#### CSE 130 Programming Languages

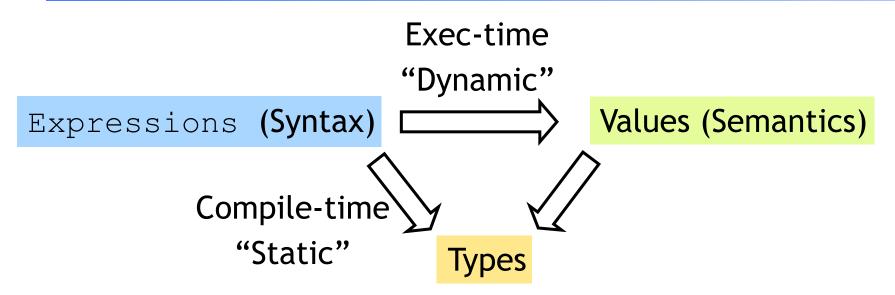
#### Datatypes

Ranjit Jhala

UC San Diego



# Recap: ML's Holy Trinity

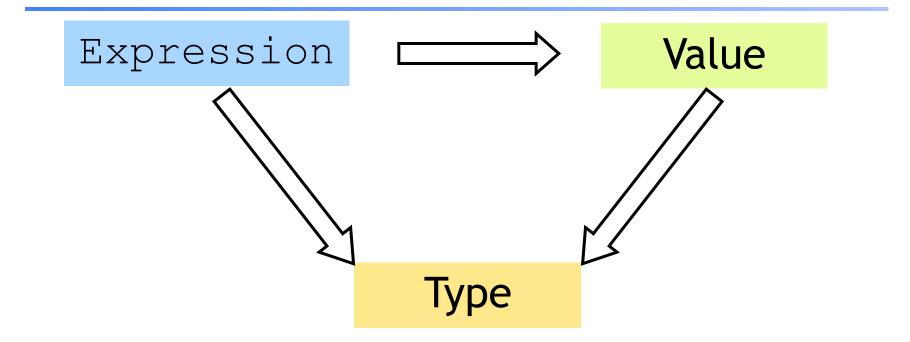


- 1. Programmer enters expression
- 2. ML checks if expression is "well-typed"
  - Using a precise set of rules, ML tries to find a unique type for the expression meaningful type for the expr
- 3. ML evaluates expression to compute value
  - Of the same "type" found in step 2

# Story So Far...

- Simple Expressions
- Branches
- Let-Bindings ...
- Today:
  - Finish Crash Course
  - Datatypes

#### Next: functions, but remember ...



Everything is an expression Everything has a value Everything has a type

#### A function is a value!

#### A shorthand for function binding

# let neg = fun f -> fun x -> not (f x); ...
# let neg f x = not (f x); val neg : int -> int -> bool = fn
# let is5gte = neg is5lt; val is5gte : int -> bool = fn; # is5gte 10; val it : bool = false; # is5gte 2; val it : bool = true;

#### Put it together: a "filter" function

If arg "matches" ...then use this pattern... this Body Expr

val filter : ('a->bool) -> 'a list-> 'a lisi) = fn

```
# let list1 = [1;31;12;4;7;2;10];;
# filter is5lt list1 ;;
val it : int list = [31;12;7;10]
# filter is5gte list1;;
val it : int list = [1;4;2]
# filter even list1;;
val it : int list = [12;4;2;10]
```

#### Put it together: a "partition" function

# let partition f l = (filter f l, filter (neg f) l);
val partition : ('a->bool)-> 'a list-> 'a lisi \* 'a list =
fn

```
# let list1 = [1,31,12,4,7,2,10];
- ...
# partition is5lt list1 ;
val it : (int list * int list) = ([31,12,7,10],[1,2,10]
```

# partition even list1;
val it : (int list \* int list) = ([12,4,2,10],[1,31,7])

#### A little trick ...

```
# 2 <= 3;; ...
val it : bool = true
# "ba" <= "ab";;
val it : bool = false
# let lt = (<) ;;
val it : 'a \rightarrow 'a \rightarrow bool = fn
# 1t 2 3;;
val it : bool = true;
# lt "ba" "ab" ;;
val it : bool = false;
# let is5Lt = lt 5;
val is5lt : int -> bool = fn;
# is5lt 10;
val it : bool = true;
# is5lt 2;
val it : bool = false;
```

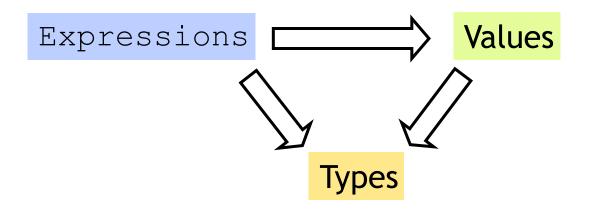
#### Put it together: a "quicksort" function

## Now, lets begin at the beginning ...



- Ocaml-top issues?
- Please post questions to Piazza
- Seating: Don't Worry!

