

Foundations of Software Fall 2023

Week 2

Part I

Modeling programming languages

Where we're going

Going Meta...

The functional programming style used in OCaml and Scala is based on treating *programs as data* — i.e., on writing functions that manipulate other functions as their inputs and outputs.

Everything in this course is based on treating *programs as mathematical objects* — i.e., we will be building mathematical theories whose basic objects of study are programs (and whole programming languages).

Jargon: We will be studying the *metatheory* of programming languages.

Warning!

The material in the next couple of lectures is more slippery than it may first appear.

“I believe it when I hear it” is not a sufficient test of understanding.

A much better test is “I can explain it so that someone else believes it.”

“You never really misunderstand something
until you try to teach it...”
— Anon.

Basics of Induction (Review)

Induction

Principle of *ordinary induction* on natural numbers:

Suppose that P is a predicate on the natural numbers.

Then:

If $P(0)$

and, for all i , $P(i)$ implies $P(i + 1)$,

then $P(n)$ holds for all n .

Example

Theorem: $2^0 + 2^1 + \dots + 2^n = 2^{n+1} - 1$, for every n .

Proof: Let $P(i)$ be " $2^0 + 2^1 + \dots + 2^i = 2^{i+1} - 1$."

- Show $P(0)$:

$$2^0 = 1 = 2^1 - 1$$

- Show that $P(i)$ implies $P(i+1)$:

$$\begin{aligned} 2^0 + 2^1 + \dots + 2^{i+1} &= (2^0 + 2^1 + \dots + 2^i) + 2^{i+1} \\ &= (2^{i+1} - 1) + 2^{i+1} && \text{by IH} \\ &= 2 \cdot (2^{i+1}) - 1 \\ &= 2^{i+2} - 1 \end{aligned}$$

- The result ($P(n)$ for all n) follows by the principle of (ordinary) induction.

Shorthand form

Theorem: $2^0 + 2^1 + \dots + 2^n = 2^{n+1} - 1$, for every n .

Proof: By induction on n .

► Base case ($n = 0$):

$$2^0 = 1 = 2^1 - 1$$

► Inductive case ($n = i + 1$):

$$\begin{aligned} 2^0 + 2^1 + \dots + 2^{i+1} &= (2^0 + 2^1 + \dots + 2^i) + 2^{i+1} \\ &= (2^{i+1} - 1) + 2^{i+1} && \text{IH} \\ &= 2 \cdot (2^{i+1}) - 1 \\ &= 2^{i+2} - 1 \end{aligned}$$

Complete Induction

Principle of *complete induction* on natural numbers:

Suppose that P is a predicate on the natural numbers.

Then:

If, for each natural number n ,

given $P(i)$ for all $i < n$

we can show $P(n)$,

then $P(n)$ holds for all n .

Example of complete induction (shorthand form)

Theorem: Every natural $n > 1$ is the product of (one or more) prime numbers.

Proof: By complete induction on n .

- ▶ IH: Every natural $1 < m < n$ is the product of prime numbers.
- ▶ If n is a prime number, then it is the product of itself.
- ▶ Otherwise,
 - ▶ By definition, there exist $1 < m_1, m_2 < n$ such that $n = m_1 m_2$.
 - ▶ By the IH, m_1 and m_2 are both the product of prime numbers.
 - ▶ Therefore, $n = m_1 m_2$ is also the product of prime numbers.

Complete versus ordinary induction

Ordinary and complete induction are *interderivable* — assuming one, we can prove the other.

Thus, the choice of which to use for a particular proof is purely a question of style.

We'll see some other (equivalent) styles as we go along.

Clicker questions

Clicker question: Why do you think we use clicker questions?
(several possible answers)

- A. Check attendance to lectures
- B. Help you effectively learn
- C. Encourage interactivity during lectures
- D. Evaluate (with grades)

URL: ttpoll.eu

Session ID: [cs452](#)

Do not use any login – use guest mode only

Legal information regarding the use of clickers

Goal = to encourage active participation (not to assess individuals)

You may choose between using the *website*, the *smartphone app* or a *hardward clicker*.

