Computer-Supported Modeling and Reasoning

Jan-Georg Smaus

WS06/07

How to Use Screen Notes

These screen notes are generated from sources that were originally intended for hypermedia, as lecture slides or online course. Frequently, the slides contain highlighted terms that come with an annotation, i.e., a more detailed explanation that would usually be given by the lecturer during the lecture. When one looks at the slides on the screen, one can click on such a term and will be linked to the annotation. However, there is a danger that one gets lost. In the present rendering, Screen Notes, the annotations are realised as footnotes. Thus the thread of the lecture can be followed without any jumping within the document, while forward and backward references are still realised as hyperlinks. Screen Notes are not suitable for being printed! For printing use Lecture Notes.

1 General Introduction

What this Course is about

- Mechanizing and using logic

  – program verification[1]:
    input: theories, programs, proper-
    ties
    output: correctness guarantees
  – Hilbert's program[2]:
    input: arithmetic
    output: theorems

- Technically: mechanization and appli-
  cation[3]

- Generally: making logic come to life
  and useful as a general tool.

---

[1]Verification is the process of formally proving that a program has the desired properties. To this end, it is necessary to define a specification language in which the desired properties can be formulated, i.e. specified. One must define a semantics for this language as well as for the program. These semantics must be linked in such a way that it is meaningful to say: "Program X makes formula Φ true".

[2]In the 1920's, David Hilbert attempted a single rigorous formalization of all of mathematics, named Hilbert's program. He was concerned with the following three questions:

1. Is mathematics complete in the sense that every statement can be proved or disproved?
2. Is mathematics consistent in the sense that no statement can be proved both true and false?
3. Is mathematics decidable in the sense that there exists a definite method to determine the truth or falsity of any mathematical statement?

Hilbert believed that the answer to all three questions was 'yes'.

Thanks to the the incompleteness theorem of Gödel (1931) and the undecidability of first-order logic shown by Church and Turing (1936–37) we know now that his dream will never

Overview: Four Parts

1. Logics[4] (propositional, first-order, higher-order): appr. 6 units

---

[3]We will learn to make logic run on a computer by using the Isabelle system.

[4]The word logic is used in a wider and a narrower sense.

In a wider sense, logic is the science of reasoning. In fact, it is the science that reasons about reasoning itself.

In a narrower sense, a logic is just a precisely defined language allowing to write down statements, together with a predefined meaning for some of the syntactic entities of this language. Propositional logic, first-order logic, and higher-order logic are three different logics.

[5]A metalogic is a logic that allows us to express properties of another logic.

[6]Intuitively, whenever you do computer-supported modeling and reasoning, you have to formalize a tiny portion of the "world", the portion that your problem lives in. For example, rational numbers may or may not exist in this portion. A theory is such a formalization of a tiny portion of the "world". A theory extends a logic by axioms that describe that portion

Overview: Four Parts

1. Logics[4] (propositional, first-order, higher-order): appr. 6 units

2. Metalogics[5] (Isabelle): appr. 2 units

---

[3]We will learn to make logic run on a computer by using the Isabelle system.

[4]The word logic is used in a wider and a narrower sense.

In a wider sense, logic is the science of reasoning. In fact, it is the science that reasons about reasoning itself.

In a narrower sense, a logic is just a precisely defined language allowing to write down statements, together with a predefined meaning for some of the syntactic entities of this language. Propositional logic, first-order logic, and higher-order logic are three different logics.

[5]A metalogic is a logic that allows us to express properties of another logic.

[6]Intuitively, whenever you do computer-supported modeling and reasoning, you have to formalize a tiny portion of the "world", the portion that your problem lives in. For example, rational numbers may or may not exist in this portion. A theory is such a formalization of a tiny portion of the "world". A theory extends a logic by axioms that describe that portion

Overview: Four Parts

1. Logics[4] (propositional, first-order, higher-order): appr. 6 units

2. Metalogics[5] (Isabelle): appr. 2 units

3. Modeling mathematics and computer science (programming languages) in higher-order logic: appr. 6 units

---

[3]We will learn to make logic run on a computer by using the Isabelle system.

[4]The word logic is used in a wider and a narrower sense.

In a wider sense, logic is the science of reasoning. In fact, it is the science that reasons about reasoning itself.

In a narrower sense, a logic is just a precisely defined language allowing to write down statements, together with a predefined meaning for some of the syntactic entities of this language. Propositional logic, first-order logic, and higher-order logic are three different logics.

[5]A metalogic is a logic that allows us to express properties of another logic.

[6]Intuitively, whenever you do computer-supported modeling and reasoning, you have to formalize a tiny portion of the "world", the portion that your problem lives in. For example, rational numbers may or may not exist in this portion. A theory is such a formalization of a tiny portion of the "world". A theory extends a logic by axioms that describe that portion

Overview: Four Parts

1. Logics[4] (propositional, first-order, higher-order): appr. 6 units

2. Metalogics[5] (Isabelle): appr. 2 units

3. Modeling mathematics and computer science (programming languages) in higher-order logic: appr. 6 units

4. Some case study in formalizing a theory[6] (functional or imperative programming, or the specification language Z): appr. 2 units

Prsentation roughly follows this structure.

---

[3]We will learn to make logic run on a computer by using the Isabelle system.

[4]The word logic is used in a wider and a narrower sense.

In a wider sense, logic is the science of reasoning. In fact, it is the science that reasons about reasoning itself.

In a narrower sense, a logic is just a precisely defined language allowing to write down statements, together with a predefined meaning for some of the syntactic entities of this language. Propositional logic, first-order logic, and higher-order logic are three different logics.

[5]A metalogic is a logic that allows us to express properties of another logic.

[6]Intuitively, whenever you do computer-supported modeling and reasoning, you have to formalize a tiny portion of the "world", the portion that your problem lives in. For example, rational numbers may or may not exist in this portion. A theory is such a formalization of a tiny portion of the "world". A theory extends a logic by axioms that describe that portion

Why this Course Matters

Academic motivation: deepen knowledge of logic and formal reasoning

___

of the "world".

Theories will be considered in more detail later.

Why this Course Matters

Academic motivation: deepen knowledge of logic and formal reasoning

Practical motivation: verification and formal methods

- The last decade has seen spectacular hardware and software failures and the birth of a new discipline: the verification engineer

- Exciting positions at companies like Intel, Gemplus, . . .

---

of the "world".

Theories will be considered in more detail later.

Why this Course Matters (2)

In general:

- Understanding formal reasoning improves understanding of how to build correct systems

- Mechanization provides formal guarantees

Want to see some Isabelle/HOL applications?

Relationship to other Courses

Logic: deduction, foundations, and applications

Software engineering: specification, refinement, verification

Hardware: formalizing and reasoning about circuit models

Artificial Intelligence: knowledge representation, reasoning, deduction

In general, you will develop a deeper understanding of mathematical and logical reasoning, which is central to computer science.

Requirements

Some knowledge of logic[7] is useful for this course.

---

[7]We will introduce different logics and formal systems (so-called calculi) used to deduce formulas in a logic. We will neglect other aspects that are usually treated in classes or textbooks on logic, e.g.:

- semantics (interpretations) of logics; and

- correctness and completeness of calculi.

As an introduction we recommend [vD80].

Requirements

Some knowledge of logic[7] is useful for this course.

We will try to accommodate different backgrounds, e.g. with pointers to additional material. Your feedback is essential!

You must be willing to participate in the labs and get your hands dirty using a proof development system:

- further develop course material

- present material on pragmatics of mechanized reasoning

- hands-on experience.

---

[7]We will introduce different logics and formal systems (so-called calculi) used to deduce formulas in a logic. We will neglect other aspects that are usually treated in classes or textbooks on logic, e.g.:

- semantics (interpretations) of logics; and

- correctness and completeness of calculi.

As an introduction we recommend [vD80].

Requirements

Some knowledge of logic[7] is useful for this course.

We will try to accommodate different backgrounds, e.g. with pointers to additional material. Your feedback is essential!

You must be willing to participate in the labs and get your hands dirty using a proof development system:

- further develop course material

- present material on pragmatics of mechanized reasoning

- hands-on experience.

Experience shows that it makes no sense to follow just a little bit. It is hard in the beginning but the rewards are large.

---

[7]We will introduce different logics and formal systems (so-called calculi) used to deduce formulas in a logic. We will neglect other aspects that are usually treated in classes or textbooks on logic, e.g.:

- semantics (interpretations) of logics; and

- correctness and completeness of calculi.

As an introduction we recommend [vD80].

What's Happening in Freiburg?

Harald Hiss, Stefan Maus, and Stefan Wölfl work with Isabelle here at Freiburg:

- There is a trend to use XML (a generalization of HTML) for database applications. However, this gives rise to possible inconsistencies. Harald uses Isabelle to prove formally that such inconsistencies cannot occur.

- There are various formal theories that allow to reason about about the relationship of objects in space and time. Stefan uses Isabelle for proving consequences of such theories, dependencies between theories etc.

Also, David Basin occasionally seeks PhD students.

2 Propositional Logic

2.1 Propositional Logic: Overview

- System for formalizing certain valid patterns of reasoning

- Expressions built by combining "atomic propositions" using not, if...then..., and, or, etc.

- Validity[8] means: no counterexample. Validity independent of content. Depends on form of the expressions $\Rightarrow$ can make patterns explicit by replacing words by symbols

  From if A then B and A it follows that B.

2 Propositional Logic

2.1 Propositional Logic: Overview

- System for formalizing certain valid patterns of reasoning

- Expressions built by combining "atomic propositions" using not, if...then..., and, or, etc.

- Validity[8] means: no counterexample. Validity independent of content. Depends on form of the expressions $\Rightarrow$ can make patterns explicit by replacing words by symbols

$$\frac{A \to B \quad A}{B}$$

2 Propositional Logic

2.1 Propositional Logic: Overview

- System for formalizing certain valid patterns of reasoning

- Expressions built by combining "atomic propositions" using not, if...then..., and, or, etc.

- Validity[8] means: no counterexample. Validity independent of content. Depends on form of the expressions $\Rightarrow$ can make patterns explicit by replacing words by symbols
$$\frac{A \to B \quad A}{B}$$

- What about[9]

    From if A then B and B it follows that A?

---

[8]A and B are symbols whose meaning is not "hard-wired" into propositional logic.

    From if A then B and A it follows that B

is valid because it is true regardless of what A and B "mean", and in particular, regardless of whether A and B stand for true or false propositions.
[9]

    From if A then B and B it follows that A

is invalid because there is a counterexample:
Let A be "Kim is a man" and B be "Kim is a person".

More Examples

1. If it is Sunday, then I don't need to work.
   It is Sunday.
   Therefore I don't need to work.

2. It will rain or snow.
   It will not snow.
   Therefore it will rain.

3. The Butler is guilty or the Maid is guilty.
   The Maid is guilty or the Cook is guilty.
   Therefore either the Butler is guilty or the Cook is guilty.

---

1. If it is Sunday, then I don't need to work.
   It is Sunday.
   Therefore I don't need to work. VALID

2. It will rain or snow.
   It is too warm for snow.
   Therefore it will rain. VALID

3. The Butler is guilty or the Maid is guilty.
   The Maid is guilty or the Cook is guilty.
   Therefore either the Butler is guilty or the Cook is guilty.
   NOT VALID

More Examples (Which are Valid?)[10]

1. If it is Sunday, then I don't need to work.
   It is Sunday.
   Therefore I don't need to work.

2. It will rain or snow.
   It will not snow.
   Therefore it will rain.

3. The Butler is guilty or the Maid is guilty.
   The Maid is guilty or the Cook is guilty.
   Therefore either the Butler is guilty or the Cook is guilty.

---
[10]

1. If it is Sunday, then I don't need to work.
   It is Sunday.
   Therefore I don't need to work. VALID

2. It will rain or snow.
   It is too warm for snow.
   Therefore it will rain. VALID

3. The Butler is guilty or the Maid is guilty.
   The Maid is guilty or the Cook is guilty.
   Therefore either the Butler is guilty or the Cook is guilty.
   NOT VALID

History

- Propositional logic was developed to make this all precise.

- Laws for valid reasoning were known to the Stoic philosophers (about 300 BC).

- The formal system is often attributed to George Boole (1815-1864).

Further reading: [vD80], [Tho91, chapter 1].

More Formal Examples

Formalization allows us to "turn the crank"[11].

---

[11]By formalizing patterns of reasoning, we make it possible for such reasoning to be checked or even carried out by a computer.

From known patterns of reasoning new patterns of reasoning can be constructed.

[12]At this stage, we are content with a formalization that builds on geometrical notions like "above" or "to the right of". In other words, our formalization consists of geometrical objects like trees.

We study formalization in more detail later.

[13]A proof system or deductive system is characterized by a particular set of rules plus the general principles of how rules are grafted together to trees in natural deduction. We will see this shortly, but note that natural deduction is just one style of proof systems.

We call the rules in that particular set basic rules. Later we will see one can also derive rules.

More Formal Examples

Formalization allows us to "turn the crank"[11].

Phrases like "from . . . it follows" or "therefore" are formalized[12] as derivation rules, e.g.

$$\frac{A \to B \quad A}{B} \to\text{-}E$$

---

[11]By formalizing patterns of reasoning, we make it possible for such reasoning to be checked or even carried out by a computer.

From known patterns of reasoning new patterns of reasoning can be constructed.

[12]At this stage, we are content with a formalization that builds on geometrical notions like "above" or "to the right of". In other words, our formalization consists of geometrical objects like trees.

We study formalization in more detail later.

[13]A proof system or deductive system is characterized by a particular set of rules plus the general principles of how rules are grafted together to trees in natural deduction. We will see this shortly, but note that natural deduction is just one style of proof systems.

We call the rules in that particular set basic rules. Later we will see one can also derive rules.

More Formal Examples

Formalization allows us to "turn the crank"[11].

Phrases like "from . . . it follows" or "therefore" are formalized[12] as derivation rules, e.g.

$$\frac{A \to B \quad A}{B} \ \to\text{-}E$$

Rules are grafted together to build trees called derivations.

This defines a proof system[13] in the style of natural deduction.

---

[11]By formalizing patterns of reasoning, we make it possible for such reasoning to be checked or even carried out by a computer.

From known patterns of reasoning new patterns of reasoning can be constructed.

[12]At this stage, we are content with a formalization that builds on geometrical notions like "above" or "to the right of". In other words, our formalization consists of geometrical objects like trees.

We study formalization in more detail later.

[13]A proof system or deductive system is characterized by a particular set of rules plus the general principles of how rules are grafted together to trees in natural deduction. We will see this shortly, but note that natural deduction is just one style of proof systems.

We call the rules in that particular set basic rules. Later we will see one can also derive rules.

2.2 Formalizing Propositional Logic

- We must formalize
    1. Language[14] and semantics
    2. Deductive system

2.2 Formalizing Propositional Logic

- We must formalize
    1. Language[14] and semantics
    2. Deductive system

- Here we will focus on formalizing the deductive machinery and say little about metatheorems[15] (soundness and completeness[16]).

2.2 Formalizing Propositional Logic

- We must formalize

    1. Language[14] and semantics
    2. Deductive system

- Here we will focus on formalizing the deductive machinery and say little about metatheorems[15] (soundness and completeness[16]).

- For labs we will carry out proofs using the Isabelle System. Isabelle supports a Natural Deduction deductive system.

---

[14] By language we mean the language of formulae. We can also say that we define the (object) logic. Here "logic" is used in the narrower sense.

[15] A metatheorem is a theorem about a proof system, as opposed to a theorem derived within the proof system. The statement "proof system XYZ is sound" is a metatheorem.

[16] A proof system is sound if only valid propositions can be derived in it.

A proof system is complete if all valid propositions can be derived in it.

2.3 Propositional Logic: Language and Semantics

Propositions are built from a collection of (propositional) variables[17] and closed under disjunction, conjunction, implication, ...

---

[17]In mathematics, logic and computer science, there are various notions of variable. In propositional logic, a variable stands for a proposition, i.e., a variable can be interpreted as $True$ or $False$.

This will be different in logics that we will learn about later.

Propositional Logic: Language (2)

More formally: Let a set V of variables be given. L_P, the[18]
language of propositional logic, is the smallest set[19] where:

- X in L_P if X in V.

---

[18]Strictly speaking, the definition of L_P depends on V. A
different choice of variables leads to a different language of
propositional logic, and so we should not speak of the lan-
guage of propositional logic, but rather of a language of propo-
sitional logic. However, for propositional logic, one usually
does not care much about the names of the variables, or about
the fact that their number could be insufficient to write down
a certain formula of interest. We usually assume that there
are countably infinitely many variables.

Later, we will be more fussy about this point.

[19]The language of propositional logic is a set of formulae,
defined by induction. Note the following points about the
definition, which are important characteristics of any inductive
definition:

- By the second item in the definition, L_P is non-empty
  (also, one would usually have that V is non-empty, since
  otherwise L_P is not very interesting);

- L_P is required to be the smallest set meeting the above

- $\perp$ [20] in L_p.

- $(A \wedge$ [21] $B)$ in L_P if A in L_P and B in L_P.

- $(A \vee B)$ in L_P if A in L_P and B in L_P.

- $(A \to B)$ in L_P if A in L_P and B in L_P.

  conditions. Otherwise, anything (a number, a dog, the pope) could be a propositional formula.

- All conditions (or rules) defining L_P have the form: if ψ_1 and ... and ψ_n are in L_P, then some formula built from ψ_1 and ... and ψ_n is in L_P.

  It is crucial that no negation is involved here. If for example, there was a rule stating: if A is in L_P then A is not in L_P, then there could be no L_P fulfilling such a rule.

More detail on inductive definitions can be found in an article by Aczel [Acz77].

[20]

The symbol $\perp$ stands for "false".

[21]The connectives are called conjunction ($\wedge$), disjunction ($\vee$), implication ($\to$) and negation ($\neg$).

The connectives $\wedge, \vee, \to$ are binary since they connect two formulas, the connective $\neg$ is unary (most of the time, one

- $((\neg A)^{22}$ in L_P if A in L_P.)

The elements of L_P are called (propositional) formulas[24].

   We omit unnecessary brackets[25].

---

only uses the word connective for binary connective).

   [22]"Officially", negation does not exist in our language and proof system. Negation is only used as a shorthand, or syntactic sugar[23], for reasons of convenience. In paper-and-pencil proofs, we are allowed to erase any occurrence of $\neg P$ and replace it with $P \rightarrow \bot$, or vice versa, at any time. However, we shall see that when proofs are automated, this process must be made explicit.

   [24]In logic, the word "formula" has a specific meaning. Formulae are a syntactic category, namely the expressions that stand for a statement. So formulas are syntactic expressions that are interpreted (on the semantic level) as $True$ or $False$.

   We will later learn about another syntactic category, that of terms.

   [25]To save brackets, we use standard associativity and precedences. All binary connectives are right-associative:

$$A \circ B \circ C \equiv A \circ (B \circ C)$$

The precedences are $\neg$ before $\wedge$ before $\vee$ before $\rightarrow$. So for

Propositional Logic: Semantics

An assignment is a function $\mathcal{A} : V \to \{0, 1\}$. We say that $\mathcal{A}$ assigns a truth value to each propositional variable. We identify 1 with $True$ and 0 with $False$.

$\mathcal{A}$ is lifted (=extended) to formulas in L_P as follows . . .

---

example

$$A \to B \wedge \neg C \vee D \equiv A \to ((B \wedge (\neg C)) \vee D)$$

Propositional Logic: Semantics (2)

$$\mathcal{A}(\bot) = 0$$

$$\mathcal{A}(\neg\phi) = \begin{cases} 1 & \text{if } \mathcal{A}(\phi) = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{A}(\phi \wedge \psi) = \begin{cases} 1 & \text{if } \mathcal{A}(\phi) = 1 \text{ and } \mathcal{A}(\psi) = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{A}(\phi \vee \psi) = \begin{cases} 1 & \text{if } \mathcal{A}(\phi) = 1 \text{ or } \mathcal{A}(\psi) = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{A}(\phi \rightarrow \psi) = \begin{cases} 1 & \text{if } \mathcal{A}(\phi) = 0 \text{ or}^{26} \mathcal{A}(\psi) = 1 \\ 0 & \text{otherwise} \end{cases}$$

Propositional Logic: Semantics (3)

If $\mathcal{A}(\phi) = 1$, we write $\mathcal{A} \models \phi$.

Two formulae are equivalent if they yield the same truth value for any assignment of the propositional variables.

The semantics will be generalised later.

2.4 Deductive System: Natural Deduction

Developed by Gentzen [Gen35] and Prawitz [Pra65].
  Designed to support 'natural' logical arguments:

- we make (temporary) assumptions;

- we derive new formulas by applying rules;

- there is also a mechanism for "getting rid of" assumptions.

Natural Deduction (2)

Derivations are trees

$$\frac{\dfrac{A \to (B \to C) \quad A}{B \to C} \ \to\text{-}E \quad B}{C} \ \to\text{-}E$$

where the leaves are called assumptions.

---

[27]For the moment, the way to understand it is as follows: by writing $A \to (B \to C), A, B \vdash C$, we assert that C can be derived in this proof system under the assumptions $A \to (B \to C), A, B$.

We will say more about the $\vdash$ notation later.

Natural Deduction (2)

Derivations are trees

$$\frac{\dfrac{A \to (B \to C) \quad A}{B \to C} \to\text{-}E \quad B}{C} \to\text{-}E$$

where the leaves are called assumptions.

Write $A_1, ..., A_n \vdash A$ if there exists a derivation of A with assumptions $A_1, ..., A_n$, e.g. $A \to (B \to C), A, B \vdash C$[27].

---

[27]For the moment, the way to understand it is as follows: by writing $A \to (B \to C), A, B \vdash C$, we assert that C can be derived in this proof system under the assumptions $A \to (B \to C), A, B$.

We will say more about the $\vdash$ notation later.

Natural Deduction (2)

Derivations are trees

$$\frac{\dfrac{A \to (B \to C) \quad A}{B \to C} \to\text{-}E \quad B}{C} \to\text{-}E$$

where the leaves are called assumptions.

Write $A_1, ..., A_n \vdash A$ if there exists a derivation of A with assumptions $A_1, ..., A_n$, e.g. $A \to (B \to C), A, B \vdash C$[27].

A proof is a derivation where we "got rid" of all assumptions.

---

[27]For the moment, the way to understand it is as follows: by writing $A \to (B \to C), A, B \vdash C$, we assert that C can be derived in this proof system under the assumptions $A \to (B \to C), A, B$.

We will say more about the $\vdash$ notation later.

Natural Deduction: an Abstract Example[28]

- Language $\mathcal{L} = \{\heartsuit, \clubsuit, \spadesuit, \diamondsuit\}$.

---

[28]Natural deduction is not just about propositional logic! We explain here the general principles of natural deduction, not just the application to propositional logic.

In order to emphasize that applying natural deduction is a completely mechanical process, we give an example that is void of any intuition.

It is important that you understand this process. Applying rules mechanically is one thing. Understanding why this process is semantically justified is another.

[29]The first rule reads: if at some root of a tree in the forest you have constructed so far, there is a $\diamondsuit$, then you are allowed to draw a line underneath that $\diamondsuit$ and write $\clubsuit$ underneath that line.

The third rule reads: if the forest you have constructed so far contains two neighboring trees, where the left tree has root $\clubsuit$ and the right tree has root $\spadesuit$, then you are allowed to draw a line underneath those two roots and write $\heartsuit$ underneath that line.

[30]The last rule reads: if at some root of a tree in the forest

Natural Deduction: an Abstract Example[28]

- Language $\mathcal{L} = \{\heartsuit, \clubsuit, \spadesuit, \diamondsuit\}$.

- Deductive system given by rules of proof:

$$\frac{\diamondsuit}{\clubsuit}\,\alpha \qquad \frac{\diamondsuit}{\spadesuit}\,\beta \qquad \frac{\clubsuit \quad \spadesuit}{\heartsuit}\,\gamma$$

How do you read these rules?[29]

---

[28]Natural deduction is not just about propositional logic! We explain here the general principles of natural deduction, not just the application to propositional logic.

In order to emphasize that applying natural deduction is a completely mechanical process, we give an example that is void of any intuition.

It is important that you understand this process. Applying rules mechanically is one thing. Understanding why this process is semantically justified is another.
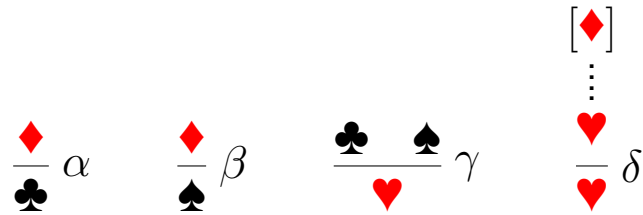
[29]The first rule reads: if at some root of a tree in the forest you have constructed so far, there is a $\diamondsuit$, then you are allowed to draw a line underneath that $\diamondsuit$ and write $\clubsuit$ underneath that line.

The third rule reads: if the forest you have constructed so far contains two neighboring trees, where the left tree has root $\clubsuit$ and the right tree has root $\spadesuit$, then you are allowed to draw a line underneath those two roots and write $\heartsuit$ underneath that line.

[30]The last rule reads: if at some root of a tree in the forest

Natural Deduction: an Abstract Example[28]

- Language $\mathcal{L} = \{\heartsuit, \clubsuit, \spadesuit, \diamondsuit\}$.

- Deductive system given by rules of proof:

$$\frac{\diamondsuit}{\clubsuit}\,\alpha \qquad \frac{\diamondsuit}{\spadesuit}\,\beta \qquad \frac{\clubsuit \quad \spadesuit}{\heartsuit}\,\gamma \qquad \frac{\heartsuit}{\heartsuit}\,\delta \;\; {\begin{matrix}[\diamondsuit]\\ \vdots\\ \heartsuit\end{matrix}}$$

How about this one?[30]

---

[28]Natural deduction is not just about propositional logic! We explain here the general principles of natural deduction, not just the application to propositional logic.

In order to emphasize that applying natural deduction is a completely mechanical process, we give an example that is void of any intuition.

It is important that you understand this process. Applying rules mechanically is one thing. Understanding why this process is semantically justified is another.
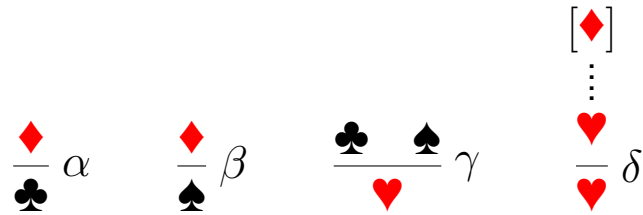
[29]The first rule reads: if at some root of a tree in the forest you have constructed so far, there is a $\diamondsuit$, then you are allowed to draw a line underneath that $\diamondsuit$ and write $\clubsuit$ underneath that line.

The third rule reads: if the forest you have constructed so far contains two neighboring trees, where the left tree has root $\clubsuit$ and the right tree has root $\spadesuit$, then you are allowed to draw a line underneath those two roots and write $\heartsuit$ underneath that line.

[30]The last rule reads: if at some root of a tree in the forest

Natural Deduction: an Abstract Example[28]

- Language $\mathcal{L} = \{\heartsuit, \clubsuit, \spadesuit, \diamondsuit\}$.

- Deductive system given by rules of proof:

$$\frac{\diamondsuit}{\clubsuit}\,\alpha \qquad \frac{\diamondsuit}{\spadesuit}\,\beta \qquad \frac{\clubsuit \quad \spadesuit}{\heartsuit}\,\gamma \qquad \begin{array}{c}[\diamondsuit]\\ \vdots \\ \heartsuit \end{array}\frac{}{\heartsuit}\,\delta$$

How about this one?[30]

$\alpha, \beta, \gamma, \delta$ are just names for the rules.

---

[28]Natural deduction is not just about propositional logic! We explain here the general principles of natural deduction, not just the application to propositional logic.

In order to emphasize that applying natural deduction is a completely mechanical process, we give an example that is void of any intuition.

It is important that you understand this process. Applying rules mechanically is one thing. Understanding why this process is semantically justified is another.

[29]The first rule reads: if at some root of a tree in the forest you have constructed so far, there is a $\diamondsuit$, then you are allowed to draw a line underneath that $\diamondsuit$ and write $\clubsuit$ underneath that line.

The third rule reads: if the forest you have constructed so far contains two neighboring trees, where the left tree has root $\clubsuit$ and the right tree has root $\spadesuit$, then you are allowed to draw a line underneath those two roots and write $\heartsuit$ underneath that line.

[30]The last rule reads: if at some root of a tree in the forest

Proof of ♥

The proof:

$$\frac{♦}{♣}\,\alpha \qquad \frac{♦}{♠}\,\beta \qquad \frac{♣\quad♠}{♥}\,\gamma \qquad \frac{\begin{array}{c}[♦]\\\vdots\\♥\end{array}}{♥}\,\delta$$
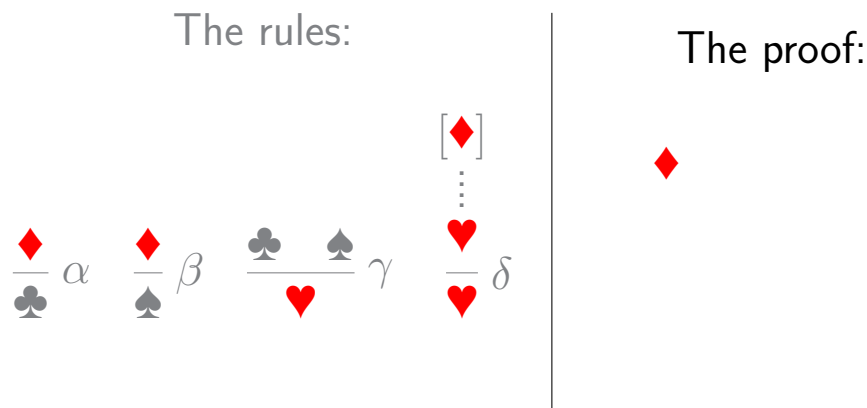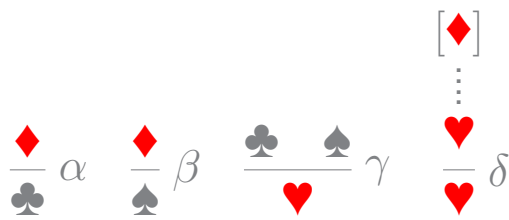
you have constructed so far, there is a ♥, then you are allowed to draw a line underneath that ♥ and write ♦ underneath that line. Moreover you are allowed to discharge (eliminate, close) 0 or more occurrences of ♦ at the leaves of the tree.

Discharging is marked by writing [] around the discharged formula.

Note that generally, the tree may contain assumptions other than ♦ at the leaves. However, these must not be discharged in this rule application. They will remain open until they might be discharged by some other rule application later.

Proof of ♥

The proof:

$$\dfrac{\color{red}\blacklozenge}{\color{gray}\clubsuit}\,\alpha \qquad \dfrac{\color{red}\blacklozenge}{\color{gray}\spadesuit}\,\beta \qquad \dfrac{\color{gray}\clubsuit\quad\color{gray}\spadesuit}{\color{red}\heartsuit}\,\gamma \qquad \dfrac{\begin{array}{c}[\color{red}\blacklozenge]\\ \vdots\\ \color{red}\heartsuit\end{array}}{\color{red}\heartsuit}\,\delta$$

$\color{red}\blacklozenge$

We make[31] an assumption. The assumption is now open[32].

---

you have constructed so far, there is a ♥, then you are allowed to draw a line underneath that ♥ and write ♦ underneath that line. Moreover you are allowed to discharge (eliminate, close) 0 or more occurrences of ♦ at the leaves of the tree.

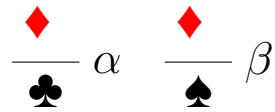Discharging is marked by writing [] around the discharged formula.

Note that generally, the tree may contain assumptions other than ♦ at the leaves. However, these must not be discharged in this rule application. They will remain open until they might be discharged by some other rule application later.

Proof of ♥

The proof:

$$\frac{♦}{♣}\,\alpha \qquad \frac{♦}{♠}\,\beta \qquad \frac{♣ \quad ♠}{♥}\,\gamma \qquad \frac{\genfrac{}{}{0pt}{}{[♦]}{\genfrac{}{}{0pt}{}{\vdots}{♥}}}{♥}\,\delta$$

$$\frac{\genfrac{}{}{0pt}{}{♦}{}}{♣}\,\alpha$$

We apply α.

---

you have constructed so far, there is a ♥, then you are allowed to draw a line underneath that ♥ and write ♦ underneath that line. Moreover you are allowed to discharge (eliminate, close) 0 or more occurrences of ♦ at the leaves of the tree.

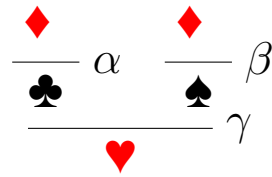Discharging is marked by writing [] around the discharged formula.

Note that generally, the tree may contain assumptions other than ♦ at the leaves. However, these must not be discharged in this rule application. They will remain open until they might be discharged by some other rule application later.

Proof of ♥

The proof:

$$\frac{♦}{♣}\,\alpha \qquad \frac{♦}{♠}\,\beta \qquad \frac{♣\quad♠}{♥}\,\gamma \qquad \frac{\begin{array}{c}[♦]\\ \vdots\\ ♥\end{array}}{♥}\,\delta \qquad\Bigg|\qquad \frac{♦}{♣}\,\alpha \qquad \frac{♦}{♠}\,\beta$$

Similarly with β.

---

you have constructed so far, there is a ♥, then you are allowed to draw a line underneath that ♥ and write ♦ underneath that line. Moreover you are allowed to discharge (eliminate, close) 0 or more occurrences of ♦ at the leaves of the tree.

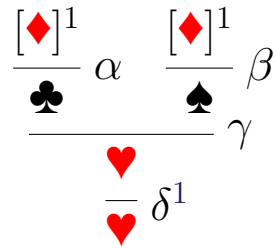Discharging is marked by writing [] around the discharged formula.

Note that generally, the tree may contain assumptions other than ♦ at the leaves. However, these must not be discharged in this rule application. They will remain open until they might be discharged by some other rule application later.

Proof of ♥

The proof:

$$\frac{♦}{♣}\,\alpha \quad \frac{♦}{♠}\,\beta \quad \frac{♣\quad♠}{♥}\,\gamma \quad \frac{\begin{array}{c}[♦]\\ \vdots\\ ♥\end{array}}{♥}\,\delta$$

$$\frac{♦}{♣}\,\alpha \quad \frac{♦}{♠}\,\beta$$
$$\frac{♣\qquad♠}{♥}\,\gamma$$

We apply γ.

you have constructed so far, there is a ♥, then you are allowed to draw a line underneath that ♥ and write ♦ underneath that line. Moreover you are allowed to discharge (eliminate, close) 0 or more occurrences of ♦ at the leaves of the tree.

Discharging is marked by writing [] around the discharged formula.

Note that generally, the tree may contain assumptions other than ♦ at the leaves. However, these must not be discharged in this rule application. They will remain open until they might be discharged by some other rule application later.

Proof of ♥

The proof:

$$\dfrac{♦}{♣}\,\alpha \qquad \dfrac{♦}{♠}\,\beta \qquad \dfrac{♣ \quad ♠}{♥}\,\gamma \qquad \dfrac{\begin{array}{c}[♦]\\ \vdots\\ ♥\end{array}}{♥}\,\delta$$

$$\dfrac{\dfrac{\dfrac{[♦]^1}{♣}\,\alpha \quad \dfrac{[♦]^1}{♠}\,\beta}{♥}\,\gamma}{♥}\,\delta^1$$

We apply δ, discharging two occurrences of ♦. We mark the brackets and the rule with a label so that it is clear which assumption is discharged in which step. The derivation is now a proof: it has no open assumptions (all discharged).

---

you have constructed so far, there is a ♥, then you are allowed to draw a line underneath that ♥ and write ♦ underneath that line. Moreover you are allowed to discharge (eliminate, close) 0 or more occurrences of ♦ at the leaves of the tree.

Discharging is marked by writing [] around the discharged formula.

Note that generally, the tree may contain assumptions other than ♦ at the leaves. However, these must not be discharged in this rule application. They will remain open until they might be discharged by some other rule application later.

2.5 Deductive System: Rules of Propositional Logic

We have rules for conjunction, implication, disjunction, falsity and negation.

Some rules introduce[33], others eliminate connectives.

---

[33]It is typical that the basic rules of a proof system can be classified as introduction or elimination rules for a particular connective.

This classification provides obvious names for the rules and may guide the search for proofs.

The rules for conjunction are pronounced and-introduction, and-elimination-left, and and-elimination-right.

Apart from the basic rules, we will later see that there are also derived rules.

Rules of Propositional Logic: Conjunction

- Rules of two kinds: introduce   connectives

$$\frac{A \quad B}{A \wedge B} \wedge\text{-}I$$

---

[34] The letters A and B in the rules are not propositional variables. Instead, they can stand for arbitrary propositional formulas. One can also say that A and B are metavariables, i.e., they are variables of the proof system as opposed to object variables, i.e., variables of the language that we reason about (here: propositional logic).

When a rule is applied, the metavariables of it must be replaced with actual formulae. We say that a rule is being instantiated.

We will see more about the use of metavariables later.

[35] A rule is valid if for any assignment under which the assumptions of the formula are true, the conclusion is true as well.

This is consistent with the earlier intuitive explanation of validity of a formula. Details can be found in any textbook on logic [vD80].

Note that while the notation $\mathcal{A} \models \ldots$ will be used again later, there $\mathcal{A}$ will not stand for an assignment, but rather for

Rules of Propositional Logic: Conjunction

- Rules of two kinds: introduce and eliminate connectives

$$\frac{A \quad B}{A \wedge B} \wedge\text{-}I \quad \frac{A \wedge B}{A} \wedge\text{-}EL \quad \frac{A \wedge B}{B} \wedge\text{-}ER$$

---

[34]The letters A and B in the rules are not propositional variables. Instead, they can stand for arbitrary propositional formulas. One can also say that A and B are metavariables, i.e., they are variables of the proof system as opposed to object variables, i.e., variables of the language that we reason about (here: propositional logic).

When a rule is applied, the metavariables of it must be replaced with actual formulae. We say that a rule is being instantiated.

We will see more about the use of metavariables later.

[35]A rule is valid if for any assignment under which the assumptions of the formula are true, the conclusion is true as well.

This is consistent with the earlier intuitive explanation of validity of a formula. Details can be found in any textbook on logic [vD80].

Note that while the notation $\mathcal{A} \models \ldots$ will be used again later, there $\mathcal{A}$ will not stand for an assignment, but rather for

Rules of Propositional Logic: Conjunction

- Rules of two kinds: introduce and eliminate connectives

$$\frac{A \quad B}{A \wedge B} \wedge\text{-}I \quad \frac{A \wedge B}{A} \wedge\text{-}EL \quad \frac{A \wedge B}{B} \wedge\text{-}ER$$

- Rules are schematic[34].

- Why valid[35]? If all assumptions are true, then so is conclusion

$$\mathcal{A} \models A \wedge B \text{ iff } \mathcal{A} \models A \text{ and } \mathcal{A} \models B$$

---

[34]The letters A and B in the rules are not propositional variables. Instead, they can stand for arbitrary propositional formulas. One can also say that A and B are metavariables, i.e., they are variables of the proof system as opposed to object variables, i.e., variables of the language that we reason about (here: propositional logic).

When a rule is applied, the metavariables of it must be replaced with actual formulae. We say that a rule is being instantiated.

We will see more about the use of metavariables later.

[35]A rule is valid if for any assignment under which the assumptions of the formula are true, the conclusion is true as well.