#### What about more complex data ?



#### Many kinds of expressions:

- 1. Simple
- 2. Variables
- 3. Functions

#### What about more complex data ?

- We've seen some base types and values:
  Integers, Floats, Bool, String etc.
- Some ways to build up types:
  - Products (tuples), records, "lists"
  - Functions
- Design Principle: Orthogonality
  - Don't clutter core language with stuff
  - Few, powerful orthogonal building techniques
  - Put "derived" types, values, functions in libraries

#### What about more complex data ?

- We've seen some base types and values: - Integers, Floats, Bool, String etc.
- Some ways to build up types:
  - Products (tuples), records, "lists"
  - Functions

# Next: Building datatypes

Three key ways to build complex types/values

1. "Each-of" types

Value of T contains value of T1 and a value of T2

2. "One-of" types Value of T contains value of T1 or a value of T2

3. "Recursive"

Value of T contains (sub)-value of same type T

# Next: Building datatypes

Three key ways to build complex types/values

1. "Each-of" types (T1 \* T2)

Value of T contains value of T1 and a value of T2

2. "One-of" types Value of T contains value of T1 or a value of T2

3. "Recursive"

Value of T contains (sub)-value of same type T

# Suppose I wanted ...

- ... a program that processed lists of attributes
- Name (string)
- Age (integer)
- DOB (int-int-int)
- Address (string)
- Height (float)
- Alive (boolean)
- Phone (int-int)
- email (string)

Many kinds of attributes (too many to put in a record)

• can have multiple names, addresses, phones, emails etc. Want to store them in a list. Can I ?

# Suppose I wanted ...

Attributes:

- Name (string)
- Age (integer)
- DOB (int-int-int)
- Address (string)
- Height (real)
- Alive (boolean)
- Phone (int-int)
- email (string)

type attrib = Name of string Age of int DOB of int\*int\*int Address of string Height **of** float Alive **of** bool Phone of int\*int Email of string;;

# Quiz: Here is a typedef ...

type attrib = Name of string I Age of int I Height of float

What is the type of: Name "Tony Stark"
(a) Syntax Error
(b) Type Error
(c) string
(d) attrib
(e) 'a

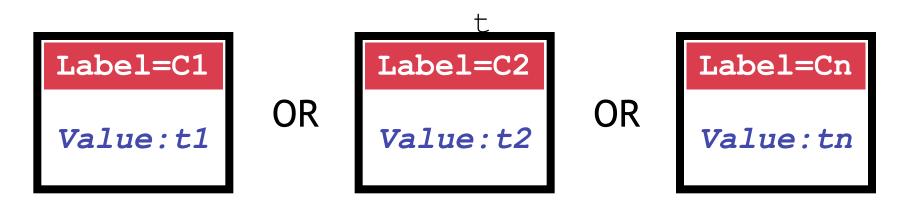
## **Constructing Datatypes**

#### type t = C1 of t1 | C2 of t2 | ... | Cn of tn

- t is a new datatype.
- A value of type  $\ensuremath{\texttt{t}}$  is either:
  - a value of type *t1* placed in a box labeled C1
- Or a value of type *t2* placed in a box labeled C2 Or ...
- Or a value of type *tn* placed in a box labeled Cn

# **Constructing Datatypes**

#### type t = C1 of t1 | C2 of t2 | ... | Cn of tn



All have the type t

## How to PUT values into box?



## Question: Here is a typedef ...

type attrib = Name of string I Age of int I Height of float

What is the type of: Age "Tony Stark"

- (a) Syntax Error
- (b) Type Error
- (C) string
- (d) attrib
- (e) **'a**

#### How to PUT values into box?

#### How to create values of type attrib?

```
# let a1 = Name "Ranjit";;
val x : attrib = Name "Ranjit"
# let a2 = Height 5.83;;
val a2 : attrib = Height 5.83
# let year = 1977 ;;
val year : int = 1977
# let a3 = DOB (9,8,year) ;;
val a3 : attrib = DOB (9,8,1977)
# let a_l = [a1;a2;a3];;
val a3 : attrib list = ...
```

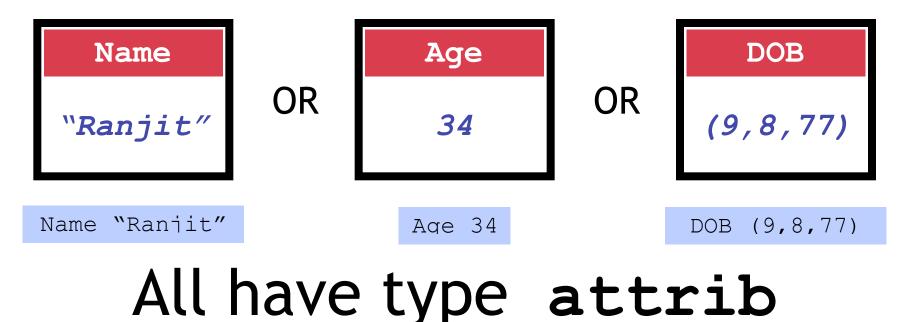
```
type attrib =
```

- Name **of** string
- | Age of int
- | DOB of int\*int\*int
- | Address of string
- | Height **of** float
- | Alive of bool
- | Phone **of** int\*int
- I Email of string;;

# **Constructing Datatypes**

#### type attrib

- = Name of string | Age of int | DOB of int\*int\*int
- | Address of string | Height of float | Alive of bool
- | Phone of int\*int | Email of string;;



# One-of types

- We've defined a "one-of" type named attrib
- Elements are one of:
  - string,
  - int,
  - int\*int\*int,
  - float,
  - bool ...

- datatype attrib =
   Name of string
   Age of int
   DOB of int\*int\*int
   Address of string
   Height of real
   Alive of bool
   Phone of int\*int
   Email of string;
- Can create uniform attrib lists
- Say I want a function to print attribs...

#### Question: Here is a typedef ...

What is the type of: [Name "Ranjit"; Age 35; Dob(9,8,77)] (a) Syntax Error (b) Type Error (C) string \* int \* (int\*int\*int) list (d)'a list (e) attrib list

# How to TEST & TAKE whats in box?



ls it a ... string? or an int? or an int\*int\*int? or ...

