Foundations of Software Winter Semester 2007



October 9, 2007 - version 1.2

Programming in the Lambda-Calculus, Continued

Recall: Church Booleans

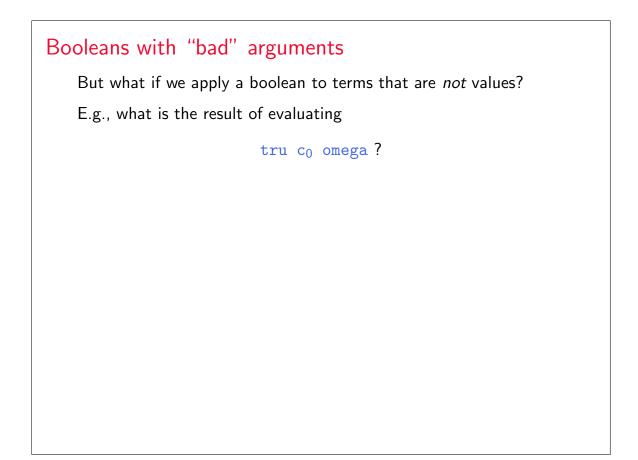
We showed last time that, if b is a boolean (i.e., it behaves like either tru or fls), then, for any values v and w, either

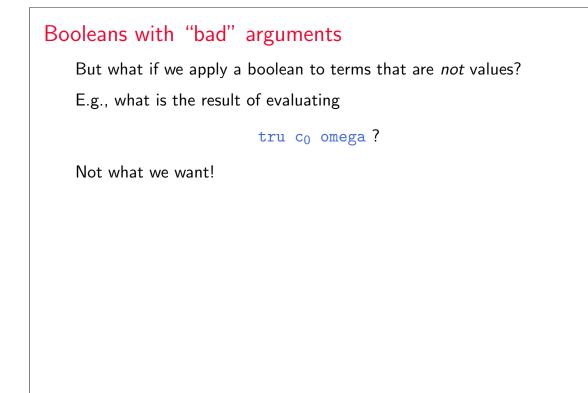
 $b v w \longrightarrow^* v$

(if b behaves like tru) or

 $b v w \longrightarrow^* w$

(if b behaves like fls).





```
A better way

Wrap the branches in an abstraction, and use a dummy "unit

value," to force evaluation of thunks:

unit = \lambda x. x

Use a "conditional function":

test = \lambda b. \lambda t. \lambda f. b t f unit

If tru' is or behaves like tru, fls' is or behaves like fls, and s

and t are arbitrary terms then

test tru' (\lambdadummy. s) (\lambdadummy. t) \rightarrow* s

test fls' (\lambdadummy. s) (\lambdadummy. t) \rightarrow* t
```

Recall: The z Operator

In the last lecture, we defined an operator z that calculates the "fixed point" of a function it is applied to:

```
z = \lambda f. \lambda y. (\lambda x. f (\lambda y. x x y)) (\lambda x. f (\lambda y. x x y)) y
```

```
That is, if z_f = z f then z_f v \longrightarrow^* f z_f v.
```

Recall: Factorial

As an example, we defined the factorial function as follows:

```
fact =
z (\lambdafct.
\lambdan.
if n=0 then 1
else n * (fct (pred n)))
```

For simplicity, we used primitive values from the calculus of numbers and booleans presented in week 2, and even used shortcuts like 1 and *.

As mentioned, this can be translated "straightforwardly" into the pure lambda-calculus. Let's do that.

Lambda calculus version of Factorial (not!)

Here is the naive translation:

```
badfact =

z (\lambdafct.

\lambdan.

iszro n

c_1

(times n (fct (prd n))))
```

Why is this not what we want?

Lambda calculus version of Factorial (not!) Here is the naive translation: $badfact = z (\lambda fct.) \\ \lambda n. \\ iszro n \\ c_1 \\ (times n (fct (prd n))))$ Why is this not what we want? (Hint: What happens when we evaluate badfact c_0 ?)