Data protection

- ▶ Data is processed outside Switzerland, which may include the USA or EU countries
- ▶ Contract with EPFL = *Turning Technologies will not reuse the data collected for any other purpose*
- ▶ Condition: you must use the website or smarphone app in *guest mode* (without entering any personal information)

Induction and non-naturals

Clicker question: What other sets of mathematical objects can we do valid complete induction on? (several possible answers)

- A. Integers (\mathbb{Z})
- B. Non-negative rationals (\mathbb{Q}^+)
- C. Non-negative reals (\mathbb{R}^+)
- D. Strings of characters (as in Java/Scala/Python)

URL: ttpoll.eu

Session ID: [cs452](#)

Syntax

Simple Arithmetic Expressions

Here is a BNF grammar for a very simple language of arithmetic expressions:

<code>t ::=</code>	<i>terms</i>
<code> true</code>	<i>constant true</i>
<code> false</code>	<i>constant false</i>
<code> if t then t else t</code>	<i>conditional</i>
<code> 0</code>	<i>constant zero</i>
<code> succ t</code>	<i>successor</i>
<code> pred t</code>	<i>predecessor</i>
<code> iszero t</code>	<i>zero test</i>

Terminology:

- ▶ `t` here is a *metavariable*

Abstract vs. concrete syntax

Q: Does this grammar define a set of *character strings*, a set of *token lists*, or a set of *abstract syntax trees*?

Abstract vs. concrete syntax

Q: Does this grammar define a set of *character strings*, a set of *token lists*, or a set of *abstract syntax trees*?

A: In a sense, all three. But we are primarily interested, here, in abstract syntax trees.

For this reason, grammars like the one on the previous slide are sometimes called *abstract grammars*. An abstract grammar *defines* a set of abstract syntax trees and *suggests* a mapping from character strings to trees.

We then *write* terms as linear character strings rather than trees simply for convenience. If there is any potential confusion about what tree is intended, we use parentheses to disambiguate.

Q: So, are

```
succ 0  
succ (0)  
(((succ (((((0))))))))
```

“the same term”?

What about

```
succ 0  
pred (succ (succ 0))
```

?

A more explicit form of the definition

The set \mathcal{T} of *terms* is the smallest set such that

1. $\{\text{true}, \text{false}, 0\} \subseteq \mathcal{T}$;
2. if $t_1 \in \mathcal{T}$, then $\{\text{succ } t_1, \text{pred } t_1, \text{iszero } t_1\} \subseteq \mathcal{T}$;
3. if $t_1 \in \mathcal{T}$, $t_2 \in \mathcal{T}$, and $t_3 \in \mathcal{T}$, then
if t_1 then t_2 else $t_3 \in \mathcal{T}$.

Inference rules

An alternate notation for the same definition:

$$\begin{array}{c} \text{true} \in \mathcal{T} \\ \hline \text{succ } t_1 \in \mathcal{T} \end{array} \quad \begin{array}{c} \text{false} \in \mathcal{T} \\ \hline \text{pred } t_1 \in \mathcal{T} \end{array} \quad \begin{array}{c} 0 \in \mathcal{T} \\ \hline \text{iszero } t_1 \in \mathcal{T} \end{array}$$
$$\frac{t_1 \in \mathcal{T} \quad t_2 \in \mathcal{T} \quad t_3 \in \mathcal{T}}{\text{if } t_1 \text{ then } t_2 \text{ else } t_3 \in \mathcal{T}}$$

Note that “the smallest set closed under...” is implied (but often not stated explicitly).

Terminology:

- ▶ axiom vs. rule
- ▶ concrete rule vs. rule schema

Terms, concretely

Define an infinite sequence of sets, $\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \dots$, as follows:

$$\mathcal{S}_0 = \emptyset$$

$$\begin{aligned}\mathcal{S}_{i+1} = & \{\text{true}, \text{false}, 0\} \\ & \cup \{\text{succ } t_1, \text{pred } t_1, \text{iszero } t_1 \mid t_1 \in \mathcal{S}_i\} \\ & \cup \{\text{if } t_1 \text{ then } t_2 \text{ else } t_3 \mid t_1, t_2, t_3 \in \mathcal{S}_i\}\end{aligned}$$

Now let

$$\mathcal{S} = \bigcup_i \mathcal{S}_i$$

Comparing the definitions

We have seen two different presentations of terms:

1. as the *smallest* set that is *closed* under certain rules (\mathcal{T})
 - ▶ explicit inductive definition
 - ▶ BNF shorthand
 - ▶ inference rule shorthand
2. as the *limit* (\mathcal{S}) of a series of sets (of larger and larger terms)

Comparing the definitions

We have seen two different presentations of terms:

1. as the *smallest* set that is *closed* under certain rules (\mathcal{T})
 - ▶ explicit inductive definition
 - ▶ BNF shorthand
 - ▶ inference rule shorthand
2. as the *limit* (\mathcal{S}) of a series of sets (of larger and larger terms)

What does it mean to assert that “these presentations are equivalent”?