## How to TEST & TAKE whats in box?



#### Look at TAG!

#### Question: Here is a typedef ...

#### type attrib = Name of string | Age of int | ... What does this evaluate to?

let welcome a = match a with | Name s -> s in welcome (Name "Ranjit") (a) Name "Ranjit" : `a (b) Type Error (C) Name "Ranjit" : attrib (d) "Ranjit" : string (e) Runtime Error

# How to tell whats in the box ?

type attrib =
 Name of string
 Age of int
 DOB of int\*int\*int
 Address of string
 Height of float
 Alive of bool
 Phone of int\*int

| match e with |                   |      |                                |
|--------------|-------------------|------|--------------------------------|
| I            | Name s ->         | (*s: | string *)                      |
| I            | Age i ->          | (*i: | int *)                         |
| I            | DOB(d,m,y)->      | (*d: | <pre>int,m: int,y: int*)</pre> |
| I            | Address a ->      | (*a: | <pre>string*)</pre>            |
| Ι            | Height h ->       | (*h: | int *)                         |
| I            | Alive <b>b</b> -> | (*b: | bool*)                         |
| I            | Phone(y,r)->      | (*a: | int, r: int*)                  |

#### Pattern-match expression: check if e is of the form ...

- On match:
  - value in box bound to pattern variable
  - matching result expression is evaluated
- Simultaneously test and extract contents of box

# How to tell whats in the box ?

match e with
| Name s -> printf "%s" s
| Age i -> printf "%d" i
| DOB(d,m,y) -> printf "%d/%d/%d" d m y
| Address s -> printf "%s" s
| Height h -> printf "%f" h
| Alive b -> printf "%b" b s
| Phone(a,r) -> printf "(%d)-%d" a r

Pattern-match expression: check if e is of the form ...

- On match:
  - value in box bound to pattern variable
  - matching result expression is evaluated
- Simultaneously test and extract contents of box

#### Question: Here is a typedef ...

#### type attrib = Name of string | Age of int | ... What does this evaluate to?

let welcome a = match a with

| Name s -> s

```
in welcome (Age 34)
```

```
(a) Name "Ranjit" : 'a
```

```
(b) Type Error
```

```
(C) Name "Ranjit" : attrib
```

- (d) "Ranjit" : string
- (e) Runtime Error

# How to tell whats in the box

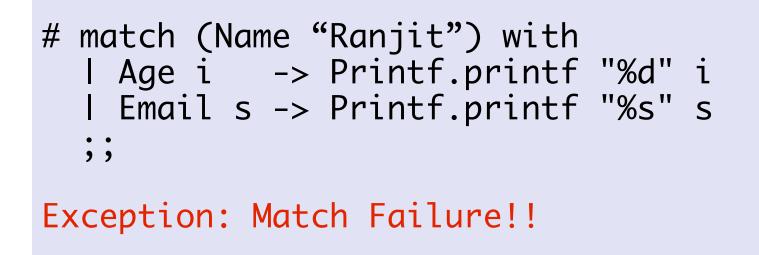
First case matches the tag (Name) Evals branch with <mark>S</mark> "bound" to string contents

## How to TEST & TAKE whats in box?



BEWARE!! Be sure to handle all TAGS!

## Beware! Handle All TAGS!



None of the cases matched the tag (Name) Causes nasty *Run-Time Error* 

## Compiler To The Rescue!!

```
# let printAttrib a =
    match a with
        Name s -> Printf.printf "%s" s
        Age i -> Printf.printf "%d" i
        DOB (d,m,y) -> Printf.printf "%d / %d / %d" d m y
        Address addr -> Printf.printf "%s" addr
        Height h -> Printf.printf "%f" h
        Alive b -> Printf.printf "%b" b
        Email e -> Printf.printf "%s" e
    ;;
Warning P: this pattern-matching is not exhaustive.
Here is an example of a value that is not matched:
Phone (_, _)
```

#### Compile-time checks for: missed cases: ML warns if you miss a case!

# Q: What does this evaluate to ?

type attrib = Name of string | ...

- (a) Type Error
- (b) "Welcome!Mickey" : string
- (c) Runtime Error
- (d) "Hello!Mickey" : string
- (e) "Hello!MickeyWelcome!Mickey"Ranjit" : string

# Compiler To The Rescue!!

```
# let printAttrib a =
    match a with
        Name s -> Printf.printf "%s" s
        Age i -> Printf.printf "%d" i
        DOB (d,m,y) -> Printf.printf "%d / %d / %d" d m y
        ...
        Age i -> Printf.printf "%d" i
    ;;
Warning U: this match case is unused.
```

#### Compile-time checks for: redundant cases: ML warns if a case never matches

# Benefits of match-with

match e with
 C1 x1 -> e1
 C2 x2 -> e2
 I ...
 I Cn xn -> en
 I Cn of tn

type t =
 C1 of t1
I Cn of tn

1. Simultaneous test-extract-bind

2. Compile-time checks for: missed cases: ML warns if you miss a t value redundant cases: ML warns if a case never matches

# match-with is an Expression

| match e |    |    | with |           |
|---------|----|----|------|-----------|
| (       | 21 | x1 | ->   | <b>e1</b> |
| (       | 22 | x2 | ->   | e2        |
| .       | •• |    |      |           |
| (       | Cn | xn | ->   | en        |
|         |    |    |      |           |

# Q: What does this evaluate to ?

type attrib = Name of string | Age of int |  $\dots$ 

let welcome a = match a with

I Name s -> s

∣ Age i -> i

in welcome (Name "Ranjit")

- (a) "Ranjit" : string
- (b) Type Error
- (C) Name "Ranjit" : attrib
- (d) Runtime Error