This is consistent with the earlier intuitive explanation of validity of a formula. Details can be found in any textbook on logic [vD80].

Note that while the notation $\mathcal{A} \models \ldots$ will be used again later, there $\mathcal{A}$ will not stand for an assignment, but rather for

Example Derivation with Conjunction

The rules:

$$\frac{A \quad B}{A \wedge B} \wedge\text{-}I$$

$$\frac{A \wedge B}{A} \wedge\text{-}EL$$

$$\frac{A \wedge B}{B} \wedge\text{-}ER$$

a construct having an assignment as one constituent. This is because we will generalize, and in the new setting we need something more complex than just an assignment. But in spirit $\mathcal{A} \models \ldots$ will still mean the same thing.

[36]All three rules have a non-empty sequence of assumptions. Thus to build a tree using these rules, we must first make some assumptions.

None of the rules involves discharging an assumption.

We have said earlier that a proof is a derivation with no open assumptions.

Consequently, the answer is no.

Example Derivation with Conjunction

$$\frac{A \quad B}{A \wedge B} \text{ } \wedge\text{-}I$$

$$\frac{A \wedge B}{A} \text{ } \wedge\text{-}EL \qquad \frac{A \wedge (B \wedge C)}{A} \text{ } \wedge\text{-}EL$$

$$\frac{A \wedge B}{B} \text{ } \wedge\text{-}ER$$

36

---

a construct having an assignment as one constituent. This is because we will generalize, and in the new setting we need something more complex than just an assignment. But in spirit $\mathcal{A} \models \ldots$ will still mean the same thing.

[36]All three rules have a non-empty sequence of assumptions. Thus to build a tree using these rules, we must first make some assumptions.

None of the rules involves discharging an assumption.

We have said earlier that a proof is a derivation with no open assumptions.

Consequently, the answer is no.

Example Derivation with Conjunction

The rules:

$$\frac{A \quad B}{A \wedge B} \; \wedge\text{-}I$$

$$\frac{A \wedge B}{A} \; \wedge\text{-}EL$$

$$\frac{A \wedge B}{B} \; \wedge\text{-}ER$$

$$\frac{A \wedge (B \wedge C)}{A} \; \wedge\text{-}EL \qquad \frac{A \wedge (B \wedge C)}{B \wedge C} \; \wedge\text{-}ER$$

---

a construct having an assignment as one constituent. This is because we will generalize, and in the new setting we need something more complex than just an assignment. But in spirit $\mathcal{A} \models \ldots$ will still mean the same thing.

[36]All three rules have a non-empty sequence of assumptions. Thus to build a tree using these rules, we must first make some assumptions.

None of the rules involves discharging an assumption.

We have said earlier that a proof is a derivation with no open assumptions.

Consequently, the answer is no.

Example Derivation with Conjunction

The rules:

$$\frac{A \quad B}{A \wedge B} \wedge\text{-}I$$

$$\frac{A \wedge B}{A} \wedge\text{-}EL$$

$$\frac{A \wedge B}{B} \wedge\text{-}ER$$

$$\frac{A \wedge (B \wedge C)}{A} \wedge\text{-}EL \qquad \frac{\dfrac{A \wedge (B \wedge C)}{B \wedge C} \wedge\text{-}ER}{C} \wedge\text{-}ER$$

a construct having an assignment as one constituent. This is because we will generalize, and in the new setting we need something more complex than just an assignment. But in spirit $\mathcal{A} \models \ldots$ will still mean the same thing.

[36]All three rules have a non-empty sequence of assumptions. Thus to build a tree using these rules, we must first make some assumptions.

None of the rules involves discharging an assumption.

We have said earlier that a proof is a derivation with no open assumptions.

Consequently, the answer is no.

Example Derivation with Conjunction

The rules:

$$\frac{A \quad B}{A \wedge B} \wedge\text{-}I$$

$$\frac{A \wedge B}{A} \wedge\text{-}EL$$

$$\frac{A \wedge B}{B} \wedge\text{-}ER$$

$$\frac{\dfrac{A \wedge (B \wedge C)}{A} \wedge\text{-}EL \qquad \dfrac{\dfrac{A \wedge (B \wedge C)}{B \wedge C} \wedge\text{-}ER}{C} \wedge\text{-}ER}{A \wedge C} \wedge\text{-}I$$

---

a construct having an assignment as one constituent. This is because we will generalize, and in the new setting we need something more complex than just an assignment. But in spirit $\mathcal{A} \models \dots$ will still mean the same thing.

[36]All three rules have a non-empty sequence of assumptions. Thus to build a tree using these rules, we must first make some assumptions.

None of the rules involves discharging an assumption.

We have said earlier that a proof is a derivation with no open assumptions.

Consequently, the answer is no.

Example Derivation with Conjunction

The rules:

$$\frac{A \quad B}{A \wedge B} \wedge\text{-}I$$

$$\frac{A \wedge B}{A} \wedge\text{-}EL$$

$$\frac{A \wedge B}{B} \wedge\text{-}ER$$

$$\frac{\dfrac{A \wedge (B \wedge C)}{A} \wedge\text{-}EL \qquad \dfrac{\dfrac{A \wedge (B \wedge C)}{B \wedge C} \wedge\text{-}ER}{C} \wedge\text{-}ER}{A \wedge C} \wedge\text{-}I$$

Can we **prove** anything with just these three rules?[36]

a construct having an assignment as one constituent. This is because we will generalize, and in the new setting we need something more complex than just an assignment. But in spirit $\mathcal{A} \models \ldots$ will still mean the same thing.

[36]All three rules have a non-empty sequence of assumptions. Thus to build a tree using these rules, we must first make some assumptions.

None of the rules involves **discharging** an assumption.

We have said earlier that a **proof** is a derivation with no open assumptions.

Consequently, the answer is **no**.

Rules of Propositional Logic: Implication

- Rules

$$\frac{\begin{array}{c}[A]\\ \vdots\\ B\end{array}}{A \to B} \to\text{-}I \qquad \frac{A \to B \quad A}{B} \to\text{-}E$$

Rules of Propositional Logic: Implication

- Rules

$$\frac{\begin{array}{c} [A] \\ \vdots \\ B \end{array}}{A \to B} \to\text{-}I \quad \frac{A \to B \quad A}{B} \to\text{-}E$$

- $\to$-E is also called modus ponens.

Rules of Propositional Logic: Implication

- Rules

$$\frac{\begin{array}{c}[A]\\ \vdots\\ B\end{array}}{A \to B} \to\text{-}I \quad \frac{A \to B \quad A}{B} \to\text{-}E$$

- $\to$-*E* is also called modus ponens.

- $\to$-*I* formalizes strategy:
  To derive $A \to B$, derive B under the additional assumption A.

A very Simple Proof

The simplest proof we can think of is the proof of $P \to P$.

$$P$$

---

[37]When we make the assumption P, we obtain a forest consisting of one tree. In this tree, P is at the same time a leaf and the root. Thus the tree P is a degenerate example of the schema

$[A]$
$\vdots$
B

where both A and B are replaced with P.

Therefore we may apply rule $\to$-*I*, similarly as in our abstract example.

A very Simple Proof

The simplest proof we can think of is the proof of $P \to P$.

$$\frac{[P]^1}{P \to P} \to\text{-}I^1$$

Do you find this strange?[37]

---

[37]When we make the assumption P, we obtain a forest consisting of one tree. In this tree, P is at the same time a leaf and the root. Thus the tree P is a degenerate example of the schema

$$\begin{array}{c}[A]\\ \vdots\\ B\end{array}$$

where both A and B are replaced with P.

Therefore we may apply rule $\to$-I, similarly as in our abstract example.

Examples with Conjunction and Implication

1. $A \to B \to A$[38]

2. $A \wedge (B \wedge C) \to A \wedge C$[39]

The rule(s):

[38]
$$\dfrac{\begin{array}{c}[A]\\ \vdots \\ B\end{array}}{A \to B} \to\text{-}I$$

The proof:

$$\dfrac{\dfrac{[A]^1}{B \to A} \to\text{-}I}{A \to B \to A} \to\text{-}I^1$$

The rules:

$$\dfrac{A \quad B}{A \wedge B} \wedge\text{-}I$$

$$\dfrac{A \wedge B}{A} \wedge\text{-}EL$$

[39] $$\dfrac{A \wedge B}{B} \wedge\text{-}ER$$

$$\dfrac{\begin{array}{c}[A]\\ \vdots \\ B\end{array}}{A \to B} \to\text{-}I$$

The proof:

$$\dfrac{\dfrac{\dfrac{[A \wedge (B \wedge C)]^2}{A} \wedge\text{-}EL \qquad \dfrac{\dfrac{[A \wedge (B \wedge C)]^2}{B \wedge C} \wedge\text{-}ER}{C} \wedge\text{-}ER}{A \wedge C} \wedge\text{-}I}{(A \wedge (B \wedge C)) \to (A \wedge C)} \to\text{-}I^2$$

3. $(A \to B \to C) \to (A \to B) \to A \to C$[40]

Are these object or metavariables here?[41]

The rules:	The proof:

The rules:

$$[A]$$
$$\vdots$$
$$\dfrac{B}{A \to B} \to\text{-}I$$

[40]

$$\dfrac{A \to B \quad A}{B} \to\text{-}E$$

The proof:

$$\dfrac{\dfrac{[(A \to B \to C)]^3 \quad [A]^5}{B \to C} \to\text{-}E \quad \dfrac{[(A \to B)]^4 \quad [A]^5}{B} \to\text{-}E}{\dfrac{\dfrac{\dfrac{C}{A \to C} \to\text{-}I^5}{(A \to B) \to A \to C} \to\text{-}I^4}{(A \to B \to C) \to (A \to B) \to A \to C} \to\text{-}I^3}}$$

[41]In these examples, you may regard A, B, C as propositional variables. On the other hand, the proofs are schematic, i.e., they go through for any formula replacing A, B, and C.

Disjunction

- Rules

$$\frac{A}{A \vee B} \ \vee\text{-}IL \qquad \frac{B}{A \vee B} \ \vee\text{-}IR \qquad \frac{A \vee B \quad \overset{[A]}{\underset{\vdots}{C}} \quad \overset{[B]}{\underset{\vdots}{C}}}{C} \ \vee\text{-}E$$

Disjunction

- Rules

$$\frac{A}{A \vee B} \vee\text{-}IL \qquad \frac{B}{A \vee B} \vee\text{-}IR \qquad \frac{A \vee B \quad \overset{[A]}{\overset{\vdots}{C}} \quad \overset{[B]}{\overset{\vdots}{C}}}{C} \vee\text{-}E$$

- Formalizes case-split strategy for using $A \vee B$.

Disjunction: Example

- Rules

$$\frac{A}{A \vee B} \; \vee\text{-}IL \qquad \frac{B}{A \vee B} \; \vee\text{-}IR \qquad \frac{A \vee B \quad \overset{[A]}{\underset{\vdots}{C}} \quad \overset{[B]}{\underset{\vdots}{C}}}{C} \; \vee\text{-}E$$

- Example: formalize and prove

  When it rains then I wear my jacket.
  When it snows then I wear my jacket.
  It is raining or snowing.
  Therefore I wear my jacket.

Falsity and Negation

- Falsity

$$\frac{\bot}{A} \ \bot\text{-}E$$

No introduction rule![42]

Falsity and Negation

- Falsity

$$\frac{\bot}{A} \ \bot\text{-}E$$

No introduction rule![42]

- Negation: define $\neg A$ as $A \to \bot$. Rules for $\neg$ just special cases[43] of rules for $\to$. Convenient to have

---

[42] The symbol $\bot$ stands for "false".

It should be intuitively clear that since the purpose of a proof system is to derive true formulae, there is no introduction rule for falsity. One may wonder: what is the role of $\bot$ then? We will see this soon. The main role is linked to negation. We quote from [And02, p. 152]:

> $\bot$ plays the role of a contradiction in indirect proofs.

[43] The rule

$$\frac{\neg A \quad A}{\bot}$$

is simply an instance of $\to$-E (since $\neg A$ is shorthand for $A \to \bot$).

Likewise, the rule

$$\frac{\begin{array}{c} [A] \\ \vdots \\ \bot \end{array}}{\neg A}$$

$$\frac{\neg A \quad A}{B} \text{¬-}E^{44} \quad \textcolor{blue}{\text{derived by}} \quad \frac{\dfrac{\neg A \quad A}{\bot} \text{→-}E}{B} \text{⊥-}E$$

is simply an instance of *→-I*. Therefore, we will not introduce these as special rules. But there is a special rule *¬-E*.

[44]For negation, it is common to have a rule

$$\frac{\neg A \quad A}{B} \text{¬-}E$$

We have seen how this rule can be derived. The concept of deriving rules will be explained more systematically <u>later</u>.

This rule is also called <u>ex falso quod libet</u> (from the false whatever you like).

Intuitionistic versus Classical Logic

- Peirce's Law: $((A \rightarrow B) \rightarrow A) \rightarrow A$.
  Is this valid[45]? Provable[46]?

Intuitionistic versus Classical Logic

- Peirce's Law: $((A \to B) \to A) \to A$.
  Is this valid[45]? Provable[46]?

---

[45]Yes, simply check the truth table:

A	B	$((A \to B) \to A) \to A$
$True$	$True$	$True$
$True$	$False$	$True$
$False$	$True$	$True$
$False$	$False$	$True$

[46]In the proof system given so far, this is not provable. To prove that it is not provable requires an analysis of so-called normal forms of proofs. However, we do not do this here.

- It is provable in classical logic[47], obtained by adding

$$A \vee \neg A^{48} \quad \text{or} \quad \cfrac{\begin{array}{c}[\neg A]\\ \vdots\\ \bot\end{array}}{A}\,RAA_{49} \quad \text{or} \quad \cfrac{\begin{array}{c}[\neg A]\\ \vdots\\ A\end{array}}{A}\,\textit{classical}_{50}.$$

---

[47] The proof system we have given so far is a proof system for intuitionistic logic. The main point about intuitionistic logic is that one cannot claim that every statement is either true or false, but rather, evidence must be given for every statement.

In classical reasoning, the law of the excluded middle holds.

One also says that proofs in intuitionistic logic are constructive whereas proofs in classical logic are not necessarily constructive.

We quote the first sentence from [Min00]:

> Intuitionistic logic is studied here as part of familiar classical logic which allows an effective interpretation and mechanical extraction of programs from proofs.

The difference between intuitionistic and classical logic has been the topic of a fundamental discourse in the literature on logic [PM68] [Tho91, chapter 3]. Often proofs contain case distinctions, assuming that for any statement ψ, either ψ or $\neg\psi$ holds. This reasoning is classical; it does not apply in intuitionistic logic.

[48]$A \lor \neg A$ is called axiom of the excluded middle.

[49]The rule

$$\frac{\overset{\displaystyle [\neg A]}{\underset{\displaystyle \vdots}{\quad}}{\perp}}{A} \; RAA$$

is called reduction ad absurdum.

[50]The rule

$$\frac{\overset{\displaystyle [\neg A]}{\underset{\displaystyle \vdots}{A}}}{A} \; \textit{classical}$$

corresponds to the formulation is Isabelle.

Example of Classical Reasoning

There exist irrational numbers a and b such that a^b is rational.

Example of Classical Reasoning

There exist irrational numbers a and b such that a^b is rational.

Proof: Let b be $\sqrt{2}$ and consider whether or not b^b is rational.

Case 1: If rational, let $a = b = \sqrt{2}$

Case 2: If irrational, let $a = \sqrt{2}^{\sqrt{2}}$, and then

$$a^b = \sqrt{2}^{\sqrt{2}^{\sqrt{2}}} = \sqrt{2}^{(\sqrt{2}*\sqrt{2})} = \sqrt{2}^2 = 2$$

Example of Classical Reasoning

There exist irrational numbers a and b such that a^b is rational.

Proof: Let b be $\sqrt{2}$ and consider whether or not b^b is rational.

Case 1: If rational, let $a = b = \sqrt{2}$

Case 2: If irrational, let $a = \sqrt{2}^{\sqrt{2}}$, and then

$$a^b = \sqrt{2}^{\sqrt{2}^{\sqrt{2}}} = \sqrt{2}^{(\sqrt{2}*\sqrt{2})} = \sqrt{2}^2 = 2$$

We still don't know how to choose a and b so that a^b is rational. Hence the proof if non-constructive.

Overview of Rules

$$\frac{A \quad B}{A \wedge B} \ \wedge\text{-}I \qquad \frac{A \wedge B}{A} \ \wedge\text{-}EL \qquad \frac{A \wedge B}{B} \ \wedge\text{-}ER$$

$$\frac{A}{A \vee B} \ \vee\text{-}IL \qquad \frac{B}{A \vee B} \ \vee\text{-}IR \qquad \frac{A \vee B \quad \overset{[A]}{\overset{\vdots}{C}} \quad \overset{[B]}{\overset{\vdots}{C}}}{C} \ \vee\text{-}E$$

$$\frac{\overset{[A]}{\overset{\vdots}{B}}}{A \rightarrow B} \ \rightarrow\text{-}I \qquad \frac{A \rightarrow B \quad A}{B} \ \rightarrow\text{-}E \qquad \frac{\bot}{A} \ \bot\text{-}E$$

2.6 Deductive System: Derived Rules

Using the basic rules, we can derive new rules.
Example: Resolution rule.

2.6 Deductive System: Derived Rules

Using the basic rules, we can derive new rules.
  Example: Resolution rule.

$$\frac{R \lor S \quad \neg S}{R}$$

It looks like this.

2.6 Deductive System: Derived Rules

Using the basic rules, we can derive new rules.
    Example: Resolution rule.

$$\neg S$$

$$\frac{R \vee S \quad \neg S}{R} \qquad \frac{\begin{array}{c} R \vee S \end{array}}{R}$$

We build a fragment of a derivation by writing the conclusion R and the assumptions $R \vee S$ and $\neg S$.

2.6 Deductive System: Derived Rules

Using the basic rules, we can derive new rules.
    Example: Resolution rule.

$$\neg S$$

$$\frac{R \vee S \quad \neg S}{R} \qquad \frac{R \vee S \quad R}{R} \quad \vee\text{-}E$$

Since we have assumption $R \vee S$, using $\vee$-*E* seems a good idea. So we should make assumptions R and S. First R. But that is a derivation of R from R!

2.6 Deductive System: Derived Rules

Using the basic rules, we can derive new rules.

 Example: Resolution rule.

$$\neg S \quad S$$

$$\frac{R \vee S \quad \neg S}{R} \qquad \frac{R \vee S \quad R}{R} \text{ $\vee$-}E$$

So now S.

2.6 Deductive System: Derived Rules

Using the basic rules, we can derive new rules.
  Example: Resolution rule.

$$\frac{\neg S \quad S}{\bot} \rightarrow\text{-}E$$

$$\frac{R \vee S \quad \neg S}{R} \qquad \frac{R \vee S \quad R}{R} \ \vee\text{-}E$$

$\neg S$ and S allow us to apply $\rightarrow$-E.

2.6 Deductive System: Derived Rules

Using the basic rules, we can derive new rules.

Example: Resolution rule.

$$
\cfrac{R \lor S \quad \neg S}{R}
\qquad
\cfrac{R \lor S \quad R \quad \cfrac{\cfrac{\cfrac{\neg S \quad S}{\bot}\ \to\text{-}E}{R}\ \bot\text{-}E}{}}{R}\ \lor\text{-}E
$$

To apply $\lor$-*E* in the end, we need to derive R. But that's easy using $\bot$-*E*!

2.6 Deductive System: Derived Rules

Using the basic rules, we can derive new rules.

Example: Resolution rule.

$$\frac{R \vee S \quad \neg S}{R} \qquad \frac{R \vee S \quad [R]^1 \quad \dfrac{\dfrac{\neg S \quad [S]^1}{\bot} \to\text{-}E}{R} \; \bot\text{-}E}{R} \; \vee\text{-}E^1$$

Finally, we can apply $\vee$-*E*. The derivation with open assumptions is a new rule that can be used like any other rule.

A Variation of Natural Deduction: Boxes

We have seen just one deductive system.

One variation of natural deduction is the following: A derivation is not a tree, but a sequence of numbered lines. Instead of subtrees relying on open assumptions, a subderivation relying on an assumption is enclosed in a box.

You find this explained in [HR04].

2.7 Alternative Deductive System Using Sequent Notation

One can base the deductive system around the derivability judgement[51], i.e., reason about $\Gamma \vdash A$ where $\Gamma \equiv A_1, \ldots, A_n$ instead of individual formulae.

---

[51]An object like $A \rightarrow (B \rightarrow C), A, B \vdash C$ is called a derivability judgement. We explained it earlier as simply asserting the fact that there exists a derivation tree with C at its root and open assumptions $A \rightarrow (B \rightarrow C), A, B$.

However, it is also possible to make such judgements the central objects of the deductive system, i.e., have rules involving such objects.

The notation $\Gamma \vdash A$ is called sequent notation. However, this should not be confused with the sequent calculus (we will consider it later). The sequent calculus is based on sequents, which are syntactic entities of the form $A_1, \ldots, A_n \vdash B_1, \ldots, B_m$, where the $A_1, \ldots, A_n, B_1, \ldots, B_m$ are all formulae. You see that this definition is more general than the derivability judgements we consider here.

What we are about to present is a kind of hybrid between natural deduction and the sequent calculus, which we might call natural deduction using a sequent notation.

Sequent Rules (for $\rightarrow$ /$\wedge$ Fragment)

Rules for assumptions[52] and weakening[53]:

$$\Gamma \vdash A^{54} \quad (\text{where } A \in \Gamma) \qquad \frac{\Gamma \vdash B}{A, \Gamma \vdash B} \; \textit{weaken}$$

Sequent Rules (for $\rightarrow /\wedge$ Fragment)

Rules for assumptions[52] and weakening[53]:

$$\Gamma \vdash A^{54} \quad (\text{where } A \in \Gamma) \qquad \frac{\Gamma \vdash B}{A, \Gamma \vdash B} \ \textit{weaken}$$

Rules for $\wedge$ and $\rightarrow$:

$$\frac{\Gamma \vdash A \quad \Gamma \vdash B}{\Gamma \vdash A \wedge B} \ \textit{$\wedge$-I} \qquad \frac{\Gamma \vdash A \wedge B}{\Gamma \vdash A} \ \textit{$\wedge$-EL} \qquad \frac{\Gamma \vdash A \wedge B}{\Gamma \vdash B} \ \textit{$\wedge$-ER}$$

$$\frac{A, \Gamma \vdash B}{\Gamma \vdash A \rightarrow B} \ \textit{$\rightarrow$-I} \qquad \frac{\Gamma \vdash A \rightarrow B \quad \Gamma \vdash A}{\Gamma \vdash B} \ \textit{$\rightarrow$-E}$$

---

[52]The special rule for assumptions takes the role in this sequent style notation that the process of making and discharging assumptions had in natural deduction based on trees.

It is not so obvious that the two ways of writing proofs are equivalent, but we shall become familiar with this in the exercises by doing proofs on paper as well as in Isabelle.

[53]The rule *weaken* is

$$\frac{\Gamma \vdash B}{A, \Gamma \vdash B} \ \textit{weaken}$$

Intuitively, the soundness of rule *weaken* should be clear: having an additional assumption in the context cannot hurt since there is no proof rule that requires the absence of some assumption.

We will see an application of that rule later.

[54]An axiom is a rule without premises. We call a rule with premises proper.

More rules can be derived[55].

One can write an axiom A as

$$\frac{}{A}$$

to emphasise that it is a rule with an empty set of premises.

Note that the natural deduction rules for propositional logic contain no axioms. In the sequent style formalization, having the assumption rule (axiom) is essential for being able to prove anything, but in the natural deduction style we learned first, we can construct proofs without having any axioms.

[55] As an example, consider

$$\frac{A, B, \Gamma \vdash C \quad \Gamma \vdash A \wedge B}{\Gamma \vdash C} \wedge\text{-}E$$

Example: Refinement Style with Metavariables

$$\frac{}{\vdash A \wedge (B \wedge C) \to A \wedge C} \to\text{-}I$$

We want to show that $A \wedge (B \wedge C) \to A \wedge C$ is a tautology, i.e., that it is derivable without any assumptions.

This rule can be derived as follows:

$$\cfrac{\cfrac{\cfrac{\cfrac{A, B, \Gamma \vdash C}{A, \Gamma \vdash B \to C} \to\text{-}I}{\Gamma \vdash A \to B \to C} \to\text{-}I \qquad \cfrac{\Gamma \vdash A \wedge B}{\Gamma \vdash A} \wedge\text{-}EL}{\Gamma \vdash B \to C} \to\text{-}E \qquad \cfrac{\Gamma \vdash A \wedge B}{\Gamma \vdash B} \wedge\text{-}ER}{\Gamma \vdash C} \to\text{-}E$$

Example: Refinement Style with Metavariables

$$\cfrac{\cfrac{}{A \wedge (B \wedge C) \vdash A \wedge C} \wedge\text{-}I}{\vdash A \wedge (B \wedge C) \to A \wedge C} \to\text{-}I$$

The topmost connective of the formula is $\to$, so the best rule[56] to choose is $\to$-I.

This rule can be derived as follows:

$$\cfrac{\cfrac{\cfrac{\cfrac{A, B, \Gamma \vdash C}{A, \Gamma \vdash B \to C} \to\text{-}I}{\Gamma \vdash A \to B \to C} \to\text{-}I \quad \cfrac{\Gamma \vdash A \wedge B}{\Gamma \vdash A} \wedge\text{-}EL}{\Gamma \vdash B \to C} \to\text{-}E \quad \cfrac{\Gamma \vdash A \wedge B}{\Gamma \vdash B} \wedge\text{-}ER}{\Gamma \vdash C} \to\text{-}E$$

Example: Refinement Style with Metavariables

$$\frac{\dfrac{}{A \wedge (B \wedge C) \vdash A} \wedge\text{-}EL \qquad \dfrac{}{A \wedge (B \wedge C) \vdash C} \wedge\text{-}ER}{\dfrac{A \wedge (B \wedge C) \vdash A \wedge C}{\vdash A \wedge (B \wedge C) \rightarrow A \wedge C} \rightarrow\text{-}I} \wedge\text{-}I$$

The topmost connective of the formula is $\wedge$, so the best rule to choose is $\wedge$-*I*.

This rule can be derived as follows:

$$\frac{\dfrac{\dfrac{\dfrac{A, B, \Gamma \vdash C}{A, \Gamma \vdash B \rightarrow C} \rightarrow\text{-}I}{\Gamma \vdash A \rightarrow B \rightarrow C} \rightarrow\text{-}I \quad \dfrac{\Gamma \vdash A \wedge B}{\Gamma \vdash A} \wedge\text{-}EL}{\Gamma \vdash B \rightarrow C} \rightarrow\text{-}E \qquad \dfrac{\Gamma \vdash A \wedge B}{\Gamma \vdash B} \wedge\text{-}ER}{\Gamma \vdash C} \rightarrow\text{-}E$$

Example: Refinement Style with Metavariables

$$\cfrac{\cfrac{A \wedge (B \wedge C) \vdash A \wedge {\color{red}?X}}{A \wedge (B \wedge C) \vdash A} \wedge\text{-}EL \qquad \cfrac{}{A \wedge (B \wedge C) \vdash C} \wedge\text{-}ER}{\cfrac{A \wedge (B \wedge C) \vdash A \wedge C}{\vdash A \wedge (B \wedge C) \to A \wedge C} \to\text{-}I} \wedge\text{-}I$$

Things are becoming less obvious. To know that $\wedge$-*EL* is the best rule for the r.h.s., you need to inspect the assumption $A \wedge (B \wedge C)$.

This rule can be derived as follows:

$$\cfrac{\cfrac{\cfrac{A, B, \Gamma \vdash C}{A, \Gamma \vdash B \to C} \to\text{-}I}{\Gamma \vdash A \to B \to C} \to\text{-}I \qquad \cfrac{\cfrac{\Gamma \vdash A \wedge B}{\Gamma \vdash A} \wedge\text{-}EL}{\Gamma \vdash B \to C} \to\text{-}E \qquad \cfrac{\Gamma \vdash A \wedge B}{\Gamma \vdash B} \wedge\text{-}ER}{\Gamma \vdash C} \to\text{-}E$$

Example: Refinement Style with Metavariables

$$\cfrac{\cfrac{A \wedge (B \wedge C) \vdash A \wedge \,{\color{red}?X}}{A \wedge (B \wedge C) \vdash A} \wedge\text{-}EL \qquad \cfrac{\cfrac{}{A \wedge (B \wedge C) \vdash ({\color{red}?Y} \wedge C)} \wedge\text{-}ER}{A \wedge (B \wedge C) \vdash C} \wedge\text{-}ER}{\cfrac{A \wedge (B \wedge C) \vdash A \wedge C}{\vdash A \wedge (B \wedge C) \to A \wedge C} \to\text{-}I} \wedge\text{-}I$$

Now it's becoming even more difficult. To know that $\wedge\text{-}ER$ is the best rule for the l.h.s., you need to look deep into the assumption $A \wedge (B \wedge C)$.

This rule can be derived as follows:

$$\cfrac{\cfrac{\cfrac{A, B, \Gamma \vdash C}{A, \Gamma \vdash B \to C} \to\text{-}I}{\Gamma \vdash A \to B \to C} \to\text{-}I \qquad \cfrac{\cfrac{\Gamma \vdash A \wedge B}{\Gamma \vdash A} \wedge\text{-}EL}{\Gamma \vdash B \to C} \to\text{-}E \qquad \cfrac{\Gamma \vdash A \wedge B}{\Gamma \vdash B} \wedge\text{-}ER}{\Gamma \vdash C} \to\text{-}E$$

Example: Refinement Style with Metavariables

$$\cfrac{\cfrac{A \wedge (B \wedge C) \vdash A \wedge {\color{red}?X}}{A \wedge (B \wedge C) \vdash A} \text{ $\wedge$-EL} \quad \cfrac{\cfrac{\cfrac{A \wedge (B \wedge C) \vdash {\color{red}?Z} \wedge ({\color{red}?Y} \wedge C)}{A \wedge (B \wedge C) \vdash ({\color{red}?Y} \wedge C)} \text{ $\wedge$-ER}}{A \wedge (B \wedge C) \vdash C} \text{ $\wedge$-ER}}{A \wedge (B \wedge C) \vdash A \wedge C}}{\cfrac{A \wedge (B \wedge C) \vdash A \wedge C}{\vdash A \wedge (B \wedge C) \to A \wedge C} \text{ $\to$-I}} \text{ $\wedge$-I}$$

Again you need to look at both sides of the $\vdash$ to decide what to do.

This rule can be derived as follows:

$$\cfrac{\cfrac{\cfrac{\cfrac{A, B, \Gamma \vdash C}{A, \Gamma \vdash B \to C} \text{ $\to$-I}}{\Gamma \vdash A \to B \to C} \text{ $\to$-I} \quad \cfrac{\Gamma \vdash A \wedge B}{\Gamma \vdash A} \text{ $\wedge$-EL}}{\Gamma \vdash B \to C} \text{ $\to$-E} \quad \cfrac{\Gamma \vdash A \wedge B}{\Gamma \vdash B} \text{ $\wedge$-ER}}{\Gamma \vdash C} \text{ $\to$-E}$$

Example: Refinement Style with Metavariables

$$\cfrac{\cfrac{A \wedge (B \wedge C) \vdash A \wedge {\color{red}?X}}{A \wedge (B \wedge C) \vdash A} \text{∧-EL} \qquad \cfrac{\cfrac{\cfrac{A \wedge (B \wedge C) \vdash {\color{red}?Z} \wedge ({\color{red}?Y} \wedge C)}{A \wedge (B \wedge C) \vdash ({\color{red}?Y} \wedge C)} \text{∧-ER}}{A \wedge (B \wedge C) \vdash C} \text{∧-ER}}{\cfrac{A \wedge (B \wedge C) \vdash A \wedge C}{\vdash A \wedge (B \wedge C) \to A \wedge C} \text{→-I}} \text{∧-I}}{}$$

Solution for ${\color{red}?Z} = A$, ${\color{red}?Y} = B$ and ${\color{red}?X} = (B \wedge C)$.

This rule can be derived as follows:

$$\cfrac{\cfrac{\cfrac{\cfrac{A, B, \Gamma \vdash C}{A, \Gamma \vdash B \to C} \text{→-I}}{\Gamma \vdash A \to B \to C} \text{→-I} \qquad \cfrac{\cfrac{\Gamma \vdash A \wedge B}{\Gamma \vdash A} \text{∧-EL}}{} }{\Gamma \vdash B \to C} \text{→-E} \qquad \cfrac{\Gamma \vdash A \wedge B}{\Gamma \vdash B} \text{∧-ER}}{\Gamma \vdash C} \text{→-E}$$

Comments about Refinement

This crazy way of carrying out proofs is the (standard) Isabelle-way!

- Refinement style means we work from goals to axioms[57]

- metavariables used to delay commitments

Isabelle allows other refinements[58]/alternatives too (see labs).

---

[57]As you saw in our animation, we worked from the root of the tree to the leaves.

[58]One aspect you might have noted in the proof is that the steps at the top, where $\wedge$-*EL* and $\wedge$-*ER* were used, required non-obvious choices, and those choices were based on the assumptions in the current derivability judgement.

In Isabelle, we will apply other rules and proof techniques that allow us to manipulate assumptions explicitly. These techniques make the process of finding a proof more deterministic.

But that is just one aspect. We will give a more theoretic account of the way Isabelle constructs proofs later.

Outlook

- Computer Supported Modeling and Reasoning is about turning logic into a useful tool and bringing it to life.

- We will cover:
  - deductive aspects of logic (their proof systems)
  - metatheoretic aspects (their representation)
  - pragmatics (their use), and
  - case studies

- This is an active, hands-on course
  - The labs are as important as (if not more than!) the lectures

– Individual projects are possible. Individual initiative desired!

3 Natural Deduction: Review

Overview

- Short review: ND Systems and proofs

- First-Order Logic

  - Overview

  - Syntax

  - Semantics

  - Deduction, some derived rules, and examples

How Are ND Proofs Built?

ND proofs[59] build derivations under (possibly temporary) assumptions.

---

[59]ND stands for Natural Deduction. It was explained in the previous lecture.

ND: Example for $\rightarrow$ /$\wedge$ Fragment

Rules:

$$\frac{A \quad B}{A \wedge B} \; \wedge\text{-}I \qquad \frac{A \wedge B}{A} \; \wedge\text{-}EL$$

$$\frac{A \wedge B}{B} \; \wedge\text{-}ER \qquad \frac{\begin{array}{c} [A] \\ \vdots \\ B \end{array}}{A \rightarrow B} \; \rightarrow\text{-}I$$

$$\frac{A \rightarrow B \quad A}{B} \; \rightarrow\text{-}E$$

Proof:

$$\frac{\dfrac{\dfrac{[A \wedge B]^1}{B} \; \wedge\text{-}EL \quad \dfrac{[A \wedge B]^1}{A} \; \wedge\text{-}ER}{B \wedge A} \; \wedge\text{-}I}{A \wedge B \rightarrow B \wedge A} \; \rightarrow\text{-}I^1$$

Alternative Formalization Using Sequents[60]

Rules (for $\rightarrow/\wedge$ fragment). Here, Γ is a set of formulae.

$$\Gamma \vdash A \quad (\text{where } A \in \Gamma)$$

$$\frac{\Gamma \vdash A \quad \Gamma \vdash B}{\Gamma \vdash A \wedge B} \wedge\text{-}I \qquad \frac{\Gamma \vdash A \wedge B}{\Gamma \vdash A} \wedge\text{-}EL \qquad \frac{\Gamma \vdash A \wedge B}{\Gamma \vdash B} \wedge\text{-}ER$$

$$\frac{A, \Gamma \vdash B}{\Gamma \vdash A \rightarrow B} \rightarrow\text{-}I \qquad \frac{\Gamma \vdash A \rightarrow B \quad \Gamma \vdash A}{\Gamma \vdash B} \rightarrow\text{-}E$$

Two representations equivalent. Sequent notation seems simpler in practice[61].

---

[60] The judgement $(\Gamma \vdash \phi)$ means that we can derive ϕ from the assumptions in Γ using certain rules. As, explained in the previous lecture, one can make such judgements the central objects of the deductive system.

[61] In particular, the sequent style notation is more amenable to automation, and thus it is closer to what happens in Isabelle.

Example: Refinement Style with Metavariables

$$\cfrac{\cfrac{\cfrac{}{A \wedge (B \wedge C) \vdash A \wedge\, ?X}}{A \wedge (B \wedge C) \vdash A} \qquad \cfrac{\cfrac{\cfrac{}{A \wedge (B \wedge C) \vdash\, ?Z \wedge (?Y \wedge C)}}{A \wedge (B \wedge C) \vdash (?Y \wedge C)}}{A \wedge (B \wedge C) \vdash C}}{\cfrac{A \wedge (B \wedge C) \vdash A \wedge C}{\vdash A \wedge (B \wedge C) \to A \wedge C}}$$

Solution for $?Z = A$, $?Y = B$ and $?X = (B \wedge C)$.
We went through this example in detail last lecture.

Comments about Refinement

This crazy way of carrying out proofs is the (standard) Isabelle-way!

- Refinement style means we work from goals to axioms

- Metavariables used to delay commitments

Isabelle allows other refinements/alternatives too (see labs).

4 First-Order Logic

4.1 First-Order Logic: Overview

In propositional logic, formulae are Boolean[62] combinations of propositions. This will remain important for modeling simple patterns of reasoning.

An atomic proposition is just a letter (variable). All one can say about it is that it is true or false. E.g. it is meaningless to say "A and B state something similar". Also, infinity plays no role.

---

[62]The set (or "type") $bool$ contains the two truth values $True, False$. A propositional formula containing n variables can be viewed as a function $bool^n \rightarrow bool$. For each combination of values $True, False$ for the variables, the whole formula assumes the value $True$ or $False$.

First-Order Logic: the Essence

In first-order logic, an atom(ic proposition) says that "things" have certain "properties"[63]. Infinitely many "things" can be denoted, hence infinitely many atoms generated and distinguished. Comparisons of atoms become meaningful: "Tim is a boy" and "Carl is a boy" state something similar.

Example reasoning: "Tim is a boy"; "boys don't cry"; hence "Tim doesn't cry".

Further reading: [vD80], [Tho91, chapter 1].

---

[63]In propositional logic, there is no notation for writing "thing x has property p" or "things x and y are related as follows" or for denoting the "thing obtained from thing x by applying some operation".

In particular, no statement about all elements of a possibly infinite domain can be expressed in propositional logic, since each formula involves only finitely many different variables, and up to equivalence and for a set containing n variables, there are only finitely many (to be precise $2^{(2^n)}$) different propositional formulae.

Variables: Intuition

In first-order logic, we talk about "things" that have certain "properties".

A variable in first-order logic stands for a "thing".

Variables: Intuition

In first-order logic, we talk about "things" that have certain "properties".

A variable in first-order logic stands for a "thing".

This is in contrast to propositional logic where variables stand for propositions.

It is common to use letters x, y, z for variables.

Predicates: Intuition

A predicate denotes a property/relation.

$p(x) \equiv x$ is a prime number $d(x, y) \equiv x$ is divisible by y

Predicates: Intuition

A predicate denotes a property/relation.

$$p(x) \equiv x \text{ is a prime number} \quad d(x, y) \equiv x \text{ is divisible by } y$$

Propositional connectives are used to build statements

- x is a prime and y or z is divisible by x

Predicates: Intuition

A predicate denotes a property/relation.