Lambda calculus version of Factorial

A better version:

Displaying numbers fact $c_3 \rightarrow^*$

Displaying numbers

```
fact c_3 \longrightarrow^* (\lambda s. \lambda z.
                           s ((\lambdas. \lambdaz.
                              s ((\lambdas. \lambdaz.
                                  s ((\lambdas. \lambdaz.
                                      s ((\lambdas. \lambdaz.
                                         s ((\lambdas. \lambdaz.
                                            s ((\lambdas. \lambdaz. z)
                                                  s z))
                                              s z))
                                           s z))
                                       s z))
                                    s z))
                                s z))
```

Ugh!

Displaying numbers If we enrich the pure lambda-calculus with "regular numbers," we can display church numerals by converting them to regular numbers: realnat = λ n. n (λ m. succ m) 0 Now: realnat (times $c_2 c_2$) * succ (succ (succ (succ zero))).

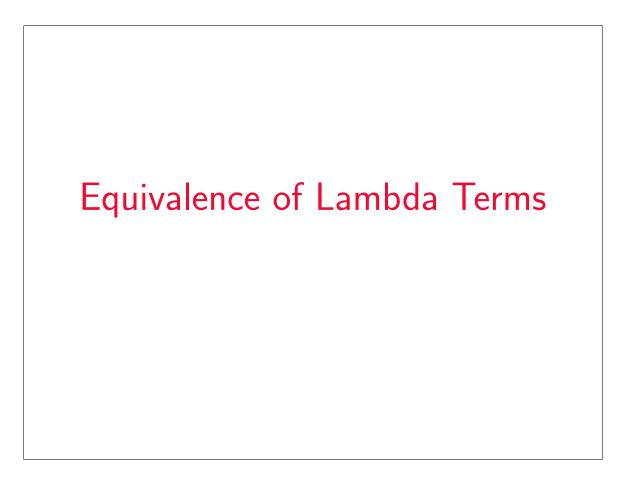
Displaying numbers

Alternatively, we can convert a few specific numbers:

whack = λ n. (equal n c₀) c₀ ((equal n c₁) c₁ ((equal n c₂) c₂ ((equal n c₃) c₃ ((equal n c₄) c₄ ((equal n c₅) c₅ ((equal n c₆) c₆ n))))))

Now:

```
whack (fact c_3)
\longrightarrow^*
\lambdas. \lambdaz. s (s (s (s (s z))))
```



Recall: Church Numerals

We have seen how certain terms in the lambda-calculus can be used to represent natural numbers.

 $c_0 = \lambda s. \quad \lambda z. \quad z$ $c_1 = \lambda s. \quad \lambda z. \quad s \quad z$ $c_2 = \lambda s. \quad \lambda z. \quad s \quad (s \quad z)$ $c_3 = \lambda s. \quad \lambda z. \quad s \quad (s \quad (s \quad z))$

Other lambda-terms represent common operations on numbers:

 $scc = \lambda n. \lambda s. \lambda z. s (n s z)$

Recall: Church Numerals

We have seen how certain terms in the lambda-calculus can be used to represent natural numbers.

 $c_0 = \lambda s. \quad \lambda z. \quad z$ $c_1 = \lambda s. \quad \lambda z. \quad s \quad z$ $c_2 = \lambda s. \quad \lambda z. \quad s \quad (s \quad z)$ $c_3 = \lambda s. \quad \lambda z. \quad s \quad (s \quad (s \quad z))$

Other lambda-terms represent common operations on numbers:

 $scc = \lambda n. \lambda s. \lambda z. s (n s z)$

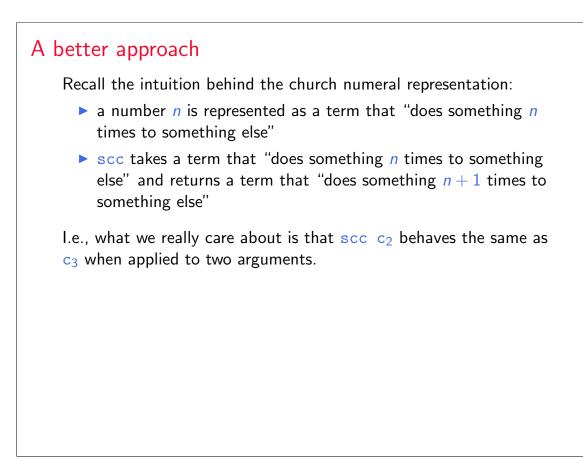
In what sense can we say this representation is "correct"? In particular, on what basis can we argue that scc on church numerals corresponds to ordinary successor on numbers?

The naive approach

One possibility:

For each *n*, the term scc c_n evaluates to c_{n+1} .

The naive approach... doesn't work One possibility: For each *n*, the term scc c_n evaluates to c_{n+1}. Unfortunately, this is false. E.g.: $scc c_2 = (\lambda n. \lambda s. \lambda z. s (n s z)) (\lambda s. \lambda z. s (s z))$ $\rightarrow \lambda s. \lambda z. s ((\lambda s. \lambda z. s (s z)) s z)$ $\neq \lambda s. \lambda z. s (s (s z))$ $= c_3$



```
scc c<sub>2</sub> v w = (\lambdan. \lambdas. \lambdaz. s (n s z)) (\lambdas. \lambdaz. s (s z)) v w

\rightarrow (\lambdas. \lambdaz. s ((\lambdas. \lambdaz. s (s z)) v z)) w

\rightarrow (\lambdaz. v ((\lambdas. \lambdaz. s (s z)) v w)

\rightarrow v ((\lambdas. \lambdaz. s (s z)) v w)

\rightarrow v (v (v w))

c<sub>3</sub> v w = (\lambdas. \lambdaz. s (s (s z))) v w

\rightarrow (\lambdaz. v (v (v z))) w

\rightarrow v (v (v w)))
```

A general question

We have argued that, although $scc c_2$ and c_3 do not evaluate to the same thing, they are nevertheless "behaviorally equivalent."

What, precisely, does behavioral equivalence mean?

Intuition

Roughly,

"terms ${\tt s}$ and ${\tt t}$ are behaviorally equivalent"

should mean:

"there is no 'test' that distinguishes s and t — i.e., no way to put them in the same context and observe different results."

Intuition

Roughly,

"terms s and t are behaviorally equivalent"

should mean:

"there is no 'test' that distinguishes s and t — i.e., no way to put them in the same context and observe different results."

To make this precise, we need to be clear what we mean by a *testing context* and how we are going to *observe* the results of a test.

Examples

```
\begin{array}{l} \mathrm{tru} = \lambda \mathrm{t.} \ \lambda \mathrm{f.} \ \mathrm{t} \\ \mathrm{tru'} = \lambda \mathrm{t.} \ \lambda \mathrm{f.} \ (\lambda \mathrm{x.x}) \ \mathrm{t} \\ \mathrm{fls} = \lambda \mathrm{t.} \ \lambda \mathrm{f.} \ \mathrm{f} \\ \mathrm{omega} = (\lambda \mathrm{x.} \ \mathrm{x} \ \mathrm{x}) \ (\lambda \mathrm{x.} \ \mathrm{x} \ \mathrm{x}) \\ \mathrm{poisonpill} = \lambda \mathrm{x.} \ \mathrm{omega} \\ \mathrm{placebo} = \lambda \mathrm{x.} \ \mathrm{tru} \\ \mathrm{Y}_{\mathrm{f}} = (\lambda \mathrm{x.} \ \mathrm{f} \ (\mathrm{x} \ \mathrm{x})) \ (\lambda \mathrm{x.} \ \mathrm{f} \ (\mathrm{x} \ \mathrm{x})) \end{array}
```

Which of these are behaviorally equivalent?

Observational equivalence

As a first step toward defining behavioral equivalence, we can use the notion of *normalizability* to define a simple notion of *test*.