Operational Semantics

Abstract Machines

An *abstract machine* consists of:

- ▶ a set of *states*
- ▶ a *transition relation* on states, written \longrightarrow

We read “ $t \longrightarrow t'$ ” as “ t evaluates to t' in one step”.

A state records *all* the information in the machine at a given moment. For example, an abstract-machine-style description of a conventional microprocessor would include the program counter, the contents of the registers, the contents of main memory, and the machine code program being executed.

Abstract Machines

For the very simple languages we are considering at the moment, however, the term being evaluated is the whole state of the abstract machine.

Nb. Often, the transition relation is actually a partial function: i.e., from a given state, there is at most one possible next state. But in general there may be many.

Operational semantics for Booleans

Syntax of terms and values

t ::=

true

false

if t then t else t

terms

constant true

constant false

conditional

v ::=

true

false

values

true value

false value

Evaluation relation for Booleans

The evaluation relation $t \longrightarrow t'$ is the smallest relation closed under the following rules:

$\text{if true then } t_2 \text{ else } t_3 \longrightarrow t_2 \quad (\text{E-IFTRUE})$

$\text{if false then } t_2 \text{ else } t_3 \longrightarrow t_3 \quad (\text{E-IFFALSE})$

$$\frac{t_1 \longrightarrow t'_1}{\text{if } t_1 \text{ then } t_2 \text{ else } t_3 \longrightarrow \text{if } t'_1 \text{ then } t_2 \text{ else } t_3} \quad (\text{E-IF})$$

Terminology

Computation rules:

`if true then t2 else t3 → t2 (E-IFTRUE)`

`if false then t2 else t3 → t3 (E-IFFALSE)`

Congruence rule:

$$\frac{t_1 \rightarrow t'_1}{\text{if } t_1 \text{ then } t_2 \text{ else } t_3 \rightarrow \text{if } t'_1 \text{ then } t_2 \text{ else } t_3} \text{ (E-IF)}$$

Computation rules perform “real” computation steps.

Congruence rules determine *where* computation rules can be applied next.

Evaluation, more explicitly

\longrightarrow is the smallest two-place relation closed under the following rules:

$$\begin{array}{l} ((\text{if true then } t_2 \text{ else } t_3), t_2) \in \longrightarrow \\ ((\text{if false then } t_2 \text{ else } t_3), t_3) \in \longrightarrow \\ \hline (t_1, t'_1) \in \longrightarrow \\ \hline ((\text{if } t_1 \text{ then } t_2 \text{ else } t_3), (\text{if } t'_1 \text{ then } t_2 \text{ else } t_3)) \in \longrightarrow \end{array}$$

The notation $t \longrightarrow t'$ is short-hand for $(t, t') \in \longrightarrow$.

An example

Let t be the term

`if true then (if false then false else true) else false`

Clicker question: What is t' in $t \rightarrow t'$?

- A. $t' = \text{true}$
- B. $t' = \text{if true then true else false}$
- C. $t' = \text{if false then false else true}$
- D. There is no such t'
- E. I don't know

URL: ttpoll.eu

Session ID: [cs452](#)

Reading for next week

- ▶ Chapter 3 – Untyped Arithmetic Expressions
- ▶ Some of it is recap; most important:
 - ▶ 3.3 Induction on terms
 - ▶ 3.5 Evaluation

Induction on Syntax

Questions

Recall: terms, concretely

Define an infinite sequence of sets, $\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \dots$, as follows:

$$\mathcal{S}_0 = \emptyset$$

$$\begin{aligned}\mathcal{S}_{i+1} = & \{\text{true}, \text{false}, 0\} \\ & \cup \{\text{succ } t_1, \text{pred } t_1, \text{iszero } t_1 \mid t_1 \in \mathcal{S}_i\} \\ & \cup \{\text{if } t_1 \text{ then } t_2 \text{ else } t_3 \mid t_1, t_2, t_3 \in \mathcal{S}_i\}\end{aligned}$$

Now let

$$\mathcal{S} = \bigcup_i \mathcal{S}_i$$

Questions

Last week, we explored the relationship between induction and *well-founded* sets and relations (W, \prec) .

