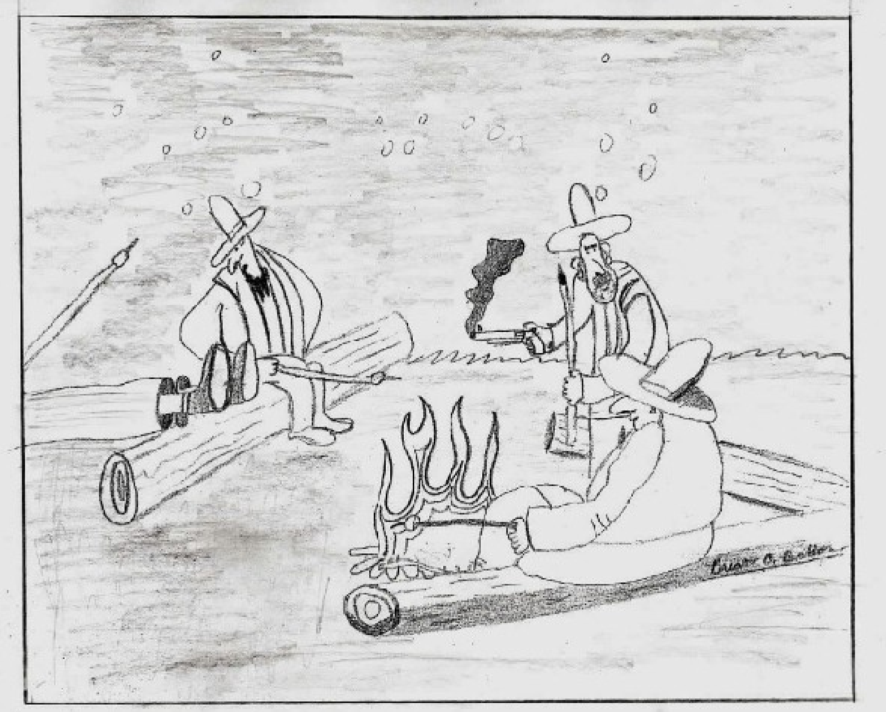
Simply-Typed Lambda Calculus



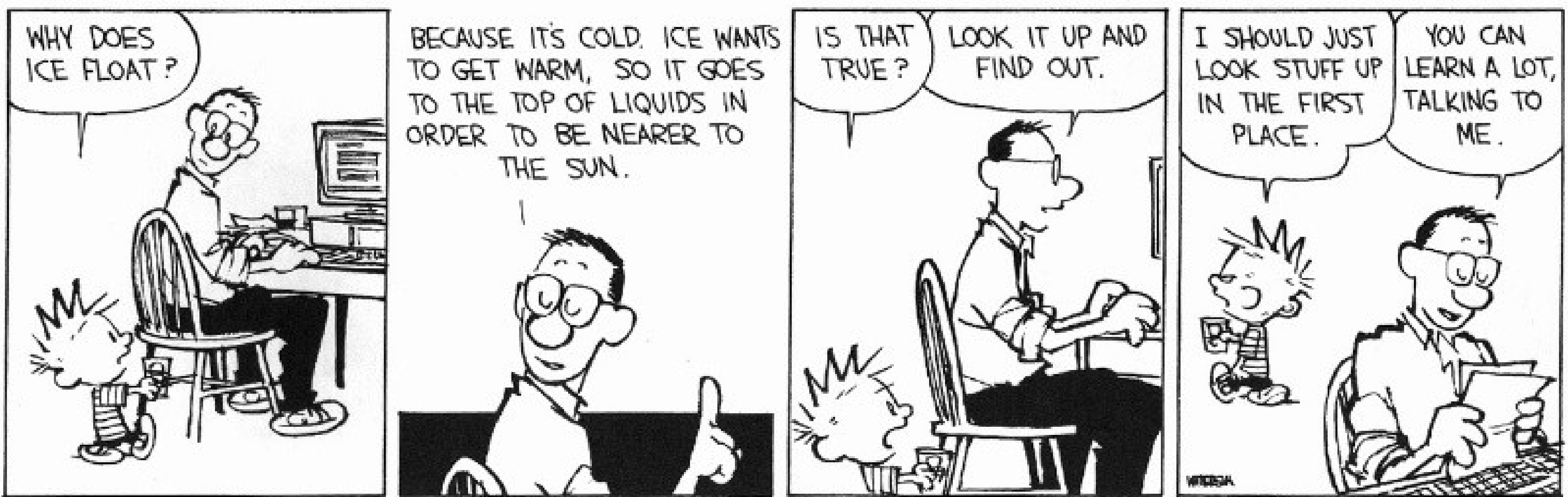
You guys are both my witnesses... He insinuated that ZFC set theory is superior to Type Theory!