$p(x) \equiv x$ is a prime number $\quad d(x, y) \equiv x$ is divisible by y

Propositional connectives are used to build statements

- x is a prime and y or z is divisible by x

$$p(x) \wedge (d(y, x) \vee d(z, x))$$

Predicates: Intuition

A predicate denotes a property/relation.

$p(x) \equiv x$ is a prime number $\quad d(x, y) \equiv x$ is divisible by y

Propositional connectives are used to build statements

- x is a prime and y or z is divisible by x

$$p(x) \wedge (d(y, x) \vee d(z, x))$$

- x is a man and y is a woman and x loves y but not vice versa

Predicates: Intuition

A predicate denotes a property/relation.

$p(x) \equiv x$ is a prime number $d(x, y) \equiv x$ is divisible by y

Propositional connectives are used to build statements

- x is a prime and y or z is divisible by x

$$p(x) \wedge (d(y, x) \vee d(z, x))$$

- x is a man and y is a woman and x loves y but not vice versa

$$m(x) \wedge w(y) \wedge l(x, y) \wedge \neg l(y, x)$$

Predicates: Intuition (2)

We can represent only "abstractions" of these in propositional logic, e.g., $p \wedge (d_1 \vee d_2)$ could be an abstraction of $p(x) \wedge (d(y,x) \vee d(z,x))$.

Here p stands for "x is a prime" and d_1 stands for "y is divisible by x".

But the sense in which $p(x)$, $d(y,x)$, $d(z,x)$ state something similar is lost. What it means to be divisible or to be a prime cannot be expressed.

Functions: Intuition

- A constant stands for a "fixed thing"[64] in a domain[65].

---

[64]As opposed to a variable which also stands for a "thing". This distinction will become clear soon.

[65]For example, the set of integers, the set of characters, the set of people, you name it!

Any set of "things" that we want to reason about.

[66]$\mathbb{N}$ denotes the natural numbers.

[67]So a function symbol f denotes an operation that takes n "things" and returns a "thing". $f(t_1, \ldots, t_n)$ is a "thing" that depends on "things" $t_1, \ldots, t_n$.

The generic notation for function application is like this: $f(t_1, \ldots, t_n)$, but the brackets are omitted for nullary functions ($=$ constants), and many common function symbols like $+$ are denoted infix, so we write $0+0$ instead of $+(0, 0)$. Another common notation is prefix notation without brackets, as in -2. There are also other notations.

Functions: Intuition

- A constant stands for a "fixed thing"[64] in a domain[65].

- More generally, a function of arity n expresses an n-ary operation over some domain, e.g.

Function	arity	expresses ...
0		
s		
$+$		

---

[64] As opposed to a variable which also stands for a "thing". This distinction will become clear soon.

[65] For example, the set of integers, the set of characters, the set of people, you name it!

Any set of "things" that we want to reason about.

[66] $\mathbb{N}$ denotes the natural numbers.

[67] So a function symbol f denotes an operation that takes n "things" and returns a "thing". $f(t_1, \ldots, t_n)$ is a "thing" that depends on "things" $t_1, \ldots, t_n$.

The generic notation for function application is like this: $f(t_1, \ldots, t_n)$, but the brackets are omitted for nullary functions ($=$ constants), and many common function symbols like $+$ are denoted infix, so we write $0 + 0$ instead of $+(0, 0)$. Another common notation is prefix notation without brackets, as in -2. There are also other notations.

Functions: Intuition

- A constant stands for a "fixed thing"[64] in a domain[65].

- More generally, a function of arity n expresses an n-ary operation over some domain, e.g.

Function	arity	expresses ...
0	nullary	
s	unary	
$+$	binary	

---

[64]As opposed to a variable which also stands for a "thing". This distinction will become clear soon.

[65]For example, the set of integers, the set of characters, the set of people, you name it!

Any set of "things" that we want to reason about.

[66]$\mathbb{N}$ denotes the natural numbers.

[67]So a function symbol f denotes an operation that takes n "things" and returns a "thing". $f(t_1, \ldots, t_n)$ is a "thing" that depends on "things" $t_1, \ldots, t_n$.

The generic notation for function application is like this: $f(t_1, \ldots, t_n)$, but the brackets are omitted for nullary functions ($=$ constants), and many common function symbols like $+$ are denoted infix, so we write $0+0$ instead of $+(0,0)$. Another common notation is prefix notation without brackets, as in -2. There are also other notations.

Functions: Intuition

- A constant stands for a "fixed thing"[64] in a domain[65].

- More generally, a function of arity n expresses an n-ary operation over some domain, e.g.

Function	arity	expresses ...
0	nullary	number "0"
s	unary	successor in $\mathbb{N}$[66]
$+$	binary	function plus in $\mathbb{N}$

  Note special notations[67]: infix, prefix, etc.

---

[64]As opposed to a variable which also stands for a "thing". This distinction will become clear soon.

[65]For example, the set of integers, the set of characters, the set of people, you name it!

Any set of "things" that we want to reason about.

[66]$\mathbb{N}$ denotes the natural numbers.

[67]So a function symbol f denotes an operation that takes n "things" and returns a "thing". $f(t_1, \ldots, t_n)$ is a "thing" that depends on "things" $t_1, \ldots, t_n$.

The generic notation for function application is like this: $f(t_1, \ldots, t_n)$, but the brackets are omitted for nullary functions ($=$ constants), and many common function symbols like $+$ are denoted infix, so we write $0 + 0$ instead of $+(0,0)$. Another common notation is prefix notation without brackets, as in -2. There are also other notations.

Quantifiers: Intuition

- A variable stands for "some[68] thing" in a domain of discourse. Quantifiers $\forall, \exists$ are used to speak about all or some members of this domain.

---

[68]Just like a constant, a variable stands for a "thing".

The most important difference between a constant and a variable is that one can quantify over a variable, so one can make statements such as "for all x ..." or "there exists x such that ...".

[69]Intuitively, satisfiable means "can be made true" and valid means "always true".

More formally, this will be defined later.

Quantifiers: Intuition

- A variable stands for "some[68] thing" in a domain of discourse. Quantifiers $\forall, \exists$ are used to speak about all or some members of this domain.

- Examples: Are they satisfiable? valid?[69]

$$\forall x.\, \exists y.\, y * 2 = x$$

---

[68]Just like a constant, a variable stands for a "thing".

The most important difference between a constant and a variable is that one can quantify over a variable, so one can make statements such as "for all x ..." or "there exists x such that ...".

[69]Intuitively, satisfiable means "can be made true" and valid means "always true".

More formally, this will be defined later.

Quantifiers: Intuition

- A variable stands for "some[68] thing" in a domain of discourse. Quantifiers $\forall, \exists$ are used to speak about all or some members of this domain.

- Examples: Are they satisfiable? valid?[69]

$$\forall x.\, \exists y.\, y * 2 = x \quad \text{true for rationals}$$

---

[68]Just like a constant, a variable stands for a "thing".

The most important difference between a constant and a variable is that one can quantify over a variable, so one can make statements such as "for all x ..." or "there exists x such that ...".

[69]Intuitively, satisfiable means "can be made true" and valid means "always true".

More formally, this will be defined later.

Quantifiers: Intuition

- A variable stands for "some[68] thing" in a domain of discourse. Quantifiers $\forall, \exists$ are used to speak about all or some members of this domain.

- Examples: Are they satisfiable? valid?[69]

$$\forall x.\, \exists y.\, y * 2 = x \quad \text{true for rationals}$$
$$x < y \rightarrow \exists z.\, x < z \wedge z < y$$

---

[68]Just like a constant, a variable stands for a "thing".

The most important difference between a constant and a variable is that one can quantify over a variable, so one can make statements such as "for all x ..." or "there exists x such that ...".

[69]Intuitively, satisfiable means "can be made true" and valid means "always true".

More formally, this will be defined later.

Quantifiers: Intuition

- A variable stands for "some[68] thing" in a domain of discourse. Quantifiers $\forall, \exists$ are used to speak about all or some members of this domain.

- Examples: Are they satisfiable? valid?[69]

$$\forall x.\, \exists y.\, y * 2 = x \quad \text{true for rationals}$$
$$x < y \rightarrow \exists z.\, x < z \wedge z < y \quad \text{true for any dense order}$$

---

[68]Just like a constant, a variable stands for a "thing".

The most important difference between a constant and a variable is that one can quantify over a variable, so one can make statements such as "for all x ..." or "there exists x such that ...".

[69]Intuitively, satisfiable means "can be made true" and valid means "always true".

More formally, this will be defined later.

Quantifiers: Intuition

- A variable stands for "some[68] thing" in a domain of discourse. Quantifiers $\forall, \exists$ are used to speak about all or some members of this domain.

- Examples: Are they satisfiable? valid?[69]

$$\forall x.\, \exists y.\, y * 2 = x \quad \text{true for rationals}$$
$$x < y \rightarrow \exists z.\, x < z \wedge z < y \quad \text{true for any dense order}$$
$$\exists x.\, x \neq 0$$

---

[68] Just like a constant, a variable stands for a "thing".

The most important difference between a constant and a variable is that one can quantify over a variable, so one can make statements such as "for all x ..." or "there exists x such that ...".

[69] Intuitively, satisfiable means "can be made true" and valid means "always true".

More formally, this will be defined later.

Quantifiers: Intuition

- A variable stands for "some[68] thing" in a domain of discourse. Quantifiers $\forall, \exists$ are used to speak about all or some members of this domain.

- Examples: Are they satisfiable? valid?[69]

$$\forall x.\, \exists y.\, y * 2 = x \quad \text{true for rationals}$$
$$x < y \rightarrow \exists z.\, x < z \wedge z < y \quad \text{true for any dense order}$$
$$\exists x.\, x \neq 0 \quad \text{true for domains with more than one element}$$

---

[68] Just like a constant, a variable stands for a "thing".

The most important difference between a constant and a variable is that one can quantify over a variable, so one can make statements such as "for all x ..." or "there exists x such that ...".

[69] Intuitively, satisfiable means "can be made true" and valid means "always true".

More formally, this will be defined later.

Quantifiers: Intuition

- A variable stands for "some[68] thing" in a domain of discourse. Quantifiers $\forall, \exists$ are used to speak about all or some members of this domain.

- Examples: Are they satisfiable? valid?[69]

$$\forall x. \exists y. y * 2 = x \quad \text{true for rationals}$$
$$x < y \rightarrow \exists z. x < z \wedge z < y \quad \text{true for any dense order}$$
$$\exists x. x \neq 0 \quad \text{true for domains with more than one element}$$
$$(\forall x. p(x, x)) \rightarrow p(a, a)$$

---

[68] Just like a constant, a variable stands for a "thing".

The most important difference between a constant and a variable is that one can quantify over a variable, so one can make statements such as "for all x ..." or "there exists x such that ...".

[69] Intuitively, satisfiable means "can be made true" and valid means "always true".

More formally, this will be defined later.

Quantifiers: Intuition

- A variable stands for "some[68] thing" in a domain of discourse. Quantifiers $\forall, \exists$ are used to speak about all or some members of this domain.

- Examples: Are they satisfiable? valid?[69]

$$\forall x.\, \exists y.\, y * 2 = x \quad \text{true for rationals}$$
$$x < y \rightarrow \exists z.\, x < z \wedge z < y \quad \text{true for any dense order}$$
$$\exists x.\, x \neq 0 \quad \text{true for domains with more than one element}$$
$$(\forall x.\, p(x, x)) \rightarrow p(a, a) \quad \text{valid}$$

---

[68] Just like a constant, a variable stands for a "thing".

The most important difference between a constant and a variable is that one can quantify over a variable, so one can make statements such as "for all x ..." or "there exists x such that ...".

[69] Intuitively, satisfiable means "can be made true" and valid means "always true".

More formally, this will be defined later.

4.2 First-Order Logic: Syntax

- Two syntactic categories: terms[70] and formulae

- A first-order language[71] is characterized by giving a finite collection of function symbols $\mathcal{F}$ and predicate symbols $\mathcal{P}$ as well as a set Var of variables.

4.2 First-Order Logic: Syntax

- Two syntactic categories: terms[70] and formulae

- A first-order language[71] is characterized by giving a finite collection of function symbols $\mathcal{F}$ and predicate symbols $\mathcal{P}$ as well as a set Var of variables.

- Sometimes write f^i (or p^i) to indicate that function symbol f (or predicate symbol p) has arity $i \in \mathbb{N}$.

4.2 First-Order Logic: Syntax

- Two syntactic categories: terms[70] and formulae

- A first-order language[71] is characterized by giving a finite collection of function symbols $\mathcal{F}$ and predicate symbols $\mathcal{P}$ as well as a set Var of variables.

- Sometimes write f^i (or p^i) to indicate that function symbol f (or predicate symbol p) has arity $i \in \mathbb{N}$.

- One often calls the pair $\langle \mathcal{F}, \mathcal{P} \rangle$ a signature.

---

[70]We have already learned about the syntactic category of formulae last lecture.

A term is an expression that stands for a "thing".

Intuitively, this is what first-order logic is about: We have terms that stand for "things" and formulae that stand for statements/propositions about those "things".

But couldn't a statement also be a "thing"? And couldn't a "thing" depend on a statement?

In first-order logic: no!

[71]There isn't simply the language of first-order logic! Rather, the definition of a first-order language is parametrised by giving a $\mathcal{F}$ and a $\mathcal{P}$. Each symbol in $\mathcal{F}$ and $\mathcal{P}$ must have an associated arity, i.e., the number of arguments the function or predicate takes. This could be formalized by saying that the elements of $\mathcal{F}$ are pairs of the form f/n, where f is the symbol itself and n, and likewise for $\mathcal{P}$. All that matters is that it is specified in some unambiguous way what the arity of each symbols is.

Terms in First-Order Logic

$Term$[72], the set of terms, is the smallest set where

1. $x \in Term$ if $x \in Var$, and

2. $f^n(t_1, \ldots, t_n) \in Term$ if $f^n \in \mathcal{F}$ and $t_j \in Term$, for all $1 \leq j \leq n$.

One often calls the pair $\langle \mathcal{F}, \mathcal{P} \rangle$ a signature. Generally, a signature specifies the "fixed symbols" (as opposed to variables) of a particular logic language.

Strictly speaking, a first-order language is also parametrised by giving a set of variables Var, but this is inessential. Var is usually assumed to be a countably infinite set of symbols, and the particular choice of names of these symbols is not relevant.

[72] $Term$ and $Form$ together make up a first-order language. Note that strictly speaking, $Term$ and $Form$ depend on the signature, but we always assume that the signature is clear from the context.

Formulae in First-Order Logic

$Form$, the set of formulae, is the smallest set where

1. $\bot \in Form$,

2. $p^n(t_1, \ldots, t_n) \in Form$ if $p^n \in \mathcal{P}$ and $t_j \in Term$, for all $1 \le j \le n$,

3. $\neg\phi \in Form$ if $\phi \in Form$,

4. $(\phi \circ \psi) \in Form$ if $\phi \in Form$, $\psi \in Form$ and $\circ \in \{\wedge, \vee, \rightarrow\}$,

5. $\mathsf{Q}x.\,\phi \in Form$ if $\phi \in Form$ and $x \in Var$ and $\mathsf{Q} \in \{\forall, \exists\}$.

Formulae as in point 2 are called atoms.
Note quantifier scoping[73].

---

[73]We adopt the convention that the scope of a quantifier extends as much as possible to the right, e.g.

$$\forall x.p(x) \vee q(x)$$

is

$$\forall x.(p(x) \vee q(x))$$

and not

$$(\forall x.p(x)) \vee q(x)$$

This is a matter of dispute and other conventions are around, but the one we adopt here corresponds to Isabelle.

Compare this to the precedences and associativity in propositional logic.

Variable Occurrences

- All occurrences of a variable in a formula[74] are bound or free or binding.

  A variable x in a formula ϕ is bound if x occurs within a subformula of ϕ of the form $\exists x.\psi$ or $\forall x.\psi$.

- Example:

  $$(q(x) \vee \exists x.\, \forall y.\, p(f(x), z) \wedge q(y)) \vee \forall x.\, r(x, z, g(x))$$

Which are bound?

---

[74]All occurrences of a variable in a term or formula are bound or free or binding. These notions are defined by induction on the structure of terms/formulae. This is why the following definition is along the lines of our definition of terms and formulae.

1. The (only) occurrence of x in the term x is a free occurrence of x in x;

2. the free occurrences of x in $f(t_1, \ldots, t_n)$ are the free occurrences of x in $t_1, \ldots, t_n$;

3. there are no free occurrences of x in $\bot$;

4. the free occurrences of x in $p(t_1, \ldots, t_n)$ are the free occurrences of x in $t_1, \ldots, t_n$;

5. the free occurrences of x in $\neg\phi$ are the free occurrences of x in ϕ;

6. the free occurrences of x in $\psi \circ \phi$ are the free occurrences of x in ψ and the free occurrences of x in ϕ ($\circ \in \{\wedge, \vee, \rightarrow\}$);

Variable Occurrences

- All occurrences of a variable in a formula[74] are bound or free or binding.

  A variable x in a formula ϕ is bound if x occurs within a subformula of ϕ of the form $\exists x.\psi$ or $\forall x.\psi$.

- Example:

$$(q(x) \vee \exists x.\, \forall y.\, p(f(x), z) \wedge q(y)) \vee \forall x.\, r(x, z, g(x))$$

Which are bound? Which are free?

---

[74]All occurrences of a variable in a term or formula are bound or free or binding. These notions are defined by induction on the structure of terms/formulae. This is why the following definition is along the lines of our definition of terms and formulae.

1. The (only) occurrence of x in the term x is a free occurrence of x in x;

2. the free occurrences of x in $f(t_1, \ldots, t_n)$ are the free occurrences of x in $t_1, \ldots, t_n$;

3. there are no free occurrences of x in $\bot$;

4. the free occurrences of x in $p(t_1, \ldots, t_n)$ are the free occurrences of x in $t_1, \ldots, t_n$;

5. the free occurrences of x in $\neg\phi$ are the free occurrences of x in ϕ;

6. the free occurrences of x in $\psi \circ \phi$ are the free occurrences of x in ψ and the free occurrences of x in ϕ ($\circ \in \{\wedge, \vee, \rightarrow\}$);

Variable Occurrences

- All occurrences of a variable in a formula[74] are bound or free or binding.

  A variable x in a formula ϕ is bound if x occurs within a subformula of ϕ of the form $\exists x.\psi$ or $\forall x.\psi$.

- Example:

$$(q(x) \vee \exists x.\, \forall y.\, p(f(x), z) \wedge q(y)) \vee \forall x.\, r(x, z, g(x))$$

Which are bound? Which are free? Which are binding?

---

[74]All occurrences of a variable in a term or formula are bound or free or binding. These notions are defined by induction on the structure of terms/formulae. This is why the following definition is along the lines of our definition of terms and formulae.

1. The (only) occurrence of x in the term x is a free occurrence of x in x;

2. the free occurrences of x in $f(t_1, \ldots, t_n)$ are the free occurrences of x in $t_1, \ldots, t_n$;

3. there are no free occurrences of x in $\bot$;

4. the free occurrences of x in $p(t_1, \ldots, t_n)$ are the free occurrences of x in $t_1, \ldots, t_n$;

5. the free occurrences of x in $\neg\phi$ are the free occurrences of x in ϕ;

6. the free occurrences of x in $\psi \circ \phi$ are the free occurrences of x in ψ and the free occurrences of x in ϕ ($\circ \in \{\wedge, \vee, \rightarrow\}$);

Variable Occurrences

- All occurrences of a variable in a formula[74] are bound or free or binding.

  A variable x in a formula ϕ is bound if x occurs within a subformula of ϕ of the form $\exists x.\psi$ or $\forall x.\psi$.

- Example:

$$(q(x) \vee \exists x.\, \forall y.\, p(f(x), z) \wedge q(y)) \vee \forall x.\, r(x, z, g(x))$$

Which are bound? Which are free? Which are binding?

There will be an exercise.

---

[74]All occurrences of a variable in a term or formula are bound or free or binding. These notions are defined by induction on the structure of terms/formulae. This is why the following definition is along the lines of our definition of terms and formulae.

1. The (only) occurrence of x in the term x is a free occurrence of x in x;

2. the free occurrences of x in $f(t_1, \ldots, t_n)$ are the free occurrences of x in $t_1, \ldots, t_n$;

3. there are no free occurrences of x in $\bot$;

4. the free occurrences of x in $p(t_1, \ldots, t_n)$ are the free occurrences of x in $t_1, \ldots, t_n$;

5. the free occurrences of x in $\neg\phi$ are the free occurrences of x in ϕ;

6. the free occurrences of x in $\psi \circ \phi$ are the free occurrences of x in ψ and the free occurrences of x in ϕ ($\circ \in \{\wedge, \vee, \rightarrow\}$);

4.3 First-Order Logic: Semantics

A structure[75] is a pair $\mathcal{A} = \langle U_{\mathcal{A}}, I_{\mathcal{A}} \rangle$ where $U_{\mathcal{A}}$ is an nonempty set, the universe, and $I_{\mathcal{A}}$ is a mapping where

1. $I_{\mathcal{A}}(f^n)$ is an n-ary (total) function on $U_{\mathcal{A}}$, for $f^n \in \mathcal{F}$,

2. $I_{\mathcal{A}}(p^n)$ is an n-ary relation on $U_{\mathcal{A}}$, for $p^n \in \mathcal{P}$, and

3. $I_{\mathcal{A}}(x)$ is an element of $U_{\mathcal{A}}$, for each $x \in Var$.

As shorthand, write $p^{\mathcal{A}}$ for $I_{\mathcal{A}}(p)$, etc.

---

7. the free occurrences of x in $\forall y.\, \psi$, where $y \neq x$, are the free occurrences of x in ψ; likewise for $\exists$;

8. x has no free occurrences in $\forall x.\, \psi$; in $\forall x.\, \psi$, the (outermost) $\forall$ binds all free occurrences of x in ψ; the occurrence of x next to $\forall$ is a binding occurrence of x; likewise for $\exists$.

A variable occurrence is bound if it is not free and not binding.

We also define

$$FV(\phi) := \{x \mid x \text{ has a free occurrence in } \phi\}$$

[75]As usual, there isn't just one way of formalizing things, and so we now explain some other notions that you may have heard in the context of semantics for first-order logic.

A universe is sometimes also called domain.

As you saw, a structure gives a meaning to functions, pred-

The Value of Terms

Let $\mathcal{A}$ be a structure. We define the value of a term t under $\mathcal{A}$, written $\mathcal{A}(t)$, as

1. $\mathcal{A}(x) = x^{\mathcal{A}}$, for $x \in Var$, and

2. $\mathcal{A}(f(t_1, \ldots, t_n)) = f^{\mathcal{A}}(\mathcal{A}(t_1), \ldots, \mathcal{A}(t_n))$.

---

icates, and variables.

An alternative formalization is to have three different mappings for this purpose:

1. an algebra gives a meaning to the function symbols (more precisely, an algebra is a pair consisting of a domain and a mapping giving a meaning to the function symbols);

2. in addition, an interpretation gives a meaning also to the predicate symbols;

3. a variable assignment, also called valuation, gives a meaning to the variables.

As before, we assume that the signature is clear from the context. Strictly speaking, we should say "structure for a particular signature".

Details can be found in any textbook on logic [vD80].

The Value of Formulae

We define the (truth-)value of the formula ϕ under $\mathcal{A}$, written $\mathcal{A}(\phi)$, as

$$\mathcal{A}(p(t_1, \ldots, t_n)) = \begin{cases} 1 & \text{if } (\mathcal{A}(t_1), \ldots \mathcal{A}(t_n)) \in p^{\mathcal{A}} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{A}(\forall x. \phi) = \begin{cases} 1 & \text{if for all } u \in U_{\mathcal{A}}, \mathcal{A}_{[x/u]}{}^{76}(\phi) = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{A}(\exists x. \phi) = \begin{cases} 1 & \text{if for some } u \in U_{\mathcal{A}}, \mathcal{A}_{[x/u]}(\phi) = 1 \\ 0 & \text{otherwise} \end{cases}$$

Rest as for propositional logic.

Models

- If $\mathcal{A}(\phi) = 1$, we write $\mathcal{A} \models \phi$ and say ϕ is true in $\mathcal{A}$ or $\mathcal{A}$ is a model of ϕ.

---

[77] A structure is suitable for ϕ if it defines meanings for the signature of ϕ, i.e., for the symbols that occur in ϕ. Of course, these meanings must also respect the arities, so an n-ary function symbols must be interpreted as an n-ary function. Without explicitly mentioning it, we always assume that structures are suitable.

[78] If you are happy with the definition of a model just given, this is fine. But if you are confused because you remember a different definition from your previous studies of logic, then these comments may help.

As explained before, it is common to distinguish an interpretation, which gives a meaning to the symbols in the signature, from an assignment, which gives a meaning to the variables. Let us use $\mathcal{I}$ to denote an interpretation and A to denote an assignment.

Recall that we wrote $\mathcal{A}(.)$ for the meaning of a term or formula. In the alternative terminology, we write $\mathcal{I}(A)(.)$ instead. This makes sense since in the alternative terminology,

Models

- If $\mathcal{A}(\phi) = 1$, we write $\mathcal{A} \models \phi$ and say ϕ is true in $\mathcal{A}$ or $\mathcal{A}$ is a model of ϕ.

- If every suitable structure[77] is a model, we write $\models \phi$ and say ϕ is valid or ϕ is a tautology.

---

[77]A structure is suitable for ϕ if it defines meanings for the signature of ϕ, i.e., for the symbols that occur in ϕ. Of course, these meanings must also respect the arities, so an n-ary function symbols must be interpreted as an n-ary function. Without explicitly mentioning it, we always assume that structures are suitable.

[78]If you are happy with the definition of a model just given, this is fine. But if you are confused because you remember a different definition from your previous studies of logic, then these comments may help.

As explained before, it is common to distinguish an interpretation, which gives a meaning to the symbols in the signature, from an assignment, which gives a meaning to the variables. Let us use $\mathcal{I}$ to denote an interpretation and A to denote an assignment.

Recall that we wrote $\mathcal{A}(.)$ for the meaning of a term or formula. In the alternative terminology, we write $\mathcal{I}(A)(.)$ instead. This makes sense since in the alternative terminology,

Models

- If $\mathcal{A}(\phi) = 1$, we write $\mathcal{A} \models \phi$ and say ϕ is true in $\mathcal{A}$ or $\mathcal{A}$ is a model of ϕ.

- If every suitable structure[77] is a model, we write $\models \phi$ and say ϕ is valid or ϕ is a tautology.

- If there is at least one model for ϕ, then ϕ is satisfiable.

---

[77]A structure is suitable for ϕ if it defines meanings for the signature of ϕ, i.e., for the symbols that occur in ϕ. Of course, these meanings must also respect the arities, so an n-ary function symbols must be interpreted as an n-ary function. Without explicitly mentioning it, we always assume that structures are suitable.

[78]If you are happy with the definition of a model just given, this is fine. But if you are confused because you remember a different definition from your previous studies of logic, then these comments may help.

As explained before, it is common to distinguish an interpretation, which gives a meaning to the symbols in the signature, from an assignment, which gives a meaning to the variables. Let us use $\mathcal{I}$ to denote an interpretation and A to denote an assignment.

Recall that we wrote $\mathcal{A}(.)$ for the meaning of a term or formula. In the alternative terminology, we write $\mathcal{I}(A)(.)$ instead. This makes sense since in the alternative terminology,

Models

- If $\mathcal{A}(\phi) = 1$, we write $\mathcal{A} \models \phi$ and say ϕ is true in $\mathcal{A}$ or $\mathcal{A}$ is a model of ϕ.

- If every suitable structure[77] is a model, we write $\models \phi$ and say ϕ is valid or ϕ is a tautology.

- If there is at least one model for ϕ, then ϕ is satisfiable.

- If there is no model for ϕ, then ϕ is contradictory.

---

[77]A structure is suitable for ϕ if it defines meanings for the signature of ϕ, i.e., for the symbols that occur in ϕ. Of course, these meanings must also respect the arities, so an n-ary function symbols must be interpreted as an n-ary function. Without explicitly mentioning it, we always assume that structures are suitable.

[78]If you are happy with the definition of a model just given, this is fine. But if you are confused because you remember a different definition from your previous studies of logic, then these comments may help.

As explained before, it is common to distinguish an interpretation, which gives a meaning to the symbols in the signature, from an assignment, which gives a meaning to the variables. Let us use $\mathcal{I}$ to denote an interpretation and A to denote an assignment.

Recall that we wrote $\mathcal{A}(.)$ for the meaning of a term or formula. In the alternative terminology, we write $\mathcal{I}(A)(.)$ instead. This makes sense since in the alternative terminology,

Models

- If $\mathcal{A}(\phi) = 1$, we write $\mathcal{A} \models \phi$ and say ϕ is true in $\mathcal{A}$ or $\mathcal{A}$ is a model of ϕ.

- If every suitable structure[77] is a model, we write $\models \phi$ and say ϕ is valid or ϕ is a tautology.

- If there is at least one model for ϕ, then ϕ is satisfiable.

- If there is no model for ϕ, then ϕ is contradictory.

There is also more differentiated terminology.[78]

---

[77]A structure is suitable for ϕ if it defines meanings for the signature of ϕ, i.e., for the symbols that occur in ϕ. Of course, these meanings must also respect the arities, so an n-ary function symbols must be interpreted as an n-ary function. Without explicitly mentioning it, we always assume that structures are suitable.

[78]If you are happy with the definition of a model just given, this is fine. But if you are confused because you remember a different definition from your previous studies of logic, then these comments may help.

As explained before, it is common to distinguish an interpretation, which gives a meaning to the symbols in the signature, from an assignment, which gives a meaning to the variables. Let us use $\mathcal{I}$ to denote an interpretation and A to denote an assignment.

Recall that we wrote $\mathcal{A}(.)$ for the meaning of a term or formula. In the alternative terminology, we write $\mathcal{I}(A)(.)$ instead. This makes sense since in the alternative terminology,

An Example

$$\forall x.\, p(x, s(x))$$

An Example

$$\forall x.\, p(x, s(x))$$

$\mathcal{I}$ and A together contain the same information as $\mathcal{A}$ in the original terminology. We define:

- For a given $\mathcal{I}$, we say that ϕ is satisfiable in $\mathcal{I}$ if there exists an A so that $\mathcal{I}(A)(\phi) = 1$;

- for a given $\mathcal{I}$, we write $\mathcal{I} \models \phi$ and say ϕ is true in $\mathcal{I}$ or $\mathcal{I}$ is a model of ϕ, if for all A, we have $\mathcal{I}(A)(\phi) = 1$;

- we say ϕ is satisfiable if there exists an $\mathcal{I}$ so that ϕ is satisfiable in $\mathcal{I}$;

- we write $\models \phi$ and say ϕ is valid if for every (suitable) $\mathcal{I}$, we have $\mathcal{I} \models \phi$.

Note that satisfiable (without "for ...") and valid mean the same thing in both terminologies, whereas true in ... means slightly different things, since a structure is not the same thing as an interpretation.

A model[79]:

$$\begin{aligned}
U_{\mathcal{A}} &= \mathbb{N} \\
p^{\mathcal{A}} &= \{(m,n) \mid m <^{80} n\} \\
s^{\mathcal{A}}(x) &= x+1
\end{aligned}$$

---

[79] It is true that for all numbers n, n is less than $n+1$.

[80] In logic, we insist on the distinction between syntax and semantics. In particular, we set up the formalism so that the syntax is fixed first and then the semantics, and so there could be different semantics for the same syntax.

But the dilemma is that once we want to give a particular semantics, we can only do so using again some kind of language, hence syntax. This is usually natural language interspersed with usual mathematical notation such as $<$, $+$ etc.

Some people try to mark the distinction between syntax and semantics somehow, e.g., by saying 0 is a constant that could mean anything, whereas $\mathbf{0}$ is the number zero as it exists in the mathematical world.

When we give semantics, the symbols $<$, $+$, and 1 have their usual mathematical meanings. The function that maps x to $x+1$ is also called successor function. Of course, when we write $m < n$, we assume that $m, n \in \mathbb{N}$, in this context.

[81] The identity function maps every object to itself.

A model[79]:　　　　　　　　　Not a model[81]:

$$U_{\mathcal{A}} = \mathbb{N} \qquad\qquad\qquad U_{\mathcal{A}} = \{\mathsf{a}, \mathsf{b}, \mathsf{c}\}$$
$$p^{\mathcal{A}} = \{(m,n) \mid m <{}^{80}n\} \quad p^{\mathcal{A}} = \{(\mathsf{a},\mathsf{b}), (\mathsf{a},\mathsf{c})\}$$
$$s^{\mathcal{A}}(x) = x+1 \qquad\qquad s^{\mathcal{A}} = \text{"the identity function"}$$

---

[79] It is true that for all numbers n, n is less than $n+1$.

[80] In logic, we insist on the distinction between syntax and semantics. In particular, we set up the formalism so that the syntax is fixed first and then the semantics, and so there could be different semantics for the same syntax.

But the dilemma is that once we want to give a particular semantics, we can only do so using again some kind of language, hence syntax. This is usually natural language interspersed with usual mathematical notation such as $<$, $+$ etc.

Some people try to mark the distinction between syntax and semantics somehow, e.g., by saying 0 is a constant that could mean anything, whereas $\mathbf{0}$ is the number zero as it exists in the mathematical world.

When we give semantics, the symbols $<$, $+$, and 1 have their usual mathematical meanings. The function that maps x to $x+1$ is also called successor function. Of course, when we write $m < n$, we assume that $m, n \in \mathbb{N}$, in this context.

[81] The identity function maps every object to itself.

4.4 Towards a Deductive System

In natural language, quantifiers are often implicit[82]: males don't cry.

4.4 Towards a Deductive System

In natural language, quantifiers are often implicit[82]: all males don't cry.

4.4 Towards a Deductive System

In natural language, quantifiers are often implicit[82]: all males don't cry.

Some phrases in natural language proofs have the flavor of introduction rules.

Take "boys are males" and "males don't cry" implies "boys don't cry": assume an arbitrary boy x; then x is a male; hence x doesn't cry; hence "x is a boy" implies "x doesn't cry" ; since x was arbitrary, we can say this for all x.

4.4 Towards a Deductive System

In natural language, quantifiers are often implicit[82]: all males don't cry.

Some phrases in natural language proofs have the flavor of introduction rules.

Take "boys are males" and "males don't cry" implies "boys don't cry": assume an arbitrary boy x; then x is a male; hence x doesn't cry; hence "x is a boy" implies "x doesn't cry" ($\rightarrow$-I); since x was arbitrary, we can say this for all x. ($\forall$-I). See later.

4.4 Towards a Deductive System

In natural language, quantifiers are often implicit[82]: all males don't cry.

Some phrases in natural language proofs have the flavor of introduction rules.

Take "boys are males" and "males don't cry" implies "boys don't cry": assume an arbitrary boy x; then x is a male; hence x doesn't cry; hence "x is a boy" implies "x doesn't cry" ($\rightarrow$-I); since x was arbitrary, we can say this for all x. ($\forall$-I). See later.

Existential statements are proven by giving a witness.

It is not true that for every character $\alpha \in \{a, b, c\}$, $(\alpha, \alpha) \in \{(a, b), (a, c)\}$. E.g., $(a, a) \notin \{(a, b), (a, c)\}$.

[82]In the statement

$$\text{if } x > 2 \text{ then } x^2 > 4$$

the $\forall$-quantifier is implicit. It should be

$$\text{for all } x, \text{ if } x > 2 \text{ then } x^2 > 4.$$

4.5 First-Order Logic: Deductive System

First-order logic is a generalization of propositional logic. All the rules of propositional logic are "inherited"[83].

But we must introduce rules for the quantifiers.

---

[83]First-order logic inherits all the rules of propositional logic. Note however that the metavariables in the rules now range over first-order formulae.

Universal Quantification ($\forall$): Rules

$$\frac{P(x)}{\forall x.\, P(x)}\ \forall\text{-}I^* \qquad \frac{\forall x.\, P(x)}{P(t)}\ \forall\text{-}E$$

where side condition (also called: proviso or eigenvariable condition) $*$ means: x must be arbitrary.

---

[84]Similarly as in the previous lecture, one should note that P is not a predicate, but rather $P(x)$ is a schematic expression: $P(x)$ stands for any formula, possibly containing occurrences of x.

In the context of $\forall$-E, $P(t)$ stands for a formula where all occurrences of x are replaced by t.

Universal Quantification ($\forall$): Rules

$$\frac{P(x)}{\forall x.\, P(x)}\ \forall\text{-}I^* \qquad \frac{\forall x.\, P(x)}{P(t)}\ \forall\text{-}E$$

where side condition (also called: proviso or eigenvariable condition) $*$ means: x must be arbitrary.

Note that rules are schematic[84].

---

[84]Similarly as in the previous lecture, one should note that P is not a predicate, but rather $P(x)$ is a schematic expression: $P(x)$ stands for any formula, possibly containing occurrences of x.

In the context of $\forall$-E, $P(t)$ stands for a formula where all occurrences of x are replaced by t.

Universal Quantification: Side Condition

What does arbitrary mean? Consider the following "proof"

$$x = 0$$

---

[85]When one has a predicate symbol $=$, it is usual to have a rule that says that $=$ is reflexive.

Don't worry about it at this stage, just take it that we have such a rule. We will look at this later.

[86]The side condition is violated in the proof since in the first $\forall$-I step, x does occur free in $x = 0$.

Universal Quantification: Side Condition

What does arbitrary mean? Consider the following "proof"

$$\frac{x = 0}{\forall x.\, x = 0}\ \forall\text{-}I$$

---

[85]When one has a predicate symbol $=$, it is usual to have a rule that says that $=$ is reflexive.

Don't worry about it at this stage, just take it that we have such a rule. We will look at this later.

[86]The side condition is violated in the proof since in the first $\forall$-I step, x does occur free in $x = 0$.

Universal Quantification: Side Condition

What does arbitrary mean? Consider the following "proof"

$$\cfrac{\cfrac{[x = 0]^1}{\forall x.\, x = 0}\; \forall\text{-}I}{x = 0 \to \forall x.\, x = 0}\; \to\text{-}I^1$$

---

[85]When one has a predicate symbol $=$, it is usual to have a rule that says that $=$ is reflexive.

Don't worry about it at this stage, just take it that we have such a rule. We will look at this later.

[86]The side condition is violated in the proof since in the first $\forall\text{-}I$ step, x does occur free in $x = 0$.

Universal Quantification: Side Condition

What does arbitrary mean? Consider the following "proof"

$$\dfrac{\dfrac{\dfrac{[x = 0]^1}{\forall x.\, x = 0}\ \forall\text{-}I}{x = 0 \to \forall x.\, x = 0}\ \to\text{-}I^1}{\forall x.\, (x = 0 \to \forall x.\, x = 0)}\ \forall\text{-}I$$

---

[85] When one has a predicate symbol $=$, it is usual to have a rule that says that $=$ is reflexive.

Don't worry about it at this stage, just take it that we have such a rule. We will look at this later.

[86] The side condition is violated in the proof since in the first $\forall$-*I* step, x does occur free in $x = 0$.

Universal Quantification: Side Condition

What does arbitrary mean? Consider the following "proof"

$$\cfrac{\cfrac{\cfrac{\cfrac{[x = 0]^1}{\forall x.\, x = 0} \;\forall\text{-}I}{x = 0 \to \forall x.\, x = 0} \;\to\text{-}I^1}{\forall x.\, (x = 0 \to \forall x.\, x = 0)} \;\forall\text{-}I}{0 = 0 \to \forall x.\, x = 0} \;\forall\text{-}E$$

---

[85]When one has a predicate symbol $=$, it is usual to have a rule that says that $=$ is reflexive.

