ListVal [] = 0
| ListVal (x::l) = DigitVal x + 10 * (ListVal l)
in ListVal(rev(explode s))
end;
> val NumOfString = fn : string -> int
NumOfString "2001";
> val it = 2001 : int

```

Armed with this string-to-number converter, it is routine to convert values of type tree to values of type lam. The fourth clause of the definition of Convert below is a little hack to make: \x1 x2 ... xn. e parse as: \x1. \x2. ... \xn. e. This hack makes use of the fact that sequences of variables parse as left-associated applications.
\begin{tabular}{|l|l|l|}
\hline ParseLam "\\
x y z. w"; \\
> val it = Node ("Abs", [Comb (Comb (Atom "x", Atom "y"), Atom "z"), Atom "w"]) \\
\hline
\end{tabular}
```

fun Convert (Atom x) =
168
if IsNumber x then Con(Num(NumOfString x)) else Var x
| Convert (Comb (a,b)) =
App(Convert a, Convert b)
| Convert (Node("Abs",[Atom x, a])) =
Abs(x,Convert a)
| Convert (Node("Abs",[Comb(a1, Atom x), a2])) =
Convert(Node("Abs",[a1, Node("Abs",[Atom x,a2])]))
| Convert (Node("ADD",[a,b])) =
App(App(Con(Op2("+",(op+))), Convert a), Convert b);
> Warning: match nonexhaustive
> val Convert = fn : tree -> lam

```

The function PL (for "Parse Lambda expression") parses a string and then converts it to a value of type lam.
```

val PL = Convert o ParseLam;
169
> val PL = fn : string -> lam
PL "x+y";
> val it = App (App (Con (Op2 fn),Var "x"),Var "y") : lam
PL "(<br>x.x+y) y";
$>$ val it $=$
App (Abs ("x",App (App (Con (Op2 fn), Var "x"), Var "y")),
Var "y")
: lam

```

Here is the fixed-point operator \(\mathbf{Y}\) (see Section 2.4):
```

PL "<br>f. (<br>x f.(<br>z.x x f)) (<br>x f.(<br>z.x x f))"; 170
val it =
Abs
("f",
App
(Abs ("x",Abs ("f",Abs ("z",App (App (Var "x",Var "x"),Var "f")))),
Abs ("x",Abs ("f",Abs ("z",App (App (Var "x",Var "x"),Var "f"))))))
: lam

```

An 'unparser' (or 'pretty-printer') will be useful for viewing elements of type lam. The one that follows (UPL) is rather crude - for example, it does not attempt to format expressions across lines, though it does at leat avoid putting brackets around variables.
The name UPL stands for "UnParse Lambda expression" and BUPL for "Bracket and UnParse Lambda expression".


\subsection*{6.2 Implementing substitution}

Recall the definition of substitution on page 10.
\begin{tabular}{|c|c|}
\hline \(E\) & \(E\left[E^{\prime} / V\right]\) \\
\hline \(V\) & \(E^{\prime}\) \\
\hline \(V^{\prime} \quad\left(\right.\) where \(\left.V \neq V^{\prime}\right)\) & \(V^{\prime}\) \\
\hline \(E_{1} E_{2}\) & \(E_{1}\left[E^{\prime} / V\right] E_{2}\left[E^{\prime} / V\right]\) \\
\hline \(\lambda V . E_{1}\) & \(\lambda V . E_{1}\) \\
\hline \(\lambda V^{\prime} . E_{1}\) (where \(V \neq V^{\prime}\) and \(V^{\prime}\) is not free in \(E^{\prime}\) ) & \(\lambda V^{\prime} . E_{1}\left[E^{\prime} / V\right]\) \\
\hline \(\lambda V^{\prime} . E_{1}\) (where \(V \neq V^{\prime}\) and \(V^{\prime}\) is free in \(E^{\prime}\) ) & \(\lambda V^{\prime \prime} . E_{1}\left[V^{\prime \prime} / V^{\prime}\right]\left[E^{\prime} / V\right]\) where \(V^{\prime \prime}\) is a variable not free in \(E^{\prime}\) or \(E_{1}\) \\
\hline
\end{tabular}

This is easily implemented in ML. Some auxiliary set-theoretic functions on lists are needed (some of which have been met before). First a test for membership.
\begin{tabular}{|c|c|}
\hline fun Mem x [] = false & 172 \\
\hline | Mem x ( \(\mathrm{x}^{\prime}: \mathbf{: s}\) ) \(=\left(\mathrm{x}=\mathrm{x}^{\prime}\right.\) ) orelse Mem x s; & \\
\hline > val Mem = fn : ', a -> ', a list -> bool & \\
\hline
\end{tabular}

Note that the union of two lists defined below does not introduce duplicates.
\begin{tabular}{|c|c|}
\hline fun Union [] l = 1 & 173 \\
\hline \begin{tabular}{l}
| Union (x::11) 12 = \\
if Mem x 12 then Union 1112 else \(\mathrm{x}::(\) Union 11 12)
\end{tabular} & \\
\hline > val Union = fn : ''a list -> ',a list -> ''a list & \\
\hline Union [1, 2, 3, 4, 5] [2, 3, 4, 5, 6, 7] ; & \\
\hline > val it \(=[1,2,3,4,5,6,7]\) : int list & \\
\hline
\end{tabular}

Subtract 1112 removes all members of 12 from 11 (i.e. is ' 11 minus 12 ').
```