Two terms s and t are said to be *observationally equivalent* if either both are normalizable (i.e., they reach a normal form after a finite number of evaluation steps) or both diverge.

I.e., we "observe" a term's behavior simply by running it and seeing if it halts.

Observational equivalence

As a first step toward defining behavioral equivalence, we can use the notion of *normalizability* to define a simple notion of *test*.

Two terms s and t are said to be *observationally equivalent* if either both are normalizable (i.e., they reach a normal form after a finite number of evaluation steps) or both diverge.

I.e., we "observe" a term's behavior simply by running it and seeing if it halts.

Aside:

Is observational equivalence a decidable property?

Observational equivalence

As a first step toward defining behavioral equivalence, we can use the notion of *normalizability* to define a simple notion of *test*.

Two terms s and t are said to be *observationally equivalent* if either both are normalizable (i.e., they reach a normal form after a finite number of evaluation steps) or both diverge.

I.e., we "observe" a term's behavior simply by running it and seeing if it halts.

Aside:

- Is observational equivalence a decidable property?
- Does this mean the definition is ill-formed?

Examples

omega and tru are not observationally equivalent

Examples

- omega and tru are not observationally equivalent
- tru and fls are observationally equivalent

Behavioral Equivalence

This primitive notion of observation now gives us a way of "testing" terms for behavioral equivalence

Terms s and t are said to be *behaviorally equivalent* if, for every finite sequence of values v_1, v_2, \ldots, v_n , the applications

 $s v_1 v_2 \ldots v_n$

and

t $v_1 v_2 \ldots v_n$

are observationally equivalent.

Examples

These terms are behaviorally equivalent:

tru = $\lambda t. \lambda f. t$ tru' = $\lambda t. \lambda f. (\lambda x.x) t$

So are these:

omega = $(\lambda x. x x) (\lambda x. x x)$ $Y_f = (\lambda x. f (x x)) (\lambda x. f (x x))$

These are not behaviorally equivalent (to each other, or to any of the terms above):

fls = λt . λf . f poisonpill = λx . omega placebo = λx . tru

Proving behavioral equivalence

Given terms s and t, how do we *prove* that they are (or are not) behaviorally equivalent?

Proving behavioral inequivalence

To prove that s and t are *not* behaviorally equivalent, it suffices to find a sequence of values $v_1 \dots v_n$ such that one of

 $s v_1 v_2 \ldots v_n$

 and

t $v_1 v_2 \ldots v_n$

diverges, while the other reaches a normal form.

Proving behavioral inequivalence

Example:

the argument sequence (λx. x) poisonpill (λx. x) demonstrate that tru is not behaviorally equivalent to fls:

tru $(\lambda x. x)$ poisonpill $(\lambda x. x)$ $\longrightarrow^{*} (\lambda x. x) (\lambda x. x)$ $\longrightarrow^{*} \lambda x. x$

fls (λx . x) poisonpill (λx . x) \longrightarrow^* poisonpill (λx . x), which diverges

Proving behavioral equivalence

To prove that s and t *are* behaviorally equivalent, we have to work harder: we must show that, for *every* sequence of values $v_1 \dots v_n$, either both

 $s v_1 v_2 \ldots v_n$

 and

t v₁ v₂ ... v_n

diverge, or else both reach a normal form.

How can we do this?

Proving behavioral equivalence

In general, such proofs require some additional machinery that we will not have time to get into in this course (so-called *applicative bisimulation*). But, in some cases, we can find simple proofs. *Theorem:* These terms are behaviorally equivalent:

tru = $\lambda t. \lambda f. t$ tru' = $\lambda t. \lambda f. (\lambda x.x) t$

Proof: Consider an arbitrary sequence of values $v_1 \dots v_n$.

- For the case where the sequence has up to one element (i.e., n ≤ 1), note that both tru / tru v₁ and tru' / tru' v₁ reach normal forms after zero / one reduction steps.
- For the case where the sequence has more than one element (i.e., n > 1), note that both tru v₁ v₂ v₃ ... v_n and tru' v₁ v₂ v₃ ... v_n reduce to v₁ v₃ ... v_n. So either both normalize or both diverge.