Don't worry about it at this stage, just take it that we have such a rule. We will look at this later.

[86]The side condition is violated in the proof since in the first $\forall$-*I* step, x does occur free in $x = 0$.

Universal Quantification: Side Condition

What does arbitrary mean? Consider the following "proof"

$$\cfrac{\cfrac{\cfrac{\cfrac{[x = 0]^1}{\forall x.\, x = 0}\ \forall\text{-}I}{x = 0 \to \forall x.\, x = 0}\ \to\text{-}I^1}{\forall x.\, (x = 0 \to \forall x.\, x = 0)}\ \forall\text{-}I}{0 = 0 \to \forall x.\, x = 0}\ \forall\text{-}E \qquad \cfrac{}{0 = 0}\ \text{refl}^{85}$$

---

[85]When one has a predicate symbol $=$, it is usual to have a rule that says that $=$ is reflexive.

Don't worry about it at this stage, just take it that we have such a rule. We will look at this later.

[86]The side condition is violated in the proof since in the first $\forall\text{-}I$ step, x does occur free in $x = 0$.

Universal Quantification: Side Condition

What does arbitrary mean? Consider the following "proof"

$$\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{[x = 0]^1}{\forall x.\, x = 0}\ \forall\text{-}I}{x = 0 \to \forall x.\, x = 0}\ \to\text{-}I^1}{\forall x.\, (x = 0 \to \forall x.\, x = 0)}\ \forall\text{-}I}{0 = 0 \to \forall x.\, x = 0}\ \forall\text{-}E \qquad \cfrac{}{0 = 0}\ \textit{refl}^{85}}{\forall x.\, x = 0}\ \to\text{-}E$$

---

[85]When one has a predicate symbol $=$, it is usual to have a rule that says that $=$ is reflexive.

Don't worry about it at this stage, just take it that we have such a rule. We will look at this later.

[86]The side condition is violated in the proof since in the first $\forall$-I step, x does occur free in $x = 0$.

Universal Quantification: Side Condition

What does arbitrary mean? Consider the following "proof"

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[x = 0]^1}{\forall x.\, x = 0}\ \forall\text{-}I}{x = 0 \to \forall x.\, x = 0}\ \to\text{-}I^1}{\forall x.\, (x = 0 \to \forall x.\, x = 0)}\ \forall\text{-}I}{0 = 0 \to \forall x.\, x = 0}\ \forall\text{-}E \qquad \dfrac{}{0 = 0}\ refl^{85}}{\forall x.\, x = 0}\ \to\text{-}E$$

Formal meaning of side condition: x not free in any open assumption on which $P(x)$ depends. Violated![86]

---

[85]When one has a predicate symbol $=$, it is usual to have a rule that says that $=$ is reflexive.

Don't worry about it at this stage, just take it that we have such a rule. We will look at this later.

[86]The side condition is violated in the proof since in the first $\forall$-I step, x does occur free in $x = 0$.

Another Proof? (1)

Is the following a proof? Is the conclusion valid?

$$\cfrac{\cfrac{[\forall x. \,\neg \forall y.\, x = y]^1}{\neg \forall y.\, y = y} \;\forall\text{-}E}{(\forall x. \,\neg \forall y.\, x = y) \rightarrow \neg \forall y.\, y = y} \;\rightarrow\text{-}I^1$$

Another Proof? (1)

Is the following a proof? Is the conclusion valid?

$$\dfrac{\dfrac{[\forall x.\, \neg \forall y.\, x = y]^1}{\neg \forall y.\, y = y}\ \forall\text{-}E}{(\forall x.\, \neg \forall y.\, x = y) \rightarrow \neg \forall y.\, y = y}\ \rightarrow\text{-}I^1$$

Conclusion is not valid.

The formula is false when $U_{\mathcal{A}}$ has at least 2 elements.[87]

Another Proof? (1)

Is the following a proof? Is the conclusion valid?

$$\cfrac{\cfrac{[\forall x.\,\neg\forall y.\,x = y]^1}{\neg\forall y.\,y = y}\;\forall\text{-}E}{(\forall x.\,\neg\forall y.\,x = y) \rightarrow \neg\forall y.\,y = y}\;\rightarrow\text{-}I^1$$

Proof is incorrect.

Reason: Substitution[88] must avoid capturing[89] variables. Replacing x with y in $\forall$-*E* is illegal because y is bound in $\neg\forall y.\,y = y$. This detail concerns substitution (and renaming of bound variables), not $\forall$-E. Exercise

Another Proof? (2)

$\forall x.\, A(x) \land B(x)$

---

[90]In both cases, x does not occur free in $\forall x.\, A(x) \land B(x)$, which is the open assumption on which $A(x)$, respectively $B(x)$, depends.

Another Proof? (2)

$$\frac{\forall x.\, A(x) \land B(x)}{A(x) \land B(x)} \ \forall\text{-}E$$

---

[90]In both cases, x does not occur free in $\forall x.\, A(x) \land B(x)$, which is the open assumption on which $A(x)$, respectively $B(x)$, depends.

Another Proof? (2)

$$\frac{\dfrac{\forall x.\, A(x) \land B(x)}{A(x) \land B(x)} \;\; \forall\text{-}E}{A(x)} \;\; \land\text{-}EL$$

---

[90]In both cases, x does not occur free in $\forall x.\, A(x) \land B(x)$, which is the open assumption on which $A(x)$, respectively $B(x)$, depends.

Another Proof? (2)

$$\frac{\dfrac{\forall x.\, A(x) \land B(x)}{\dfrac{A(x) \land B(x)}{\dfrac{A(x)}{\forall x.\, A(x)}\;\forall\text{-}I}\;\land\text{-}EL}}{}\;\forall\text{-}E$$

---

[90]In both cases, x does not occur free in $\forall x.\, A(x) \land B(x)$, which is the open assumption on which $A(x)$, respectively $B(x)$, depends.

Another Proof? (2)

$$\frac{\dfrac{\forall x.\, A(x) \wedge B(x)}{A(x) \wedge B(x)} \ \forall\text{-}E}{\dfrac{A(x)}{\forall x.\, A(x)} \ \forall\text{-}I} \ \wedge\text{-}EL \qquad \frac{\dfrac{\forall x.\, A(x) \wedge B(x)}{A(x) \wedge B(x)} \ \forall\text{-}E}{\dfrac{B(x)}{\forall x.\, B(x)} \ \forall\text{-}I} \ \wedge\text{-}ER$$

---

[90]In both cases, x does not occur free in $\forall x.\, A(x) \wedge B(x)$, which is the open assumption on which $A(x)$, respectively $B(x)$, depends.

Another Proof? (2)

$$\cfrac{\cfrac{\cfrac{\forall x.\, A(x) \wedge B(x)}{A(x) \wedge B(x)}\ \forall\text{-}E}{A(x)}\ \wedge\text{-}EL}{\forall x.\, A(x)}\ \forall\text{-}I \qquad \cfrac{\cfrac{\cfrac{\forall x.\, A(x) \wedge B(x)}{A(x) \wedge B(x)}\ \forall\text{-}E}{B(x)}\ \wedge\text{-}ER}{\forall x.\, B(x)}\ \forall\text{-}I$$

$$\cfrac{\qquad\qquad\qquad\qquad\qquad\qquad\qquad}{(\forall x.\, A(x)) \wedge (\forall x.\, B(x))}\ \wedge\text{-}I$$

---

[90]In both cases, x does not occur free in $\forall x.\, A(x) \wedge B(x)$, which is the open assumption on which $A(x)$, respectively $B(x)$, depends.

Another Proof? (2)

$$\dfrac{\dfrac{\dfrac{\dfrac{[\forall x.\, A(x) \wedge B(x)]^1}{A(x) \wedge B(x)} \text{ }\forall\text{-}E}{A(x)} \text{ }\wedge\text{-}EL}{\forall x.\, A(x)} \text{ }\forall\text{-}I \qquad \dfrac{\dfrac{\dfrac{[\forall x.\, A(x) \wedge B(x)]^1}{A(x) \wedge B(x)} \text{ }\forall\text{-}E}{B(x)} \text{ }\wedge\text{-}ER}{\forall x.\, B(x)} \text{ }\forall\text{-}I}{\dfrac{(\forall x.\, A(x)) \wedge (\forall x.\, B(x))}{(\forall x.\, A(x) \wedge B(x)) \rightarrow (\forall x.\, A(x)) \wedge (\forall x.\, B(x))} \text{ }\rightarrow\text{-}I^1} \text{ }\wedge\text{-}I$$

---

[90]In both cases, x does not occur free in $\forall x.\, A(x) \wedge B(x)$, which is the open assumption on which $A(x)$, respectively $B(x)$, depends.

82

Another Proof? (2)

$$\cfrac{\cfrac{\cfrac{\cfrac{[\forall x.\, A(x) \wedge B(x)]^1}{A(x) \wedge B(x)}\ \text{$\forall$-E}}{A(x)}\ \text{$\wedge$-EL}}{\forall x.\, A(x)}\ \text{$\forall$-I} \qquad \cfrac{\cfrac{\cfrac{[\forall x.\, A(x) \wedge B(x)]^1}{A(x) \wedge B(x)}\ \text{$\forall$-E}}{B(x)}\ \text{$\wedge$-ER}}{\forall x.\, B(x)}\ \text{$\forall$-I}}{\cfrac{(\forall x.\, A(x)) \wedge (\forall x.\, B(x))}{(\forall x.\, A(x) \wedge B(x)) \rightarrow (\forall x.\, A(x)) \wedge (\forall x.\, B(x))}\ \text{$\rightarrow$-I^1}}\ \text{$\wedge$-I}$$

Yes (check side conditions[90] of $\forall$-I).

---

[90]In both cases, x does not occur free in $\forall x.\, A(x) \wedge B(x)$, which is the open assumption on which $A(x)$, respectively $B(x)$, depends.

Boys Don't Cry

Let $\phi \equiv (\forall x.\, b(x) \to m(x)) \land (\forall x.\, m(x) \to \neg c(x))$.

$$
\cfrac{
  \cfrac{
    \cfrac{
      \cfrac{[\phi]^1}{\forall x.\, m(x) \to \neg c(x)} \land\text{-}ER
    }{m(x) \to \neg c(x)} \forall\text{-}E
    \qquad
    \cfrac{
      \cfrac{
        \cfrac{[\phi]^1}{\forall x.\, b(x) \to m(x)} \land\text{-}EL
      }{b(x) \to m(x)} \forall\text{-}E
      \qquad [b(x)]^2
    }{m(x)} \to\text{-}E
  }{\neg c(x)} \to\text{-}E
}{
  \cfrac{
    \cfrac{
      \cfrac{b(x) \to \neg c(x)}{\forall x.\, b(x) \to \neg c(x)} \forall\text{-}I
    }{\phi \to (\forall x.\, b(x) \to \neg c(x))} \to\text{-}I^1
  }{}
} \to\text{-}I^2
$$

Define[91] $A \leftrightarrow B$ as $A \to B \wedge B \to A$.

The following rule can be derived (in propositional logic, actually):

$$\frac{\begin{array}{cc} [A] & [B] \\ \vdots & \vdots \\ B & A \end{array}}{A \leftrightarrow B} \leftrightarrow\text{-}I$$

You could do this as an exercise!

---

[91]By defining we mean, use $A \leftrightarrow B$ as shorthand for $A \to B \wedge B \to A$, in the same way as we regard negation as a shorthand.

Proof?

$$\frac{\dfrac{[A]^1}{\forall x.\, A}\ \forall\text{-}I \quad \dfrac{[\forall x.\, A]^1}{A}\ \forall\text{-}E}{A \leftrightarrow \forall x.\, A}\ \leftrightarrow\text{-}I^1$$

Proof?

$$\frac{\dfrac{[A]^1}{\forall x.\, A} \ \forall\text{-}I \quad \dfrac{[\forall x.\, A]^1}{A} \ \forall\text{-}E}{A \leftrightarrow \forall x.\, A} \ \leftrightarrow\text{-}I^1$$

Yes, but only if x not free in A.

Proof?

$$\frac{\dfrac{[A]^1}{\forall x.\, A}\ \forall\text{-}I \quad \dfrac{[\forall x.\, A]^1}{A}\ \forall\text{-}E}{A \leftrightarrow \forall x.\, A}\ \leftrightarrow\text{-}I^1$$

Yes, but only if x not free in A.

Similar requirement arises in proving $(\forall x.\, A \to B(x)) \leftrightarrow (A \to \forall x.\, B(x))$.

Side Conditions and Proof Boxes

We mentioned previously a style of writing derivations where subderivations based on temporary assumptions are enclosed in boxes.

These boxes are also handy for doing derivations in first-order logic, since one can use the very clear formulation: a variable occurs inside or outside of a box. See [HR04].

Existential Quantification

- We could define[92] $\exists x. A$ as $\neg \forall x. \neg A$.

- Equivalence follows from our definition of semantics.

$$\mathcal{A}(\neg A) = \begin{cases} 1 & \text{if } \mathcal{A}(A) = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{A}(\forall x. A) = \begin{cases} 1 & \text{if for all } u \in U_{\mathcal{A}}, \mathcal{A}_{[x/u]}(A) = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{A}(\exists x. A) = \begin{cases} 1 & \text{if for some } u \in U_{\mathcal{A}}, \mathcal{A}_{[x/u]}(A) = 1 \\ 0 & \text{otherwise} \end{cases}$$

Conclude: $\mathcal{A}(\exists x. A) = \mathcal{A}(\neg \forall x. \neg A)$

---

[92]By defining we mean, use $\exists x. A$ as shorthand for $\neg \forall x. \neg A$, in the same way as we regard negation as a shorthand.

However, we have already introduced $\exists$ as syntactic entity, and also its semantics. If we now want to treat it as being defined in terms of $\forall$, for the purposes of building a deductive system, we must be sure that $\exists x. A$ is semantically equivalent to $\neg \forall x. \neg A$, i.e., that $\mathcal{A}(\exists x. A) = \mathcal{A}(\neg \forall x. \neg A)$.

Where do the Rules for $\exists$ Come from?

- We can[93] use definition $\exists x.\, A \equiv \neg\forall x.\, \neg A$ and the given rules for $\forall$ to derive ND proof rules.

93

- We can use definition $\exists x.\, A \equiv \neg\forall x.\, \neg A$ and the given rules for $\forall$ to derive ND proof rules.
  In this case, the soundness of the derived rules is guaranteed since
    * the rules for $\forall$ are sound;
    * we have proven the equivalence of $\exists x.\, A$ and $\neg\forall x.\, \neg A$ semantically.
- Alternative: give rules as part of the deduction system and prove the equivalence as a lemma, instead of by definition. In this case, the soundness must be proven by hand (however, proving rules sound is an aspect we neglect in this course). But once this is done, the equivalence of $\exists x.\, A$ and $\neg\forall x.\, \neg A$ can be proven within the deductive system, rather than by hand, provided that the deductive system is complete.

Where do the Rules for $\exists$ Come from?

- We can[93] use definition $\exists x.\,A \equiv \neg\forall x.\,\neg A$ and the given rules for $\forall$ to derive ND proof rules.

- Alternatively, we can give rules as part of the deduction system and prove equivalence as a lemma, instead of by definition.

  We will do the first here. The Isabelle formalization follows the second approach.

93

- We can use definition $\exists x.\,A \equiv \neg\forall x.\,\neg A$ and the given rules for $\forall$ to derive ND proof rules.
  In this case, the soundness of the derived rules is guaranteed since
    * the rules for $\forall$ are sound;
    * we have proven the equivalence of $\exists x.\,A$ and $\neg\forall x.\,\neg A$ semantically.

- Alternative: give rules as part of the deduction system and prove the equivalence as a lemma, instead of by definition. In this case, the soundness must be proven by hand (however, proving rules sound is an aspect we neglect in this course). But once this is done, the equivalence of $\exists x.\,A$ and $\neg\forall x.\,\neg A$ can be proven within the deductive system, rather than by hand, provided that the deductive system is complete.

∃-*I* as a Derived Rule

The rule:

$$\frac{A(t)}{\exists x.\, A(x)}\ \exists\text{-}I$$

$$\exists x.\, A(x)$$

We want to have $\exists x.\, A(x)$ as conclusion.

∃-*I* as a Derived Rule

The rule:

$$\frac{A(t)}{\exists x.\, A(x)} \; \exists\text{-}I$$

$$\neg\forall x.\, \neg A(x)$$

But by definition that's $\neg\forall x.\, \neg A(x)$.

∃-*I* as a Derived Rule

The rule:

$$\frac{A(t)}{\exists x.\, A(x)} \; \exists\text{-}I$$

$$\forall x.\, \neg A(x)$$

$$\bot$$
$$\neg\forall x.\, \neg A(x)$$

We aim for applying →-*I* in the last step (recall ¬-definition).

∃-*I* as a Derived Rule

The rule:

$$\frac{A(t)}{\exists x.\, A(x)}\ \exists\text{-}I$$

We apply ∀-*E*.

$$\frac{\forall x.\, \neg A(x)}{\neg A(t)}\ \forall\text{-}E$$

$$\bot$$

$$\neg \forall x.\, \neg A(x)$$

∃-*I* as a Derived Rule

The rule:

$$\frac{A(t)}{\exists x.\, A(x)} \ \exists\text{-}I$$

$$\frac{\dfrac{\forall x.\, \neg A(x)}{\neg A(t)} \ \forall\text{-}E \qquad A(t)}{\bot} \ \rightarrow\text{-}E$$

$$\neg\forall x.\, \neg A(x)$$

Making assumption $A(t)$ allows us to use $\rightarrow$-*E* (recall ¬-definition).

∃-*I* as a Derived Rule

The rule:

$$\frac{A(t)}{\exists x.\, A(x)} \;\exists\text{-}I$$

$$\frac{\dfrac{\dfrac{[\forall x.\, \neg A(x)]^1}{\neg A(t)}\;\forall\text{-}E \qquad A(t)}{\bot}\;\rightarrow\text{-}E}{\neg\forall x.\, \neg A(x)}\;\rightarrow\text{-}I^1$$

Finally we can apply →-*I*. Note that the assumption $A(t)$ is still open.

∃-*E* as a Derived Rule

The rule:

$$
\cfrac{\exists x.\, A(x) \qquad \begin{array}{c} [A(x)] \\ \vdots \\ B \end{array}}{B} \ \text{∃-}E
$$

$$\exists x.\, A(x)$$

We will use $\exists x.\, A(x)$ as one assumption.

∃-*E* as a Derived Rule

The rule:

$$
\frac{\exists x.\, A(x) \qquad \begin{array}{c} [A(x)] \\ \vdots \\ B \end{array}}{B} \ \exists\text{-}E
$$

$$\neg\forall x.\,\neg A(x)$$

But by definition that's $\neg\forall x.\,\neg A(x)$.

∃-*E* as a Derived Rule

The rule:

$$\cfrac{\exists x.\, A(x) \qquad \begin{array}{c}[A(x)]\\ \vdots\\ B\end{array}}{B}\ \text{∃-}E$$

$$\begin{array}{c}A(x)\\ \vdots\\ B\end{array}$$

$$\neg\forall x.\, \neg A(x)$$

We assume a hypothetical derivation[94].

$\exists$-E as a Derived Rule

The rule:

$$\frac{\exists x.\, A(x) \qquad \begin{array}{c} [A(x)] \\ \vdots \\ B \end{array}}{B} \; \exists\text{-}E$$

$$\frac{\neg B \qquad \begin{array}{c} A(x) \\ \vdots \\ B \end{array}}{\bot} \to\text{-}E$$

$$\neg \forall x.\, \neg A(x)$$

We make an additional assumption and apply $\to$-E (recall $\neg$-definition)

∃-*E* as a Derived Rule

The rule:

$$\frac{\exists x.\,A(x) \qquad \overset{[A(x)]}{\overset{\vdots}{B}}}{B} \;\exists\text{-}E$$

$$\frac{\overset{[A(x)]^2}{\vdots}}{\cfrac{\cfrac{\neg B \qquad B}{\bot} \to\text{-}E}{\neg A(x)} \to\text{-}I^2}}{\neg \forall x.\,\neg A(x)}$$

Now we can discharge the assumption $A(x)$ made in the hypothetical derivation.

$\exists$-E as a Derived Rule

The rule:

$$\frac{\exists x.\, A(x) \qquad \overset{[A(x)]}{\underset{B}{\vdots}}}{B} \; \exists\text{-}E$$

$$\frac{\neg\forall x.\, \neg A(x) \qquad \dfrac{\dfrac{\dfrac{\neg B \qquad \overset{[A(x)]^2}{\underset{B}{\vdots}}}{\bot} \to\text{-}E}{\neg A(x)} \to\text{-}I^2}{\forall x.\, \neg A(x)} \; \forall\text{-}I}{}$$

At this step, the side condition from $\forall$-I applies. $\exists$-E will inherit it![95]

90

$\exists$-E as a Derived Rule

The rule:

$$\frac{\exists x.\,A(x) \qquad \overset{[A(x)]}{\overset{\vdots}{B}}}{B} \; \exists\text{-}E$$

$$\frac{\neg\forall x.\,\neg A(x) \qquad \dfrac{\dfrac{\dfrac{\neg B \qquad \overset{[A(x)]^2}{\overset{\vdots}{B}}}{\bot} \to\text{-}E}{\neg A(x)} \to\text{-}I^2}{\forall x.\,\neg A(x)} \forall\text{-}I}{\bot} \to\text{-}E$$

We apply $\to$-E.

∃-*E* as a Derived Rule

The rule:

$$\frac{\exists x.\, A(x) \qquad \overset{[A(x)]}{\vdots} \; B}{B} \; \exists\text{-}E$$

$$\frac{\neg\forall x.\, \neg A(x) \qquad \dfrac{\dfrac{\dfrac{[\neg B]^1 \qquad \overset{[A(x)]^2}{\vdots} \; B}{\bot} \to\text{-}E}{\neg A(x)} \to\text{-}I^2}{\forall x.\, \neg A(x)} \forall\text{-}I}{\dfrac{\bot}{B} \; RAA^1} \to\text{-}E$$

We are done. Note that this proof uses classical[96] reasoning.

Example Derivation Using ∃-*E*

Assumption: x does not occur free in B

$$\exists x.\, A(x) \qquad \dfrac{\dfrac{\forall x.\, A(x) \to B}{A(x) \to B}\ \forall\text{-}E \qquad A(x)}{B}\ \to\text{-}E$$

Example Derivation Using $\exists$-E

Assumption: x does not occur free in B

$$
\cfrac{\exists x.\, A(x) \qquad \cfrac{\cfrac{\forall x.\, A(x) \to B}{A(x) \to B}\ \forall\text{-}E \qquad [A(x)]^3}{B}\ \to\text{-}E}{B}\ \exists\text{-}E^3
$$

Example Derivation Using ∃-*E*

Assumption: x does not occur free in B

$$
\cfrac{
  \cfrac{[\exists x.\ A(x)]^2 \qquad
    \cfrac{
      \cfrac{
        \cfrac{[\forall x.\ A(x) \to B]^1}{A(x) \to B}\ \forall\text{-}E \qquad [A(x)]^3
      }{B}\ \to\text{-}E
    }{B}\ \exists\text{-}E^3
  }{(\exists x.\ A(x)) \to B}\ \to\text{-}I^2
}{(\forall x.\ A(x) \to B) \to ((\exists x.\ A(x)) \to B)}\ \to\text{-}I^1
$$

4.6 Conclusion on FOL

- Propositional logic is good for modeling simple patterns of reasoning like "if ... then ... else".

4.6 Conclusion on FOL

- Propositional logic is good for modeling simple patterns of reasoning like "if . . . then . . . else".

- In first-order logic, one has "things" and relations on / properties of "things". Quantify over "things". Powerful[97]!

- Some people advocate intuitionistic, relevant, and other

---

[97]In first-order logic, one has "things" and relations/properties that may or may not hold for these "things". Quantifiers are used to speak about "all things" and "some things".

For example, one can reason:

All men are mortal, Socrates is a man, therefore Socrates is mortal.

The idea underlying first-order logic is so general, abstract, and powerful that vast portions of human (mathematical) reasoning can be modeled with it.

In fact, first-order logic is the most prominent logic of all. Many people know about it: not only mathematicians and computer scientists, but also linguists, philosophers, psychologists, economists etc. are likely to learn about first-order logic in their education.

While some applications in the fields mentioned above require other logics, e.g. modal logics, those can often be reduced to first-order logic, so that first-order logic remains the

point of reference.

On the other hand, logics that are strictly more expressive than first-order logic are only known to and studied by few specialists within mathematics and computer science.

This example about Socrates and men is a very well-known one. You may wonder: what is the history of this example?

In English, the example is commonly given using the word "man", although one also finds "human". Like many languages (e.g., French, Italian), English often uses "man" for "human being", although this use of language may be considered discriminating against women. E.g. [Tho95a]:

> **man** [. . .] **1** an adult human male, esp. as distinct from a woman or boy. **2** a human being; a person (*no man is perfect*).

While the example does not, strictly speaking, imply that "man" is used in the meaning of "human being", this is strongly suggested both by the content of the example (or should women be immortal?) and the fact that languages

that do have a word for "human being" (e.g. "Mensch" in German) usually give the example using this word. In fact, the example is originally in Old Greek, and there the word ἄνθρωπος (anthropos = human being), as opposed to ἀνήρ (anér = human male), is used.

The example is a so-called syllogism of the first figure, which the scholastics called Barbara. It was developed by Aristotle [Ari] in an abstract form, i.e., without using the concrete name "Socrates". In his terminology, ἄνθρωπος is the middle term that is used as subject in the first premise and as predicate in the second premise (this is what is called first figure). Aristotle formulated the syllogism as follows: If A of all B and B is said of all C, then A must be said of all C.

And why "Socrates"? It is not exactly clear how it came about that this particular syllogism is associated with Socrates. In any case, as far it is known, Socrates did not investigate any questions of logic. However, Aristotle frequently uses Socrates and Kallias as standard names for individuals

"deviant" logics[98].

[Ari]. Possibly there were statutes of Socrates and Kallias standing in the hall where Aristotle gave his lectures, so it was convenient for him to point to the statutes whenever he was making a point involving two individuals.

[98]There are still controversies about what the best logic is for reasoning about "things" and properties/relations, and scope (quantification). Some argue for intuitionistic, relevance, modal and other "deviant" logics.

An example where first-order logic is inappropriate might be:

> From "a dollar buys a candy bar" and "a dollar buys an ice cream" we cannot normally conclude "a dollar buys a candy bar and an ice cream".

However, such analogies should be treated with care. Depending on how ice-creams, candy bars, dollars and buying are modeled, first-order logic may very well be appropriate.

Modal logics are logics that have modality operators, usually $\Box$ and $\Diamond$.

Sometimes these denote temporal aspects, e.g., $\Box\phi$ means

- Limitation: cannot quantify over predicates[99].

- "A" world or "the" world is modeled in first-order logic using so-called first-order theories. This will be studied next lecture.

"ϕ always holds". But many other interpretations are possible, e.g., $\square_A \phi$ could mean "A knows that ϕ holds" [HC68].

In relevance logics, it is not true that $A \rightarrow B$ holds whenever A is false. Rather, A must somehow be "relevant" for B.

[99]The idea underlying first-order logic seems so general that it is not so apparent what its limitations could be. The limitations will become clear as we study more expressive logics.

For the moment, note the following: in first-order logic, we quantify over variables (hence, domain elements), not over predicates. The number of predicates is fixed in a particular first-order language. So for example, it is impossible to express the following:

> For all unary predicates p, if there exists an x such that $p(x)$ is true, then there exists a smallest x such that $p(x)$ is true,

since we would be quantifying over p.

5 First-Order Logic with Equality

Overview

Last lecture: first-order logic.
This lecture:

- first-order logic with equality and first-order theories;

- set-theoretic reasoning.

We extend language and deductive system to formalize and reason about the (mathematical) world.

FOL with Equality

Equality is a logical symbol rather than a mathematical one[100].

Speak of first-order logic with equality rather than adding equality as "just another predicate".

In logic languages, it is common to distinguish between logical and non-logical symbols. We explain this for first-order logic.

Recall that there isn't just the language of first-order logic, but rather defining a particular signature gives us a first-order language. The logical symbols are those that are part of any first-order language and whose meaning is "hard-wired" into the formalism of first-order logic, like $\wedge$ or $\forall$. The non-logical symbols are those given by a particular signature, and whose meaning must be defined "by the user" by giving a structure.

Above we say "mathematical" instead of "non-logical" because we assume that mathematics is our domain of discourse, so that the signature contains the symbols of "mathematics".

Now what status should the equality symbol $=$ have? We will assume that $=$ is a symbol whose meaning is hard-wired into the formalism. One then speaks of first-order logic with equality.

Syntax and Semantics

Syntax: $=$ is a binary infix predicate.

$$t_1 = t_2 \in \textit{Form} \text{ if } t_1, t_2 \in \textit{Term}.$$

Alternatively, one could regard $=$ as an ordinary (binary infix) predicate. However, even if one does not give $=$ a special status, anyone reading $=$ has a certain expectation. Thus it would be very confusing to have a structure that defines $=$ as a, say, non-reflexive relation.

101

$$I_{\mathcal{A}}(s\!=\!t) = \begin{cases} 1 & \text{if } I_{\mathcal{A}}(s)\!=\!I_{\mathcal{A}}(t) \\ 0 & \text{otherwise} \end{cases}$$

The first $=$ is a predicate symbol.

Syntax and Semantics

Syntax: $=$ is a binary infix predicate.

$t_1 = t_2 \in Form$ if $t_1, t_2 \in Term$.

Semantics : recall a structure is a pair $\mathcal{A} = \langle U_{\mathcal{A}}, I_{\mathcal{A}} \rangle$ and $I_{\mathcal{A}}(t)$ is the interpretation of t.

$$I_{\mathcal{A}}(s = t) = \begin{cases} 1 & \text{if } I_{\mathcal{A}}(s) = I_{\mathcal{A}}(t) \\ 0 & \text{otherwise} \end{cases}$$

Note the three completely different uses of "$=$"[101] here!

Alternatively, one could regard $=$ as an ordinary (binary infix) predicate. However, even if one does not give $=$ a special status, anyone reading $=$ has a certain expectation. Thus it would be very confusing to have a structure that defines $=$ as a, say, non-reflexive relation.

[101]

$$I_{\mathcal{A}}(s{=}t) = \begin{cases} 1 & \text{if } I_{\mathcal{A}}(s){=}I_{\mathcal{A}}(t) \\ 0 & \text{otherwise} \end{cases}$$

The first $=$ is a predicate symbol.

The second $=$ is a definitional occurrence: The expression on the left-hand side is defined to be equal to the value of the right-hand side.

Syntax and Semantics

Syntax: $=$ is a binary infix predicate.

$t_1 = t_2 \in$ *Form* if $t_1, t_2 \in$ *Term*.

Semantics : recall a structure is a pair $\mathcal{A} = \langle U_\mathcal{A}, I_\mathcal{A} \rangle$ and $I_\mathcal{A}(t)$ is the interpretation of t.

$$I_\mathcal{A}(s = t) = \begin{cases} 1 & \text{if } I_\mathcal{A}(s) = I_\mathcal{A}(t) \\ 0 & \text{otherwise} \end{cases}$$

Note the three completely different uses of "$=$"[101] here!

Alternatively, one could regard $=$ as an ordinary (binary infix) predicate. However, even if one does not give $=$ a special status, anyone reading $=$ has a certain expectation. Thus it would be very confusing to have a structure that defines $=$ as a, say, non-reflexive relation.

[101]

$$I_\mathcal{A}(s{=}t) = \begin{cases} 1 & \text{if } I_\mathcal{A}(s){=}I_\mathcal{A}(t) \\ 0 & \text{otherwise} \end{cases}$$

The first $=$ is a predicate symbol.

The second $=$ is a definitional occurrence: The expression on the left-hand side is defined to be equal to the value of the right-hand side.

The third $=$ is semantic equality, i.e., the identity relation on the domain.

Rules[102]

- Equality is an equivalence relation[103]

$$\frac{}{x = x} \; \textit{refl} \qquad \frac{x = y}{y = x} \; \textit{sym} \qquad \frac{x = y \quad y = z}{x = z} \; \textit{trans}$$

Rules[102]

- Equality is an equivalence relation[103]

$$\frac{}{x = x} \; \textit{refl} \qquad \frac{x = y}{y = x} \; \textit{sym} \qquad \frac{x = y \quad y = z}{x = z} \; \textit{trans}$$

- Equality is also a congruence[104] on terms and all rela-

---

[102]Since $=$ is a logical symbol in the formalism of first-order logic with equality, there should be derivation rules for $=$ to derive which formulas $a = b$ are true.

[103]In general mathematical terminology, a relation $\equiv$ is an equivalence relation if the following three properties hold:

Reflexivity: $a \equiv a$ for all a;

Symmetry: $a \equiv b$ implies $b \equiv a$;

Transitivity: $a \equiv b$ and $b \equiv c$ implies $a \equiv c$.

Example: being equal modulo 6.
"a is equal b modulo 6" is often written $a \equiv b \bmod 6$.

[104]In general mathematical terminology, a relation $\cong$ is a congruence w.r.t. (or: on) f, where f has arity n, if $a_1 \cong b_1, \ldots, a_n \cong b_n$ implies $f(a_1, \ldots, a_n) \cong f(b_1, \ldots, b_n)$.

Example: being equal modulo 6 is congruent w.r.t. multiplication.

$14 \equiv 8 \bmod 6$ and $15 \equiv 9 \bmod 6$, hence $14 \cdot 15 \equiv 8 \cdot 9 \bmod 6$.

tions[105]

$$\frac{x_1 = y_1 \quad \cdots \quad x_n = y_n}{t(x_1, \ldots, x_n) = t(y_1, \ldots, y_n)} \quad cong_1$$

$$\frac{x_1 = y_1 \quad \cdots \quad x_n = y_n \quad A(x_1, \ldots, x_n)}{A(y_1, \ldots, y_n)} \quad cong_2$$

This can be defined in an analogous way for a property (relation) P.

Example: being equal modulo 6 is congruent w.r.t. divisibility by 3.

$15 \equiv 9 \bmod 6$ and 15 is divisible by 3, hence 9 is divisible by 3.

$14 \equiv 8 \bmod 6$ and 14 is not divisible by 3, hence 8 is not divisible by 3.

[105]Why did we use letters t and A here?

Recall the rules for building terms and atoms.

Is $t(x_1, \ldots, x_n)$ a term, and $A(x_1, \ldots, x_n)$ and atom, obtained by one application of such a rule, i.e.: is t a function symbol in $\mathcal{F}$, applied to $x_1, \ldots, x_n$, and is A a predicate symbols in $\mathcal{P}$, applied to $x_1, \ldots, x_n$?

tions[105]

$$\frac{x_1 = y_1 \quad \cdots \quad x_n = y_n}{t(x_1, \ldots, x_n) = t(y_1, \ldots, y_n)} \; cong_1$$

$$\frac{x_1 = y_1 \quad \cdots \quad x_n = y_n \quad A(x_1, \ldots, x_n)}{A(y_1, \ldots, y_n)} \; cong_2$$

This can be defined in an analogous way for a property (relation) P.

Example: being equal modulo 6 is congruent w.r.t. divisibility by 3.

$15 \equiv 9 \bmod 6$ and 15 is divisible by 3, hence 9 is divisible by 3.

$14 \equiv 8 \bmod 6$ and 14 is not divisible by 3, hence 8 is not divisible by 3.

[105]Why did we use letters t and A here?

Recall the rules for building terms and atoms.

Is $t(x_1, \ldots, x_n)$ a term, and $A(x_1, \ldots, x_n)$ and atom, obtained by one application of such a rule, i.e.: is t a function symbol in $\mathcal{F}$, applied to $x_1, \ldots, x_n$, and is A a predicate symbols in $\mathcal{P}$, applied to $x_1, \ldots, x_n$?

In general, no! The notations $t(x_1, \ldots, x_n)$ and $A(x_1, \ldots, x_n)$ are metanotations. $t(x_1, \ldots, x_n)$ stands for any term in which $x_1, \ldots, x_n$ occur, and $A(x_1, \ldots, x_n)$ stands for any atom in which $x_1, \ldots, x_n$ occur.

Soundness of Rules

Equality in $I_\mathcal{A}$ is an equivalence relation[106] and functions/predicates/logical-

---

This is why we used letters t (term) and A (atom) here instead of f (function) and P (predicate).

And in this context, the notation $t(y_1, \ldots, y_n)$ stands for the term obtained from $t(x_1, \ldots, x_n)$ by replacing all occurrences of x_1 with y_1 and so forth. In analogy the notation $A(y_1, \ldots, y_n)$ is defined.

Note that in the schematic formulation of the rule, we use letters x and y to suggest variables, but the rule applies to arbitrary terms.

This description is not very formal, but this is not too problematic since we will be more formal once we have some useful machinery for this at hand.

[106]On the semantic level, two things are equal if they are identical. Semantic equality is an equivalence relation. This semantic fact is so fundamental that we cannot explain it any further.

So one can prove that $I_\mathcal{A}(s = s) = 1$ for all all terms s, because $I_\mathcal{A}(s) = I_\mathcal{A}(s)$ for all terms, and likewise for symmetry

operators are "truth-functional" [107].

   Adding further rules gives us an equational theory, e.g. groups.

---

and transitivity.

[107] If $t(x)$ is a term containing x and $t(y)$ is the term obtained from $t(x)$ by replacing all occurrences of x with y, and moreover $I_{\mathcal{A}}(x = y) = 1$, then $I_{\mathcal{A}}(x) = I_{\mathcal{A}}(y)$. One can show by induction on the structure of t that $I_{\mathcal{A}}(t(x)) = I_{\mathcal{A}}(t(y))$.

   So by "truth-functional" we mean that the value $I_{\mathcal{A}}(t(x))$ depends on $I_{\mathcal{A}}(x)$, not on x itself.

   This can be generalized to n variables as in the rule.

   An analogous proof can be done for rule $cong_2$.

Congruence: Alternatives

One can specialize congruence rules to replace only some term occurrences.

$$\frac{x_1 = y_1 \ \cdots \ x_n = y_n}{t[z_1 \leftarrow x_1, \ldots, z_n \leftarrow x_n] = t[z_1 \leftarrow y_1, \ldots, z_n \leftarrow y_n]} \ cong_1$$

$$\frac{x_1 = y_1 \ \cdots \ x_n = y_n \quad A[z_1 \leftarrow y_1, \ldots, z_n \leftarrow y_n]}{A[z_1 \leftarrow x_1, \ldots, z_n \leftarrow x_n]} \ cong_2$$

One time the z's are replaced with x's and one time with y's.[108]

---

[108] The notation $t[z_1 \leftarrow x_1, \ldots, z_n \leftarrow x_n]$ stands for the term obtained from t by simultaneously replacing each z_i ($i \in \{1, \ldots, n\}$) with x_i.

$[z_1 \leftarrow x_1, \ldots, z_n \leftarrow x_n]$ is called a substitution.

To have an unambiguous notation for "replacing some occurrences of $x_1, \ldots, x_n$", we start from a term t containing variable occurrences $z_1, \ldots, z_n$. On the LHS, these are replaced with $x_1, \ldots, x_n$, on the RHS they are replaced with $y_1, \ldots, y_n$. So on the RHS we have a term obtained from the one on the LHS by replacing some occurrences of $x_1, \ldots, x_n$ with $y_1, \ldots, y_n$.

One can say that the $z_1, \ldots, z_n$ are introduced to mark the occurrences of $x_1, \ldots, x_n$ that should be replaced by $y_1, \ldots, y_n$.