fun Subtract [] l = []
| Subtract (x::l1) 12 =
if Mem x l2 then Subtract l1 l2 else x::(Subtract l1 12);
> val Subtract = fn : ''a list -> ''a list -> ''a list
Subtract [1,2,3,4,5] [3,4,5,6];
> val it = [1,2] : int list

```

Using Mem, Union and Subtract the function Frees to compute a list of the free variables in a \(\lambda\)-expression is easily defined.
```

fun Frees (Var x) = [x]
| Frees (Con c) = []
| Frees (App(e1,e2)) = Union (Frees e1) (Frees e2)
| Frees (Abs(x,e)) = Subtract (Frees e) [x];
> val Frees = fn : lam -> string list
PL "<br>x.x+y";
> val it = Abs ("x",App (App (Con (Op2 fn),Var "x"),Var "y")) : lam
Frees it;
> val it = ["y"] : string list

```

Substitution needs to rename variables to avoid 'capture'. This will be done by priming them.
\begin{tabular}{|l|l|}
\hline fun Prime \(x=x^{\sim \prime \prime}, " ;\) & 176 \\
Prime "x"; & \\
\(>\) val it \(=\) "x'" : string & \\
\hline
\end{tabular}

Variant xl x primes x sufficient number of times so that the result does not occur in the list xl .
```

fun Variant xl x = 177
if Mem x xl then Variant xl (Prime x) else x;
> val Variant = fn : string list -> string -> string
Variant ["x","y","z","y'","w"] "y";
> val it = "y"," : string

```

Now, at last, substitution can be defined: Subst E E' V computes E [E'/V] according to the table above.
```

fun Subst (e as Var x') e' x = if x=x' then e' else e
| Subst (e as Con c) e' x = e
| Subst (App(e1, e2)) e' x = App(Subst e1 e' x, Subst e2 e' x)
| Subst (e as Abs (x',e1)) e' x =
if x=x' then e
else if Mem x' (Frees e')
then let val x', = Variant (Frees e' @ Frees e1) x'
in Abs(x',', Subst(Subst e1 (Var x'') x') e' x)
end
else Abs(x', Subst e1 e' x);
val Subst = fn : lam -> lam -> string -> lam

```

Here are some examples:
```

Subst (PL"(<br>x.x+y) x") (PL"1") "x";
179
> val it =
App (Abs ("x",App (App (Con (Op2 fn),Var "x"),Var "y")),
Con (Num 1))
: lam
UPL it;
> val it = "(<br>x. x+y)(1)" : string
UPL(Subst (PL"<br>x.x+y") (PL"x+1") "y");
> val it = "(<br>x'. x'+(x+1))" : string

```

A function EvalN can now be defined to do normal order reduction (sometimes called 'call-by-name' [30]). Note that the evaluation does not 'go inside' \(\lambda\)-bodies, so does not compute normal forms.
```

fun EvalN (e as Var _ ) = e
| EvalN (e as Con _) = e
| EvalN (Abs(x,e)) = Abs(x, e)
| EvalN (App(Con a1, Con a2)) = Con(ConApply(a1,a2))
| EvalN (App(e1,e2)) =
case EvalN e1
of (Abs(x,e3)) => EvalN(Subst e3 e2 x)
| (e1' as Con a1) => (case EvalN e2
of (Con a2) => Con(ConApply(a1,a2))
| e2' }\quad>\quad\operatorname{App}(e\mp@subsup{1}{}{\prime},e2')
| e1' => App(e1', EvalN e2);
> val EvalN = fn : lam -> lam

```

Here is a typical example that only terminates with normal order evaluation:
\begin{tabular}{|l|l|}
\hline EvalN (PL"(\\
x.1) \(((\backslash \backslash x . ~ x ~ x) ~(\backslash \backslash x . ~ x ~ x)) ") ; ~\) & 181 \\
\(>\) val it \(=\) Con (Num 1) \(:\) lam & \\
\hline
\end{tabular}
val true \(="\)
x y z.w"

\subsection*{6.3 The SECD machine}

Call-by-value evaluation can be programmed with the function Evalv.
```