Proving behavioral equivalence Theorem: These terms are behaviorally equivalent: $\begin{array}{l}
\text{omega} = (\lambda x. x x) (\lambda x. x x) \\
\text{Y}_{f} = (\lambda x. f (x x)) (\lambda x. f (x x))
\end{array}$

Proof: Both

omega $v_1 \ldots v_n$

and

 $Y_f v_1 \dots v_n$

diverge, for every sequence of arguments $v_1 \dots v_n$.

Inductive Proofs about the Lambda Calculus

Two induction principles

Like before, we have two ways to prove that properties are true of the untyped lambda calculus.

- Structural induction on terms
- Induction on a derivation of $t \longrightarrow t'$.

Let's look at an example of each.

Structural induction on terms

To show that a property $\mathcal P$ holds for all lambda-terms $\mathtt{t},$ it suffices to show that

- *P* holds when t is a variable;
- P holds when t is a lambda-abstraction \lambda x. t₁, assuming that P holds for the immediate subterm t₁; and
- P holds when t is an application t₁ t₂, assuming that P holds for the immediate subterms t₁ and t₂.

Structural induction on terms

To show that a property ${\mathcal P}$ holds for all lambda-terms ${\tt t},$ it suffices to show that

- *P* holds when t is a variable;
- P holds when t is a lambda-abstraction \u03c0 x. t₁, assuming that P holds for the immediate subterm t₁; and
- P holds when t is an application t₁ t₂, assuming that P holds for the immediate subterms t₁ and t₂.

N.b.: The variant of this principle where "immediate subterm" is replaced by "arbitrary subterm" is also valid. (Cf. *ordinary induction* vs. *complete induction* on the natural numbers.)

An example of structural induction on terms

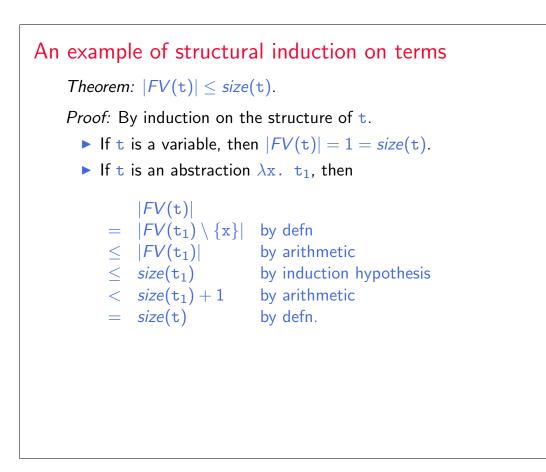
Define the set of *free variables* in a lambda-term as follows:

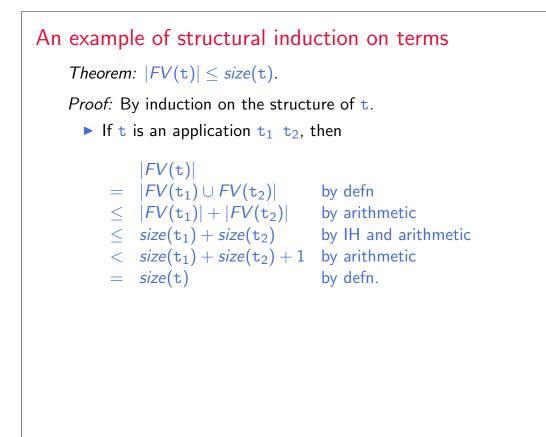
$$\begin{split} & FV(\mathtt{x}) = \{\mathtt{x}\} \\ & FV(\lambda\mathtt{x}.\mathtt{t}_1) = FV(\mathtt{t}_1) \setminus \{\mathtt{x}\} \\ & FV(\mathtt{t}_1 \ \mathtt{t}_2) = FV(\mathtt{t}_1) \cup FV(\mathtt{t}_2) \end{split}$$