Note that in the schematic formulation of the rule, we use letters x and y to suggest variables, but the rule applies to arbitrary terms. The z's however are variables (substitutions replace variables, not arbitrary terms).

Examples

How many ways are there to choose some occurrences of x in $x^2 + y^2 > 12 \cdot x$?

---

[109]The atom $x^2 + y^2 > 12 \cdot x$ contains two occurrences of x. There are four ways to choose some occurrences of x in $x^2 + y^2 > 12 \cdot x$.

Each of those ways corresponds to an atom obtained from $x^2 + y^2 > 12 \cdot x$ by replacing some occurrences of x with z. That is, there are four different A's such that $A[x/z] = x^2 + y^2 > 12 \cdot x$. Now the atom above the line in the examples is obtained by substituting x for z, and the atom below the line is obtained by substituting y for z.

Examples

How many ways are there to choose some occurrences of x in $x^2 + y^2 > 12 \cdot x$? 4, namely:

$$A = x^2 + y^2 > 12 \cdot x, \quad A = z^2 + y^2 > 12 \cdot x,$$
$$A = x^2 + y^2 > 12 \cdot z, \quad A = z^2 + y^2 > 12 \cdot z.$$

---

[109]The atom $x^2 + y^2 > 12 \cdot x$ contains two occurrences of x. There are four ways to choose some occurrences of x in $x^2 + y^2 > 12 \cdot x$.

Each of those ways corresponds to an atom obtained from $x^2 + y^2 > 12 \cdot x$ by replacing some occurrences of x with z. That is, there are four different A's such that $A[x/z] = x^2 + y^2 > 12 \cdot x$. Now the atom above the line in the examples is obtained by substituting x for z, and the atom below the line is obtained by substituting y for z.

Examples

How many ways are there to choose some occurrences of x in $x^2 + y^2 > 12 \cdot x$? 4, namely:

$$A = x^2 + y^2 > 12 \cdot x, \quad A = z^2 + y^2 > 12 \cdot x, \,_{109}$$
$$A = x^2 + y^2 > 12 \cdot z, \quad A = z^2 + y^2 > 12 \cdot z.$$

We show two ways:

$$\frac{x = 3 \quad x^2 + y^2 > 12 \cdot x}{3^2 + y^2 > 12 \cdot x} \text{ with } A = z^2 + y^2 > 12 \cdot x$$

$$\frac{x = 3 \quad x^2 + y^2 > 12 \cdot x}{x^2 + y^2 > 12 \cdot 3} \text{ with } A = x^2 + y^2 > 12 \cdot z$$

---

[109] The atom $x^2 + y^2 > 12 \cdot x$ contains two occurrences of x. There are four ways to choose some occurrences of x in $x^2 + y^2 > 12 \cdot x$.

Each of those ways corresponds to an atom obtained from $x^2 + y^2 > 12 \cdot x$ by replacing some occurrences of x with z. That is, there are four different A's such that $A[x/z] = x^2 + y^2 > 12 \cdot x$. Now the atom above the line in the examples is obtained by substituting x for z, and the atom below the line is obtained by substituting y for z.

Isabelle Rule

The Isabelle FOL rule is simply[110] (using a tree syntax)

$$\frac{x = y \quad P(x)}{P(y)} \; \textit{subst}$$

or literally

$$[\![a = b;\, P(a)]\!] \implies P(b)$$

---

[110]The Isabelle FOL rule is:

$$\frac{x = y \quad P(x)}{P(y)} \; \textit{subst}$$

In this rule, P is an Isabelle metavariable.

Why doesn't the Isabelle rule contain a z to mark which occurrences should be replaced?

We cannot understand this yet, but think of P as a formula where some positions are marked in such a way that once we apply P to t (we write $P(t)$), t will be inserted into all those positions. This is why $P(x)$ is a formula and $P(y)$ is a formula obtained by replacing some occurrences of x with y.

Proving $\exists x.\, t = x$

$$\cfrac{\cfrac{}{t = t}\ \textit{refl}}{\exists x.\, t = x}\ \textit{∃-I}$$

Proving $\exists x.\, t = x$

$$\dfrac{\dfrac{}{t = t}\ \textit{refl}}{\exists x.\, t = x}\ \textit{∃-I}$$

In the rule $\dfrac{A(t)}{\exists x.\, A(x)}\ \textit{∃-I}$, "$A(x)$" is metanotation. In the example, $A(x) = (t = x)$.

Proving $\exists x.\, t = x$

$$\frac{\dfrac{}{t = t}\ \textit{refl}}{\exists x.\, t = x}\ \textit{∃-I}$$

In the rule $\dfrac{A(t)}{\exists x.\, A(x)}\ \textit{∃-I}$, "$A(x)$" is metanotation. In the example, $A(x) = (t = x)$.

Notational confusion avoided by a precise metalanguage.

6 First-Order Theories

What Is a Theory?

Recall our intuitive explanation of theories.

A theory involves certain function and/or predicate symbols for which certain "laws" hold.

Depending on the context, these symbols may co-exist with other symbols.

Technically, the laws are added as rules (in particular, axioms) to the proof system.

A structure in which these rules are true is then called a model of the theory.

6.1 Example 1: Partial Orders

- The language of the theory of partial orders[111]: $\leq$[112]

What Is a Theory?

Recall our intuitive explanation of theories.

A theory involves certain function and/or predicate symbols for which certain "laws" hold.

Depending on the context, these symbols may co-exist with other symbols.

Technically, the laws are added as rules (in particular, axioms) to the proof system.

A structure in which these rules are true is then called a model of the theory.

6.1 Example 1: Partial Orders

- The language of the theory of partial orders[111]: $\leq$[112]

---

[111]A partial order is a binary relation that is reflexive, transitive, and anti-symmetric: $a \leq b$ and $b \leq a$ implies $a = b$.

[112]$\leq$ is (by convention) a binary infix predicate symbol.

The theory of partial orders involves only this symbol, but that does not mean that there could not be any other symbols in the context.

- Axioms

$$\forall x, y, z.\, x \le y \wedge y \le z \rightarrow x \le z^{113}$$
$$\forall x, y.\, x \le y \wedge y \le x \leftrightarrow x = y^{114}$$

- Axioms

$$\forall x, y, z.\, x \le y \wedge y \le z \to x \le z^{113}$$
$$\forall x, y.\, x \le y \wedge y \le x \leftrightarrow x = y^{114}$$

- Alternative to axioms is to use rules

$$\frac{x \le y \quad y \le z}{x \le z}\ \textit{trans} \qquad \frac{x \le y \quad y \le x}{x = y}\ \textit{antisym} \qquad \frac{x = y}{x \le y}\ \textit{$\le$-refl}$$

Such a conversion is possible since implication is the main connective.[115]

---

[113]The axiom $\forall x, y, z.\, x \le y \wedge y \le z \to x \le z$ encodes transitivity.

[114]Note that $\forall x, y.\, x \le y \wedge y \le x \leftrightarrow x = y$ encodes both antisymmetry ($\to$) and reflexivity ($\leftarrow$). Recall that $A \leftrightarrow B$ as shorthand for $A \to B \wedge B \to A$.

[115]One can see that using $\to$-I and $\to$-E, one can always convert a proof using the axioms to one using the proper rules.

More generally, an axiom of the form $\forall x_1, \ldots, x_n.\ A_1 \wedge \ldots \wedge A_n \to B$ can be converted to a rule

$$\frac{A_1 \quad \ldots A_n}{B}\ .$$

Do it in Isabelle!

A Second Transitivity Rule

One may also consider adding the rule

$$\frac{x = y}{y \leq x} \text{ $\leq$-refl2}$$

to the system. This rule can be derived as follows:

$$\frac{\dfrac{x = y}{y = x} \text{ sym}}{y \leq x} \text{ $\leq$-refl}$$

More on Orders

- A partial order $\leq$ is a linear or total order[116] when

$$\forall x, y.\, x \leq y \vee y \leq x$$

  Note: no "pure" rule formulation[117] of this disjunction.

---

[116]We define these notions according to usual mathematical terminology.

A partial order $\leq$ is a linear or total order if for all a, b, either $a \leq b$ or $b \leq a$.

A partial order $\leq$ is dense if for all a, b where $a < b$, there exists a c such that $a < c$ and $c < b$.

[117]The axiom $\forall x, y.\, x \leq y \vee y \leq x$ cannot be phrased as a proper rule in the style of, for example, the transitivity axiom.

[118]We use $s < t$ as shorthand for $s \leq t \wedge \neg s = t$.

We say that $<$ is the strict part of the partial order $\leq$.

More on Orders

- A partial order $\leq$ is a linear or total order[116] when

$$\forall x, y.\, x \leq y \vee y \leq x$$

  Note: no "pure" rule formulation[117] of this disjunction.

- A total order $\leq$ is dense when, in addition

$$\forall x, y.\, x <^{[118]} y \rightarrow \exists z.(x < z \wedge z < y)$$

  What does $<$ mean?

---

[116]We define these notions according to usual mathematical terminology.

A partial order $\leq$ is a linear or total order if for all a, b, either $a \leq b$ or $b \leq a$.

A partial order $\leq$ is dense if for all a, b where $a < b$, there exists a c such that $a < c$ and $c < b$.

[117]The axiom $\forall x, y.\, x \leq y \vee y \leq x$ cannot be phrased as a proper rule in the style of, for example, the transitivity axiom.

[118]We use $s < t$ as shorthand for $s \leq t \wedge \neg s = t$.

We say that $<$ is the strict part of the partial order $\leq$.

Structures for Orders …

Give structures for orders that are …

1. not total:

---

[119]The $\subseteq$-relation is partial but not total. As an example, consider the $\subseteq$-relation on the set of subsets of $\{1, 2\}$.

$$\{1, 2\}$$

$$\{1\} \qquad \{2\}$$

$$\emptyset$$

Depicting partial orders by a such a graph is quite common. Here, node a is below node b and connected by an arc if and only if $a < b$ and there exists no c with $a < c < b$.
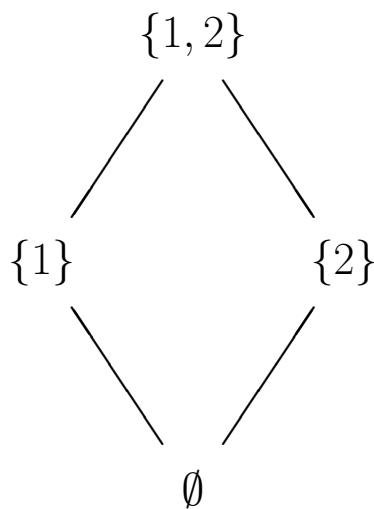
In this example, we have the partial order

$$\{(\emptyset, \emptyset), (\{1\}, \{1\}), (\{1\}, \{1\}), (\{1, 2\}, \{1, 2\}),$$
$$(\emptyset, \{1\}), (\emptyset, \{1\}), (\{1\}, \{1, 2\}), (\{1\}, \{1, 2\})\}.$$

Structures for Orders ...

Give structures for orders that are ...

1. not total: $\subseteq$-relation[119];
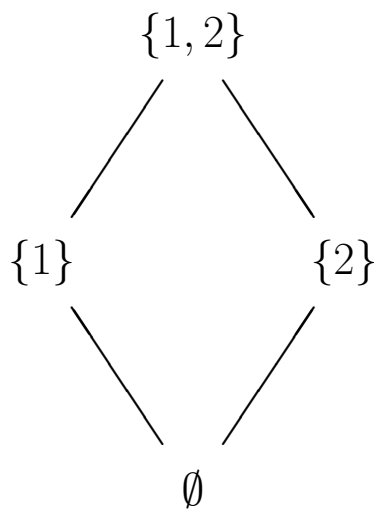
2. total but not dense:

---

[119]The $\subseteq$-relation is partial but not total. As an example, consider the $\subseteq$-relation on the set of subsets of $\{1, 2\}$.

$$\{1,2\}$$

$$\{1\} \qquad\qquad \{2\}$$

$$\emptyset$$

Depicting partial orders by a such a graph is quite common. Here, node a is below node b and connected by an arc if and only if $a < b$ and there exists no c with $a < c < b$.

In this example, we have the partial order
$$\{(\emptyset, \emptyset), (\{1\}, \{1\}), (\{1\}, \{1\}), (\{1, 2\}, \{1, 2\}),$$
$$(\emptyset, \{1\}), (\emptyset, \{1\}), (\{1\}, \{1, 2\}), (\{1\}, \{1, 2\})\}.$$

Structures for Orders ...

Give structures for orders that are ...

1. not total: $\subseteq$-relation[119];
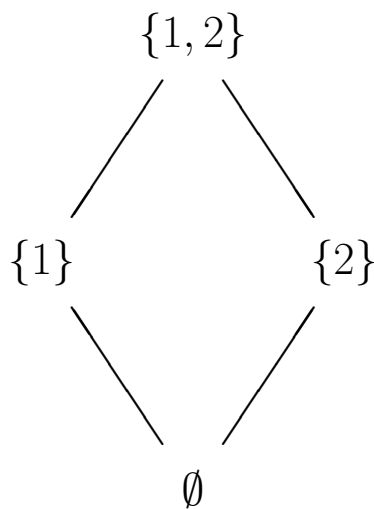
2. total but not dense: integers with $\leq$;

3. dense:

---

[119]The $\subseteq$-relation is partial but not total. As an example, consider the $\subseteq$-relation on the set of subsets of $\{1, 2\}$.

Depicting partial orders by a such a graph is quite common. Here, node a is below node b and connected by an arc if and only if $a < b$ and there exists no c with $a < c < b$.

In this example, we have the partial order
$$\{(\emptyset, \emptyset), (\{1\}, \{1\}), (\{1\}, \{1\}), (\{1, 2\}, \{1, 2\}),$$
$$(\emptyset, \{1\}), (\emptyset, \{1\}), (\{1\}, \{1, 2\}), (\{1\}, \{1, 2\})\}.$$

Structures for Orders ...

Give structures for orders that are ...

1. not total: $\subseteq$-relation[119];
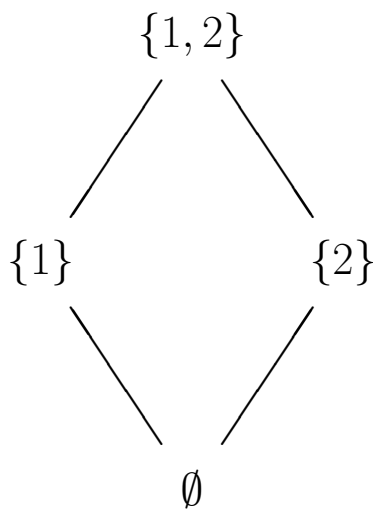
2. total but not dense: integers with $\leq$;

3. dense: reals with $\leq$.

---

[119]The $\subseteq$-relation is partial but not total. As an example, consider the $\subseteq$-relation on the set of subsets of $\{1, 2\}$.

$$\{1, 2\}$$

$$\{1\} \qquad\qquad \{2\}$$

$$\emptyset$$

Depicting partial orders by a such a graph is quite common. Here, node a is below node b and connected by an arc if and only if $a < b$ and there exists no c with $a < c < b$.

In this example, we have the partial order

$$\{(\emptyset, \emptyset), (\{1\}, \{1\}), (\{1\}, \{1\}), (\{1, 2\}, \{1, 2\}),$$
$$(\emptyset, \{1\}), (\emptyset, \{1\}), (\{1\}, \{1, 2\}), (\{1\}, \{1, 2\})\}.$$

6.2 Example 2: Groups

- Language: Function symbols $\_ \cdot \_$, $\_^{-1}$, e^{120}

6.2 Example 2: Groups

- Language: Function symbols $\_ \cdot \_$, $\_^{-1}$, e[120]

- A group is[121] a model[122] of

$$\forall x, y, z.\, (x \cdot y) \cdot z \;=\; x \cdot (y \cdot z) \qquad \text{(assoc)}$$
$$\forall x.\, x \cdot e \;=\; x \qquad\qquad\;\; \text{(r-neutr)}$$
$$\forall x.\, x \cdot x^{-1} \;=\; e \qquad\qquad\;\; \text{(r-inv)}$$

---

[120] $\_ \cdot \_$ is a binary infix function symbol (in fact, only $\cdot$ is the symbol, but the notation $\_ \cdot \_$ is used to indicate the fact that the symbol stands between its arguments).

$\_^{-1}$ is a unary function symbol written as superscript. Again, the $\_$ is used to indicate where the argument goes.

e is a nullary function symbol ($=$ constant).

Note that groups are very common in mathematics, and many different notations, i.e., function names and fixity (infix, prefix...) are used for them.

[121] In general mathematical terminology, a group consists of three function symbols $\_ \cdot \_$, $\_^{-1}$, e, obeying the following laws:

Associativity $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all a, b, c,

Right neutral $a \cdot e = a$ for all a,

Right inverse $a \cdot a^{-1} = e$ for all a.

[122] A model of the group axioms is a structure in which the group axioms are true.

It is an example of an equational theory[123].

It is an example of an equational theory[123].

Two theorems: (1) $x^{-1} \cdot x = e$ and (2) $e \cdot x = x$

We will now prove them.

However, when we say something like, "this model is a group", then this is a slight abuse of terminology, since there may be other function symbols around that are also interpreted by the structure.

So when we say "this model is a group", we mean, "this model is a model of the group axioms for function symbols $\_ \cdot \_$, $\_^{-1}$, and e clear from the context".

[123]An equational theory is a set of equations. Each equation is an axiom.

Sometimes, each equation is surrounded by several $\forall$-quantifiers binding all the free variables in the equation, but often the equation is regarded as implicitly universally quantified.

More generally, a conditional equational theory consists of proper rules where the premises are called conditions [Höl90].

Note also that sometimes, one also considers the basic rules of equality as being part of every equational theory. Whenever one has an equational theory, one implies that the basic rules

Theorem 1

$$\forall x, y, z. \, (x \cdot y) \cdot z = x \cdot (y \cdot z) \quad \text{(assoc)}$$
$$\forall x. \, x \cdot e = x \quad \text{(r-neutr)}$$
$$\forall x. \, x \cdot x^{-1} = e \quad \text{(r-inv)}$$

$$x^{-1} \cdot x = e \tag{1}$$

$$x^{-1} \cdot x =$$

are present; whether or not one assumes that they are formally elements of the equational theory is just a technical detail.

Theorem 1

$$\forall x, y, z.\, (x \cdot y) \cdot z = x \cdot (y \cdot z) \quad \text{(assoc)}$$
$$\forall x.\, x \cdot e = x \quad \text{(r-neutr)}$$
$$\forall x.\, x \cdot x^{-1} = e \quad \text{(r-inv)}$$

$$x^{-1} \cdot x = e \tag{1}$$

$$x^{-1} \cdot x = x^{-1} \cdot (x \cdot e)$$

are present; whether or not one assumes that they are formally elements of the equational theory is just a technical detail.

Theorem 1

$$\begin{aligned}
\forall x, y, z. \, (x \cdot y) \cdot z &= x \cdot (y \cdot z) \quad \text{(assoc)} \\
\forall x. \, x \cdot e &= x \quad \text{(r-neutr)} \\
\forall x. \, x \cdot x^{-1} &= e \quad \text{(r-inv)}
\end{aligned}$$

$$x^{-1} \cdot x = e \tag{1}$$

$$x^{-1} \cdot x = x^{-1} \cdot (x \cdot e)$$

are present; whether or not one assumes that they are formally elements of the equational theory is just a technical detail.

Theorem 1

$$\forall x, y, z.\, (x \cdot y) \cdot z = x \cdot (y \cdot z) \quad \text{(assoc)}$$
$$\forall x.\, x \cdot e = x \quad \text{(r-neutr)}$$
$$\forall x.\, x \cdot x^{-1} = e \quad \text{(r-inv)}$$

$$x^{-1} \cdot x = e \tag{1}$$

$$x^{-1} \cdot x = x^{-1} \cdot (x \cdot e) = x^{-1} \cdot (x \cdot (x^{-1} \cdot x^{-1^{-1}}))$$

are present; whether or not one assumes that they are formally elements of the equational theory is just a technical detail.

Theorem 1

$$\forall x, y, z. \, (x \cdot y) \cdot z \; = \; x \cdot (y \cdot z) \quad \text{(assoc)}$$
$$\forall x. \, x \cdot e \qquad\qquad = \qquad\quad x \quad \text{(r-neutr)}$$
$$\forall x. \, x \cdot x^{-1} \qquad\quad = \qquad\quad e \quad\;\; \text{(r-inv)}$$

$$x^{-1} \cdot x = e \qquad\qquad\qquad\qquad (1)$$

$$x^{-1} \cdot x = x^{-1} \cdot (x \cdot e) = x^{-1} \cdot (x \cdot (x^{-1} \cdot x^{-1^{-1}}))$$

are present; whether or not one assumes that they are formally elements of the equational theory is just a technical detail.

Theorem 1

$$\frac{\begin{array}{rcll} \forall x, y, z.\, (x \cdot y) \cdot z & = & x \cdot (y \cdot z) & \text{(assoc)} \\ \forall x.\, x \cdot e & = & x & \text{(r-neutr)} \\ \forall x.\, x \cdot x^{-1} & = & e & \text{(r-inv)} \end{array}}{x^{-1} \cdot x = e} \tag{1}$$

$$x^{-1} \cdot x = x^{-1} \cdot (x \cdot e) = x^{-1} \cdot (x \cdot (x^{-1} \cdot x^{-1^{-1}})) =$$
$$x^{-1} \cdot ((x \cdot x^{-1}) \cdot x^{-1^{-1}})$$

are present; whether or not one assumes that they are formally elements of the equational theory is just a technical detail.

Theorem 1

$$\forall x, y, z.\, (x \cdot y) \cdot z = x \cdot (y \cdot z) \quad \text{(assoc)}$$
$$\forall x.\, x \cdot e = x \quad \text{(r-neutr)}$$
$$\forall x.\, x \cdot x^{-1} = e \quad \text{(r-inv)}$$

$$x^{-1} \cdot x = e \tag{1}$$

$$x^{-1} \cdot x = x^{-1} \cdot (x \cdot e) = x^{-1} \cdot (x \cdot (x^{-1} \cdot x^{-1^{-1}})) =$$
$$x^{-1} \cdot ((x \cdot x^{-1}) \cdot x^{-1^{-1}})$$

---

are present; whether or not one assumes that they are formally elements of the equational theory is just a technical detail.

Theorem 1

$$\begin{aligned}
\forall x, y, z.\, (x \cdot y) \cdot z &= x \cdot (y \cdot z) &&\text{(assoc)} \\
\forall x.\, x \cdot e &= x &&\text{(r-neutr)} \\
\forall x.\, x \cdot x^{-1} &= e &&\text{(r-inv)}
\end{aligned}$$

$$x^{-1} \cdot x = e \qquad\qquad (1)$$

$$x^{-1} \cdot x = x^{-1} \cdot (x \cdot e) = x^{-1} \cdot (x \cdot (x^{-1} \cdot x^{-1^{-1}})) =$$
$$x^{-1} \cdot ((x \cdot x^{-1}) \cdot x^{-1^{-1}}) = x^{-1} \cdot (e \cdot x^{-1^{-1}})$$

---

are present; whether or not one assumes that they are formally elements of the equational theory is just a technical detail.

Theorem 1

$$\forall x, y, z. \, (x \cdot y) \cdot z \; = \; x \cdot (y \cdot z) \quad \text{(assoc)}$$
$$\forall x. \, x \cdot e \qquad\qquad = \qquad\qquad x \quad \text{(r-neutr)}$$
$$\forall x. \, x \cdot x^{-1} \qquad\quad = \qquad\qquad e \quad \text{(r-inv)}$$

$$x^{-1} \cdot x = e \tag{1}$$

$$x^{-1} \cdot x = x^{-1} \cdot (x \cdot e) = x^{-1} \cdot (x \cdot (x^{-1} \cdot x^{-1^{-1}})) =$$
$$x^{-1} \cdot ((x \cdot x^{-1}) \cdot x^{-1^{-1}}) = x^{-1} \cdot (e \cdot x^{-1^{-1}})$$

are present; whether or not one assumes that they are formally elements of the equational theory is just a technical detail.

Theorem 1

$$\forall x, y, z. \, (x \cdot y) \cdot z \;=\; x \cdot (y \cdot z) \quad \text{(assoc)}$$
$$\forall x. \, x \cdot e \;=\; x \quad \text{(r-neutr)}$$
$$\forall x. \, x \cdot x^{-1} \;=\; e \quad \text{(r-inv)}$$

$$x^{-1} \cdot x = e \tag{1}$$

$$x^{-1} \cdot x = x^{-1} \cdot (x \cdot e) = x^{-1} \cdot (x \cdot (x^{-1} \cdot x^{-1^{-1}})) =$$
$$x^{-1} \cdot ((x \cdot x^{-1}) \cdot x^{-1^{-1}}) = x^{-1} \cdot (e \cdot x^{-1^{-1}}) =$$
$$(x^{-1} \cdot e) \cdot x^{-1^{-1}}$$

---

are present; whether or not one assumes that they are formally
elements of the equational theory is just a technical detail.

117

Theorem 1

$$\forall x, y, z. \, (x \cdot y) \cdot z \;=\; x \cdot (y \cdot z) \quad \text{(assoc)}$$
$$\forall x. \, x \cdot e \;=\; x \quad \text{(r-neutr)}$$
$$\forall x. \, x \cdot x^{-1} \;=\; e \quad \text{(r-inv)}$$

$$x^{-1} \cdot x = e \tag{1}$$

$$x^{-1} \cdot x = x^{-1} \cdot (x \cdot e) = x^{-1} \cdot (x \cdot (x^{-1} \cdot x^{-1^{-1}})) =$$
$$x^{-1} \cdot ((x \cdot x^{-1}) \cdot x^{-1^{-1}}) = x^{-1} \cdot (e \cdot x^{-1^{-1}}) =$$
$$(x^{-1} \cdot e) \cdot x^{-1^{-1}}$$

are present; whether or not one assumes that they are formally elements of the equational theory is just a technical detail.

Theorem 1

$$\forall x, y, z.\, (x \cdot y) \cdot z = x \cdot (y \cdot z) \quad \text{(assoc)}$$
$$\forall x.\, x \cdot e \qquad\quad = \qquad\quad x \quad \text{(r-neutr)}$$
$$\forall x.\, x \cdot x^{-1} \qquad\quad = \qquad\quad e \quad \text{(r-inv)}$$

$$x^{-1} \cdot x = e \tag{1}$$

$$x^{-1} \cdot x = x^{-1} \cdot (x \cdot e) = x^{-1} \cdot (x \cdot (x^{-1} \cdot x^{-1^{-1}})) =$$
$$x^{-1} \cdot ((x \cdot x^{-1}) \cdot x^{-1^{-1}}) = x^{-1} \cdot (e \cdot x^{-1^{-1}}) =$$
$$(x^{-1} \cdot e) \cdot x^{-1^{-1}} = x^{-1} \cdot x^{-1^{-1}}$$

are present; whether or not one assumes that they are formally elements of the equational theory is just a technical detail.

Theorem 1

$$\begin{aligned}
\forall x, y, z. \, (x \cdot y) \cdot z &= x \cdot (y \cdot z) && \text{(assoc)} \\
\forall x. \, x \cdot e &= x && \text{(r-neutr)} \\
\forall x. \, x \cdot x^{-1} &= e && \text{(r-inv)}
\end{aligned}$$

$$x^{-1} \cdot x = e \tag{1}$$

$$
\begin{aligned}
x^{-1} \cdot x &= x^{-1} \cdot (x \cdot e) = x^{-1} \cdot (x \cdot (x^{-1} \cdot x^{-1^{-1}})) = \\
x^{-1} \cdot ((x \cdot x^{-1}) \cdot x^{-1^{-1}}) &= x^{-1} \cdot (e \cdot x^{-1^{-1}}) = \\
(x^{-1} \cdot e) \cdot x^{-1^{-1}} &= x^{-1} \cdot x^{-1^{-1}}
\end{aligned}
$$

are present; whether or not one assumes that they are formally elements of the equational theory is just a technical detail.

Theorem 1

$$\forall x, y, z.\, (x \cdot y) \cdot z = x \cdot (y \cdot z) \quad \text{(assoc)}$$
$$\forall x.\, x \cdot e = x \quad \text{(r-neutr)}$$
$$\forall x.\, x \cdot x^{-1} = e \quad \text{(r-inv)}$$

$$x^{-1} \cdot x = e \tag{1}$$

$$x^{-1} \cdot x = x^{-1} \cdot (x \cdot e) = x^{-1} \cdot (x \cdot (x^{-1} \cdot x^{-1^{-1}})) =$$
$$x^{-1} \cdot ((x \cdot x^{-1}) \cdot x^{-1^{-1}}) = x^{-1} \cdot (e \cdot x^{-1^{-1}}) =$$
$$(x^{-1} \cdot e) \cdot x^{-1^{-1}} = x^{-1} \cdot x^{-1^{-1}} = e$$

---

are present; whether or not one assumes that they are formally elements of the equational theory is just a technical detail.

Theorem 1

$$\forall x, y, z.\, (x \cdot y) \cdot z = x \cdot (y \cdot z) \quad \text{(assoc)}$$
$$\forall x.\, x \cdot e = x \quad \text{(r-neutr)}$$
$$\forall x.\, x \cdot x^{-1} = e \quad \text{(r-inv)}$$

$$x^{-1} \cdot x = e \tag{1}$$

$$x^{-1} \cdot x = x^{-1} \cdot (x \cdot e) = x^{-1} \cdot (x \cdot (x^{-1} \cdot x^{-1^{-1}})) =$$
$$x^{-1} \cdot ((x \cdot x^{-1}) \cdot x^{-1^{-1}}) = x^{-1} \cdot (e \cdot x^{-1^{-1}}) =$$
$$(x^{-1} \cdot e) \cdot x^{-1^{-1}} = x^{-1} \cdot x^{-1^{-1}} = e.$$

are present; whether or not one assumes that they are formally elements of the equational theory is just a technical detail.

Theorem 2

$$\forall x, y, z. \, (x \cdot y) \cdot z \; = \; x \cdot (y \cdot z) \quad \text{(assoc)}$$
$$\forall x. \, x \cdot e \qquad\qquad = \qquad\qquad x \quad \text{(r-neutr)}$$
$$\forall x. \, x \cdot x^{-1} \qquad\quad = \qquad\qquad e \quad \text{(r-inv)}$$

$$e \cdot x = x \qquad\qquad\qquad\qquad (2)$$

$e \cdot x$

Theorem 2

$$\forall x, y, z. \, (x \cdot y) \cdot z \; = \; x \cdot (y \cdot z) \quad \text{(assoc)}$$
$$\forall x. \, x \cdot e \qquad\qquad = \qquad\qquad x \quad \text{(r-neutr)}$$
$$\forall x. \, x \cdot x^{-1} \qquad\quad = \qquad\qquad e \quad \text{(r-inv)}$$

$$e \cdot x = x \tag{2}$$

$$e \cdot x = (x \cdot x^{-1}) \cdot x$$

Theorem 2

$$\forall x, y, z. (x \cdot y) \cdot z = x \cdot (y \cdot z) \quad \text{(assoc)}$$
$$\forall x. x \cdot e = x \quad \text{(r-neutr)}$$
$$\forall x. x \cdot x^{-1} = e \quad \text{(r-inv)}$$

$$e \cdot x = x \tag{2}$$

$$e \cdot x = (x \cdot x^{-1}) \cdot x$$

Theorem 2

$$\forall x, y, z. \, (x \cdot y) \cdot z = x \cdot (y \cdot z) \quad \text{(assoc)}$$
$$\forall x. \, x \cdot e = x \quad \text{(r-neutr)}$$
$$\forall x. \, x \cdot x^{-1} = e \quad \text{(r-inv)}$$

$$e \cdot x = x \qquad\qquad (2)$$

$$e \cdot x = (x \cdot x^{-1}) \cdot x = x \cdot (x^{-1} \cdot x)$$

Theorem 2

$$\forall x, y, z. \, (x \cdot y) \cdot z = x \cdot (y \cdot z) \quad \text{(assoc)}$$
$$\forall x. \, x \cdot e = x \quad \text{(r-neutr)}$$
$$\forall x. \, x \cdot x^{-1} = e \quad \text{(r-inv)}$$

$$e \cdot x = x \tag{2}$$

$$e \cdot x = (x \cdot x^{-1}) \cdot x = x \cdot (x^{-1} \cdot x) \qquad \text{(Theorem 1)}$$

Theorem 2

$$
\begin{aligned}
\forall x, y, z.\, (x \cdot y) \cdot z &= x \cdot (y \cdot z) && \text{(assoc)} \\
\forall x.\, x \cdot e &= x && \text{(r-neutr)} \\
\forall x.\, x \cdot x^{-1} &= e && \text{(r-inv)}
\end{aligned}
$$

$$
e \cdot x = x \tag{2}
$$

$$
e \cdot x = (x \cdot x^{-1}) \cdot x = x \cdot (x^{-1} \cdot x) = x \cdot e
$$

Theorem 2

$$\forall x, y, z.\, (x \cdot y) \cdot z \;=\; x \cdot (y \cdot z) \quad \text{(assoc)}$$
$$\forall x.\, x \cdot e \;=\; x \quad \text{(r-neutr)}$$
$$\forall x.\, x \cdot x^{-1} \;=\; e \quad \text{(r-inv)}$$

$$e \cdot x = x \tag{2}$$

$$e \cdot x = (x \cdot x^{-1}) \cdot x = x \cdot (x^{-1} \cdot x) = x \cdot e$$

Theorem 2

$$\forall x, y, z.\, (x \cdot y) \cdot z \;=\; x \cdot (y \cdot z) \quad \text{(assoc)}$$
$$\forall x.\, x \cdot e \qquad\qquad =\qquad\qquad x \quad \text{(r-neutr)}$$
$$\forall x.\, x \cdot x^{-1} \qquad\quad =\qquad\qquad e \quad \text{(r-inv)}$$

$$e \cdot x = x \tag{2}$$

$$e \cdot x = (x \cdot x^{-1}) \cdot x = x \cdot (x^{-1} \cdot x) = x \cdot e = x$$

Theorem 2

$$\begin{array}{rcll}
\forall x, y, z.\, (x \cdot y) \cdot z & = & x \cdot (y \cdot z) & \text{(assoc)} \\
\forall x.\, x \cdot e & = & x & \text{(r-neutr)} \\
\forall x.\, x \cdot x^{-1} & = & e & \text{(r-inv)}
\end{array}$$

$$e \cdot x = x \qquad\qquad\qquad (2)$$

$$e \cdot x = (x \cdot x^{-1}) \cdot x = x \cdot (x^{-1} \cdot x) = x \cdot e = x.$$

6.3 Lessons Learned from these Examples

Equational proofs are often tricky!

- Equalities used in different directions, "eureka"[124] terms, etc.

- In some cases (the word problem[125] is) decidable.

---

[124]By "eureka" terms we mean terms that have to be guessed in order to find a proof. At least at first sight, it seems like these terms simply fall from the sky.

The Greek heureka is 1st person singular perfect of heuriskō, "to find". It was exclaimed by Archimedes upon discovering how to test the purity of Hiero's crown.

[125]The word problem w.r.t. an equational theory (here: the group axioms) is the problem of deciding whether two terms s and t are equal in the theory, that is to say, whether the formula $s = t$ is true in any model of the theory.

Equational versus ND Proofs

- Above proofs were of a particular, equational form[126].

---

[126]An equational proof consists simply of a sequence of equations, written as $t_1 = t_2 = \ldots = t_n$, where each t_{i+1} is obtained from t_i by replacing some subterm s with a term s', provided the equality $s = s'$ holds.

This style of proof can be justified by the rules given for equality, in particular the congruences. However, it looks very different from the natural deduction style.

[127]

Most steps use the congruence rule $cong_2$.

Each framed box in the derivation tree stands for a sub-tree consisting of a group axiom and possibly several applications

Equational versus ND Proofs

- Above proofs were of a particular, equational form[126].

- In Isabelle this is accomplished by term rewriting.

  Term rewriting is a process for replacing equals by equals (see later).

---

[126]An equational proof consists simply of a sequence of equations, written as $t_1 = t_2 = \ldots = t_n$, where each t_{i+1} is obtained from t_i by replacing some subterm s with a term s', provided the equality $s = s'$ holds.

This style of proof can be justified by the rules given for equality, in particular the congruences. However, it looks very different from the natural deduction style.

127

$$e \cdot x = x$$

Most steps use the congruence rule $cong_2$.

Each framed box in the derivation tree stands for a sub-tree consisting of a group axiom and possibly several applications

Equational versus ND Proofs

- Above proofs were of a particular, equational form[126].

- In Isabelle this is accomplished by term rewriting.

  Term rewriting is a process for replacing equals by equals (see later).

- Alternative is natural deduction:

  - requires explicit proofs using equality rules;
  - tedious in practice. Try it on above examples![127]

---

[126]An equational proof consists simply of a sequence of equations, written as $t_1 = t_2 = \ldots = t_n$, where each t_{i+1} is obtained from t_i by replacing some subterm s with a term s', provided the equality $s = s'$ holds.

This style of proof can be justified by the rules given for equality, in particular the congruences. However, it looks very different from the natural deduction style.

[127]

$$e \cdot x = x$$

Most steps use the congruence rule $cong_2$.

Each framed box in the derivation tree stands for a sub-tree consisting of a group axiom and possibly several applications

Equational versus ND Proofs

- Above proofs were of a particular, equational form[126].

- In Isabelle this is accomplished by term rewriting.

  Term rewriting is a process for replacing equals by equals (see later).

- Alternative is natural deduction:

  - requires explicit proofs using equality rules;
  - tedious in practice. Try it on above examples![127]

---

[126]An equational proof consists simply of a sequence of equations, written as $t_1 = t_2 = \ldots = t_n$, where each t_{i+1} is obtained from t_i by replacing some subterm s with a term s', provided the equality $s = s'$ holds.

This style of proof can be justified by the rules given for equality, in particular the congruences. However, it looks very different from the natural deduction style.

[127]

$$\frac{\overline{\quad\quad}}{x \cdot e = x} \qquad \frac{\overline{\quad\quad\quad\quad\quad\quad\quad\quad}}{e \cdot x = x \cdot e}$$
$$e \cdot x = x$$

Most steps use the congruence rule $cong_2$.

Each framed box in the derivation tree stands for a sub-tree consisting of a group axiom and possibly several applications

Equational versus ND Proofs

- Above proofs were of a particular, equational form[126].

- In Isabelle this is accomplished by term rewriting.

  Term rewriting is a process for replacing equals by equals (see later).

- Alternative is natural deduction:

  - requires explicit proofs using equality rules;
  - tedious in practice. Try it on above examples![127]

---

[126]An equational proof consists simply of a sequence of equations, written as $t_1 = t_2 = \ldots = t_n$, where each t_{i+1} is obtained from t_i by replacing some subterm s with a term s', provided the equality $s = s'$ holds.

This style of proof can be justified by the rules given for equality, in particular the congruences. However, it looks very different from the natural deduction style.

[127]

$$\frac{\boxed{\text{r-neutr}}}{x \cdot e = x} \qquad \frac{}{e \cdot x = x \cdot e}$$
$$e \cdot x = x$$

Most steps use the congruence rule $cong_2$.

Each framed box in the derivation tree stands for a sub-tree consisting of a group axiom and possibly several applications

Equational versus ND Proofs

- Above proofs were of a particular, equational form[126].

- In Isabelle this is accomplished by term rewriting.

  Term rewriting is a process for replacing equals by equals (see later).