One-Slide Summary

- A **type** is an upper bound on the range of values a program expression could take on at run-time.
- A formal **type system**, also known as a static semantics, describes rules for checking types.
- A **typing judgment** typically associates a **typing environment** and an expression with a type.
- The **simply-typed lambda calculus** adds type annotations for function abstractions.
- A type system is **sound** iff every expression evaluates to a value in that expression's static type.

Review!

- What is *operational semantics*? When would you use *contextual (small-step)* semantics?
- What is *satisfiability modulo theories*?
- What is *axiomatic semantics*? What is a *verification condition*?



Today's (Short?) Cunning Plan

- Type System Overview
- First-Order Type Systems
- Typing Rules
- Typing Derivations
- Type Safety

Types

- A program variable can assume a *range of values* during the execution of a program

- An upper bound of such a range is called a type of the variable
  - A variable of type “bool” is supposed to assume only boolean values
  - If x has type “bool” then the boolean expression “not(x)” has a sensible meaning during every run of the program

Typed and Untyped Languages

- Untyped languages
  - Do *not* restrict the range of values for a given variable
  - Operations might be applied to inappropriate arguments. The behavior in such cases might be unspecified
  - The pure λ-calculus is an extreme case of an untyped language (however, its behavior is completely specified)

- (Statically) Typed languages
  - Variables are assigned (non-trivial) types
  - A type system keeps track of types
  - Types might or might not appear in the program itself
  - Languages can be explicitly typed or implicitly typed

The Purpose Of Types

- The foremost purpose of types is *to prevent certain types of run-time execution errors*
- Traditional trapped execution errors
  - Cause the computation to stop immediately
  - And are thus well-specified behavior
  - Usually enforced by hardware
  - e.g., Division by zero, floating point op with a NaN
  - e.g., Dereferencing the address 0 (on most systems)
- Untrapped execution errors
  - Behavior is unspecified (depends on the state of the machine = this is very bad!)
  - e.g., accessing past the end of an array
  - e.g., jumping to an address in the data segment

Why Typed Languages?

- Development
  - *Type checking catches early many mistakes*
  - Reduced debugging time
  - Typed signatures are a powerful basis for design
  - Typed signatures enable separate compilation
- Maintenance
  - Types act as checked specifications
  - Types can enforce abstraction
- Execution
  - Static checking reduces the need for dynamic checking
  - Safe languages are easier to analyze statically
    - the compiler can generate better code

Properties of Type Systems

- How do types differ from other program annotations?
  - Types are more precise than comments
  - Types are more easily mechanizable than program specifications
- Expected properties of type systems:
  - Types should be enforceable
  - Types should be checkable algorithmically
  - Typing rules should be transparent
    - Should be easy to see why a program is not well-typed

Why Formal Type Systems?

- Many typed languages have informal descriptions of the type systems (e.g., in language reference manuals)
- A fair amount of careful analysis is required to avoid false claims of type safety
- A formal presentation of a type system is a precise specification of the type checker
  - And allows formal proofs of type safety
- But even informal knowledge of the principles of type systems help

Formalizing a Language

1. Syntax
   - Of expressions (programs)
   - Of types
   - Issues of binding and scoping
2. **Static semantics (typing rules)**
   - Define the typing judgment and its derivation rules
3. Dynamic Semantics (e.g., operational)
   - Define the evaluation judgment and its derivation rules
4. **Type soundness**
   - Relates the static and dynamic semantics
   - State and prove the soundness theorem