If we consider induction on terms in that context, with $W = \mathcal{T}$, what could be the associated well-founded relation?

Clicker question: How can we define $t_1 \prec t_2$? (several possible answers)

- A. $t_1 < t_2$
- B. $\text{depth}(t_1) < \text{depth}(t_2)$
- C. t_1 is an immediate subterm of t_2
- D. $t_2 \longrightarrow t_1$

URL: ttpoll.eu

Session ID: [cs452](#)

Induction on Terms

Definition: The *depth* of a term t is the smallest i such that $t \in \mathcal{S}_i$.

From the definition of \mathcal{S} , it is clear that, if a term t is in \mathcal{S}_i , then all of its immediate subterms must be in \mathcal{S}_{i-1} , i.e., they must have strictly smaller depths.

This observation justifies the *principle of induction on terms*.
Let P be a predicate on terms.

*If, for each term s ,
 given $P(r)$ for all immediate subterms r of s
 we can show $P(s)$,
then $P(t)$ holds for all t .*

Inductive Function Definitions

The set of constants appearing in a term t , written $Consts(t)$, is defined as follows:

$$\begin{aligned} Consts(true) &= \{true\} \\ Consts(false) &= \{false\} \\ Consts(0) &= \{0\} \\ Consts(succ\ t_1) &= Consts(t_1) \\ Consts(pred\ t_1) &= Consts(t_1) \\ Consts(iszero\ t_1) &= Consts(t_1) \\ Consts(if\ t_1\ then\ t_2\ else\ t_3) &= Consts(t_1) \cup Consts(t_2) \\ &\quad \cup Consts(t_3) \end{aligned}$$

Simple, right?

First question:

Normally, a “definition” just assigns a convenient name to a previously-known thing. But here, the “thing” on the right-hand side involves the very name that we are “defining”!

So in what sense is this a definition??

Second question:

Suppose we had written this instead...

The set of constants appearing in a term t , written $BadConsts(t)$, is defined as follows:

$$\begin{aligned}BadConsts(true) &= \{true\}\\BadConsts(false) &= \{false\}\\BadConsts(0) &= \{0\}\\BadConsts(0) &= \{\}\\BadConsts(succ\ t_1) &= BadConsts(t_1)\\BadConsts(pred\ t_1) &= BadConsts(t_1)\\BadConsts(iszero\ t_1) &= BadConsts(iszero\ (iszero\ t_1))\end{aligned}$$

What is the essential difference between these two definitions?

How do we tell the difference between well-formed inductive definitions and ill-formed ones?

What, exactly, does a well-formed inductive definition mean?

What is a function?

Recall that a *function* f from A (its domain) to B (its co-domain) can be viewed as a two-place *relation* (called the “graph” of the function) with certain properties:

- ▶ It is *total*: Every element of its domain occurs at least once in its graph. More precisely:

For every $a \in A$, there exists some $b \in B$ such that $(a, b) \in f$.

- ▶ It is *deterministic*: every element of its domain occurs at most once in its graph. More precisely:

If $(a, b_1) \in f$ and $(a, b_2) \in f$, then $b_1 = b_2$.

We have seen how to define relations inductively. E.g....
Let *Consts* be the smallest two-place relation closed under the following rules:

$$(\text{true}, \{\text{true}\}) \in \text{Consts}$$

$$(\text{false}, \{\text{false}\}) \in \text{Consts}$$

$$(0, \{0\}) \in \text{Consts}$$

$$\frac{(t_1, C) \in \text{Consts}}{(\text{succ } t_1, C) \in \text{Consts}}$$

$$\frac{(t_1, C) \in \text{Consts}}{(\text{pred } t_1, C) \in \text{Consts}}$$

$$\frac{(t_1, C) \in \text{Consts}}{(\text{iszero } t_1, C) \in \text{Consts}}$$

$$\frac{(t_1, C_1) \in \text{Consts} \quad (t_2, C_2) \in \text{Consts} \quad (t_3, C_3) \in \text{Consts}}{(\text{if } t_1 \text{ then } t_2 \text{ else } t_3, C_1 \cup C_2 \cup C_3) \in \text{Consts}}$$

This definition certainly defines a *relation* (i.e., the smallest one with a certain closure property).