fun EvalV (e as Var _) = e 182
| EvalV (e as Con _) = e
| EvalV (e as Abs(_,_)) = e
| EvalV (App(e1,e2)) =
let val e2' = EvalV e2
in
(case EvalV e1
of (Abs(x,e3)) => EvalV(Subst e3 e2' x)
| (e1' as Con a) => (case e2'
of (Con a2) => Con(ConApply(a1,a2))
| _ }\quad>\operatorname{App}(e\mp@subsup{1}{}{\prime},e2')
| e1' }\quad>\quad\operatorname{App}(e\mp@subsup{1}{}{\prime},e\overline{2},)
end;

```

The SECD machine is a classical virtual machine for reducing \(\lambda\)-expressions using call-by-value. It was developed in the 1960's by Peter Landin and has been analysed by Gordon Plotkin [30]. Various more recent practical vitual machines for ML are descendents of the SECD machine.
The name SECD comes from Stack, Environment, Control and Dump which are the four components of the machine state.
The stack of an SECD machine holds a sequence (represented by a list) of atoms and closures. The environment provides values of variables. A closure is an abstraction paired with an environment. The mutually recursive datatypes of item and env represent items and environments, respectively. Environments are represented by association lists of variables (represented by strings) and and items.
\begin{tabular}{|ll|l|}
\hline datatype item & \(=\) Atomic of atom & \\
& | Closure of (lam * env) & 183 \\
and & env \(=\) EmptyEnv \\
& & | Env of string \(*\) item \(*\) env;
\end{tabular}

The function Lookup looks up the value of a variable in an environment (and raises an exception is the variable doesn't have a variable).
\begin{tabular}{|ll|l|}
\hline exception LookupErr; & & 184 \\
\begin{tabular}{cl} 
fun Lookup(s,EmptyEnv) \\
| Lookup(s,Env(s',i,env)) & = raise LookupErr \\
if s=s' then i else Lookup(s,env);
\end{tabular} & \\
\hline
\end{tabular}

The control of an SECD is a sequence (represented by a list) of instructions which are either the special operation Ap or a \(\lambda\)-expression.
```

datatype instruction = Ap | Exp of lam; }18

```

A datatype state that represents SECD machine states can now be defined.
\begin{tabular}{rl|l|}
\hline type stack \(=\) & item list \\
and control \(=\) & instruction list; & \\
datatype state \(=\) & NullState \\
& \(\mid\) State of (stack \(*\) env * control * state);
\end{tabular}

The transitions of the SECD machine are given by the function Step.
```

fun Step(State(v::S, E, [], State(S', E', C', D'))) = $\quad 187$
State(v::S', E', C', D')
| Step(State(S, E, Exp(Var x)::C, D)) =
State (Lookup (x, E): : S, E, C, D)
| Step(State (S, E, Exp(Con v): C, D)) =
State(Atomic v::S, E, C, D)
| Step(State(S, E, Exp(Abs(x,e))::C,D )) =
State (Closure (Abs(x,e), E): :S, E, C, D)
| Step (State (Closure (Abs (x,e), E') : : (v::S), E, Ap: C, D) ) =
State([], Env(x,v,E'), [Exp e], State(S,E,C,D))
| Step(State(Atomic v1::(Atomic v2::S), E, Ap::C, D )) =
State (Atomic (ConApply (v1, v2)) : : S, E, C, D)
| Step $(S t a t e(S, E, \operatorname{Exp}(\operatorname{App}(e 1, e 2)):: C, D))=$
State(S, E, Exp e2::(Exp e1:: (Ap::C)), D);

```

The function Run iterates Step until a final state is reached and then returns a list of all the intermediate states.
```

fun Run(state as State([_],EmptyEnv,[],NullState)) = [state] 188
| Run state = state::Run(Step state);
> val Run = fn : state -> state list

```

The function Eval takes a \(\lambda\)-expression and evaluates it using the SECD machine.
```