Typing Judgments

- Judgment (recall from prior lectures)
  - A statement J about certain formal entities
  - Has a truth value $\models$ J
  - Has a derivation $\vdash$ J (= "a proof")
- A common form of typing judgment:

  $\Gamma \vdash \mathbf{e} : \tau$ (e is an expression and τ is a type)
- Γ (Gamma) is a set of type assignments for the free variables of e
  - Defined by the grammar $\Gamma ::= \cdot \mid \Gamma, x : \tau$
  - Type assignments for variables not free in e are not relevant
  - e.g., **x : int, y : int $\vdash$ x + y : int**

Typing rules

- Typing rules are used to derive typing judgments
- Examples:

$$\frac{}{\Gamma \vdash 1 : \texttt{int}}$$

$$\frac{x : \tau \in \Gamma}{\Gamma \vdash x : \tau}$$

$$\frac{\Gamma \vdash e_1 : \texttt{int} \quad \Gamma \vdash e_2 : \texttt{int}}{\Gamma \vdash e_1 + e_2 : \texttt{int}}$$

Typing Derivations

- A typing derivation is a derivation of a typing judgment (big surprise there …)
- Example:

$$\dfrac{\dfrac{}{x : \texttt{int} \vdash x : \texttt{int}} \qquad \dfrac{\dfrac{}{x : \texttt{int} \vdash x : \texttt{int}} \quad \dfrac{}{x : \texttt{int} \vdash 1 : \texttt{int}}}{x : \texttt{int} \vdash x + 1 : \texttt{int}}}{x : \texttt{int} \vdash x + (x + 1) : \texttt{int}}$$

- We say $\Gamma \vdash e : \tau$ to mean there exists a derivation of this typing judgment (= "we can prove it")
- Type checking: given Γ, e and τ find a derivation
- Type inference: given Γ and e, find τ and a derivation

Proving Type Soundness

- A typing judgment is either true or false
- Define what it means for a value to have a type

$$v \in \| \tau \|$$

(e.g. $5 \in \| \text{int} \|$ and $\text{true} \in \| \text{bool} \|$)

- Define what it means for an expression to have a type

$$e \in | \tau | \quad \text{iff} \quad \forall v. (e \Downarrow v \Rightarrow v \in \| \tau \|)$$

- Prove type soundness

$$\text{If } \cdot \vdash e : \tau \qquad \text{then } e \in | \tau |$$

or equivalently

$$\text{If } \cdot \vdash e : \tau \text{ and } e \Downarrow v \qquad \text{then } v \in \| \tau \|$$

- This implies safe execution (since the result of a unsafe execution is not in $\| \tau \|$ for any τ)

Upcoming Exciting Episodes

- We will give formal description of first-order type systems (no type variables)
  - Function types (simply typed λ-calculus)
  - Simple types (integers and booleans)
  - Structured types (products and sums)
  - Imperative types (references and exceptions)
  - Recursive types (linked lists and trees)
- The type systems of most common languages are first-order
- Then we move to second-order type systems
  - Polymorphism and abstract types

Q: Movies (378 / 842)

- This 1988 animated movie written and directed by Isao Takahata for Studio Ghibli was considered by Roger Ebert to be one of the most powerful anti-war films ever made. It features Seita and his sister Setsuko and their efforts to survive outside of society during the firebombing of Tokyo.

Computer Science

- This American-Canadian Turing-award winner is known for major contributions to the fields of complexity theory and proof complexity. He is known for formalizing the polynomial-time reduction, NP-completeness, P vs. NP, and showing that SAT is NP-complete. This was all done in the seminal 1971 paper *The Complexity of Theorem Proving Procedures*.

Q: Student

- This piece of diving equipment with an air-inflatable bladder changes its average density for use in SCUBA diving. It typically requires manual adjustment throughout the dive and can be augmented by breath control.

Q: Games (504 / 842)

- This 1985 falling-blocks computer game was invented by Alexey Pajitnov (Алексей Пажитнов) and inspired by pentominoes.