Q: How can we be sure that this relation is a *function*?

This definition certainly defines a *relation* (i.e., the smallest one with a certain closure property).

Q: How can we be sure that this relation is a *function*?

A: *Prove it!*

Theorem:

The relation *Consts* defined by the inference rules a couple of slides ago is total and deterministic.

I.e., for each term t there is exactly one set of terms C such that $(t, C) \in \textit{Consts}$.

Proof:

Theorem:

The relation *Consts* defined by the inference rules a couple of slides ago is total and deterministic.

I.e., for each term t there is exactly one set of terms C such that $(t, C) \in \textit{Consts}$.

Proof: By induction on t .

Theorem:

The relation *Consts* defined by the inference rules a couple of slides ago is total and deterministic.

I.e., for each term t there is exactly one set of terms C such that $(t, C) \in \text{Consts}$.

Proof: By induction on t .

To apply the induction principle for terms, we must show, for an arbitrary term t , that if

for each immediate subterm s of t , there is exactly one set of terms C_s such that $(s, C_s) \in \text{Consts}$

then

there is exactly one set of terms C such that $(t, C) \in \text{Consts}$.

Proceed by cases on the form of t .

- ▶ If t is 0 , $true$, or $false$, then we can immediately see from the definition of *Consts* that there is exactly one set of terms C (namely $\{t\}$) such that $(t, C) \in \text{Consts}$.

Proceed by cases on the form of t .

- ▶ If t is 0 , $true$, or $false$, then we can immediately see from the definition of *Consts* that there is exactly one set of terms C (namely $\{t\}$) such that $(t, C) \in \text{Consts}$.
- ▶ If t is $\text{succ } t_1$, then the induction hypothesis tells us that there is exactly one set of terms C_1 such that $(t_1, C_1) \in \text{Consts}$. But then it is clear from the definition of *Consts* that there is exactly one set C (namely C_1) such that $(t, C) \in \text{Consts}$.

Proceed by cases on the form of t .

- ▶ If t is `0`, `true`, or `false`, then we can immediately see from the definition of *Consts* that there is exactly one set of terms C (namely $\{t\}$) such that $(t, C) \in \text{Consts}$.
- ▶ If t is `succ` t_1 , then the induction hypothesis tells us that there is exactly one set of terms C_1 such that $(t_1, C_1) \in \text{Consts}$. But then it is clear from the definition of *Consts* that there is exactly one set C (namely C_1) such that $(t, C) \in \text{Consts}$.

Similarly when t is `pred` t_1 or `iszero` t_1 .

- ▶ If t is `if s_1 then s_2 else s_3` , then the induction hypothesis tells us
 - ▶ there is exactly one set of terms C_1 such that $(t_1, C_1) \in \text{Consts}$
 - ▶ there is exactly one set of terms C_2 such that $(t_2, C_2) \in \text{Consts}$
 - ▶ there is exactly one set of terms C_3 such that $(t_3, C_3) \in \text{Consts}$

But then it is clear from the definition of *Consts* that there is exactly one set C (namely $C_1 \cup C_2 \cup C_3$) such that $(t, C) \in \text{Consts}$.

How about the bad definition?

$$(\text{true}, \{\text{true}\}) \in \text{BadConsts}$$

$$(\text{false}, \{\text{false}\}) \in \text{BadConsts}$$

$$(0, \{0\}) \in \text{BadConsts}$$

$$(0, \{\}) \in \text{BadConsts}$$

$$(\text{t}_1, C) \in \text{BadConsts}$$

$$\frac{(\text{t}_1, C) \in \text{BadConsts}}{(\text{succ } \text{t}_1, C) \in \text{BadConsts}}$$

$$(\text{t}_1, C) \in \text{BadConsts}$$

$$\frac{(\text{t}_1, C) \in \text{BadConsts}}{(\text{pred } \text{t}_1, C) \in \text{BadConsts}}$$

$$(\text{iszero } (\text{iszero } \text{t}_1), C) \in \text{BadConsts}$$

$$\frac{(\text{iszero } (\text{iszero } \text{t}_1), C) \in \text{BadConsts}}{(\text{iszero } \text{t}_1, C) \in \text{BadConsts}}$$

This set of rules defines a perfectly good *relation* — it's just that this relation does not happen to be a function!