# match-with is an Expression



#### Type Rule

- e1, e2, ..., en must have same type T
- Type of whole expression is T

# Next: Building datatypes

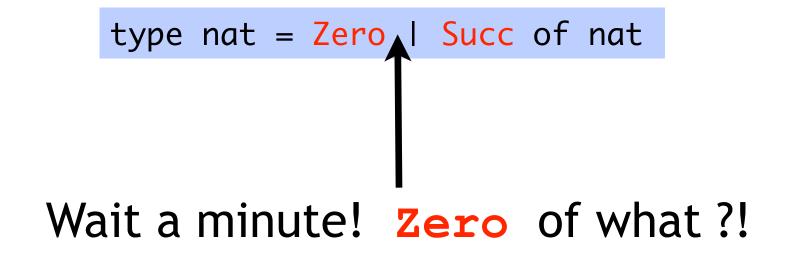
Three key ways to build complex types/values

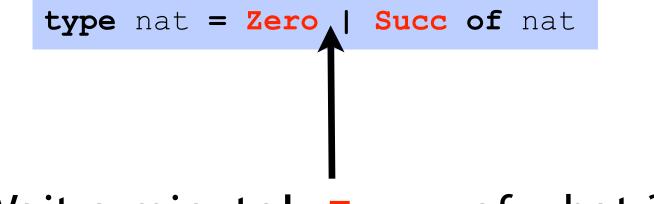
"Each-of" types t1 \* t2
 Value of T contains value of T1 and a value of T2

2. "One-of" types type t = C1 of t1 | C2 of t2
Value of T contains value of T1 or a value of T2

"Recursive" type
 Value of T contains (sub)-value of same type T

#### type nat = Zero | Succ of nat





#### Wait a minute! **Zero** of what ?!

Relax.

Means "empty box with label Zero"

type nat = Zero | Succ of nat

#### What are values of **nat**?

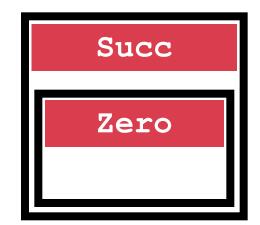
type nat = Zero | Succ of nat

#### What are values of **nat**?



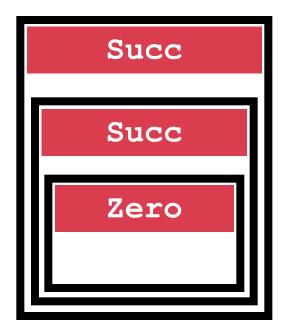
type nat = Zero | Succ of nat

#### What are values of **nat**? One **nat** contains another!



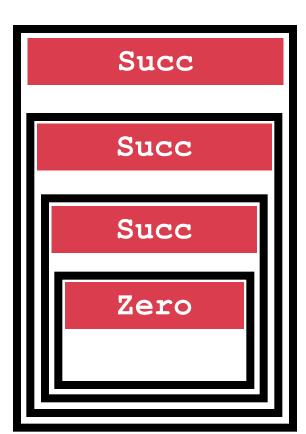
type nat = Zero | Succ of nat

#### What are values of **nat**? One **nat** contains another!



type nat = Zero | Succ of nat

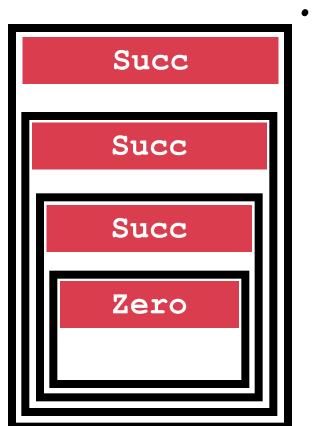
#### What are values of **nat**? One **nat** contains another!



type nat = Zero | Succ of nat

What are values of **nat**? One **nat** contains another!

nat = recursive type



# Next: Building datatypes

Three key ways to build complex types/values

"Each-of" types t1 \* t2
 Value of T contains value of T1 and a value of T2

2. "One-of" types type t = C1 of t1 | C2 of t2
Value of T contains value of T1 or a value of T2

3. "Recursive" type type t = ...I C of (...\*t)
Value of T contains (sub)-value of same type T

# Next: Lets get cosy with Recursion

#### **Recursive Code Mirrors Recursive Data**

# Next: Lets get cosy with Recursion

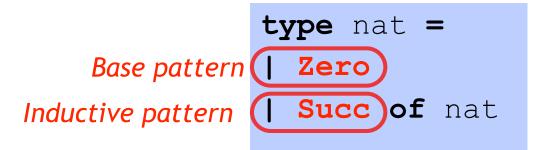
# Code Structure = Type Structure!!!

type nat =

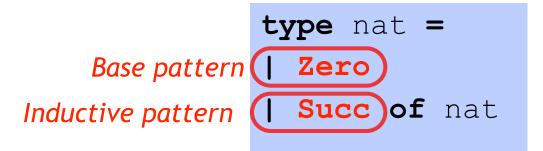
Zero

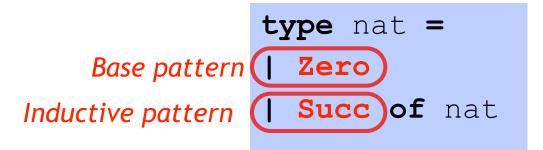
| Succ of nat

#### let rec to int n =



#### let rec to\_int n =





# Q: What does this evaluate to ?

# let rec foo n = if n<=0 then Zero else Succ(foo(n-1)) in foo 2</pre>

(a) Zero : nat
(b) Type Error
(c) 2 : nat
(c) Succ(Zero): nat
(c) Succ(Succ(Zero)) : nat