- Alternative is natural deduction:

  - requires explicit proofs using equality rules;
  - tedious in practice. Try it on above examples![127]

---

[126] An equational proof consists simply of a sequence of equations, written as $t_1 = t_2 = \ldots = t_n$, where each t_{i+1} is obtained from t_i by replacing some subterm s with a term s', provided the equality $s = s'$ holds.

This style of proof can be justified by the rules given for equality, in particular the congruences. However, it looks very different from the natural deduction style.

[127]

$$\frac{\boxed{\text{r-neutr}}}{x \cdot e = x} \qquad \frac{\dfrac{x^{-1} \cdot x = e}{e \cdot x = x \cdot e} \qquad \dfrac{}{e \cdot x = x \cdot (x^{-1} \cdot x)}}{e \cdot x = x}$$

Most steps use the congruence rule $cong_2$.

Each framed box in the derivation tree stands for a sub-tree consisting of a group axiom and possibly several applications

120

Equational versus ND Proofs

- Above proofs were of a particular, equational form[126].

- In Isabelle this is accomplished by term rewriting.

  Term rewriting is a process for replacing equals by equals (see later).

- Alternative is natural deduction:

  - requires explicit proofs using equality rules;
  - tedious in practice. Try it on above examples![127]

---

[126]An equational proof consists simply of a sequence of equations, written as $t_1 = t_2 = \ldots = t_n$, where each t_{i+1} is obtained from t_i by replacing some subterm s with a term s', provided the equality $s = s'$ holds.

This style of proof can be justified by the rules given for equality, in particular the congruences. However, it looks very different from the natural deduction style.

[127]

$$
\cfrac{\boxed{\text{r-neutr}} \quad \cfrac{\text{Theorem 1}}{x^{-1} \cdot x = e}}{\cfrac{x \cdot e = x \qquad \cfrac{e \cdot x = x \cdot (x^{-1} \cdot x)}{e \cdot x = x \cdot e}}{e \cdot x = x}}
$$

Most steps use the congruence rule $cong_2$.

Each framed box in the derivation tree stands for a sub-tree consisting of a group axiom and possibly several applications

120

Equational versus ND Proofs

- Above proofs were of a particular, equational form[126].

- In Isabelle this is accomplished by term rewriting.

  Term rewriting is a process for replacing equals by equals (see later).

- Alternative is natural deduction:

  - requires explicit proofs using equality rules;
  - tedious in practice. Try it on above examples![127]

---

[126] An equational proof consists simply of a sequence of equations, written as $t_1 = t_2 = \ldots = t_n$, where each t_{i+1} is obtained from t_i by replacing some subterm s with a term s', provided the equality $s = s'$ holds.

This style of proof can be justified by the rules given for equality, in particular the congruences. However, it looks very different from the natural deduction style.

[127]

$$
\cfrac{
  \boxed{\text{r-neutr}} \quad
  \cfrac{
    \cfrac{\text{Theorem 1}}{x^{-1} \cdot x = e} \qquad
    \cfrac{
      \cfrac{}{(x \cdot x^{-1}) \cdot x = x \cdot (x^{-1} \cdot x)} \qquad \cfrac{}{e \cdot x = (x \cdot x^{-1}) \cdot x}
    }{e \cdot x = x \cdot (x^{-1} \cdot x)}
  }{e \cdot x = x \cdot e}
}{
  \begin{array}{c} x \cdot e = x \\[2pt] \end{array} \quad e \cdot x = x
}
$$

Most steps use the congruence rule $cong_2$.

Each framed box in the derivation tree stands for a sub-tree consisting of a group axiom and possibly several applications

Equational versus ND Proofs

- Above proofs were of a particular, equational form[126].

- In Isabelle this is accomplished by term rewriting.

  Term rewriting is a process for replacing equals by equals (see later).

- Alternative is natural deduction:

  – requires explicit proofs using equality rules;

  – tedious in practice. Try it on above examples![127]

---

[126]An equational proof consists simply of a sequence of equations, written as $t_1 = t_2 = \ldots = t_n$, where each t_{i+1} is obtained from t_i by replacing some subterm s with a term s', provided the equality $s = s'$ holds.

This style of proof can be justified by the rules given for equality, in particular the congruences. However, it looks very different from the natural deduction style.

[127]

$$
\cfrac{
  \boxed{\text{r-neutr}}
}{x \cdot e = x}
\quad
\cfrac{
  \cfrac{
    \boxed{\text{Theorem 1}}
  }{x^{-1} \cdot x = e}
  \qquad
  \cfrac{
    \cfrac{\boxed{\text{assoc}}}{(x \cdot x^{-1}) \cdot x = x \cdot (x^{-1} \cdot x)}
    \qquad
    \cfrac{}{e \cdot x = (x \cdot x^{-1}) \cdot x}
  }{e \cdot x = x \cdot (x^{-1} \cdot x)}
}{\cfrac{e \cdot x = x \cdot e}{e \cdot x = x}}
$$

Most steps use the congruence rule $cong_2$.

Each framed box in the derivation tree stands for a sub-tree consisting of a group axiom and possibly several applications

Equational versus ND Proofs

- Above proofs were of a particular, equational form[126].

- In Isabelle this is accomplished by term rewriting.

  Term rewriting is a process for replacing equals by equals (see later).

- Alternative is natural deduction:

  - requires explicit proofs using equality rules;
  - tedious in practice. Try it on above examples![127]

---

[126]An equational proof consists simply of a sequence of equations, written as $t_1 = t_2 = \ldots = t_n$, where each t_{i+1} is obtained from t_i by replacing some subterm s with a term s', provided the equality $s = s'$ holds.

This style of proof can be justified by the rules given for equality, in particular the congruences. However, it looks very different from the natural deduction style.

[127]

$$
\cfrac{
\boxed{\text{r-neutr}}
\quad
\cfrac{
\cfrac{\text{Theorem 1}}{x^{-1} \cdot x = e}
\quad
\cfrac{
\boxed{\text{assoc}}
\quad
\cfrac{(x \cdot x^{-1}) \cdot x = x \cdot (x^{-1} \cdot x)}{}
\quad
\cfrac{e = x \cdot x^{-1} \qquad \overline{e \cdot x = e \cdot x}}{e \cdot x = (x \cdot x^{-1}) \cdot x}
}{e \cdot x = x \cdot (x^{-1} \cdot x)}
}{e \cdot x = x \cdot e}
}{
\cfrac{x \cdot e = x \qquad e \cdot x = x \cdot e}{e \cdot x = x}
}
$$

Most steps use the congruence rule $cong_2$.

Each framed box in the derivation tree stands for a sub-tree consisting of a group axiom and possibly several applications

Equational versus ND Proofs

- Above proofs were of a particular, equational form[126].

- In Isabelle this is accomplished by term rewriting.

  Term rewriting is a process for replacing equals by equals (see later).

- Alternative is natural deduction:

  - requires explicit proofs using equality rules;
  - tedious in practice. Try it on above examples![127]

---

[126]An equational proof consists simply of a sequence of equations, written as $t_1 = t_2 = \ldots = t_n$, where each t_{i+1} is obtained from t_i by replacing some subterm s with a term s', provided the equality $s = s'$ holds.

This style of proof can be justified by the rules given for equality, in particular the congruences. However, it looks very different from the natural deduction style.

[127]

$$
\cfrac{
  \boxed{\text{r-neutr}}
  \cfrac{x \cdot e = x}{}
  \quad
  \cfrac{
    \cfrac{\text{Theorem 1}}{x^{-1} \cdot x = e}
    \quad
    \cfrac{
      \boxed{\text{assoc}}
      \cfrac{(x \cdot x^{-1}) \cdot x = x \cdot (x^{-1} \cdot x)}{}
      \quad
      \cfrac{
        \cfrac{\text{sym}}{e = x \cdot x^{-1}}
        \quad
        \cfrac{}{e \cdot x = e \cdot x}
      }{e \cdot x = (x \cdot x^{-1}) \cdot x}
    }{e \cdot x = x \cdot (x^{-1} \cdot x)}
  }{e \cdot x = x \cdot e}
}{e \cdot x = x}
$$

Most steps use the congruence rule $cong_2$.

Each framed box in the derivation tree stands for a sub-tree consisting of a group axiom and possibly several applications

120

Equational versus ND Proofs

- Above proofs were of a particular, equational form[126].

- In Isabelle this is accomplished by term rewriting.

  Term rewriting is a process for replacing equals by equals (see later).

- Alternative is natural deduction:

  - requires explicit proofs using equality rules;
  - tedious in practice. Try it on above examples![127]

---

[126]An equational proof consists simply of a sequence of equations, written as $t_1 = t_2 = \ldots = t_n$, where each t_{i+1} is obtained from t_i by replacing some subterm s with a term s', provided the equality $s = s'$ holds.

This style of proof can be justified by the rules given for equality, in particular the congruences. However, it looks very different from the natural deduction style.

[127]

$$
\cfrac{
\boxed{\text{r-neutr}} \quad
\cfrac{
\cfrac{\boxed{\text{Theorem 1}}}{x^{-1} \cdot x = e} \quad
\cfrac{
\boxed{\text{assoc}} \quad
\cfrac{(x \cdot x^{-1}) \cdot x = x \cdot (x^{-1} \cdot x)}{} \quad
\cfrac{
\cfrac{\cfrac{x \cdot x^{-1} = e}{}}{e = x \cdot x^{-1}}\ sym \quad
\cfrac{}{e \cdot x = e \cdot x}
}{e \cdot x = (x \cdot x^{-1}) \cdot x}
}{e \cdot x = x \cdot (x^{-1} \cdot x)}
}{e \cdot x = x \cdot e}
}{e \cdot x = x}
$$

Most steps use the congruence rule $cong_2$.

Each framed box in the derivation tree stands for a sub-tree consisting of a group axiom and possibly several applications

Equational versus ND Proofs

- Above proofs were of a particular, equational form[126].

- In Isabelle this is accomplished by term rewriting.

  Term rewriting is a process for replacing equals by equals (see later).

- Alternative is natural deduction:

  - requires explicit proofs using equality rules;
  - tedious in practice. Try it on above examples![127]

---

[126]An equational proof consists simply of a sequence of equations, written as $t_1 = t_2 = \ldots = t_n$, where each t_{i+1} is obtained from t_i by replacing some subterm s with a term s', provided the equality $s = s'$ holds.

This style of proof can be justified by the rules given for equality, in particular the congruences. However, it looks very different from the natural deduction style.

[127]

Most steps use the congruence rule $cong_2$.

Each framed box in the derivation tree stands for a sub-tree consisting of a group axiom and possibly several applications

Equational versus ND Proofs

- Above proofs were of a particular, equational form[126].

- In Isabelle this is accomplished by term rewriting.

  Term rewriting is a process for replacing equals by equals (see later).

- Alternative is natural deduction:

  - requires explicit proofs using equality rules;
  - tedious in practice. Try it on above examples![127]

---

[126]An equational proof consists simply of a sequence of equations, written as $t_1 = t_2 = \ldots = t_n$, where each t_{i+1} is obtained from t_i by replacing some subterm s with a term s', provided the equality $s = s'$ holds.

This style of proof can be justified by the rules given for equality, in particular the congruences. However, it looks very different from the natural deduction style.

[127]

$$
\cfrac{
  \cfrac{\boxed{\text{r-neutr}}}{x \cdot e = x}
  \quad
  \cfrac{
    \cfrac{\boxed{\text{Theorem 1}}}{x^{-1} \cdot x = e}
    \quad
    \cfrac{
      \cfrac{\boxed{\text{assoc}}}{(x \cdot x^{-1}) \cdot x = x \cdot (x^{-1} \cdot x)}
      \quad
      \cfrac{
        \cfrac{\cfrac{\boxed{\text{r-inv}}}{x \cdot x^{-1} = e}}{e = x \cdot x^{-1}}\ \textit{sym}
        \quad
        \cfrac{}{e \cdot x = e \cdot x}\ \textit{refl}
      }{e \cdot x = (x \cdot x^{-1}) \cdot x}
    }{e \cdot x = x \cdot (x^{-1} \cdot x)}
  }{e \cdot x = x \cdot e}
}{e \cdot x = x}
$$

Most steps use the congruence rule $cong_2$.

Each framed box in the derivation tree stands for a sub-tree consisting of a group axiom and possibly several applications

7 Naïve Set Theory

7.1 Naïve Set Theory: Basics

- A set is a collection of objects where order and repetition are unimportant.

  Sets are central in mathematical reasoning [Vel94]. E.g., set of prime numbers.

7 Naïve Set Theory

7.1 Naïve Set Theory: Basics

- A set is a collection of objects where order and repetition are unimportant.

  Sets are central in mathematical reasoning [Vel94]. E.g., set of prime numbers.

- In what follows we consider a simple, intuitive formalization: "naïve set theory".

  We will be somewhat less formal than usual. Our goal is to understand standard mathematical practice.

  Later, in HOL, we will be completely formal.

of $\forall$-*E*.

Sets: Language

Assuming any first-order language with equality, we add:

- set-comprehension $\{x|P(x)\}$[128] and a binary membership predicate $\in$.

---

[128]Set comprehension is a way of defining sets. $\{x|P(x)\}$ stands for the set of elements of the universe for which $P(x)$ (some formula usually containing x) holds.

[129]It is more adequate to regard a set as a term than as a formula. A set is a "thing", not a statement about "things".

After all, we have the predicate $\in$ expecting a set on the RHS (and even the LHS may be a set!), and predicates take terms as arguments.

However, the syntax used in set comprehensions is not legal syntax for terms, since $P(x)$ is a formula.

This is why we introduce a special syntactic category for sets.

Sets: Language

Assuming any first-order language with equality, we add:

- set-comprehension $\{x|P(x)\}$[128] and a binary membership predicate $\in$.

- Term/formula distinction inadequate[129]: need a syntactic category for sets.

---

[128]Set comprehension is a way of defining sets. $\{x|P(x)\}$ stands for the set of elements of the universe for which $P(x)$ (some formula usually containing x) holds.

[129]It is more adequate to regard a set as a term than as a formula. A set is a "thing", not a statement about "things".

After all, we have the predicate $\in$ expecting a set on the RHS (and even the LHS may be a set!), and predicates take terms as arguments.

However, the syntax used in set comprehensions is not legal syntax for terms, since $P(x)$ is a formula.

This is why we introduce a special syntactic category for sets.

Sets: Language

Assuming any first-order language with equality, we add:

- set-comprehension $\{x|P(x)\}$[128] and a binary membership predicate $\in$.

- Term/formula distinction inadequate[129]: need a syntactic category for sets.

- We will be more formal about syntax later (HOL).

---

[128]Set comprehension is a way of defining sets. $\{x|P(x)\}$ stands for the set of elements of the universe for which $P(x)$ (some formula usually containing x) holds.

[129]It is more adequate to regard a set as a term than as a formula. A set is a "thing", not a statement about "things".

After all, we have the predicate $\in$ expecting a set on the RHS (and even the LHS may be a set!), and predicates take terms as arguments.

However, the syntax used in set comprehensions is not legal syntax for terms, since $P(x)$ is a formula.

This is why we introduce a special syntactic category for sets.

Sets: Language

Assuming any first-order language with equality, we add:

- set-comprehension $\{x|P(x)\}$[128] and a binary membership predicate $\in$.

- Term/formula distinction inadequate[129]: need a syntactic category for sets.

- We will be more formal about syntax later (HOL).

- Comprehension is a binding operator: x bound in $\{x|P(x)\}$.

---

[128]Set comprehension is a way of defining sets. $\{x|P(x)\}$ stands for the set of elements of the universe for which $P(x)$ (some formula usually containing x) holds.

[129]It is more adequate to regard a set as a term than as a formula. A set is a "thing", not a statement about "things".

After all, we have the predicate $\in$ expecting a set on the RHS (and even the LHS may be a set!), and predicates take terms as arguments.

However, the syntax used in set comprehensions is not legal syntax for terms, since $P(x)$ is a formula.

This is why we introduce a special syntactic category for sets.

Examples

- $\forall x. \, x \in \{y | y \bmod 6 = 0\} \rightarrow (x \bmod 2 = 0 \wedge x \bmod 3 = 0)$.

Examples

- $\forall x.\, x \in \{y | y \bmod 6 = 0\} \to (x \bmod 2 = 0 \land x \bmod 3 = 0)$.

- What does the following say?

$$2 \in \{w | 6 \notin \{x | x \text{ is divisible by } w\}\}$$

Examples

- $\forall x.\, x \in \{y | y \bmod 6 = 0\} \rightarrow (x \bmod 2 = 0 \wedge x \bmod 3 = 0)$.

- What does the following say?

$$2 \in \{w | 6 \notin \{x | x \text{ is divisible by } w\}\}$$

Answer: $6 \notin \{x | x \text{ divisible by } 2\}$ i.e., 6 not divisible by 2.

Proof Rules for Sets

Introduction, elimination, extensional equality[130]

$$\frac{P(t)}{t \in \{x|P(x)\}} \; compr\text{-}I \qquad \frac{t \in \{x|P(x)\}}{P(t)} \; compr\text{-}E$$

$$\frac{\forall x.\, x \in A \leftrightarrow x \in B}{A = B} \; =\text{-}I \qquad \frac{A = B}{\forall x.\, x \in A \leftrightarrow x \in B} \; =\text{-}E$$

The following equivalence is derivable[131]:

$$\forall x.\, P(x) \leftrightarrow x \in \{y|P(y)\}$$

---

[130]Two things are extensionally equal if they are "equal in their effects". Thus two sets are equal if they have the same members, regardless of what syntactic expressions are used to define those sets.

Note that extensional equality may be undecidable.

[131]

$$\cfrac{\cfrac{\dfrac{[P(x)]^1}{x \in \{y|P(y)\}} \; compr\text{-}I}{P(x) \to x \in \{y|P(y)\}} \to\text{-}I^1 \quad \cfrac{\dfrac{[x \in \{y|P(y)\}]^2}{P(x)} \; compr\text{-}E}{x \in \{y|P(y)\} \to P(x)} \to\text{-}I^2}{\cfrac{\cfrac{P(x) \to x \in \{y|P(y)\} \wedge x \in \{y|P(y)\} \to P(x)}{P(x) \leftrightarrow x \in \{y|P(y)\}} \; iff}{\forall x.\, P(x) \leftrightarrow x \in \{y|P(y)\}} \; \forall\text{-}I} \wedge\text{-}I$$

Rules $\wedge$-I, $\to$-I, $\forall$-I were defined in previous lectures. The step marked with iff is not a proof step in the technical sense. We only make the expansion of a shorthand notation explicit.

7.2 Digression: Sorted Reasoning

- In mathematical arguments we often (implicitly) assume that variables are restricted to some universe of discourse[132]. E.g., $x^2 < 9$ (universe either $\mathbb{R}$, $\mathbb{N}$, ...)

- To avoid ambiguity[133] we can include sort information in formulae:

$$\text{members } x \text{ of } U \text{ where } P(x) \equiv \{x \in U | P(x)\}$$

  Formally

$$\{x \in U | P(x)\} \equiv \{x \mid U(x) \wedge P(x)\}.$$

---

[132]We already know what a universe or domain is. To interpret a particular language, we have a structure interpreting all function symbols as functions on the universe.

However, it is often adequate to subdivide the universe into several "sub-universes". Those are called sorts. Note that a sort is a set.

For example, in a usual mathematical context, one may distinguish $\mathbb{R}$ (the real numbers) and $\mathbb{N}$ (the natural numbers) to say that $\sqrt{x}$ requires x to be of sort $\mathbb{R}$ and $x!$ requires x to be of sort $\mathbb{N}$.

[133]We want to make explicit the sort of the variable in question. So we do not want the set of all x such that $P(x)$ holds, but only the ones of the right sort, so the ones for which $x \in U$ (U being the sort/universe) holds.

Note there is a certain confusion here, since we write $x \in U$ in one place (so U should be a set) and $U(x)$ in another (so U should be a predicate. This confusion is deliberate and quite common. One can identify a set (sort) U with a unary

Sorted Reasoning in an Unsorted Logic

We may introduce the additional set comprehension syntax $\{x \in U | P(x)\}$, but our logic is still unsorted[134]. We have

$$y \in \{x \in U | P(x)\} \leftrightarrow y \in \{x \mid U(x) \wedge P(x)\} \leftrightarrow U(y) \wedge P(y)$$

---

predicate U such that $U(t)$ is interpreted as $True$ iff t is a member of U.

The whole expression $\{x \in U | P(x)\}$ is a special kind of syntax. Therefore, you must look at it as a whole: it makes no sense to see any meaning just in, say, the bit $x \in U$ in this expression. It is called set comprehension, and it is defined by

$$\{x \in U | P(x)\} \equiv \{x \mid U(x) \wedge P(x)\}.$$

[134] In sorted logic, sorts are part of the syntax. So the signature contains a fixed set of sorts. For each constant, it is specified what its sort is. For each function symbol, it is specified what the sort of each argument is, and what the sort of the result is. For each predicate symbol, it is specified what the sort of each argument is.

Terms and formulas that do not respect the sorts are not well-formed, and so they are not assigned a meaning.

In contrast, our logic is unsorted. The special syntax we

Sorted Quantification

$$\forall x \in U.\, P(x)^{135} \;\equiv\; \forall x.\, U(x) \rightarrow P(x)$$
$$\exists x \in U.\, P(x) \;\equiv\; \exists x.\, U(x) \wedge P(x)$$

---

provide for sorted reasoning is just syntactic sugar, i.e., we use it as shorthand and since it has an intuitive reasoning, but it has no impact on how expressive our logic is.

[135]So $\forall x \in U.\, P(x)$ is simply a shorthand or syntactic sugar for $\forall x.\, x \in U \rightarrow P(x)$, and analogously for $\exists x \in U.\, P(x)$.

7.3 Operations on Sets

- Functions on sets

$$
\begin{aligned}
A \cap {}^{136}B &\equiv \{x | x \in A \wedge x \in B\} \\
A \cup B &\equiv \{x | x \in A \vee x \in B\} \\
A \setminus B &\equiv \{x | x \in A \wedge x \notin B\}
\end{aligned}
$$

- Predicates on sets

$$
A \subseteq B \equiv \forall x.\, x \in A \rightarrow x \in B
$$

---

[136]

$\cap$ is called intersection.
$\cup$ is called union.
$\setminus$ is called set difference.
$\subseteq$ is called inclusion.

Examples of Operations on Sets

One often depicts sets as circles or bubbles.
What are $A \cap B$, $A \cup B$, $A \setminus B$?

A $\qquad\qquad\qquad\qquad$ B

Examples of Operations on Sets

One often depicts sets as circles or bubbles.
What are $A \cap B$, $A \cup B$, $A \setminus B$?

Examples of Operations on Sets

One often depicts sets as circles or bubbles.
What are $A \cap B$, $A \cup B$, $A \setminus B$?

$$A \cap B$$

Examples of Operations on Sets

One often depicts sets as circles or bubbles.
What are $A \cap B$, $A \cup B$, $A \setminus B$?

$A \cup B$
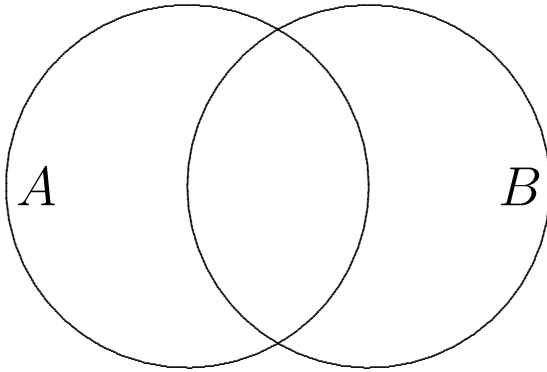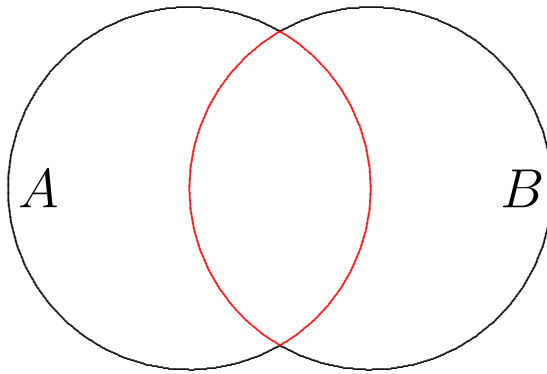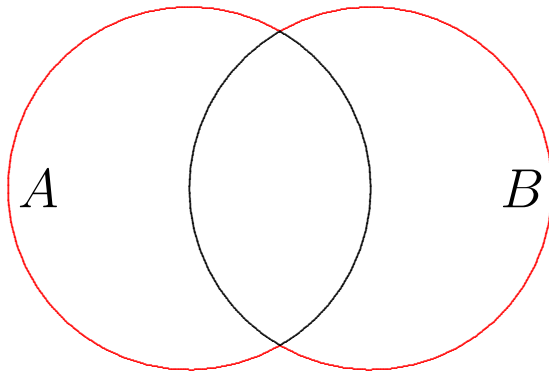
Examples of Operations on Sets

One often depicts sets as circles or bubbles.
What are $A \cap B$, $A \cup B$, $A \setminus B$?
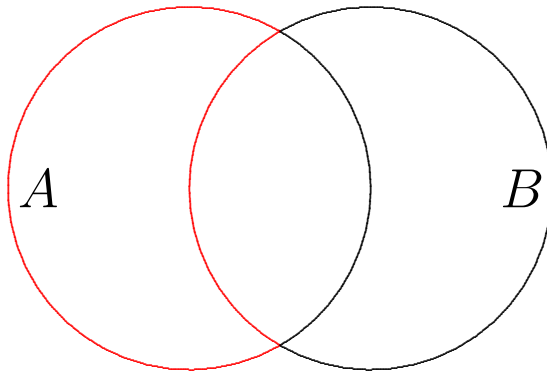
$A \setminus B$

Correspondence between Set-Theoretic and Logical Operators

$$x \in A \cap B \; \leftrightarrow \; x \in A \wedge x \in B$$
$$x \in A \cup B \; \leftrightarrow \; x \in A \vee x \in B$$
$$x \in A \setminus B \; \leftrightarrow \; x \in A \wedge x \notin B$$

These correspondences follow from the definitions of the set-theoretic operators and $\forall x. \, P(x) \leftrightarrow x \in \{y | P(y)\}$.

---

[137]When we transform an expression containing set operators $\cap, \cup, \setminus, \subseteq$ into an expression using $\wedge, \vee, \neg, \rightarrow$, we call the latter the logical form of the expression.

Correspondence between Set-Theoretic and Logical Operators

$$x \in A \cap B \;\leftrightarrow\; x \in A \wedge x \in B$$
$$x \in A \cup B \;\leftrightarrow\; x \in A \vee x \in B$$
$$x \in A \setminus B \;\leftrightarrow\; x \in A \wedge x \notin B$$

These correspondences follow from the definitions of the set-theoretic operators and $\forall x.\, P(x) \leftrightarrow x \in \{y | P(y)\}$.

 Example: what is the logical form[137] of $x \in ((A \cap B) \cup (A \cap C))$?

---

[137]When we transform an expression containing set operators $\cap, \cup, \setminus, \subseteq$ into an expression using $\wedge, \vee, \neg, \rightarrow$, we call the latter the logical form of the expression.

Correspondence between Set-Theoretic and Logical Operators

$$x \in A \cap B \;\leftrightarrow\; x \in A \land x \in B$$
$$x \in A \cup B \;\leftrightarrow\; x \in A \lor x \in B$$
$$x \in A \setminus B \;\leftrightarrow\; x \in A \land x \notin B$$

These correspondences follow from the definitions of the set-theoretic operators and $\forall x.\, P(x) \leftrightarrow x \in \{y \,|\, P(y)\}$.

Example: what is the logical form[137] of $x \in ((A \cap B) \cup (A \cap C))$? $(x \in A \land x \in B) \lor (x \in A \land x \in C)$

---

[137]When we transform an expression containing set operators $\cap, \cup, \setminus, \subseteq$ into an expression using $\land, \lor, \neg, \rightarrow$, we call the latter the logical form of the expression.

Proof of $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

---

[138]A Venn diagram draws sets as bubbles. Intersecting sets are drawn as overlapping bubbles, and the overlapping area is meant to depict the intersection of the sets.

A Venn diagram is not a proof in the sense defined earlier.

Moreover, it would not even be acceptable as a proof according to usual mathematical practice. If it is unknown whether two sets have a non-empty intersection, how are we supposed to draw them? Trying to make a case distinctions (drawing several diagrams depending on the cases) is error-prone.

Venn diagrams are useful for illustration purposes, but they are not proofs.

Proof of $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ **(1)**

Venn diagram (Is this a proof?)[138]

---

[138]A Venn diagram draws sets as bubbles. Intersecting sets are drawn as overlapping bubbles, and the overlapping area is meant to depict the intersection of the sets.

A Venn diagram is not a proof in the sense defined earlier.

Moreover, it would not even be acceptable as a proof according to usual mathematical practice. If it is unknown whether two sets have a non-empty intersection, how are we supposed to draw them? Trying to make a case distinctions (drawing several diagrams depending on the cases) is error-prone.

Venn diagrams are useful for illustration purposes, but they are not proofs.

Proof of $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ **(2)**

Natural deduction (natural language[139])

---

[139]We intersperse formal notation with natural language here in order to give an intuitive and short proof.

We can also do this more formally in Isabelle.

Proof of $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ **(2)**

Natural deduction (natural language[139])

By extensionality, suffices to show

$$\forall x.\, x \in A \cap (B \cup C) \leftrightarrow x \in (A \cap B) \cup (A \cap C).$$

---

[139]We intersperse formal notation with natural language here in order to give an intuitive and short proof.

We can also do this more formally in Isabelle.

Proof of $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ **(2)**

Natural deduction (natural language[139])

  By extensionality, suffices to show

$$\forall x.\, x \in A \cap (B \cup C) \leftrightarrow x \in (A \cap B) \cup (A \cap C).$$

For an arbitrary x, this is equivalent to establishing

$$(x \in A \wedge (x \in B \vee x \in C)) \leftrightarrow$$
$$(x \in A \wedge x \in B) \vee (x \in A \wedge x \in C)$$

---

[139]We intersperse formal notation with natural language here in order to give an intuitive and short proof.
  We can also do this more formally in Isabelle.

Proof of $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ **(2)**

Natural deduction (natural language[139])

By extensionality, suffices to show

$$\forall x.\, x \in A \cap (B \cup C) \leftrightarrow x \in (A \cap B) \cup (A \cap C).$$

For an arbitrary x, this is equivalent to establishing

$$(x \in A \land (x \in B \lor x \in C)) \leftrightarrow$$
$$(x \in A \land x \in B) \lor (x \in A \land x \in C)$$

But that is a propositional tautology.

---

[139]We intersperse formal notation with natural language here
in order to give an intuitive and short proof.
We can also do this more formally in Isabelle.

Same in Isabelle

Last proof carries over to Isabelle: extensionality, rewriting, tautology checking. Do it!

Prove: for all Sets A and B, $((A \cup B) \setminus B) \subseteq A$

Let's try a similar semi-formal proof:

Let A and B be arbitrary sets. $\hfill (\forall\text{-}I)$

Prove: for all Sets A and B, $((A \cup B) \setminus B) \subseteq A$

Let's try a similar semi-formal proof:

Let A and B be arbitrary sets.

Let A and B be arbitrary sets.   ($\forall$-I)
Let x be an element of $(A \cup B) \setminus B$ (temporary assumption)

Prove: for all Sets A and B, $((A \cup B) \setminus B) \subseteq A$

Let's try a similar semi-formal proof:

Let A and B be arbitrary sets.
Let x be element of $(A \cup B) \setminus B$.

Let A and B be arbitrary sets.	($\forall$-*I*)
Let x be an element of $(A \cup B) \setminus B$	(temporary assumption)
So $(x \in A \lor x \in B) \land \neg x \in B$	(equivalent proposition)

Prove: for all Sets A and B, $((A \cup B) \setminus B) \subseteq A$

Let's try a similar semi-formal proof:

Let A and B be arbitrary sets.
Let x be element of $(A \cup B) \setminus B$.
So $(x \in A \vee x \in B) \wedge \neg x \in B$.

Let A and B be arbitrary sets. $\qquad\qquad\qquad\qquad\qquad\qquad$ ($\forall$-I)
Let x be an element of $(A \cup B) \setminus B$ $\qquad\qquad$ (temporary assumption)
So $(x \in A \vee x \in B) \wedge \neg x \in B$ $\qquad\qquad$ (equivalent proposition)
Therefore $x \in A$ $\qquad\qquad$ (P follows from $(P \vee Q) \wedge \neg Q$)

Prove: for all Sets A and B, $((A \cup B) \setminus B) \subseteq A$

Let's try a similar semi-formal proof:

Let A and B be arbitrary sets.
Let x be element of $(A \cup B) \setminus B$.
So $(x \in A \vee x \in B) \wedge \neg x \in B$.
Therefore $x \in A$.

Let A and B be arbitrary sets. $(\forall\text{-}I)$
Let x be an element of $(A \cup B) \setminus B$ (temporary assumption)
So $(x \in A \vee x \in B) \wedge \neg x \in B$ (equivalent proposition)
Therefore $x \in A$ (P follows from $(P \vee Q) \wedge \neg Q$)
Therefore $x \in (A \cup B) \setminus B \rightarrow x \in A$ ($\rightarrow\text{-}I$)

Prove: for all Sets A and B, $((A \cup B) \setminus B) \subseteq A$

Let's try a similar semi-formal proof:

Let A and B be arbitrary sets.
Let x be element of $(A \cup B) \setminus B$.
So $(x \in A \lor x \in B) \land \neg x \in B$.
Therefore $x \in A$.
Therefore $x \in (A \cup B) \setminus B \rightarrow x \in A$.

Let A and B be arbitrary sets. $(\forall\text{-}I)$
Let x be an element of $(A \cup B) \setminus B$ (temporary assumption)
So $(x \in A \lor x \in B) \land \neg x \in B$ (equivalent proposition)
Therefore $x \in A$ (P follows from $(P \lor Q) \land \neg Q$)
Therefore $x \in (A \cup B) \setminus B \rightarrow x \in A$ $(\rightarrow\text{-}I)$
Therefore $((A \cup B) \setminus B) \subseteq A$ (def of $\subseteq$)

Prove: for all Sets A and B, $((A \cup B) \setminus B) \subseteq A$

Let's try a similar semi-formal proof:

Let A and B be arbitrary sets.
Let x be element of $(A \cup B) \setminus B$.
So $(x \in A \vee x \in B) \wedge \neg x \in B$.
Therefore $x \in A$.
Therefore $x \in (A \cup B) \setminus B \to x \in A$.
Therefore $((A \cup B) \setminus B) \subseteq A$.

Let A and B be arbitrary sets.	($\forall$-*I*)
Let x be an element of $(A \cup B) \setminus B$	(temporary assumption)
So $(x \in A \vee x \in B) \wedge \neg x \in B$	(equivalent proposition)
Therefore $x \in A$	(P follows from $(P \vee Q) \wedge \neg Q$)
Therefore $x \in (A \cup B) \setminus B \to x \in A$	($\to$-*I*)
Therefore $((A \cup B) \setminus B) \subseteq A$	(def of $\subseteq$)

Concerning forward and backwards reasoning, one may look at it as follows: we first construct the derivation step at the root of the proof tree ($\forall$-*I*), and then we jump to a leaf (by making the temporary assumption) and work downwards from there.

Prove: for all Sets A and B, $((A \cup B) \setminus B) \subseteq A$

Let's try a similar semi-formal proof:

Let A and B be arbitrary sets.
Let x be element of $(A \cup B) \setminus B$.
So $(x \in A \vee x \in B) \wedge \neg x \in B$.
Therefore $x \in A$.
Therefore $x \in (A \cup B) \setminus B \rightarrow x \in A$.
Therefore $((A \cup B) \setminus B) \subseteq A$.

Combination[140] of forward reasoning with backward reasoning. This is common in practice and usually easy to unscramble.

140

Let A and B be arbitrary sets.	$(\forall\text{-}I)$
Let x be an element of $(A \cup B) \setminus B$	(temporary assumption)
So $(x \in A \vee x \in B) \wedge \neg x \in B$	(equivalent proposition)
Therefore $x \in A$	(P follows from $(P \vee Q) \wedge \neg Q$)
Therefore $x \in (A \cup B) \setminus B \rightarrow x \in A$	$(\rightarrow\text{-}I)$
Therefore $((A \cup B) \setminus B) \subseteq A$	(def of $\subseteq$)

Concerning forward and backwards reasoning, one may look at it as follows: we first construct the derivation step at the root of the proof tree ($\forall$-I), and then we jump to a leaf (by making the temporary assumption) and work downwards from there.

7.4 Extending Set Comprehensions

Recall set comprehensions $\{x|P(x)\}$.

7.4 Extending Set Comprehensions

Recall set comprehensions $\{x|P(x)\}$.
Now what do you think this is?

$$\{f(x)|P(x)\}$$

7.4 Extending Set Comprehensions

Recall set comprehensions $\{x|P(x)\}$.
Now what do you think this is?

$$\{f(x)|P(x)\} \equiv \{y|\exists x.\, P(x) \wedge y = f(x)\}$$

7.4 Extending Set Comprehensions

Recall set comprehensions $\{x|P(x)\}$.
Now what do you think this is?

$$\{f(x)|P(x)\} \equiv \{y|\exists x.\, P(x) \wedge y = f(x)\}$$

Example: $t \in \{x^2|x > 5\}$ equivalent to

7.4 Extending Set Comprehensions

Recall set comprehensions $\{x|P(x)\}$.
Now what do you think this is?

$$\{f(x)|P(x)\} \equiv \{y|\exists x.\, P(x) \wedge y = f(x)\}$$

Example: $t \in \{x^2|x > 5\}$ equivalent to $\exists x.\, x > 5 \wedge t = x^2$.
True for $t \in \{36, 49, \ldots\}$

Indexing

Sometimes, it is natural to denote a function f applied to an argument x as "f indexed by x", so f_x, rather than $f(x)$.

Indexing

Sometimes, it is natural to denote a function f applied to an argument x as "f indexed by x", so f_x, rather than $f(x)$.

Example: let $S =$ set of students and let m_s stand for "the mother of s", for s a student. Call S an index set.

$$
\begin{aligned}
x \in \{m_s \mid s \in S\} \ &\leftrightarrow\ x \in \{y \mid \exists s.\, s \in S \wedge y = m_s\} \\
&\leftrightarrow\ \exists s.\, s \in S \wedge x = m_s \\
&\leftrightarrow\ \exists s \in S.\, x = m_s
\end{aligned}
$$

Uses extended comprehensions, indexing syntax, and sorted quantification.

Logical Forms of the New Notation

Question: what is the logical form of $\{x_i | i \in I\} \subseteq A$?

$$\{x_i | i \in I\} \subseteq A \equiv \forall x.\, x \in \{x_i | i \in I\} \to x \in A$$

follows from the definition of $\subseteq$.

We want to show

$$\forall x.\, x \in \{x_i | i \in I\} \to x \in A \equiv \forall x.\, (\exists i \in I.\, x = x_i) \to x \in A$$

$$
\begin{aligned}
x \in \{x_i | i \in I\} &\equiv & \text{(def. of notation)} \\
x \in \{y | \exists i.\, i \in I \land y = x_i\} &\equiv & \textit{compr-I} \\
\exists i.\, i \in I \land x = x_i &\equiv & \text{(Sorted quantification)} \\
\exists i \in I.\, x = x_i & &
\end{aligned}
$$

[143]It may be helpful to pronounce both forms out loud in natural language to get an intuitive feeling that they are equivalent.