Simply-Typed Lambda Calculus

- Syntax:

Notice $:\tau$

Terms	$e ::=$	x	$\mid \lambda x{:}\tau.\ e$	$\mid e_1\ e_2$
		$\mid n$	$\mid e_1 + e_2$	$\mid$ iszero e
		$\mid$ true	$\mid$ false	$\mid$ not e
		$\mid$ if e_1 then e_2 else e_3		
Types	$\tau ::=$ int $\mid$ bool $\mid \tau_1 \to \tau_2$			

- $\tau_1 \to \tau_2$ is the function type
- $\to$ associates to the right
- Arguments have typing annotations $:\tau$
- This language is also called F_1

Static Semantics of F_1

- The typing judgment

$$\Gamma \vdash e : \tau$$

- Some (simpler) typing rules:

$$\frac{x : \tau \in \Gamma}{\Gamma \vdash x : \tau} \qquad \frac{\Gamma, x : \tau \vdash e : \tau'}{\Gamma \vdash \lambda x : \tau . e : \tau \to \tau'}$$

$$\frac{\Gamma \vdash e_1 : \tau_2 \to \tau \qquad \Gamma \vdash e_2 : \tau_2}{\Gamma \vdash e_1 e_2 : \tau}$$

More Static Semantics of F_1

$$\frac{}{\Gamma \vdash n : \texttt{int}} \qquad \frac{\Gamma \vdash e_1 : \texttt{int} \qquad \Gamma \vdash e_2 : \texttt{int}}{\Gamma \vdash e_1 + e_2 : \texttt{int}}$$

Why did I leave this mysterious gap? I don't know either!

$$\frac{}{\Gamma \vdash \texttt{true} : \texttt{bool}} \qquad \frac{\Gamma \vdash e : \texttt{bool}}{\Gamma \vdash \texttt{not}\ e : \texttt{bool}}$$

$$\frac{\Gamma \vdash e_1 : \texttt{bool} \qquad \Gamma \vdash e_t : \tau \qquad \Gamma \vdash e_f : \tau}{\Gamma \vdash \texttt{if}\ e_1\ \texttt{then}\ e_t\ \texttt{else}\ e_f : \tau}$$

Typing Derivation in F_1

- Consider the term (also underlined below)

  λx : int. λb : bool. if b then f x else x

  - With the initial typing assignment f : int → Int
  - Where Γ = f : int → int, x : int, b : bool

$$
\dfrac{\dfrac{\dfrac{\Gamma \vdash b : \texttt{bool} \qquad \dfrac{\Gamma \vdash f : \texttt{int} \to \texttt{int} \qquad \Gamma \vdash x : \texttt{int}}{\Gamma \vdash f\ x : \texttt{int}} \qquad \Gamma \vdash x : \texttt{int}}{f : \texttt{int} \to \texttt{int}, x : \texttt{int}, b : \texttt{bool} \vdash \texttt{if}\ b\ \texttt{then}\ f\ x\ \texttt{else}\ x : \texttt{int}}}{f : \texttt{int} \to \texttt{int}, x : \texttt{int} \vdash \lambda b : \texttt{bool}.\ \texttt{if}\ b\ \texttt{then}\ f\ x\ \texttt{else}\ x : \texttt{bool} \to \texttt{int}}}{f : \texttt{int} \to \texttt{int} \vdash \underbrace{\lambda x : \texttt{int}. \lambda b : \texttt{bool}.\ \texttt{if}\ b\ \texttt{then}\ f\ x\ \texttt{else}\ x} : \texttt{int} \to \texttt{bool} \to \texttt{int}}
$$

Type Checking in F_1

- Type checking is *easy* because
  - Typing rules are syntax directed
  - Typing rules are compositional (what does this mean?)
  - All local variables are annotated with types
  - "Easy" = deterministic polynomial time, etc.
- In fact, type inference is *also easy* for F_1
- Without type annotations an expression may have no unique type

$$\cdot \vdash \lambda x.\ x : \text{int} \to \text{int}$$

$$\cdot \vdash \lambda x.\ x : \text{bool} \to \text{bool}$$

Operational Semantics of F_1

- Judgment:

$$e \Downarrow v$$

- Values:

$$v ::= n \mid \text{true} \mid \text{false} \mid \lambda x{:}\tau.\ e$$

- The evaluation rules …
  - Audience participation time: "raise your hand" and give me an opsem evaluation rule.
    - Function application, Lambda abstraction, if, plus, …

Opsem of F_1 (Cont.)