type nat =

Zero

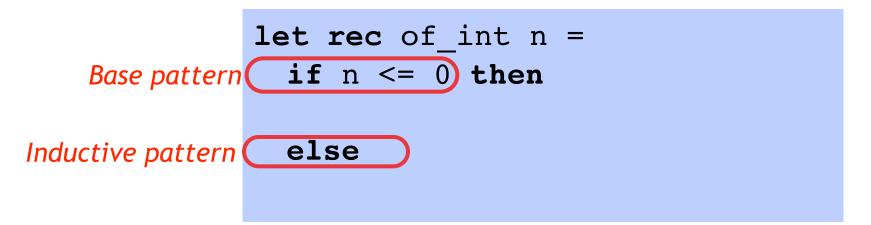
| Succ of nat

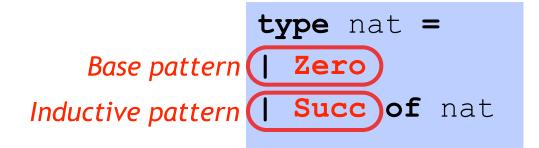
#### let rec of\_int n =

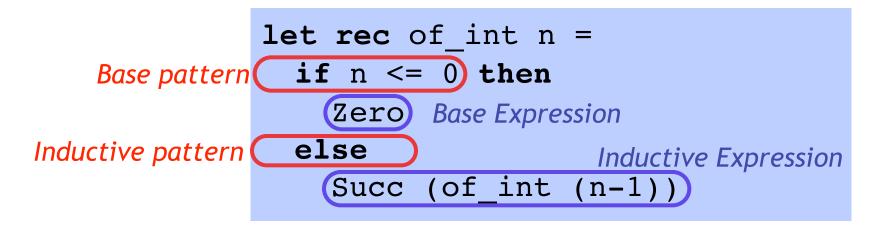


#### let rec of\_int n =







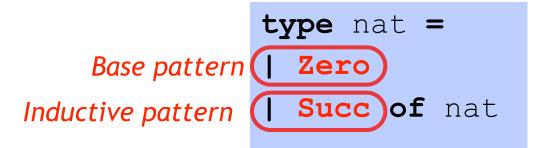


type nat =

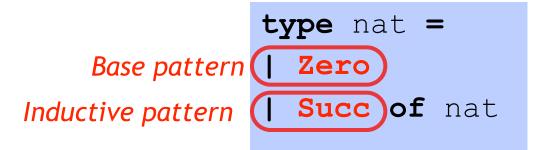
Zero

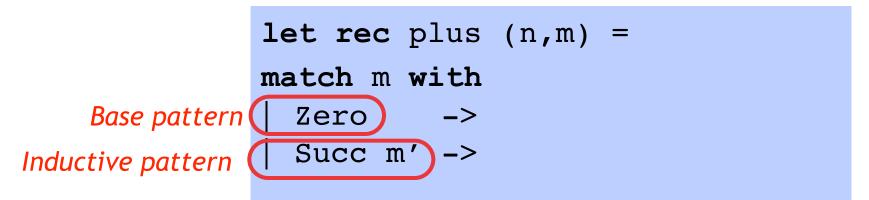
| Succ of nat

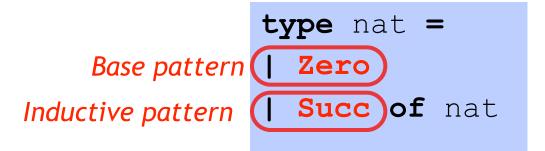
let rec plus (n,m) =

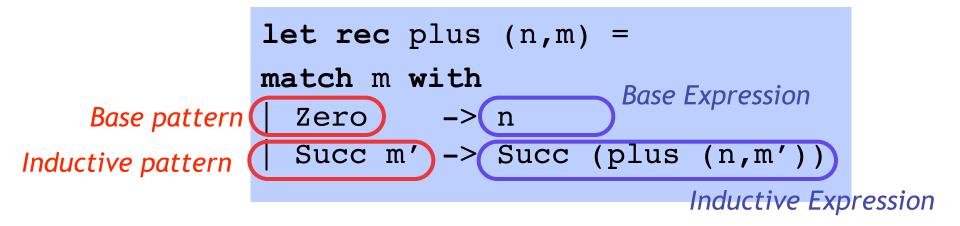


#### let rec plus (n,m) =









#### times: nat\*nat -> nat

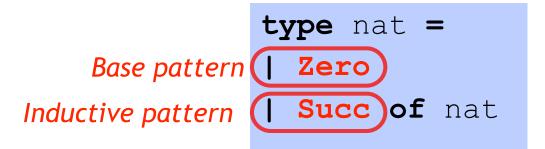
type nat =

Zero

| Succ of nat

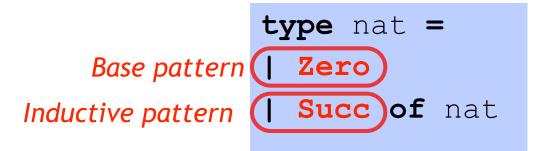
let rec times (n,m) =

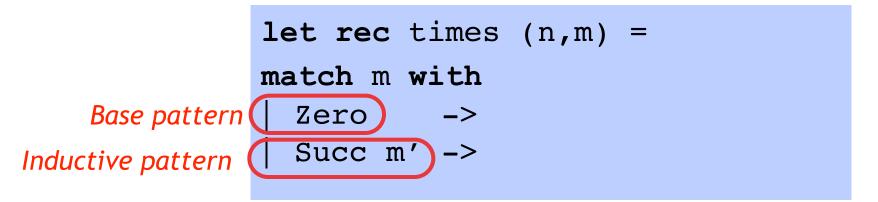
#### times: nat\*nat -> nat

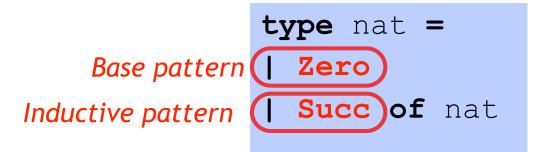


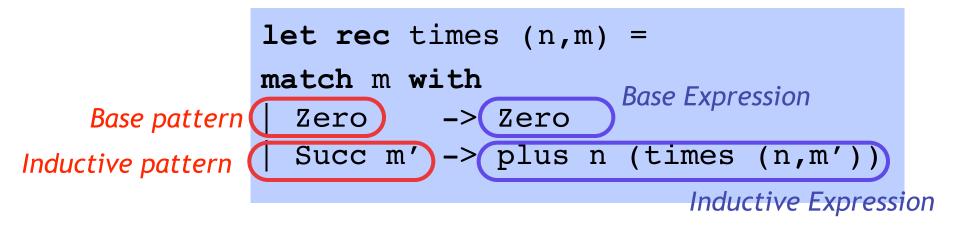
# let rec times (n,m) =

#### times: nat\*nat -> nat









#### minus: nat\*nat -> nat

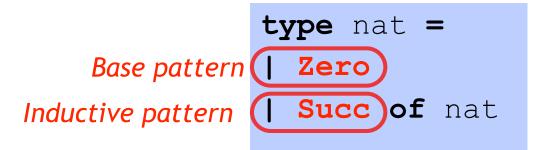
type nat =

Zero

| Succ of nat

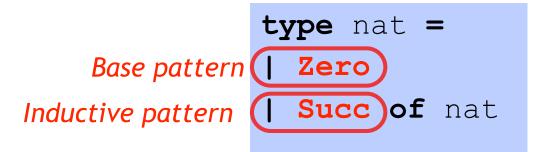
let rec minus (n,m) =

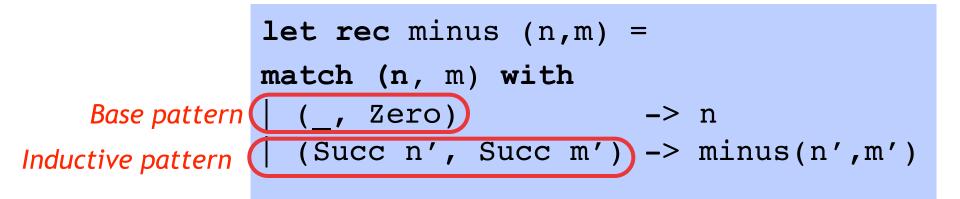
#### times: nat\*nat -> nat



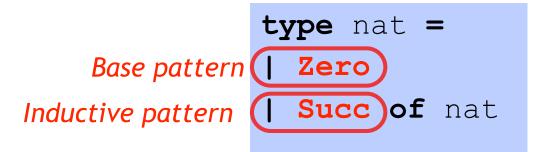