[144]Want to prove

$$(\forall x.(\exists i \in I.\, x = x_i) \to x \in A) \leftrightarrow (\forall i \in I.\, x_i \in A)$$

Logical Forms of the New Notation

Question: what is the logical form of $\{x_i | i \in I\} \subseteq A$?

$$\forall x.\, x \in \{x_i | i \in I\} \rightarrow x \in A^{141}, \quad \text{i.e.,}$$
$$\forall x.\, (\exists i \in I.\, x = x_i) \rightarrow x \in A^{142}.$$

141

$$\{x_i | i \in I\} \subseteq A \equiv \forall x.\, x \in \{x_i | i \in I\} \rightarrow x \in A$$

follows from the definition of $\subseteq$.

142

We want to show

$$\forall x.\, x \in \{x_i | i \in I\} \rightarrow x \in A \equiv \forall x.\, (\exists i \in I.\, x = x_i) \rightarrow x \in A$$

$$
\begin{array}{lll}
x \in \{x_i | i \in I\} & \equiv & \text{(def. of notation)} \\
x \in \{y | \exists i.\, i \in I \wedge y = x_i\} & \equiv & \textit{compr-I} \\
\exists i.\, i \in I \wedge x = x_i & \equiv & \text{(Sorted quantification)} \\
\exists i \in I.\, x = x_i
\end{array}
$$

[143]It may be helpful to pronounce both forms out loud in natural language to get an intuitive feeling that they are equivalent.

[144]Want to prove

$$(\forall x.(\exists i \in I.\, x = x_i) \rightarrow x \in A) \leftrightarrow (\forall i \in I.\, x_i \in A)$$

Logical Forms of the New Notation

Question: what is the logical form of $\{x_i | i \in I\} \subseteq A$?

$$\forall x.\, x \in \{x_i | i \in I\} \rightarrow x \in A^{141}, \quad \text{i.e.,}$$
$$\forall x.\, (\exists i \in I.\, x = x_i) \rightarrow x \in A^{142}.$$

Intuition[143] suggests that $\forall i \in I.\, x_i \in A$ is also correct, i.e.,

$$(\forall x.(\exists i \in I.\, x = x_i) \rightarrow x \in A) \leftrightarrow (\forall i \in I.\, x_i \in A).$$

Proving this would be another exercise[144] on using extended comprehensions, indexing syntax, and sorted quantification.

---

[141]

$$\{x_i | i \in I\} \subseteq A \equiv \forall x.\, x \in \{x_i | i \in I\} \rightarrow x \in A$$

follows from the definition of $\subseteq$.

[142]

We want to show

$$\forall x.\, x \in \{x_i | i \in I\} \rightarrow x \in A \equiv \forall x.\, (\exists i \in I.\, x = x_i) \rightarrow x \in A$$

$$
\begin{array}{lcl}
x \in \{x_i | i \in I\} & \equiv & \text{(def. of notation)} \\
x \in \{y | \exists i.\, i \in I \wedge y = x_i\} & \equiv & \textit{compr-I} \\
\exists i.\, i \in I \wedge x = x_i & \equiv & \text{(Sorted quantification)} \\
\exists i \in I.\, x = x_i & &
\end{array}
$$

[143]It may be helpful to pronounce both forms out loud in natural language to get an intuitive feeling that they are equivalent.

[144]Want to prove

$$(\forall x.(\exists i \in I.\, x = x_i) \rightarrow x \in A) \leftrightarrow (\forall i \in I.\, x_i \in A)$$

Powersets

$$\wp(A) = \{x | x \subseteq A\}.$$

What is the logical form of:

1. $x \in \wp(A)$?

- "$\rightarrow$"

  Let $i \in I$ be arbitrary. Now from assumption (for the instance x_i) we have $(\exists j \in I. \, x_i = x_j) \rightarrow x_i \in A$. But premise is true for $i = j$, so $x_i \in A$.

Powersets

$$\wp(A) = \{x \mid x \subseteq A\}.$$

What is the logical form of:

1. $x \in \wp(A)$?

   $x \subseteq A$, i.e., $\forall y. (y \in x \rightarrow y \in A)$

2. $\wp(A) \subseteq \wp(B)$?

- "$\rightarrow$"

  Let $i \in I$ be arbitrary. Now from assumption (for the instance x_i) we have $(\exists j \in I. x_i = x_j) \rightarrow x_i \in A$. But premise is true for $i = j$, so $x_i \in A$.

- "$\leftarrow$"

  Let x be arbitrary and assume $\exists i \in I. x = x_i$. So for some $i \in I$, we have $x = x_i$. Now $\forall i \in I. x_i \in A$. Hence $x \in A$.

"$\rightarrow$" in more detail: Want to prove

$$(\forall x. (\exists i \in I. x = x_i) \rightarrow x \in A) \leftrightarrow (\forall i \in I. x_i \in A)$$

Powersets

$$\wp(A) = \{x | x \subseteq A\}.$$

What is the logical form of:

1. $x \in \wp(A)$?

   $x \subseteq A$, i.e., $\forall y. (y \in x \rightarrow y \in A)$

2. $\wp(A) \subseteq \wp(B)$?

   $\forall x. x \in \wp(A) \rightarrow x \in \wp(B)$, i.e.,

- "$\rightarrow$"

  Let $i \in I$ be arbitrary. Now from assumption (for the instance x_i) we have $(\exists j \in I. x_i = x_j) \rightarrow x_i \in A$. But premise is true for $i = j$, so $x_i \in A$.

- "$\leftarrow$"

  Let x be arbitrary and assume $\exists i \in I. x = x_i$. So for some $i \in I$, we have $x = x_i$. Now $\forall i \in I. x_i \in A$. Hence $x \in A$.

"$\rightarrow$" in more detail: Want to prove

$$(\forall x.(\exists i \in I. x = x_i) \rightarrow x \in A) \leftrightarrow (\forall i \in I. x_i \in A)$$

We show $\forall i \in I. x_i \in A$ assuming $\forall x.(\exists i \in I. x = x_i) \rightarrow x \in A$.

So we show that for arbitrary $i \in I$, assuming $\forall x.(\exists i \in I. x = x_i) \rightarrow x \in A$, we have $x_i \in A$. So let $i \in I$ be arbitrary.

Powersets

$$\wp(A) = \{x | x \subseteq A\}.$$

What is the logical form of:

1. $x \in \wp(A)$?

   $x \subseteq A$, i.e., $\forall y. (y \in x \rightarrow y \in A)$

2. $\wp(A) \subseteq \wp(B)$?

   $\forall x. x \in \wp(A) \rightarrow x \in \wp(B)$, i.e.,
   $\forall x. x \subseteq A \rightarrow x \subseteq B$, i.e.,

---

- "$\rightarrow$"

  Let $i \in I$ be arbitrary. Now from assumption (for the instance x_i) we have $(\exists j \in I. x_i = x_j) \rightarrow x_i \in A$. But premise is true for $i = j$, so $x_i \in A$.

- "$\leftarrow$"

  Let x be arbitrary and assume $\exists i \in I. x = x_i$. So for some $i \in I$, we have $x = x_i$. Now $\forall i \in I. x_i \in A$. Hence $x \in A$.

"$\rightarrow$" in more detail: Want to prove

$$(\forall x.(\exists i \in I. x = x_i) \rightarrow x \in A) \leftrightarrow (\forall i \in I. x_i \in A)$$

We show $\forall i \in I. x_i \in A$ assuming $\forall x.(\exists i \in I. x = x_i) \rightarrow x \in A$.

So we show that for arbitrary $i \in I$, assuming $\forall x.(\exists i \in I. x = x_i) \rightarrow x \in A$, we have $x_i \in A$. So let $i \in I$ be arbitrary.

Powersets

$$\wp(A) = \{x | x \subseteq A\}.$$

What is the logical form of:

1. $x \in \wp(A)$?

   $x \subseteq A$, i.e., $\forall y. (y \in x \rightarrow y \in A)$

2. $\wp(A) \subseteq \wp(B)$?

   $\forall x. x \in \wp(A) \rightarrow x \in \wp(B)$, i.e.,
   $\forall x. x \subseteq A \rightarrow x \subseteq B$, i.e.,
   $\forall x. (\forall y. y \in x \rightarrow y \in A) \rightarrow (\forall y. y \in x \rightarrow y \in B)$

Exercise: prove that the last answer is equivalent to $A \subseteq B$, i.e., $\forall x. x \in A \rightarrow x \in B$.

- "$\rightarrow$"

  Let $i \in I$ be arbitrary. Now from assumption (for the instance x_i) we have $(\exists j \in I. x_i = x_j) \rightarrow x_i \in A$. But premise is true for $i = j$, so $x_i \in A$.

- "$\leftarrow$"

  Let x be arbitrary and assume $\exists i \in I. x = x_i$. So for some $i \in I$, we have $x = x_i$. Now $\forall i \in I. x_i \in A$. Hence $x \in A$.

"$\rightarrow$" in more detail: Want to prove

$$(\forall x. (\exists i \in I. x = x_i) \rightarrow x \in A) \leftrightarrow (\forall i \in I. x_i \in A)$$

We show $\forall i \in I. x_i \in A$ assuming $\forall x. (\exists i \in I. x = x_i) \rightarrow x \in A$.

So we show that for arbitrary $i \in I$, assuming $\forall x. (\exists i \in I. x = x_i) \rightarrow x \in A$, we have $x_i \in A$. So let $i \in I$ be arbitrary.

7.5 Outlook

Sets can have other sets as elements.

Since we have $\forall x.(\exists i \in I. x = x_i) \to x \in A$, by rule $\forall$-E we can specialize to $(\exists j \in I. x_i = x_j) \to x_i \in A$. But premise $(\exists j \in I. x_i = x_j)$ is true for $i = j$, and so $x_i \in A$, which is what was to be proven.

This proof could be made more formal by drawing a proof tree or using Isabelle.

"$\leftarrow$" in more Detail: Want to prove

$$(\forall x.(\exists i \in I. x = x_i) \to x \in A) \leftrightarrow (\forall i \in I. x_i \in A)$$

7.5 Outlook

Sets can have other sets as elements.

Since we have $\forall x.(\exists i \in I.\, x = x_i) \to x \in A$, by rule $\forall\text{-}E$ we can specialize to $(\exists j \in I.\, x_i = x_j) \to x_i \in A$. But premise $(\exists j \in I.\, x_i = x_j)$ is true for $i = j$, and so $x_i \in A$, which is what was to be proven.

This proof could be made more formal by drawing a proof tree or using Isabelle.

"$\leftarrow$" in more Detail: Want to prove

$$(\forall x.(\exists i \in I.\, x = x_i) \to x \in A) \leftrightarrow (\forall i \in I.\, x_i \in A)$$

We show $\forall x.(\exists i \in I.\, x = x_i) \to x \in A$, assuming $\forall i \in I.\, x_i \in A$.

So we show that for arbitrary x, assuming $\forall i \in I.\, x_i \in A$, we have $(\exists i \in I.\, x = x_i) \to x \in A$. So let x be arbitrary.

To show $(\exists i \in I.\, x = x_i) \to x \in A$, assume $\exists i \in I.\, x = x_i$. So for some $i \in I$, we have $x = x_i$. Now by our earlier assumption $\forall i \in I.\, x_i \in A$, and so it follows that $x \in A$. thus we have shown $x \in A$ under the assumption $(\exists i \in I.\, x = x_i)$, thus we have shown $(\exists i \in I.\, x = x_i) \to x \in A$,

Implicitly assume that universe of discourse is collection[145] of all sets.

which is what was to be proven.

This proof could be made more formal by drawing a proof tree or using Isabelle.

[145] We speak of collection of all sets rather than set of all sets in order to pretend that we are being careful since we are not sure if there is such a thing as a set of all sets. Therefore we use the "neutral" word collection whose meaning is obvious. . .

Implicitly assume that universe of discourse is collection[145] of all sets.

which is what was to be proven.

This proof could be made more formal by drawing a proof tree or using Isabelle.

[145]We speak of collection of all sets rather than set of all sets in order to pretend that we are being careful since we are not sure if there is such a thing as a set of all sets. Therefore we use the "neutral" word collection whose meaning is obvious. . .

Is it?

Implicitly assume that universe of discourse is collection[145] of all sets.

which is what was to be proven.

This proof could be made more formal by drawing a proof tree or using Isabelle.

[145]We speak of collection of all sets rather than set of all sets in order to pretend that we are being careful since we are not sure if there is such a thing as a set of all sets. Therefore we use the "neutral" word collection whose meaning is obvious...

Is it?

Recall that we have defined set as collection of objects in the first place. So it is rather futile to suggest now that there should be some difference between collections and sets.

The fact of the matter is: the approach of allowing arbitrary collections of "objects" and regarding such collections as "objects" themselves is naïve. We will see this shortly.

Russell's Paradox

Suppose $U := \{x \mid \top^{146}\}$. Then[147] $U \in U$.

    Quite strange but no contradiction yet.

---

[146]Assume that $\top$ is syntactic sugar for a proposition that is always true, say $\top \equiv \bot \to \bot$. We have not introduced this, but it is convenient.

    So semantically, we have $I_{\mathcal{A}}(\top) = 1$ for all $I_{\mathcal{A}}$.

[147]Recall that a set comprehension has the form $\{x \mid P(x)\}$, where $P(x)$ is a formula usually containing x.

    The set comprehension $U := \{x \mid \top\}$ is strange since $\top$ does not contain x.

    But by the introduction rule for set comprehensions, this means that $x \in U$ for any x. Thus in particular, $U \in U$.

148

    Recall $R := \{A \mid A \notin A\}$ and recall the notion of logical form.

    Let A be arbitrary (for the formal reasoning applied here, arbitrary means: it could be a set, a number, a dog, the pope, anything whatsoever).

    By the rules for set comprehension, we can prove $A \in \{A \mid A \notin A\} \to A \notin A$ and $A \notin A \to A \in \{A \mid A \notin A\}$, and

Russell's Paradox

Suppose $U := \{x \mid \top^{146}\}$. Then[147] $U \in U$.

Quite strange but no contradiction yet.

Now split sets into two categories:

1. unusual sets like U that are elements of themselves, and

2. more typical sets that are not.

---

[146]Assume that $\top$ is syntactic sugar for a proposition that is always true, say $\top \equiv \bot \to \bot$. We have not introduced this, but it is convenient.

So semantically, we have $I_{\mathcal{A}}(\top) = 1$ for all $I_{\mathcal{A}}$.

[147]Recall that a set comprehension has the form $\{x \mid P(x)\}$, where $P(x)$ is a formula usually containing x.

The set comprehension $U := \{x \mid \top\}$ is strange since $\top$ does not contain x.

But by the introduction rule for set comprehensions, this means that $x \in U$ for any x. Thus in particular, $U \in U$.

[148]

Recall $R := \{A \mid A \notin A\}$ and recall the notion of logical form.

Let A be arbitrary (for the formal reasoning applied here, arbitrary means: it could be a set, a number, a dog, the pope, anything whatsoever).

By the rules for set comprehension, we can prove $A \in \{A \mid A \notin A\} \to A \notin A$ and $A \notin A \to A \in \{A \mid A \notin A\}$, and

Russell's Paradox

Suppose $U := \{x \mid \top^{146}\}$. Then[147] $U \in U$.

   Quite strange but no contradiction yet.

   Now split sets into two categories:

1. unusual sets like U that are elements of themselves, and

2. more typical sets that are not.

   Let $R := \{A \mid A \notin A\}$.

   Using logical form[148] we derive: $\forall A.\,(A \in R \leftrightarrow A \notin A)$

---

[146]Assume that $\top$ is syntactic sugar for a proposition that is always true, say $\top \equiv \bot \to \bot$. We have not introduced this, but it is convenient.

   So semantically, we have $I_{\mathcal{A}}(\top) = 1$ for all $I_{\mathcal{A}}$.

[147]Recall that a set comprehension has the form $\{x \mid P(x)\}$, where $P(x)$ is a formula usually containing x.

   The set comprehension $U := \{x \mid \top\}$ is strange since $\top$ does not contain x.

   But by the introduction rule for set comprehensions, this means that $x \in U$ for any x. Thus in particular, $U \in U$.

[148]

   Recall $R := \{A \mid A \notin A\}$ and recall the notion of logical form.

   Let A be arbitrary (for the formal reasoning applied here, arbitrary means: it could be a set, a number, a dog, the pope, anything whatsoever).

   By the rules for set comprehension, we can prove $A \in \{A \mid A \notin A\} \to A \notin A$ and $A \notin A \to A \in \{A \mid A \notin A\}$, and

Russell's Paradox

Suppose $U := \{x \mid \top^{146}\}$. Then[147] $U \in U$.

Quite strange but no contradiction yet.
Now split sets into two categories:

1. unusual sets like U that are elements of themselves, and

2. more typical sets that are not.

Let $R := \{A \mid A \notin A\}$.
Using logical form[148] we derive: $\forall A.\, (A \in R \leftrightarrow A \notin A)$
Substituting R for A ($\forall$-E) yields $R \in R \leftrightarrow R \notin R$, which is a logical contradiction. Or even more succinctly...

---

[146]Assume that $\top$ is syntactic sugar for a proposition that is always true, say $\top \equiv \bot \to \bot$. We have not introduced this, but it is convenient.

So semantically, we have $I_{\mathcal{A}}(\top) = 1$ for all $I_{\mathcal{A}}$.

[147]Recall that a set comprehension has the form $\{x \mid P(x)\}$, where $P(x)$ is a formula usually containing x.

The set comprehension $U := \{x \mid \top\}$ is strange since $\top$ does not contain x.

But by the introduction rule for set comprehensions, this means that $x \in U$ for any x. Thus in particular, $U \in U$.

[148]

Recall $R := \{A \mid A \notin A\}$ and recall the notion of logical form.

Let A be arbitrary (for the formal reasoning applied here, arbitrary means: it could be a set, a number, a dog, the pope, anything whatsoever).

By the rules for set comprehension, we can prove $A \in \{A \mid A \notin A\} \to A \notin A$ and $A \notin A \to A \in \{A \mid A \notin A\}$, and

Russell's Paradox (Cont.)

Recall $R := \{A | A \notin A\}$.

Assume $R \in R$. By the definition of R, this means $R \in \{A | A \notin A\}$. Using *compr-E*, this implies $R \notin R$.

Now assume $R \notin R$. Using *compr-I*, this implies $R \in \{A | A \notin A\}$. By the definition of R, this means $R \in R$.

What does this tell us about sets?[149]

---

so by definition of $\leftrightarrow$, we have $A \in R \leftrightarrow A \notin A$, and since A was arbitrary, by $\forall$-*I*, we have $\forall A.\,(A \in R \leftrightarrow A \notin A)$.

[149]It tells us that there can be no such thing as the set of all sets.

The fundamental flaw of naïve set theory is in saying that a set is a collection of "objects" without worrying what an object is. If we make no restriction as to what an object is, then a set is obviously also an object. But then we effectively base the definition of the new concept set on the existence of sets, so the definition is circular.

The intuition for the solution to this dilemma is not difficult: A set is a collection of objects of which we are already sure that they exist. In particular, since we are only just about to define sets, these objects may not themselves be sets.

Once we have such sets, we can introduce "sets of second order", that is, sets that contain sets of the first kind. This process can be continued ad infinitum.

The formal details will come later.

Where Do We Go from here?

- The λ-calculus as basis for a metalanguage to avoid notational confusion

---

[150]Higher-order logic is a solution to the dilemma posed by Russell's paradox.

It is a surprisingly simple formalism which can be extended conservatively: this means that it can be ensured that the extensions cannot compromise the truth or falsity of statements that were already expressible before the extension.

Where Do We Go from here?

- The λ-calculus as basis for a metalanguage to avoid notational confusion

- Resolution and other deduction techniques: understanding Isabelle better and achieving a higher level of automation

---

[150]Higher-order logic is a solution to the dilemma posed by Russell's paradox.

It is a surprisingly simple formalism which can be extended conservatively: this means that it can be ensured that the extensions cannot compromise the truth or falsity of statements that were already expressible before the extension.

Where Do We Go from here?

- The λ-calculus as basis for a metalanguage to avoid notational confusion

- Resolution and other deduction techniques: understanding Isabelle better and achieving a higher level of automation

- Higher-order logic: a formalism for (among other things) non-naïve set theory[150]

---

[150]Higher-order logic is a solution to the dilemma posed by Russell's paradox.

It is a surprisingly simple formalism which can be extended conservatively: this means that it can be ensured that the extensions cannot compromise the truth or falsity of statements that were already expressible before the extension.

8 The λ-Calculus

The λ-Calculus: Motivation

A way of writing functions. E.g., $\lambda x.\, x + 5$ is the function taking any number n to $n + 5$. Theory underlying functional programming.

Turing-complete model of computation.

One of the most important formalisms of (theoretical) computer science!

The λ-Calculus: Motivation

A way of writing functions. E.g., $\lambda x.\, x + 5$ is the function taking any number n to $n + 5$. Theory underlying functional programming.

Turing-complete model of computation.

One of the most important formalisms of (theoretical) computer science!

Why is it interesting for us? The λ-calculus is used for representing object logics in Isabelle. It is the core of Isabelle's metalogic!

Further reading: [Tho91, chapter 2], [HS90, chapter 1].

Outline of this Lecture

- The untyped λ-calculus

- The simply typed λ-calculus ($\lambda^{\rightarrow}$)

- An extension of the typed λ-calculus

- Higher-order unification

8.1 Untyped λ-Calculus

From functional programming , you may be familiar with function definitions such as

$$f \ x = x + 5$$

The λ-calculus is a formalism for writing nameless functions.
The function $\lambda x.\, x + 5$ corresponds to f.

The λ-calculus is a formalism for writing nameless functions. The function $\lambda x.\, x + 5$ corresponds to f.

The application to say, 3, is written $(\lambda x.\, x+5)(3)$. Its result is computed by substituting 3 for x, yielding $3 + 5$, which in usual arithmetic evaluates to 8[151].

---

[151]As you might guess, the formalism of the λ-calculus is not directly related to usual arithmetic and so it is not built into this formalism that $3 + 5$ should evaluate to 8. However, it may be a reasonable choice, depending on the context, to extend the λ-calculus in this way, but this is not our concern at the moment.

Syntax

$(x \in \mathit{Var}, c \in \mathit{Const}^{152})$

$$e ::= x \mid c \mid (ee) \mid (\lambda x.\, e)^{153}$$

---

[152]Similarly as for first-order logic, a language of the untyped λ-calculus is characterized by giving a set of variables and a set of constants.

One can think of Const as a signature.

Note that Const could be empty.

Note also that the word constant has a different meaning in the λ-calculus from that of first-order logic. In both formalisms, constants are just symbols.

In first-order logic, a constant is a special case of a function symbol, namely a function symbol of arity 0.

In the λ-calculus, one does not speak of function symbols. In the untyped λ-calculus, any λ-term (including a constant) can be applied to another term, and so any λ-term can be called a "unary function". A constant being applied to a term is something which would contradict the intuition about constants in first-order logic. So for the λ-calculus, think of constant as opposed to a variable, an application, or an abstraction.

[153]A λ-term can either be

The objects generated by this grammar[154] are called λ-terms or simply terms.

- a variable (case x), or

- a constant (case c), or

- an application of a λ-term to another λ-term (case (ee)), or

- an abstraction over a variable x (case $(\lambda x.\, e)$).

[154]A notation like

$$
\begin{aligned}
e &::= x \mid c \mid (ee) \mid (\lambda x.\, e) \\
\tau &::= T \mid \tau \to \tau \\
e &::= x \mid c \mid (ee) \mid (\lambda x^\tau.\, e) \\
P &::= x \mid \neg P \mid P \wedge P \mid P \to P \ldots
\end{aligned}
$$

for specifying syntax is called Backus-Naur form (BNF) for expressing grammars. For example, the first BNF-clause reads: a λ-term can be
a variable, or
a constant, or
a λ-term applied to a λ-term, or

Conventions: iterated λ & left-associated application[155]

$$(\lambda x.\,(\lambda y.\,(\lambda z.\,((xz)(yz))))) \;\equiv\; (\lambda xyz.\,((xz)(yz)))$$
$$\equiv\; \lambda xyz.\,xz(yz)$$

Is $\lambda x.\,x + 5$ a λ-term?[156]

---

a λ-abstraction, which is a λ-term of the form $\lambda x.\,e$, where e is a λ-term.

The BNF is a very common formalism for specifying syntax, e.g., of programming languages. See here or here.

[155]We write $\lambda x_1 x_2 \ldots x_n.e$ instead of $\lambda x_1.(\lambda x_2.(\ldots e)\ldots)$.

$e_1\ e_2 \ldots e_n$ is equivalent to $(\ldots (e_1\ e_2)\ldots e_n)\ldots$, not $(e_1(e_2 \ldots e_n)\ldots)$. Note that this is in contrast to the associativity of logical operators. There are some good reasons for these conventions.

[156]Strictly speaking, $\lambda x.\,x + 5$ does not adhere to the definition of syntax of λ-terms, at least if we parse it in the usual way: $+$ is an infix constant applied to arguments x and 5.

If we parse $x+5$ as $((x+)5)$, i.e., x applied to (the constant) $+$, and the resulting term applied to (the constant) 5, then $\lambda x.\,x + 5$ would indeed adhere to the definition of syntax of λ-terms, but of course, this is pathological and not intended here.

It is convenient to allow for extensions of the syntax of λ-

Substitution

- Reduction[157] based on substitutions
$$(\lambda x.\, g\, x\, 3)(5) = (g\, x\, 3)[x \leftarrow 5]^{158} = g\, 5\, 3$$

- Must respect free and bound variables,
$$(\lambda x.\, x(\lambda x.\, xy))(e) = ((x(\lambda x.\, xy))[x \leftarrow e] = e(\lambda x.\, xy)$$

- Same problems as with quantifiers
$$\frac{\forall x.\, (P(x) \wedge \exists x.\, Q(x,y))}{P(e) \wedge \exists x.\, Q(x,y)}\ \forall\text{-}E \qquad \frac{\forall x.\, (P(x) \wedge \exists y.\, Q(x,y))}{P(y) \wedge \exists z.\, Q(y,z)}\ \forall\text{-}E$$

---

terms, allowing for:

- application to several arguments rather than just one;

- infix notation.

Such an extension is inessential for the expressive power of the λ-calculus. Instead of having a binary infix constant $+$ and writing $\lambda x.\, x + 5$, we could have a constant $plus$ according to the original syntax and write $\lambda x.\, ((plus\, x)\, 5)$ (i.e., write $+$ in a Curryed way).

[157]Reduction is the notion of "computing", or "evaluation", in the λ-calculus.

[158]Here we use the notation $e[x \leftarrow t]$ for the term obtained from e by replacing x with t. There is also the notation $e[t/x]$, and confusingly, also $e[x/t]$. We will attempt to be consistent within this course, but be aware that you may find such different notations in the literature.

Bound, Free, Binding Occurrences

Recall the notions of bound, free, and binding occurrences of variables in a term. Same thing here:

λ-calculus	FOL
$FV(x) :=$	

Bound, Free, Binding Occurrences

Recall the notions of bound, free, and binding occurrences of variables in a term. Same thing here:

λ-calculus	FOL
$FV(x) := \{x\}$	$= FV(x)$
$FV(c) :=$	

Bound, Free, Binding Occurrences

Recall the notions of bound, free, and binding occurrences of variables in a term. Same thing here:

λ-calculus	FOL
$FV(x) := \{x\}$	$= FV(x)$
$FV(c) := \emptyset$	$= FV(c)$
$FV(MN) :=$	

Bound, Free, Binding Occurrences

Recall the notions of bound, free, and binding occurrences of variables in a term. Same thing here:

λ-calculus	FOL
$FV(x) := \{x\}$	$= FV(x)$
$FV(c) := \emptyset$	$= FV(c)$
$FV(MN) := FV(M) \cup FV(N)$	$= FV(M \wedge N)$
$FV(\lambda x.\, M) :=$	

Bound, Free, Binding Occurrences

Recall the notions of bound, free, and binding occurrences of variables in a term. Same thing here:

λ-calculus	FOL
$FV(x) := \{x\}$	$= FV(x)$
$FV(c) := \emptyset$	$= FV(c)$
$FV(MN) := FV(M) \cup FV(N)$	$= FV(M \wedge N)$
$FV(\lambda x.\, M) := FV(M) \setminus \{x\}$	$= FV(\forall x.\, M)$

Example: $FV(xy(\lambda yz.\, xyz)) = \{x, y\}$

Definition of Substitution

$M[x \leftarrow N]$ means substitute N for x in M

1. $x[x \leftarrow N] =$

2. $a[x \leftarrow N] =$

3. $(PQ)[x \leftarrow N] =$

4. $(\lambda x.\, P)[x \leftarrow N] =$

5. $(\lambda y.\, P)[x \leftarrow N] =$

6. $(\lambda y.\, P)[x \leftarrow N] =$

Definition of Substitution

$M[x \leftarrow N]$ means substitute N for x in M

1. $x[x \leftarrow N] = N$

2. $a[x \leftarrow N] = a$ if a is a constant or variable other than x

3. $(PQ)[x \leftarrow N] = (P[x \leftarrow N]Q[x \leftarrow N])$

4. $(\lambda x.\, P)[x \leftarrow N] = \lambda x.\, P$

5. $(\lambda y.\, P)[x \leftarrow N] = \lambda y.\, P[x \leftarrow N]$ if $y \neq x$ and $y \notin FV(N)$

6. $(\lambda y.\, P)[x \leftarrow N] = \lambda z.\, P[y \leftarrow z][x \leftarrow N]$ if $y \neq x$ and $y \in FV(N)$ where z is a variable such that $z \notin FV(NP)$

Definition of Substitution

$M[x \leftarrow N]$ means substitute N for x in M

1. $x[x \leftarrow N] = N$

2. $a[x \leftarrow N] = a$ if a is a constant or variable other than x

3. $(PQ)[x \leftarrow N] = (P[x \leftarrow N]Q[x \leftarrow N])$

4. $(\lambda x.\, P)[x \leftarrow N] = \lambda x.\, P$

5. $(\lambda y.\, P)[x \leftarrow N] = \lambda y.\, P[x \leftarrow N]$ if $y \neq x$ and $y \notin FV(N)$

6. $(\lambda y.\, P)[x \leftarrow N] = \lambda z.\, P[y \leftarrow z][x \leftarrow N]$ if $y \neq x$ and $y \in FV(N)$ where z is a variable such that $z \notin FV(NP)$

Cases similar to those for quantifiers: λ binding is 'generic'[159].

---

[159]Recall the definition of substitution for first-order logic.

We observe that binding and substitution are some very general concepts. So far, we have seen four binding operators: $\exists$, $\forall$ and λ, and set comprehensions. The λ operator is the most generic of those operators, in that it does not have a fixed meaning hard-wired into it in the way that the quantifiers do. In fact, it is possible to have it as the only operator on the level of the metalogic. We will see this later.

Substitution: Example

$$(x(\lambda x.\, xy))[x \leftarrow \lambda z.\, z]$$

---

[160]If it wasn't for clause 6, i.e., if we applied clause 5 ignoring the requirement on freeness, then $(\lambda x.\, xy)[y \leftarrow x]$ would be $\lambda x.\, xx$.

Substitution: Example

$$(x(\lambda x.\, xy))[x \leftarrow \lambda z.\, z] \overset{3}{=} x[x \leftarrow \lambda z.\, z](\lambda x.\, xy)[x \leftarrow \lambda z.\, z]$$
$$\overset{1,4}{=} (\lambda z.\, z)\lambda x.\, xy$$

---

[160]If it wasn't for clause 6, i.e., if we applied clause 5 ignoring the requirement on freeness, then $(\lambda x.\, xy)[y \leftarrow x]$ would be $\lambda x.\, xx$.

Substitution: Example

$$(x(\lambda x.\, xy))[x \leftarrow \lambda z.\, z] \;\;\stackrel{3}{=}\;\; x[x \leftarrow \lambda z.\, z](\lambda x.\, xy)[x \leftarrow \lambda z.\, z]$$

$$\stackrel{1,4}{=}\;\; (\lambda z.\, z)\lambda x.\, xy$$

$$(\lambda x.\, xy)[y \leftarrow x]$$

---

[160]If it wasn't for clause 6, i.e., if we applied clause 5 ignoring the requirement on freeness, then $(\lambda x.\, xy)[y \leftarrow x]$ would be $\lambda x.\, xx$.

Substitution: Example

$$(x(\lambda x.\, xy))[x \leftarrow \lambda z.\, z] \stackrel{3}{=} x[x \leftarrow \lambda z.\, z](\lambda x.\, xy)[x \leftarrow \lambda z.\, z]$$
$$\stackrel{1,4}{=} (\lambda z.\, z)\lambda x.\, xy$$

$$(\lambda x.\, xy)[y \leftarrow x] \stackrel{6}{=} \lambda z.\, ((xy)[x \leftarrow z][y \leftarrow x])$$
$$\stackrel{3,1,2}{=} \lambda z.\, (zy[y \leftarrow x])$$
$$\stackrel{3,2,1}{=} \lambda z.\, zx$$

In the last example, clause 6 avoids capture, i.e., $\lambda x.\, xx$[160].

---

[160]If it wasn't for clause 6, i.e., if we applied clause 5 ignoring the requirement on freeness, then $(\lambda x.\, xy)[y \leftarrow x]$ would be $\lambda x.\, xx$.

Reduction: Intuition

Reduction is the notion of "computing", or "evaluation", in the λ-calculus.

$f\ x = x + 5 \;\rightsquigarrow\; f = \lambda x.\, x + 5$

$f\ 3 = 3 + 5 \;\rightsquigarrow\; (\lambda x.\, x + 5)(3) \rightarrow_\beta (x + 5)[x \leftarrow 3] = 3 + 5$

β-reduction replaces a parameter by an argument[161].

This should propagate into contexts[162], e.g.

$$\lambda x.(\underline{(\lambda x.\, x + 5)(3)}) \rightarrow_\beta \lambda x.(3 + 5).$$

---

[161]In the λ-term $(\lambda x.M)N$, we say that N is an argument (and the function $\lambda x.M$ is applied to this argument), and every occurrence of x in M is a parameter (we say this because x is bound by the λ).

This terminology may be familiar to you if you have experience in functional programming, but actually, it is also used in the context of function and procedure declarations in imperative programming.

[162]In

$$\lambda x.(\underline{(\lambda x.\, x + 5)(3)}),$$

the underlined part is a subterm occurring in a context. β-reduction should be applicable to this subterm.

Reduction: Definition

- β-reduction: $(\lambda x.M)N \rightarrow_\beta M[x \leftarrow N]^{163}$

- Rules for contraction (of redices[164]) in contexts:
$$\frac{M \rightarrow_\beta M'}{NM \rightarrow_\beta NM'} \qquad \frac{M \rightarrow_\beta M'}{MN \rightarrow_\beta M'N} \qquad \frac{M \rightarrow_\beta M'}{\lambda z.M \rightarrow_\beta \lambda z.M'}$$

- Reduction is reflexive-transitive closure
$$\frac{M \rightarrow_\beta N}{M \rightarrow_\beta^* N} \qquad \frac{}{M \rightarrow_\beta^* M} \qquad \frac{M \rightarrow_\beta^* N \quad N \rightarrow_\beta^* P}{M \rightarrow_\beta^* P}$$

- A term without redices is in β-normal form.

---

[163] As you see, β-reduction is defined using rules (two of them being axioms, the rest proper rules) in the same way that we have defined proof systems for logic before. Note that we wrote the first axiom defining β-reduction without a horizontal bar.

[164] In a λ-term, a subterm of the form $(\lambda x.\, M)N$ is called a redex (plural redices). It is a subterm to which β-reduction can be applied.

Reduction: Examples

$$(\lambda x.\, \lambda y.\, g\, x\, y) a\, b \rightarrow_\beta$$

---

[165]You may be familiar with functions taking several arguments, or equivalently, a tuple of arguments, rather than just one argument.

In the λ-calculus, but also in functional programming, it is common not to have tuples and instead use a technique called Currying (Schönfinkeln in German). So instead of writing $g(a, b)$, we write $g\, a\, b$, which is read as follows: g is a function which takes an argument a and returns a function which then takes an argument b.

Recall that application associates to the left, so $g\, a\, b$ is read $(g\, a)\, b$.

Currying will become even clearer once we introduce the typed λ-calculus.
[166]We say that a β-reduction sequence diverges if it is infinite.

Note that for $(\lambda xy.\, y)((\lambda x.\, xx)(\lambda x.\, xx))$, there is a finite β-reduction sequence

$$(\lambda xy.\, y)((\lambda x.\, xx)(\lambda x.\, xx)) \rightarrow_\beta \lambda y.\, y$$

Reduction: Examples

$$(\lambda x.\,\lambda y.\,g\,x\,y)a\,b \to_\beta \underline{(\lambda y.\,(g\,a\,y))}b \to_\beta$$

---

[165]You may be familiar with functions taking several arguments, or equivalently, a tuple of arguments, rather than just one argument.

In the λ-calculus, but also in functional programming, it is common not to have tuples and instead use a technique called Currying (Schönfinkeln in German). So instead of writing $g(a,b)$, we write $g\,a\,b$, which is read as follows: g is a function which takes an argument a and returns a function which then takes an argument b.

Recall that application associates to the left, so $g\,a\,b$ is read $(g\,a)\,b$.

Currying will become even clearer once we introduce the typed λ-calculus.

[166]We say that a β-reduction sequence diverges if it is infinite.

Note that for $(\lambda xy.\,y)((\lambda x.\,xx)(\lambda x.\,xx))$, there is a finite β-reduction sequence

$$(\lambda xy.\,y)((\lambda x.\,xx)(\lambda x.\,xx)) \to_\beta \lambda y.\,y$$

Reduction: Examples

$$\frac{(\lambda x.\,\lambda y.\,g\,x\,y)a\,b \rightarrow_\beta \underline{(\lambda y.\,(g\,a\,y))b} \rightarrow_\beta g\,a\,b}{\text{So } (\lambda x.\,\lambda y.\,g\,x\,y)a\,b \rightarrow_\beta^* g\,a\,b}$$

Shows Currying[165]

---

[165]You may be familiar with functions taking several arguments, or equivalently, a tuple of arguments, rather than just one argument.

In the λ-calculus, but also in functional programming, it is common not to have tuples and instead use a technique called Currying (Schönfinkeln in German). So instead of writing $g(a, b)$, we write $g\,a\,b$, which is read as follows: g is a function which takes an argument a and returns a function which then takes an argument b.

Recall that application associates to the left, so $g\,a\,b$ is read $(g\,a)\,b$.

Currying will become even clearer once we introduce the typed λ-calculus.

[166]We say that a β-reduction sequence diverges if it is infinite.

Note that for $(\lambda xy.\,y)((\lambda x.\,xx)(\lambda x.\,xx))$, there is a finite β-reduction sequence

$$(\lambda xy.\,y)((\lambda x.\,xx)(\lambda x.\,xx)) \rightarrow_\beta \lambda y.\,y$$

Reduction: Examples

$$(\lambda x.\, \lambda y.\, g\, x\, y)a\, b \rightarrow_\beta (\lambda y.\, (g\, a\, y))b \rightarrow_\beta g\, a\, b$$
$$\text{So } (\lambda x.\, \lambda y.\, g\, x\, y)a\, b \rightarrow_\beta^* g\, a\, b$$

Shows Currying[165]

$$(\lambda x.\, xx)(\lambda x.\, xx) \rightarrow_\beta$$

---

[165]You may be familiar with functions taking several arguments, or equivalently, a tuple of arguments, rather than just one argument.

In the λ-calculus, but also in functional programming, it is common not to have tuples and instead use a technique called Currying (Schönfinkeln in German). So instead of writing $g(a, b)$, we write $g\, a\, b$, which is read as follows: g is a function which takes an argument a and returns a function which then takes an argument b.