fun Eval e =
189
let fun EvalAux(State([v],EmptyEnv,[],NullState)) = v
| EvalAux state = EvalAux(Step state)
in EvalAux(State([],EmptyEnv,[Exp e],NullState)) end;
> val Eval = fn : lam -> item

```

Load loads a lambda-expression into an SECD state ready for running.
\begin{tabular}{|l|l|}
\hline fun Load e \(=\) State([], EmptyEnv, [Exp e], NullState); & 190 \\
\(>\) val Load \(=\) fn : lam \(\rightarrow\) state & \\
\hline
\end{tabular}

SECDRun parses a string, loads the resulting \(\lambda\)-expression into an SECD state and then runs the result. SECDEval is similar, but it just Evals the result.
```

fun SECDRun s = Run(Load s); 191
> val SECDRun = fn : lam -> state list
fun SECDEval s = Eval(PL s);
> val SECDEval = fn : string -> item
SECDEval "(<br>x.<br>y. x+y) 1 2";
> val it = Atomic (Num 3) : item

```

\section*{Bibliography}
[1] Augustsson, L., 'A compiler for lazy ML', in Proceedings of the ACM Symposium on LISP and Functional Programming, Austin, pp. 218-227, 1984.
[2] Barendregt, H.P., The Lambda Calculus (revised edition), Studies in Logic 103, North-Holland, Amsterdam, 1984.
[3] Barron, D.W. and Strachey, C. 'Programming', in Fox, L. (ed.), Advances in Programming and Non-numerical Computation (Chapter 3), Pergamon Press, 1966.
[4] Bird, R. and Wadler, P., An Introduction to Functional Programming, Prentice Hall, 1988.
[5] Boyer, R.S. and Moore, J S., A Computational Logic, Academic Press, 1979.
[6] De Bruijn, N.G., 'Lambda calculus notation with nameless dummies, a tool for automatic formula manipulation', Indag. Math., 34, pp. 381-392, 1972.
[7] Burge, W., Recursive Programming Techniques, Addison-Wesley, 1975.
[8] Clarke, T.J.W., et al., 'SKIM - the S, K, I Reduction Machine', in Proceedings of the 1980 ACM LISP Conference, pp. 128-135, 1980.
[9] Curry, H.B. and Feys, R., Combinatory Logic, Vol. I, North Holland, Amsterdam, 1958.
[10] Curry, H.B., Hindley, J.R. and Seldin, J.P. Combinatory Logic, Vol. II, Studies in Logic 65, North Holland, Amsterdam, 1972.
[11] Fairbairn, J. and Wray, S.C., 'Code generation techniques for functional languages', in Proceedings of the 1986 ACM Conference on LISP and Functional Programming, Cambridge, Mass., pp. 94-104, 1986.
[12] Gordon, M.J.C., 'On the power of list iteration', The Computer Journal, 22, No. 4, 1979.
[13] Gordon, M.J.C.,The Denotational Description of Programming Languages, Springer-Verlag, 1979.
[14] Gordon, M.J.C., Programming Language Theory and its Implementation, Prentice Hall International Series in Computer Science, 1988 (out of print).
[15] Gordon, M.J.C., Milner, A.J.R.G. and Wadsworth, C.P., Edinburgh LCF: a mechanized logic of computation, Springer Lecture Notes in Computer Science, Springer-Verlag, 1979.
[16] Henderson, P., Functional Programming, Application and Implementation, Prentice Hall, 1980.
[17] Henderson, P. and Morris, J.M., 'A lazy evaluator', in Proceedings of The Third Symposium on the Principles of Programming Languages, Atlanta, Georgia, pp. 95-103, 1976.
[18] Hindley, J.R., 'Combinatory reductions and lambda-reductions compared', Zeitschrift für Mathematische Logik und Grundlagen der Mathematik, 23, pp. 169-180, 1977.
[19] Hindley, J.R. and Seldin, J.P., Introduction to Combinators and \(\lambda\)-Calculus, London Mathematical Society Student Texts, 1, Cambridge University Press, 1986.
[20] Hughes, R.J.M., 'Super combinators: a new implementation method for applicative languages', in Proceedings of the 1982 ACM Symposium on LISP and Functional Programming, Pittsburgh, 1982.
[21] Kleene, S.C., ' \(\lambda\)-definability and recursiveness', Duke Math. J., pp. 340-353, 1936.
[22] Krishnamurthy, E.V. and Vickers, B.P., 'Compact numeral representation with combinators', The Journal of Symbolic Logic, 52, No. 2, pp. 519-525, June 1987.
[23] Lamport, L., LA \(_{E} X\) : A Document Preparation System, Addison-Wesley, 1986.
[24] Landin, P.J., 'The next 700 programming languages', Comm. Assoc. Comput. Mach., 9, pp. 157-164, 1966.
[25] Levy, J.-J., 'Optimal reductions in the lambda calculus', in Hindley, J.R. and Seldin, J.P. (eds), To H.B. Curry: Essays on Combinatory Logic, LambdaCalculus and Formalism, Academic Press, New York and London, 1980.