# let rec minus (n,m) =

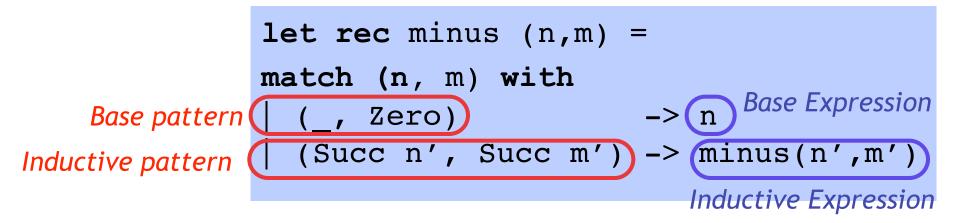
#### times: nat\*nat -> nat





#### times: nat\*nat -> nat





## Next: Lets get cosy with Recursion

#### **Recursive Code Mirrors Recursive Data**

## Lists are recursive types!

#### Think about this! What are values of int\_list ?

Cons(1,Cons(2,Cons(3,Nil))) Cons(2,Cons(3,Nil)) Cons(3,Nil) Nil

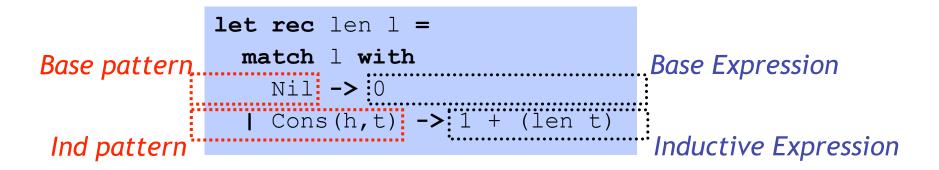
## Lists aren't built-in !

```
datatype int_list =
   Nil
   Cons of int * int_list
```

Lists are a derived type: built using elegant core!

- 1. Each-of
- 2. One-of
- 3. Recursive
- :: is just a pretty way to say "Cons"
- [] is just a pretty way to say "Nil"

## Some functions on Lists : Length

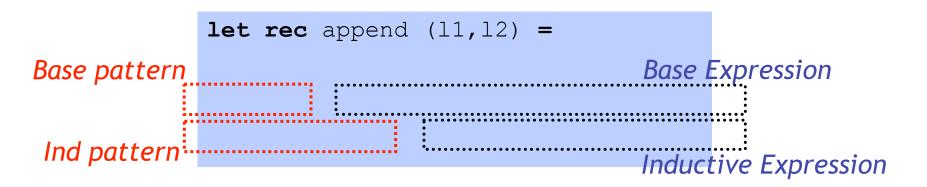


Matches everything, no binding

Pattern-matching in order

- Must match with Nil

## Some functions on Lists : Append



- Find the right induction strategy
  - Base case: pattern + expression
  - Induction case: pattern + expression

#### Well designed datatype gives strategy

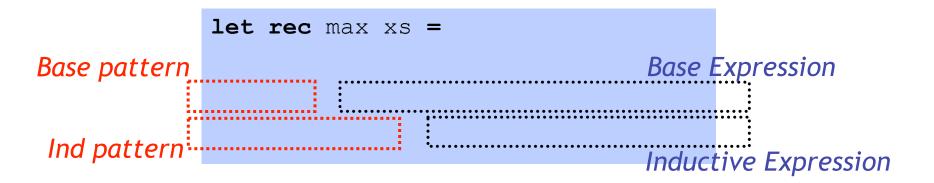
# null, hd, tl are all functions ...