Just for fun, let's calculate some cases of this relation...

- ▶ For what values of C do we have $(\text{false}, C) \in \text{BadConsts}$?

This set of rules defines a perfectly good *relation* — it's just that this relation does not happen to be a function!

Just for fun, let's calculate some cases of this relation...

- ▶ For what values of C do we have $(\text{false}, C) \in \text{BadConsts}$?
- ▶ For what values of C do we have $(\text{succ } 0, C) \in \text{BadConsts}$?

This set of rules defines a perfectly good *relation* — it's just that this relation does not happen to be a function!

Just for fun, let's calculate some cases of this relation...

- ▶ For what values of C do we have $(\text{false}, C) \in \text{BadConsts}$?
- ▶ For what values of C do we have $(\text{succ } 0, C) \in \text{BadConsts}$?
- ▶ For what values of C do we have $(\text{if false then } 0 \text{ else } 0, C) \in \text{BadConsts}$?

This set of rules defines a perfectly good *relation* — it's just that this relation does not happen to be a function!

Just for fun, let's calculate some cases of this relation...

- ▶ For what values of C do we have $(\text{false}, C) \in \text{BadConsts}$?
- ▶ For what values of C do we have $(\text{succ } 0, C) \in \text{BadConsts}$?
- ▶ For what values of C do we have $(\text{if false then } 0 \text{ else } 0, C) \in \text{BadConsts}$?
- ▶ For what values of C do we have $(\text{iszero } 0, C) \in \text{BadConsts}$?

Another Inductive Definition

$size(true)$	$=$	1
$size(false)$	$=$	1
$size(0)$	$=$	1
$size(succ\ t_1)$	$=$	$size(t_1) + 1$
$size(pred\ t_1)$	$=$	$size(t_1) + 1$
$size(iszero\ t_1)$	$=$	$size(t_1) + 1$
$size(if\ t_1\ then\ t_2\ else\ t_3)$	$=$	$size(t_1) + size(t_2) + size(t_3) + 1$

Another proof by induction

Theorem: The number of distinct constants in a term is at most the size of the term. I.e., $|Consts(t)| \leq size(t)$.

Proof:

Another proof by induction

Theorem: The number of distinct constants in a term is at most the size of the term. I.e., $|Consts(t)| \leq size(t)$.

Proof: By induction on t .

Another proof by induction

Theorem: The number of distinct constants in a term is at most the size of the term. I.e., $|Consts(t)| \leq size(t)$.

Proof: By induction on t .

Assuming the desired property for immediate subterms of t , we must prove it for t itself.

Another proof by induction

Theorem: The number of distinct constants in a term is at most the size of the term. I.e., $|Consts(t)| \leq size(t)$.

Proof: By induction on t .

Assuming the desired property for immediate subterms of t , we must prove it for t itself.

There are “three” cases to consider:

Case: t is a constant

Immediate: $|Consts(t)| = |\{t\}| = 1 = size(t)$.

Another proof by induction

Theorem: The number of distinct constants in a term is at most the size of the term. I.e., $|Consts(t)| \leq size(t)$.

Proof: By induction on t .

Assuming the desired property for immediate subterms of t , we must prove it for t itself.

There are “three” cases to consider:

Case: t is a constant

Immediate: $|Consts(t)| = |\{t\}| = 1 = size(t)$.

Case: $t = succ\ t_1, pred\ t_1, \text{ or } iszero\ t_1$

By the induction hypothesis, $|Consts(t_1)| \leq size(t_1)$. We now calculate as follows:

$|Consts(t)| = |Consts(t_1)| \leq size(t_1) < size(t_1) + 1 = size(t)$.

Case: $t = \text{if } t_1 \text{ then } t_2 \text{ else } t_3$

By the induction hypothesis, $|Consts(t_1)| \leq size(t_1)$,
 $|Consts(t_2)| \leq size(t_2)$, and $|Consts(t_3)| \leq size(t_3)$. We now
calculate as follows:

$$\begin{aligned} |Consts(t)| &= |Consts(t_1) \cup Consts(t_2) \cup Consts(t_3)| \\ &\leq |Consts(t_1)| + |Consts(t_2)| + |Consts(t_3)| \\ &\leq size(t_1) + size(t_2) + size(t_3) \\ &< size(t_1) + size(t_2) + size(t_3) + 1 \\ &= size(t). \end{aligned}$$