Define the *size* of a lambda-term as follows:

 $\begin{aligned} & \textit{size}(\mathtt{x}) = 1 \\ & \textit{size}(\lambda\mathtt{x}.\mathtt{t}_1) = \textit{size}(\mathtt{t}_1) + 1 \\ & \textit{size}(\mathtt{t}_1 \ \mathtt{t}_2) = \textit{size}(\mathtt{t}_1) + \textit{size}(\mathtt{t}_2) + 1 \end{aligned}$

Theorem: $|FV(t)| \leq size(t)$.





Induction on derivations

Recall that the reduction relation is defined as the smallest binary relation on terms satisfying the following rules:

 $\begin{array}{ll} (\lambda \mathbf{x}.\mathbf{t}_{1}) & \mathbf{v}_{2} \longrightarrow [\mathbf{x} \mapsto \mathbf{v}_{2}]\mathbf{t}_{1} & \text{(E-APPABS)} \\ \\ & \frac{\mathbf{t}_{1} \longrightarrow \mathbf{t}_{1}'}{\mathbf{t}_{1} & \mathbf{t}_{2} \longrightarrow \mathbf{t}_{1}' & \mathbf{t}_{2} \end{array} & \text{(E-APP1)} \\ \\ & \frac{\mathbf{t}_{2} \longrightarrow \mathbf{t}_{2}'}{\mathbf{v}_{1} & \mathbf{t}_{2} \longrightarrow \mathbf{v}_{1} & \mathbf{t}_{2}'} & \text{(E-APP2)} \end{array}$

Induction on derivations

Induction principle for the small-step evaluation relation.

To show that a property $\mathcal P$ holds for all derivations of $t\longrightarrow t',$ it suffices to show that

- \blacktriangleright \mathcal{P} holds for all derivations that use the rule E-AppAbs;
- *P* holds for all derivations that end with a use of E-App1 assuming that *P* holds for all subderivations; and
- *P* holds for all derivations that end with a use of E-App2 assuming that *P* holds for all subderivations.

An example of induction on derivations

Theorem: if $t \longrightarrow t'$ then $FV(t) \supseteq FV(t')$.

We must prove, for all derivations of $t \longrightarrow t'$, that $FV(t) \supseteq FV(t')$.

An example of induction on derivations

Theorem: if $t \longrightarrow t'$ then $FV(t) \supseteq FV(t')$.

Proof: by induction on the derivation of $t \longrightarrow t'$. There are three cases:

An example of induction on derivations

Theorem: if $t \longrightarrow t'$ then $FV(t) \supseteq FV(t')$.

Proof: by induction on the derivation of $t \longrightarrow t'$. There are three cases:

▶ If the derivation of $t \longrightarrow t'$ is just a use of E-AppAbs, then t is $(\lambda x.t_1)v$ and t' is $[x \mapsto v]t_1$. Reason as follows:

```
\begin{aligned} FV(t) &= FV((\lambda x.t_1)v) \\ &= FV(t_1) \setminus \{x\} \cup FV(v) \\ &\supseteq FV([x \mapsto v]t_1) \\ &= FV(t') \end{aligned}
```

An example of induction on derivations

Theorem: if $t \longrightarrow t'$ then $FV(t) \supseteq FV(t')$.

Proof: by induction on the derivation of $t \longrightarrow t'$. There are three cases:

If the derivation ends with a use of E-App1, then t has the form t₁ t₂ and t' has the form t'₁ t₂, and we have a subderivation of t₁ → t'₁

By the induction hypothesis, $FV(t_1) \supseteq FV(t'_1)$. Now calculate:

 $FV(t) = FV(t_1 t_2)$ = $FV(t_1) \cup FV(t_2)$ $\supseteq FV(t'_1) \cup FV(t_2)$ = $FV(t'_1 t_2)$ = FV(t')

E-App2 is treated similarly.