Bad ML style: More than aesthetics !

Pattern-matching better than test-extract:

- ML checks all cases covered
- ML checks no redundant cases
- ...at compile-time:
  - fewer errors (crashes) during execution
  - get the bugs out ASAP!

## Some functions on Lists : Max



- Find the right induction strategy
  - Base case: pattern + expression
  - Induction case: pattern + expression

#### Well designed datatype gives strategy

## Next: Lets get cosy with Recursion

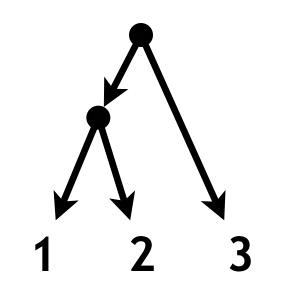
#### **Recursive Code Mirrors Recursive Data**

# Q: How is this tree represented ?

type tree =
 Leaf of int
 Node of tree\*tree

(a) (1, 2), 3

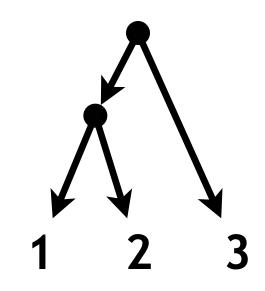
- (b) (Leaf 1, Leaf 2), Leaf 3
- (C) Node (Node (Leaf 1, Leaf 2), Leaf 3)
- (d) Node ((Leaf 1, Leaf 2), Leaf 3)
- (e) None of the above





type tree =
 Leaf of int
 Node of tree\*tree

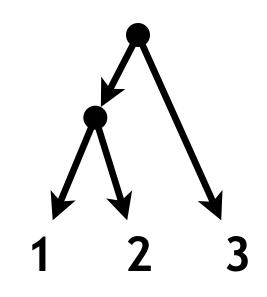
Leaf 1

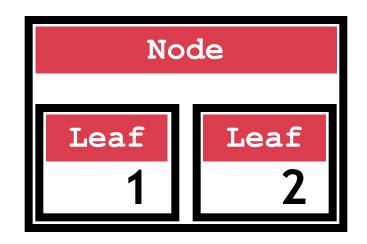




type tree =
 Leaf of int
 Node of tree\*tree

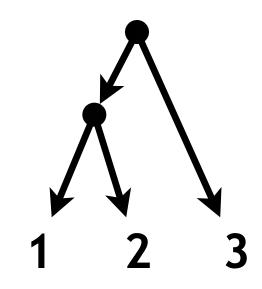
Leaf 2





type tree =
 Leaf of int
 Node of tree\*tree

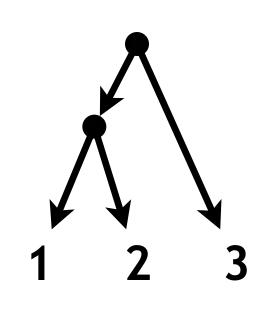
Node(Leaf 1, Leaf 2)

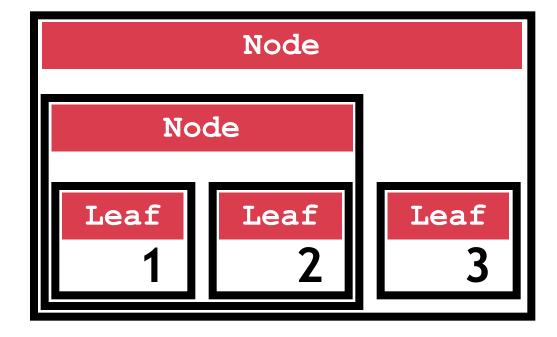


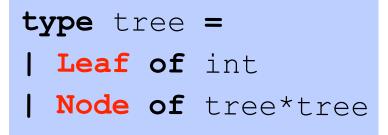
type tree =
 Leaf of int
 Node of tree\*tree



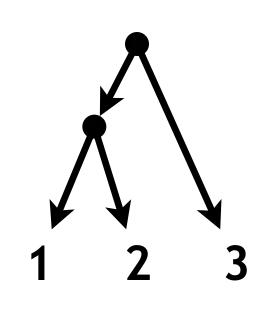
Leaf 3

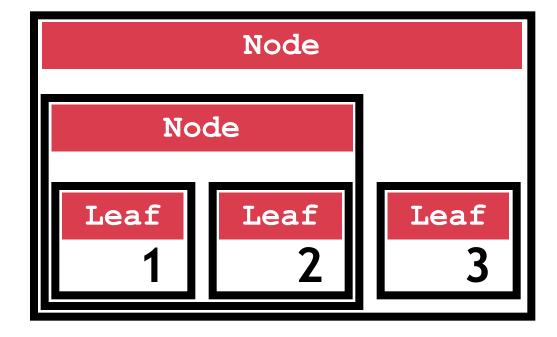


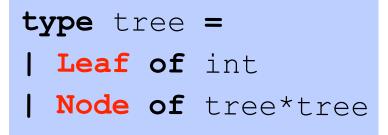




Node (Node (Leaf 1, Leaf 2), Leaf 3)







Node (Node (Leaf 1, Leaf 2), Leaf 3)