- Call-by-value evaluation rules (sample)

$$\frac{}{\lambda x : \tau . e \Downarrow \lambda x : \tau . e}$$

$$\frac{e_1 \Downarrow \lambda x : \tau . e_1' \qquad e_2 \Downarrow v_2 \qquad [v_2/x]e_1' \Downarrow v}{e_1\ e_2 \Downarrow v}$$

Where is the Call-By-Value? How might we change it?

$$\frac{}{n \Downarrow n} \qquad \frac{e_1 \Downarrow n_1 \qquad e_2 \Downarrow n_2 \qquad n = n_1 + n_2}{e_1 + e_2 \Downarrow n}$$

$$\frac{e_1 \Downarrow \texttt{true} \qquad e_t \Downarrow v}{\texttt{if } e_1 \texttt{ then } e_t \texttt{ else } e_f \Downarrow v}$$

$$\frac{e_1 \Downarrow \texttt{false} \qquad e_f \Downarrow v}{\texttt{if } e_1 \texttt{ then } e_t \texttt{ else } e_f \Downarrow v}$$

Evaluation is undefined for ill-typed programs !

Type Soundness for F_1

- Thm: **If $\cdot \vdash e : \tau$ and $e \Downarrow v$ then $\cdot \vdash v : \tau$**
  - Also called, subject reduction theorem, type preservation theorem
- This is one of the most important sorts of theorems in PL
- Whenever you make up a new safe language you are expected to prove this
  - Examples: Vault, TAL, CCured, …
- Proof: next time!

Practice Paper Reading: SIGPLAN PEPM 2025

- A Type Safe Calculus for Generating Syntax-Directed Editors

The typed lambda calculi that we consider all satisfy the same subject reduction property.

Theorem 3.1 (Subject reduction). *If* $\Gamma \vdash e : \tau$ *and* $e \rightarrow e'$, *then* $\Gamma \vdash e' : \tau$.

$$(\textsc{Seq}) \quad \frac{\langle E_1, a\rangle \overset{\alpha}{\Rightarrow} \langle E_1', a'\rangle}{\langle E_1 \ggg E_2, a\rangle \overset{\alpha}{\Rightarrow} \langle E_1' \ggg E_2, a'\rangle}$$

$$(\textsc{Seq-Trivial}) \quad \frac{}{\langle \mathsf{nil} \ggg E_2, a\rangle \overset{\epsilon}{\Rightarrow} \langle E_2, a\rangle}$$

Practice Paper Reading: POPL 2025

- Consistency of a Dependent Calculus of Indistinguishability

4.2 Type Preservation

Like level checking, type checking admits substitution (Wt-subst) and subsumption (Wt-sub), as stated in Figure 11. Due to rule Wt-Conv, deducing the well-typedness of subterms of a well-typed term doesn't follow immediately from inversion and requires separate generation lemmas, which are standard and omitted here. One notable consequence is that if **refl** is a propositional equality between a and b at some observer level ℓ_0, then they are definitionally equal at that level. Finally, we are able to prove type preservation, which proceeds by induction on the typing derivation.

Lemma 4.4. *If* $\Gamma \vdash \mathbf{refl} :^{\ell} a =^{\ell_0} b$*, then* $|\Gamma| \vdash a \Leftrightarrow^{\ell_0} b$.

Theorem 4.5 (Type preservation). *If* $a \Rightarrow b$ *(or* $a \Rightarrow^* b$*) and* $\Gamma \vdash a :^{\ell} A$ *then* $\Gamma \vdash b :^{\ell} A$.

The main judgment is the typing relation, which has the form $\boxed{\Gamma \vdash a :^{\ell} A}$ for a term a that is well typed under context Γ at observer level ℓ with type A. Its rules are given in Figure 2. To

Homework

- Read actually-exciting Leroy paper
- Homework continues
- HW4 Discussion