[26] Mauny, M. and Suárez, A., 'Implementing functional languages in the categorical abstract machine', in Proceedings of the 1986 ACM Conference on LISP and Functional Programming, pp. 266-278, Cambridge, Mass., 1986.
[27] Milner, A.J.R.G., 'A proposal for Standard ML', in Proceedings of the ACM Symposium on LISP and Functional Programming, Austin, 1984.
[28] Morris, J.H., Lambda Calculus Models of Programming Languages, Ph.D. Dissertation, M.I.T., 1968.
[29] Peyton Jones, S.L., The Implementation of Functional Programming Languages, Prentice Hall, 1987.
[30] Plotkin, G.D., 'Call-by-name, call-by-value and the \(\lambda\)-calculus', Theoretical Computer Science, 1, pp 125-159, 1975.
[31] Schönfinkel, M., 'Über die Bausteine der mathematischen Logik', Math. Annalen 92, pp. 305-316, 1924. Translation printed as 'On the building blocks of mathematical logic', in van Heijenoort, J. (ed.), From Frege to Gödel, Harvard University Press, 1967.
[32] Scott, D.S., 'Models for various type free calculi', in Suppes, P. et al. (eds), Logic, Methodology and Philosophy of ScienceIV, Studies in Logic 74, NorthHolland, Amsterdam, 1973.
[33] Stoy, J.E., Denotational Semantics: The Scott-Strachey Approach to Programming Language Theory, M.I.T. Press, 1977.
[34] Turner, D.A., 'A new implementation technique for applicative languages', Software Practice and Experience, 9, pp. 31-49, 1979.
[35] Turner, D.A., 'Another algorithm for bracket abstraction', The Journal of Symbolic Logic, 44, No. 2, pp. 267-270, June 1979.
[36] Turner, D.A., 'Functional programs as executable specifications', in Hoare, C.A.R. and Shepherdson, J.C. (eds), Mathematical Logic and Programming Languages, Prentice Hall, 1985.
[37] Wadsworth, C.P., 'The relation between computational and denotational properties for Scott's \(D_{\infty}\)-models of the lambda-calculus', S.I.A.M. Journal of Computing, 5, pp. 488-521, 1976.
[38] Wadsworth, C.P., 'Some unusual \(\lambda\)-calculus numeral systems', in Hindley, J.R. and Seldin, J.P. (eds), To H.B. Curry: Essays on Combinatory Logic, LambdaCalculus and Formalism, Academic Press, New York and London, 1980.```


[^0]:    ${ }^{1}$ Note that sum is a $\lambda$-expression, whereas + is a mathematical symbol in the 'metalanguage' (i.e. English) that we are using for talking about the $\lambda$-calculus.

[^1]:    ${ }^{1}$ Note that $\lambda x . \theta(x)$ is an expression of the $\lambda$-calculus whereas $x \mapsto \theta(x)$ is a notation of informal mathematics.

[^2]:    ${ }^{2}$ The kind of primitive recursion defined in Section 2.7.1 is first-order primitive recursion.

[^3]:    ${ }^{3}$ The mathematical characterization of the function denoted by $\mathbf{Y}$ can be found in Stoy's book [33].

[^4]:    ${ }^{1}$ The two-volume treatise Combinatory Logic $[9,10]$ is the definitive reference, but the more recent textbooks [19, 2] are better places to start.

[^5]:    ${ }^{2}$ The most relevant paper I could find is one by Hindley [18]. This compares $\lambda$-reduction with combinatory reduction, but not in a way that is prima facie relevant to the termination of combinator machines.

[^6]:    ${ }^{1}$ Readers interested in Caml should consult the Web page http://pauillac.inria.fr/caml/. Caml is a lightweight language better suited than Standard ML for use on small machines. All the constructs described in this course are in Caml, though the syntactic details differ slightly from Standard ML
    ${ }^{2}$ This overview has evolved from the description of the original ML in Section 2.1 of:
    M.J.C. Gordon, A.J.R.G Milner and C.P. Wadsworth Edinburgh LCF: A Mechanized

    Logic of Computation, Lecture Notes in Computer Science 78, Springer-Verlag 1979.
    ${ }^{3}$ PolyML was originally developed at the Cambridge Computer Laboratory and then licenced first to Imperial Software Technology and then to Abstract Hardware Limited. It has an integrated persistant storage system (database) and is less memory hungry than Standard ML of New Jersey.