## Next: Lets get cosy with Recursion

#### **Recursive Code Mirrors Recursive Data**

## Q: What does this evaluate to ?

- foo (Node(Node(Leaf 1,Leaf 2),Leaf 3))
- (a) Type Error
- (b) 1 : int
- (C) 3 : int
- (d) 6 : int

## "Sum up the leaf values". E.g.

# let t0 = Node(Node(Leaf 1, Leaf 2), Leaf 3);;

# sum\_leaf t0 ;;

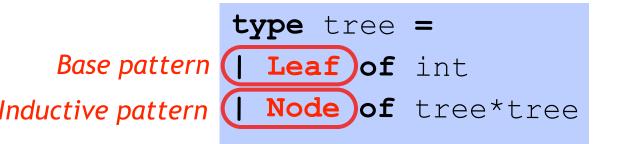
-: int = 6

type tree =

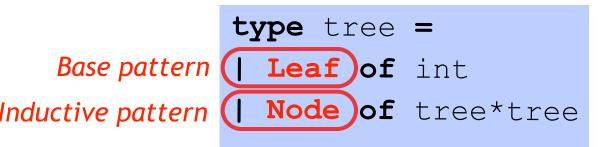
| Leaf of int

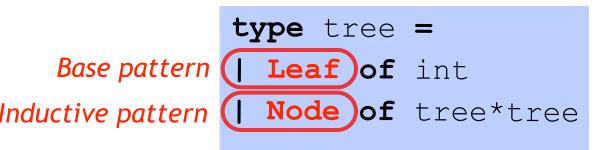
| Node of tree\*tree

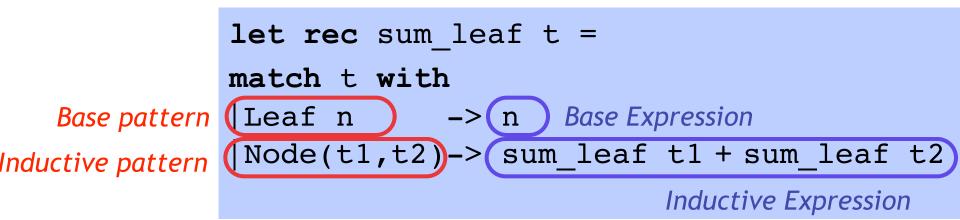
let rec sum leaf t =



#### let rec sum\_leaf t =







#### **Recursive Code Mirrors Recursive Data**

#### Code almost writes itself!

Want an arithmetic calculator to evaluate expressions like:

- 4.0 + 2.9
- 3.78 5.92
- (4.0 + 2.9) \* (3.78 5.92)

Want an arithmetic calculator to evaluate expressions like:

- 4.0 + 2.9 ====> **6**.9
- 3.78 5.92 ====> -2.14
- (4.0 + 2.9) \* (3.78 5.92) ====> -14.766

Whats a ML TYPE for REPRESENTING expressions ?

Want an arithmetic calculator to evaluate expressions like:

- 4.0 + 2.9 ====> **6.9**
- 3.78 5.92 ====> -2.14
- (4.0 + 2.9) \* (3.78 5.92) ====> -14.766

Whats a ML TYPE for REPRESENTING expressions ?

| type expr = |     |    |           |  |
|-------------|-----|----|-----------|--|
| I           | Num | of | float     |  |
| Ι           | Add | of | expr*expr |  |
| Ι           | Sub | of | expr*expr |  |
| Ι           | Mul | of | expr*expr |  |

Want an arithmetic calculator to evaluate expressions like:

- 4.0 + 2.9 ====> **6**.9
- 3.78 5.92 ====> -2.14
- (4.0 + 2.9) \* (3.78 5.92) ====> -14.766

Whats a ML FUNCTION for EVALUATING expressions ?

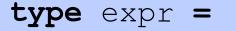
type expr =

- | Num of float
- Add of expr\*expr
- | Sub of expr\*expr
- Mul of expr\*expr

Want an arithmetic calculator to evaluate expressions like:

- 4.0 + 2.9 ====> **6**.9
- 3.78 5.92 ====> -2.14
- (4.0 + 2.9) \* (3.78 5.92) ====> -14.766

Whats a ML FUNCTION for EVALUATING expressions ?



- Num of float
- Add of expr\*expr
- | Sub of expr\*expr
- | Mul of expr\*expr

Want an arithmetic calculator to evaluate expressions like:

- 4.0 + 2.9 ====> **6**.9
- 3.78 5.92 ====> -2.14
- (4.0 + 2.9) \* (3.78 5.92) ====> -14.766

Whats a ML FUNCTION for EVALUATING expressions ?

| type expr =      | <pre>let rec eval e = match e with</pre> |
|------------------|--|
| Num of float     | Num f ->                                 |
| Add of expr*expr | <b> Add</b> (e1,e2)->                    |
| Sub of expr*expr | <b> Sub</b> (e1,e2)->                    |
| Mul of expr*expr | <b> Mul</b> (e1,e2)->                    |

Want an arithmetic calculator to evaluate expressions like:

- 4.0 + 2.9 ====> **6**.9
- 3.78 5.92 ====> -2.14
- (4.0 + 2.9) \* (3.78 5.92) ====> -14.766

Whats a ML FUNCTION for EVALUATING expressions ?

| type expr =      | <pre>let rec eval e = match e with</pre>       |
|------------------|--|
| Num of float     | Num f -> f                                     |
| Add of expr*expr | <b> Add</b> (e1,e2)-> eval e1+. eval e2        |
| Sub of expr*expr | <pre> Sub(e1,e2)-&gt; eval e1eval e2</pre>     |
| Mul of expr*expr | <pre>[Mul(e1,e2)-&gt; eval e1 *. eval e2</pre> |