Recall that application associates to the left, so $g\, a\, b$ is read $(g\, a)\, b$.

Currying will become even clearer once we introduce the typed λ-calculus.

[166]We say that a β-reduction sequence diverges if it is infinite.

Note that for $(\lambda xy.\, y)((\lambda x.\, xx)(\lambda x.\, xx))$, there is a finite β-reduction sequence

$$(\lambda xy.\, y)((\lambda x.\, xx)(\lambda x.\, xx)) \rightarrow_\beta \lambda y.\, y$$

Reduction: Examples

$$\underline{(\lambda x.\, \lambda y.\, g\, x\, y)a\, b} \rightarrow_\beta \underline{(\lambda y.\, (g\, a\, y))b} \rightarrow_\beta g\, a\, b$$
$$\text{So } (\lambda x.\, \lambda y.\, g\, x\, y)a\, b \rightarrow_\beta^* g\, a\, b$$

Shows Currying[165]

$$\underline{(\lambda x.\, xx)(\lambda x.\, xx)} \rightarrow_\beta \underline{(\lambda x.\, xx)(\lambda x.\, xx)} \rightarrow_\beta \ldots$$

Shows divergence[166]

---

[165]You may be familiar with functions taking several arguments, or equivalently, a tuple of arguments, rather than just one argument.

In the λ-calculus, but also in functional programming, it is common not to have tuples and instead use a technique called Currying (Schönfinkeln in German). So instead of writing $g(a, b)$, we write $g\, a\, b$, which is read as follows: g is a function which takes an argument a and returns a function which then takes an argument b.

Recall that application associates to the left, so $g\, a\, b$ is read $(g\, a)\, b$.

Currying will become even clearer once we introduce the typed λ-calculus.

[166]We say that a β-reduction sequence diverges if it is infinite.

Note that for $(\lambda xy.\, y)((\lambda x.\, xx)(\lambda x.\, xx))$, there is a finite β-reduction sequence

$$(\lambda xy.\, y)((\lambda x.\, xx)(\lambda x.\, xx)) \rightarrow_\beta \lambda y.\, y$$

Reduction: Examples

$$(\lambda x.\, \lambda y.\, g\, x\, y)a\, b \rightarrow_\beta (\lambda y.\, (g\, a\, y))b \rightarrow_\beta g\, a\, b$$
$$\text{So } (\lambda x.\, \lambda y.\, g\, x\, y)a\, b \rightarrow_\beta^* g\, a\, b$$

Shows Currying[165]

$$(\lambda x.\, xx)(\lambda x.\, xx) \rightarrow_\beta (\lambda x.\, xx)(\lambda x.\, xx) \rightarrow_\beta \ldots$$

Shows divergence[166]

$$\text{But } (\lambda xy.\, y)((\lambda x.\, xx)(\lambda x.\, xx)) \rightarrow_\beta \lambda y.\, y$$

---

[165]You may be familiar with functions taking several arguments, or equivalently, a tuple of arguments, rather than just one argument.

In the λ-calculus, but also in functional programming, it is common not to have tuples and instead use a technique called Currying (Schönfinkeln in German). So instead of writing $g(a, b)$, we write $g\, a\, b$, which is read as follows: g is a function which takes an argument a and returns a function which then takes an argument b.

Recall that application associates to the left, so $g\, a\, b$ is read $(g\, a)\, b$.

Currying will become even clearer once we introduce the typed λ-calculus.

[166]We say that a β-reduction sequence diverges if it is infinite.

Note that for $(\lambda xy.\, y)((\lambda x.\, xx)(\lambda x.\, xx))$, there is a finite β-reduction sequence

$$(\lambda xy.\, y)((\lambda x.\, xx)(\lambda x.\, xx)) \rightarrow_\beta \lambda y.\, y$$

Conversion

- β-conversion: "symmetric closure" of β-reduction

$$\frac{M \to_\beta^* N}{M =_\beta N} \qquad \frac{M =_\beta N}{N =_\beta M}$$

---

but there is also a diverging sequence

$$(\lambda xy.\, y)((\lambda x.\, xx)(\lambda x.\, xx)) \to_\beta (\lambda xy.\, y)((\lambda x.\, xx)(\lambda x.\, xx)) \to_\beta \ldots$$

[167]α-conversion is usually applied implicitly, i.e., without making it an explicit step. So for example, one would simply write:

$$\lambda z.\, z =_\beta \lambda x.\, x$$

[168]η-conversion is defined as

$$M =_\eta \lambda x.\, (M x) \quad \text{if } x \notin FV(M)$$

It is needed for reasoning about normal forms.

$$g\,x =_\eta \lambda y.\, g\,x\,y \quad \text{reflects} \quad g\,x\,b =_\beta (\lambda y.\, g\,x\,y)b$$

More specifically: if we did not have the η-conversion rule, then $g\,x$ and $\lambda y.\, g\,x\,y$ would not be "equivalent" up to conversion. But that seems unreasonable, because they behave

Conversion

- β-conversion: "symmetric closure" of β-reduction

$$\frac{M \to_\beta^* N}{M =_\beta N} \qquad \frac{M =_\beta N}{N =_\beta M}$$

- α-conversion: bound variable renaming (usually implicit[167])

$$\lambda x.M =_\alpha \lambda z.M[x \leftarrow z] \quad \textit{where } z \notin FV(M)$$

---

but there is also a diverging sequence

$$(\lambda xy.\, y)((\lambda x.\, xx)(\lambda x.\, xx)) \to_\beta (\lambda xy.\, y)((\lambda x.\, xx)(\lambda x.\, xx)) \to_\beta \ldots$$

[167]α-conversion is usually applied implicitly, i.e., without making it an explicit step. So for example, one would simply write:

$$\lambda z.\, z =_\beta \lambda x.\, x$$

[168]η-conversion is defined as

$$M =_\eta \lambda x.\, (M\,x) \quad \textit{if } x \notin FV(M)$$

It is needed for reasoning about normal forms.

$$g\,x =_\eta \lambda y.\, g\,x\,y \quad \textcolor{red}{\text{reflects}} \quad g\,x\,b =_\beta (\lambda y.\, g\,x\,y)b$$

More specifically: if we did not have the η-conversion rule, then $g\,x$ and $\lambda y.\, g\,x\,y$ would not be "equivalent" up to conversion. But that seems unreasonable, because they behave

Conversion

- β-conversion: "symmetric closure" of β-reduction

$$\frac{M \to_\beta^* N}{M =_\beta N} \qquad \frac{M =_\beta N}{N =_\beta M}$$

- α-conversion: bound variable renaming (usually implicit[167])

$$\lambda x.M =_\alpha \lambda z.M[x \leftarrow z] \quad \text{where } z \notin FV(M)$$

- η-conversion: for normal-form analysis[168]

$$M =_\eta \lambda x.\,(M x) \quad \text{if } x \notin FV(M)$$

---

but there is also a diverging sequence

$$(\lambda xy.\,y)((\lambda x.\,xx)(\lambda x.\,xx)) \to_\beta (\lambda xy.\,y)((\lambda x.\,xx)(\lambda x.\,xx)) \to_\beta \ldots$$

[167]α-conversion is usually applied implicitly, i.e., without making it an explicit step. So for example, one would simply write:

$$\lambda z.\,z =_\beta \lambda x.\,x$$

[168]η-conversion is defined as

$$M =_\eta \lambda x.\,(M x) \quad \text{if } x \notin FV(M)$$

It is needed for reasoning about normal forms.

$$g\,x =_\eta \lambda y.\,g\,x\,y \quad \text{reflects} \quad g\,x\,b =_\beta (\lambda y.\,g\,x\,y)b$$

More specifically: if we did not have the η-conversion rule, then $g\,x$ and $\lambda y.\,g\,x\,y$ would not be "equivalent" up to conversion. But that seems unreasonable, because they behave

the same way when applied to b. Applied to b, both terms can be converted to $g\,x\,b$. This is why it is reasonable to introduce a rule such that $g\,x$ and $\lambda y.\,g\,x\,y$ are "equivalent" up to conversion.

One also says that the η-conversion expresses the idea of extensionality [HS90, chapter 7].

Note that with the help of contraction and transitivity, η-conversion can be generalized to more than one variable, i.e. $M =_{\beta\eta} \lambda x_1 \ldots x_n.\,M\,x_1 \ldots x_n$. E.g. we can derive $\lambda xyz.\,M\,x\,y\,z =_{\beta\eta} M$:

$$
\cfrac{
\cfrac{
\cfrac{\lambda z.\,M\,x\,y\,z =_\eta M\,x\,y}{\lambda yz.\,M\,x\,y\,z =_{\beta\eta} \lambda y.\,M\,x\,y} \quad \lambda y.\,M\,x\,y =_\eta M\,x
}{
\cfrac{\lambda yz.\,M\,x\,y\,z =_{\beta\eta} M\,x}{\lambda xyz.\,M\,x\,y\,z =_{\beta\eta} \lambda x.\,M\,x} \quad\quad\quad \lambda x.\,M\,x =_\eta M
}
}{
\lambda xyz.\,M\,x\,y\,z =_{\beta\eta} M
}
$$

For any n, we call $\lambda x_1 \ldots x_n.\,M\,x_1 \ldots x_n$ an η-expansion of M.

λ-**Calculus Meta-Properties**[169]

Confluence (equivalently[170], Church-Rosser): reduction is order-independent.

For all M, N_1, N_2, if $M \rightarrow_\beta^* N_1$ and $M \rightarrow_\beta^* N_2$, then there exists a P where $N_1 \rightarrow_\beta^* P$ and $N_2 \rightarrow_\beta^* P$.

---

[169]By metaproperties, we mean properties about reduction and conversion sequences in general.

[170]A reduction $\rightarrow$ is called confluent if

for all M, N_1, N_2, if $M \rightarrow^* N_1$ and $M \rightarrow^* N_2$, then there exists a P where $N_1 \rightarrow^* P$ and $N_2 \rightarrow^* P$.

A reduction is called Church-Rosser if

for all N_1, N_2, if $N_1 \overset{*}{\leftrightarrow} N_2$, then there exists a P where $N_1 \rightarrow^* P$ and $N_2 \rightarrow^* P$.

Here, $\leftarrow := (\rightarrow)^{-1}$ is the inverse of $\rightarrow$, and $\leftrightarrow := \leftarrow \cup \rightarrow$ is the symmetric closure of $\rightarrow$, and $\overset{*}{\leftrightarrow} := (\leftrightarrow)^*$ is the reflexive transitive symmetric closure of $\rightarrow$.

So for example, if we have

$$M_1 \rightarrow M_2 \rightarrow M_3 \rightarrow M_4 \leftarrow M_5 \leftarrow M_6 \rightarrow M_7 \leftarrow M_8 \leftarrow M_9$$

then we would write $M_1 \overset{*}{\leftrightarrow} M_9$.

Confluence is equivalent to the Church-Rosser property [BN98, page 10].
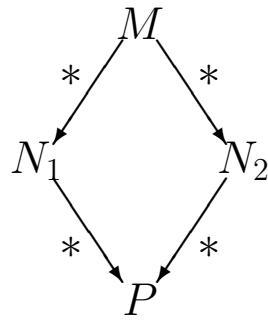
$$
\begin{array}{ccc}
 & M & \\
{\scriptstyle *}\swarrow & & \searrow{\scriptstyle *} \\
N_1 & & N_2 \\
{\scriptstyle *}\searrow & & \swarrow{\scriptstyle *} \\
 & P &
\end{array}
$$

Uniqueness of Normal Forms

Corollary of the Church-Rosser property:

If $M \to_\beta^* N_1$ and $M \to_\beta^* N_2$ where N_1 and N_2 in normal form, then

$$
\begin{array}{ccc}
 & M & \\
{}^{*}\swarrow & & \searrow{}^{*} \\
N_1 & & N_2 \\
{}^{*}\searrow & & \swarrow{}^{*} \\
 & P &
\end{array}
$$

Uniqueness of Normal Forms

Corollary of the Church-Rosser property:

    If $M \to_\beta^* N_1$ and $M \to_\beta^* N_2$ where N_1 and N_2 in normal form, then $N_1 =_\alpha N_2$.

$$M \overset{*}{\searrow} \quad \overset{*}{\swarrow}$$
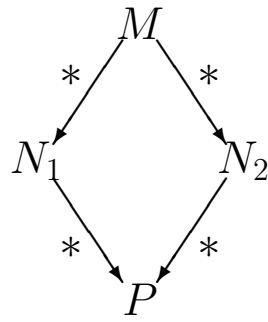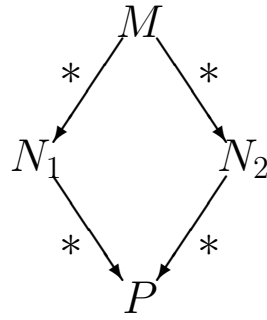
Uniqueness of Normal Forms

Corollary of the Church-Rosser property:

If $M \to_\beta^* N_1$ and $M \to_\beta^* N_2$ where N_1 and N_2 in normal form, then $N_1 =_\alpha N_2$.

Example:

$$(\lambda xy.\, y)(\underline{(\lambda x.\, xx)a}) \to_\beta \underline{(\lambda xy.\, y)(aa)} \to_\beta \lambda y.\, y$$

$$\underline{(\lambda xy.\, y)((\lambda x.\, xx)a)} \to_\beta \lambda y.\, y$$

Turing Completeness

The λ-calculus can represent all computable functions.[171]

---

[171]The untyped λ-calculus is Turing complete. This is usually shown not by mimicking a Turing machine in the λ-calculus, but rather by exploiting the fact that the Turing computable functions are the same class as the μ-recursive functions [HS90, chapter 4]. In a lecture on theory of computation, you have probably learned that the μ-recursive functions are obtained from the primitive recursive functions by so-called unbounded minimalization, while the primitive recursive functions are built from the 0-place zero function, projection functions and the successor function using composition and primitive recursion [LP81].

The proof that the untyped λ-calculus can compute all μ-recursive functions is thus based on showing that each of the mentioned ingredients can be encoded in the untyped λ-calculus. While we are not going to study this, one crucial point is that it should be possible to encode the natural numbers and the arithmetic operations in the untyped λ-calculus.

8.2 Simple Type Theory $\lambda^\rightarrow$

Motivation: Suppose you have constants 1, 2 with usual meaning. Is it sensible to write $1\,2$ (1 applied to 2)?

8.2 Simple Type Theory $\lambda^{\rightarrow}$

Motivation: Suppose you have constants 1, 2 with usual meaning. Is it sensible to write $1\,2$ (1 applied to 2)?

$\lambda^{\rightarrow}$ (simply typed λ-calculus, simple type theory) restricts syntax to "meaningful expressions".

In untyped λ-calculus, we have syntactic objects[172] called terms.

8.2 Simple Type Theory $\lambda^{\rightarrow}$

Motivation: Suppose you have constants 1, 2 with usual meaning. Is it sensible to write $1\ 2$ (1 applied to 2)?

$\lambda^{\rightarrow}$ (simply typed λ-calculus, simple type theory) restricts syntax to "meaningful expressions".

In untyped λ-calculus, we have syntactic objects[172] called terms.

We now introduce syntactic objects called types[173].

8.2 Simple Type Theory $\lambda^{\rightarrow}$

Motivation: Suppose you have constants 1, 2 with usual meaning. Is it sensible to write $1\,2$ (1 applied to 2)?

$\lambda^{\rightarrow}$ (simply typed λ-calculus, simple type theory) restricts syntax to "meaningful expressions".

In untyped λ-calculus, we have syntactic objects[172] called terms.

We now introduce syntactic objects called types[173].

We will say "a term has a type" or "a term is of a type".

---

[172]We also say that we have defined a term language. A particular language is given by a signature, although for the untyped λ-calculus this is simply the set of constants $Const$.

[173]We can say that we define a type language, i.e., a language consisting of types. A particular type language is characterized by a giving a set of base types $\mathcal{B}$. One might also call $\mathcal{B}$ a type signature.

A typical example of a set of base types would be $\{\mathbb{N}, bool\}$, where $\mathbb{N}$ represents the natural numbers and $bool$ the Boolean values $\bot$ and $\top$.

All that matters is that $\mathcal{B}$ is some fixed set "defined by the user".

Two Syntaxes

- Syntax for types ($\mathcal{B}$ a set of base types, $T \in \mathcal{B}$)

$$\tau ::= T \mid \tau \to \tau$$

Two Syntaxes

- Syntax for types ($\mathcal{B}$ a set of base types, $T \in \mathcal{B}$)

$$\tau ::= T \mid \tau \to \tau$$

Examples: $\mathbb{N}$, $\mathbb{N} \to {}^{174}\mathbb{N}$, $(\mathbb{N} \to \mathbb{N}) \to \mathbb{N}$, $\mathbb{N} \to \mathbb{N} \to \mathbb{N}^{175}$

Two Syntaxes

- Syntax for types ($\mathcal{B}$ a set of base types, $T \in \mathcal{B}$)

$$\tau ::= T \mid \tau \to \tau$$

  Examples: $\mathbb{N}$, $\mathbb{N} \to$ [174]$\mathbb{N}$, $(\mathbb{N} \to \mathbb{N}) \to \mathbb{N}$, $\mathbb{N} \to \mathbb{N} \to \mathbb{N}$[175]

- Syntax for (raw[176]) terms: λ-calculus augmented with types[177]

$$e ::= x \mid c \mid (ee) \mid (\lambda x^{\tau}.\,e)$$

---

[174]The type $\mathbb{N} \to \mathbb{N}$ is the type of a function that takes a natural number and returns a natural number.

The type $(\mathbb{N} \to \mathbb{N}) \to \mathbb{N}$ is the type of a function that takes a function, which takes a natural number and returns a natural number, and returns a natural number.

[175]To save parentheses, we use the following convention: types associate to the right, so $\mathbb{N} \to \mathbb{N} \to \mathbb{N}$ stands for $\mathbb{N} \to (\mathbb{N} \to \mathbb{N})$.

Recall that application associates to the left. This may seem confusing at first, but actually, it turns out that the two conventions concerning associativity fit together very neatly.

[176]In the context of typed versions of the λ-calculus, raw terms are terms built ignoring any typing conditions. So raw terms are simply terms as defined for the untyped λ-calculus, possibly augmented with type superscripts.

[177]So far, this is just syntax!

The notation $(\lambda x^{\tau}.\,e)$ simply specifies that binding occurrences of variables in simple type theory are tagged with a

$$(x \in Var,\ c \in Const^{178})$$

---

superscript, where the use of the letter τ makes it clear (in this particular context) that the superscript must be some type, defined by the grammar we just gave.

[178] Var and $Const$ are the sets of variables and constants, respectively, as for the untyped λ-calculus.

Signatures and Contexts

Generally (in various logic-related formalisms) a signature defines the "fixed" symbols of a language, and a context defines the "variable" symbols of a language.

---

[179]We call an expression of the form $x : \tau$ or $c : \tau$ a type binding.

The use of the letter τ makes it clear (in this particular context) that the superscript must be some type, defined by the grammar we just gave.

[180]For propositional logic, we did not use the notion of signature, although we mentioned that strictly speaking, there is not just the language of propositional logic, but rather a language of propositional logic which depends on the choice of the variables.

In first-order logic, a signature was a pair $(\mathcal{F}, \mathcal{P})$ defining the function and predicate symbols, although strictly speaking, the signature should also specify the arities of the symbols in some way. Recall that we did not bother to fix a precise technical way of specifying those arities. We were content with saying that they are specified in "some unambiguous way".

In sorted logic, the signature must also specify the sorts of

Signatures and Contexts

Generally (in various logic-related formalisms) a signature defines the "fixed" symbols of a language, and a context defines the "variable" symbols of a language. In $\lambda^{\rightarrow}$,

- a signature Σ is a sequence ($c \in Const$)

$$\Sigma ::= \langle\rangle \mid \Sigma, c : \tau^{179}$$

- a context Γ is a sequence ($x \in Var$)

$$\Gamma ::= \langle\rangle \mid \Gamma, x : \tau$$

What's the difference to signatures you have seen so far?[180]

---

[179]We call an expression of the form $x : \tau$ or $c : \tau$ a type binding.

The use of the letter τ makes it clear (in this particular context) that the superscript must be some type, defined by the grammar we just gave.

[180]For propositional logic, we did not use the notion of signature, although we mentioned that strictly speaking, there is not just the language of propositional logic, but rather a language of propositional logic which depends on the choice of the variables.

In first-order logic, a signature was a pair $(\mathcal{F}, \mathcal{P})$ defining the function and predicate symbols, although strictly speaking, the signature should also specify the arities of the symbols in some way. Recall that we did not bother to fix a precise technical way of specifying those arities. We were content with saying that they are specified in "some unambiguous way".

In sorted logic, the signature must also specify the sorts of

Type Assignment Calculus

We now define type judgements: "a term has a type" or "a term is of a type". Generally this depends on a signature Σ and a context Γ. For example

$$\Gamma \vdash_\Sigma c\, x : \sigma^{181}$$

where $\Sigma = c : \tau \to \sigma$ and $\Gamma = x : \tau$.

---

all symbols. But we did not study sorted logic in any detail.

In the untyped λ-calculus, the signature is simply the set of constants.

Summarizing, we have not been very precise about the notion of a signature so far.

For $\lambda^\to$, the rules for "legal" terms become more tricky, and it is important to be formal about signatures.

In $\lambda^\to$, a signature associates a type with each constant symbol by writing $c : \tau$.

Usually, we will assume that $Const$ is clear from the context, and that Σ contains an expression of the form $c : \tau$ for each $c \in Const$, and in fact, that Σ is clear from the context as well. Since Σ contains an expression of the form $c : \tau$ for each $c \in Const$, it is redundant to give $Const$ explicitly. It is sufficient to give Σ.

[181] The expression

$$\Gamma \vdash_\Sigma c\, x : \sigma$$

is called a type judgement. It says that given the signature

Type Assignment Calculus

We now define type judgements: "a term has a type" or "a term is of a type". Generally this depends on a signature Σ and a context Γ. For example

$$\Gamma \vdash_\Sigma c\, x : \sigma^{181}$$

where $\Sigma = c : \tau \to \sigma$ and $\Gamma = x : \tau$.

We usually leave Σ implicit and write $\vdash$ instead of $\vdash_\Sigma$.

If Γ is empty it is omitted.

all symbols. But we did not study sorted logic in any detail.

In the untyped λ-calculus, the signature is simply the set of constants.

Summarizing, we have not been very precise about the notion of a signature so far.

For $\lambda^\to$, the rules for "legal" terms become more tricky, and it is important to be formal about signatures.

In $\lambda^\to$, a signature associates a type with each constant symbol by writing $c : \tau$.

Usually, we will assume that $Const$ is clear from the context, and that Σ contains an expression of the form $c : \tau$ for each $c \in Const$, and in fact, that Σ is clear from the context as well. Since Σ contains an expression of the form $c : \tau$ for each $c \in Const$, it is redundant to give $Const$ explicitly. It is sufficient to give Σ.

---

[181] The expression

$$\Gamma \vdash_\Sigma c\, x : \sigma$$

is called a type judgement. It says that given the signature

Type Assignment Calculus: Rules[182]

$\Sigma = x : \tau$ and the context $\Gamma = c : \tau \to \sigma$, the term

$c\ x$ has type σ or

$c\ x$ is of type σ or

$c\ x$ is assigned type σ.

Recall that you have seen other judgements before.

[182]Type assignment is defined as a system of rules for deriving type judgements, in the same way that we have defined derivability judgements for logics, and β-reduction for the untyped λ-calculus.

$$\frac{c : \tau \in {}^{183}\Sigma}{\Gamma \vdash c : \tau} \; assum \qquad \Gamma, x : \tau, \Delta \vdash x : \tau \quad hyp^{184}$$

$$\frac{\Gamma \vdash e : \sigma \rightarrow \tau \quad \Gamma \vdash e' : \sigma}{\Gamma \vdash ee' : \tau} \; app \qquad \frac{\Gamma, x : \sigma^{185} \vdash e : \tau}{\Gamma \vdash \lambda x^{\sigma}.\, e : \sigma \rightarrow \tau} \; abs$$

$$\frac{c : \tau \in {}^{183}\Sigma}{\Gamma \vdash c : \tau} \; assum \qquad \Gamma, x : \tau, \Delta \vdash x : \tau \quad hyp^{184}$$

$$\frac{\Gamma \vdash e : \sigma \to \tau \quad \Gamma \vdash e' : \sigma}{\Gamma \vdash ee' : \tau} \; app \qquad \frac{\Gamma, x : \sigma^{185} \vdash e : \tau}{\Gamma \vdash \lambda x^{\sigma}. e : \sigma \to \tau} \; abs$$

[183]Recall that Σ is a sequence. By abuse of notation, we sometimes identify this sequence with a set and allow ourselves to write $c : \tau \in \Sigma$.

We may also write $\Sigma \subseteq \Sigma'$ meaning that $c : \tau \in \Sigma$ implies $c : \tau \in \Sigma'$.

[184]One could also formulate *hyp* as follows:

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

That would be in close analogy to LF, a system not tretaed here.

[185]A sequence is a collection of objects which differs from sets in that a sequence contains the objects in a certain order, and there can be multiple occurrences of an object.

We write a sequence containing the objects $o_1, \ldots, o_n$ as $\langle o_1, \ldots, o_n \rangle$, or sometimes simply $o_1, \ldots, o_n$.

If Ω is the sequence $o_1, \ldots, o_n$, then we write Ω, o for the sequence $\langle o_1, \ldots, o_n, o \rangle$ and o, Ω for the sequence

Note analogy to minimal logic over $\to$[186].

β-reduction defined as before, has subject reduction prop-

$\langle o, o_1, \ldots, o_n \rangle$.

An empty sequence is denoted by $\langle\,\rangle$.

[186]Recall the sequent rules of the "$\to / \wedge$" fragment of propositional logic. Consider now only the "$\to$" fragment. We call this fragment minimal logic over $\to$.

If you take the rule

$$\Gamma, x : \tau, \Delta \vdash x : \tau \quad \textit{hyp}$$

of $\lambda^{\to}$ and throw away the terms (so you keep only the types), you obtain essentially the rule for assumptions

$$\Gamma \vdash A \quad (\text{where } A \in \Gamma)$$

of propositional logic.

Likewise, if you do the same with the rule

$$\frac{\Gamma \vdash e : \sigma \to \tau \quad \Gamma \vdash e' : \sigma}{\Gamma \vdash ee' : \tau} \; \textit{app}$$

of $\lambda^{\to}$, you obtain essentially the rule

$$\frac{\Gamma \vdash A \to B \quad \Gamma \vdash A}{\Gamma \vdash B} \; \textit{$\to$-E}$$

of propositional logic.

Finally, if you do the same with the rule

$$\frac{\Gamma, x : \sigma \vdash e : \tau}{\Gamma \vdash \lambda x^{\sigma}.\, e \; : \; \sigma \to \tau} \; abs$$

of $\lambda^{\to}$, you obtain essentially the rule

$$\frac{A, \Gamma \vdash B}{\Gamma \vdash A \to B} \; \to\text{-}I$$

of propositional logic.

Note that in this setting, there is no analogous propositional logic rule for

$$\frac{c : \tau \in \Sigma}{\Gamma \vdash c : \tau} \; assum$$

So for the moment, we can observe a close analogy between $\lambda^{\to}$, for Σ being empty, and the $\to$ fragment of propositional logic, which is also called minimal logic over $\to$.

Such an analogy between a type theory (of which $\lambda^{\to}$ is an example) and a logic is referred to in the literature as Curry-

erty[187] and is strongly normalizing[188].

Howard isomorphism [Tho91]. One also speaks of propositions as types [GLT89]. The isomorphism is so fundamental that it is common to characterize type theories by the logic they represent, so for example, one might say:

$\lambda^{\rightarrow}$ is the type theory of minimal logic over $\rightarrow$.

[187]Subject reduction is the following property: reduction does not change the type of a term, so if $\vdash_\Sigma M : \tau$ and $M \rightarrow_\beta N$, then $\vdash_\Sigma N : \tau$.

[188]The simply-typed λ-calculus, unlike the untyped λ-calculus, is normalizing, that is to say, every term has a normal form. Even more, it is strongly normalizing, that is, this normal form is reached regardless of the reduction order.

Example 1

$$\vdash \lambda x^\sigma. \lambda y^\tau. x :$$

---

[189]In this example, you may regard σ and τ as base types (this would require that $\sigma, \tau \in \mathcal{B}$), but in fact, it is more natural to regard them as metavariables standing for arbitrary types. Whatever types you substitute for σ and τ, you obtain a derivation of a type judgement.

This is in analogy to schematic derivations in a logic.

Note also that Σ is irrelevant for the example and hence arbitrary.

Example 1

$$\vdash \lambda x^{\sigma}.\, \lambda y^{\tau}.\, x : \sigma \to (\tau \to \sigma)$$

---

[189]In this example, you may regard σ and τ as base types (this would require that $\sigma, \tau \in \mathcal{B}$), but in fact, it is more natural to regard them as metavariables standing for arbitrary types. Whatever types you substitute for σ and τ, you obtain a derivation of a type judgement.

This is in analogy to schematic derivations in a logic.

Note also that Σ is irrelevant for the example and hence arbitrary.

Example 1

$$\frac{}{\vdash \lambda x^{\sigma}.\,\lambda y^{\tau}.\,x : \sigma \rightarrow (\tau \rightarrow \sigma)} \text{ abs}$$

[189]In this example, you may regard σ and τ as base types (this would require that $\sigma, \tau \in \mathcal{B}$), but in fact, it is more natural to regard them as metavariables standing for arbitrary types. Whatever types you substitute for σ and τ, you obtain a derivation of a type judgement.

This is in analogy to schematic derivations in a logic.

Note also that Σ is irrelevant for the example and hence arbitrary.

Example 1

$$\frac{x : \sigma \vdash \lambda y^{\tau}.\, x : \tau \to \sigma}{\vdash \lambda x^{\sigma}.\, \lambda y^{\tau}.\, x : \sigma \to (\tau \to \sigma)} \text{ abs}$$

---

[189] In this example, you may regard σ and τ as base types (this would require that $\sigma, \tau \in \mathcal{B}$), but in fact, it is more natural to regard them as metavariables standing for arbitrary types. Whatever types you substitute for σ and τ, you obtain a derivation of a type judgement.

This is in analogy to schematic derivations in a logic.

Note also that Σ is irrelevant for the example and hence arbitrary.

Example 1

$$\frac{\dfrac{}{x : \sigma \vdash \lambda y^\tau. x : \tau \to \sigma} \text{ abs}}{\vdash \lambda x^\sigma. \lambda y^\tau. x : \sigma \to (\tau \to \sigma)} \text{ abs}$$

---

[189]In this example, you may regard σ and τ as base types (this would require that $\sigma, \tau \in \mathcal{B}$), but in fact, it is more natural to regard them as metavariables standing for arbitrary types. Whatever types you substitute for σ and τ, you obtain a derivation of a type judgement.

This is in analogy to schematic derivations in a logic.

Note also that Σ is irrelevant for the example and hence arbitrary.

Example 1

$$\frac{\dfrac{x : \sigma, \; y : \tau \vdash x : \sigma}{x : \sigma \vdash \lambda y^\tau . \, x : \tau \to \sigma} \; \text{abs}}{\vdash \lambda x^\sigma . \, \lambda y^\tau . \, x : \sigma \to (\tau \to \sigma)} \; \text{abs}$$

---

[189]In this example, you may regard σ and τ as base types (this would require that $\sigma, \tau \in \mathcal{B}$), but in fact, it is more natural to regard them as metavariables standing for arbitrary types. Whatever types you substitute for σ and τ, you obtain a derivation of a type judgement.

This is in analogy to schematic derivations in a logic.

Note also that Σ is irrelevant for the example and hence arbitrary.

Example 1

$$\frac{\dfrac{\overline{\rule{0pt}{1em}}}{x : \sigma, \; y : \tau \vdash x : \sigma} \; \textit{hyp}}{\dfrac{x : \sigma \vdash \lambda y^\tau. \, x : \tau \to \sigma}{\vdash \lambda x^\sigma. \, \lambda y^\tau. \, x : \sigma \to (\tau \to \sigma)} \; \text{abs}} \; \text{abs}$$

---

[189]In this example, you may regard σ and τ as base types (this would require that $\sigma, \tau \in \mathcal{B}$), but in fact, it is more natural to regard them as metavariables standing for arbitrary types. Whatever types you substitute for σ and τ, you obtain a derivation of a type judgement.

This is in analogy to schematic derivations in a logic.

Note also that Σ is irrelevant for the example and hence arbitrary.

Example 1

$$\cfrac{\cfrac{\rule{4cm}{0.4pt}}{x:\sigma,\ y:\tau \vdash x:\sigma} \ hyp}{\cfrac{x:\sigma \vdash \lambda y^\tau.\,x:\tau \to \sigma}{\vdash \lambda x^\sigma.\,\lambda y^\tau.\,x:\sigma \to (\tau \to \sigma)} \ \text{abs}} \ \text{abs}$$

Note the use of schematic types[189]!

---

[189]In this example, you may regard σ and τ as base types (this would require that $\sigma, \tau \in \mathcal{B}$), but in fact, it is more natural to regard them as metavariables standing for arbitrary types. Whatever types you substitute for σ and τ, you obtain a derivation of a type judgement.

This is in analogy to schematic derivations in a logic.

Note also that Σ is irrelevant for the example and hence arbitrary.

Example 1

$$\frac{\dfrac{x : \sigma,\ y : \tau \vdash x : \sigma}{x : \sigma \vdash \lambda y^{\tau}.\, x : \tau \to \sigma}\ \text{abs}}{\vdash \lambda x^{\sigma}.\, \lambda y^{\tau}.\, x : \sigma \to (\tau \to \sigma)}\ \text{abs}$$

Note the use of schematic types[189]!

For simplicity, applications of *hyp* are usually not explicitly marked in proof.

---

[189]In this example, you may regard σ and τ as base types (this would require that $\sigma, \tau \in \mathcal{B}$), but in fact, it is more natural to regard them as metavariables standing for arbitrary types. Whatever types you substitute for σ and τ, you obtain a derivation of a type judgement.

This is in analogy to schematic derivations in a logic.

Note also that Σ is irrelevant for the example and hence arbitrary.

Example 2

$$\Gamma = f : \sigma \to \sigma \to \tau, x : \sigma$$

$$\vdash \lambda f^{\sigma \to \sigma \to \tau}. \lambda x^{\sigma}. f\, x\, x :$$

Example 2

$$\Gamma = f : \sigma \rightarrow \sigma \rightarrow \tau, x : \sigma$$

$$\vdash \lambda f^{\sigma \rightarrow \sigma \rightarrow \tau}. \lambda x^{\sigma}. f\, x\, x : (\sigma \rightarrow \sigma \rightarrow \tau) \rightarrow \sigma \rightarrow \tau$$

Example 2

$$\Gamma = f : \sigma \rightarrow \sigma \rightarrow \tau, x : \sigma$$

$$\frac{}{\vdash \lambda f^{\sigma \rightarrow \sigma \rightarrow \tau} . \lambda x^{\sigma} . f\, x\, x : (\sigma \rightarrow \sigma \rightarrow \tau) \rightarrow \sigma \rightarrow \tau}\ \textit{abs}$$

Example 2

$$\Gamma = f : \sigma \rightarrow \sigma \rightarrow \tau, x : \sigma$$

$$\frac{f : \sigma \rightarrow \sigma \rightarrow \tau \vdash \lambda x^{\sigma}. f\, x\, x : \sigma \rightarrow \tau}{\vdash \lambda f^{\sigma \rightarrow \sigma \rightarrow \tau}. \lambda x^{\sigma}. f\, x\, x : (\sigma \rightarrow \sigma \rightarrow \tau) \rightarrow \sigma \rightarrow \tau} \; abs$$

Example 2

$$\Gamma = f : \sigma \to \sigma \to \tau, x : \sigma$$

$$\frac{\dfrac{}{f : \sigma \to \sigma \to \tau \vdash \lambda x^{\sigma}.\, f\, x\, x : \sigma \to \tau}\ abs}{\vdash \lambda f^{\sigma \to \sigma \to \tau}.\, \lambda x^{\sigma}.\, f\, x\, x : (\sigma \to \sigma \to \tau) \to \sigma \to \tau}\ abs$$

Example 2

$$\Gamma = f : \sigma \to \sigma \to \tau, x : \sigma$$

$$\cfrac{\cfrac{\Gamma \vdash f\,x\,x : \tau}{f : \sigma \to \sigma \to \tau \vdash \lambda x^{\sigma}.\,f\,x\,x : \sigma \to \tau}\ abs}{\vdash \lambda f^{\sigma \to \sigma \to \tau}.\,\lambda x^{\sigma}.\,f\,x\,x : (\sigma \to \sigma \to \tau) \to \sigma \to \tau}\ abs$$

Example 2

$$\Gamma = f : \sigma \to \sigma \to \tau, x : \sigma$$

$$\cfrac{\cfrac{\cfrac{}{\Gamma \vdash f\,x\,x : \tau}\ app}{f : \sigma \to \sigma \to \tau \vdash \lambda x^\sigma.\,f\,x\,x : \sigma \to \tau}\ abs}{\vdash \lambda f^{\sigma \to \sigma \to \tau}.\,\lambda x^\sigma.\,f\,x\,x : (\sigma \to \sigma \to \tau) \to \sigma \to \tau}\ abs$$

Example 2

$$\Gamma = f : \sigma \to \sigma \to \tau, x : \sigma$$

$$\cfrac{\cfrac{\cfrac{\Gamma \vdash f\,x : \sigma \to \tau \qquad \Gamma \vdash x : \sigma}{\Gamma \vdash f\,x\,x : \tau}\ app}{f : \sigma \to \sigma \to \tau \vdash \lambda x^\sigma.\,f\,x\,x : \sigma \to \tau}\ abs}{\vdash \lambda f^{\sigma \to \sigma \to \tau}.\,\lambda x^\sigma.\,f\,x\,x : (\sigma \to \sigma \to \tau) \to \sigma \to \tau}\ abs$$

Example 2

$$\Gamma = f : \sigma \to \sigma \to \tau, x : \sigma$$

$$\cfrac{\cfrac{\cfrac{\rule{6cm}{0.4pt}}{\Gamma \vdash f\,x : \sigma \to \tau}\ app \qquad \Gamma \vdash x : \sigma}{\Gamma \vdash f\,x\,x : \tau}\ app}{\cfrac{f : \sigma \to \sigma \to \tau \vdash \lambda x^\sigma.\,f\,x\,x : \sigma \to \tau}{\vdash \lambda f^{\sigma \to \sigma \to \tau}.\,\lambda x^\sigma.\,f\,x\,x : (\sigma \to \sigma \to \tau) \to \sigma \to \tau}\ abs}\ abs$$

Example 2

$$\Gamma = f : \sigma \to \sigma \to \tau, x : \sigma$$

$$\cfrac{\cfrac{\cfrac{\Gamma \vdash f : \sigma \to \sigma \to \tau \quad \Gamma \vdash x : \sigma}{\Gamma \vdash f\,x : \sigma \to \tau}\ app \quad \Gamma \vdash x : \sigma}{\Gamma \vdash f\,x\,x : \tau}\ app}{\cfrac{f : \sigma \to \sigma \to \tau \vdash \lambda x^\sigma.\,f\,x\,x : \sigma \to \tau}{\vdash \lambda f^{\sigma \to \sigma \to \tau}.\,\lambda x^\sigma.\,f\,x\,x : (\sigma \to \sigma \to \tau) \to \sigma \to \tau}\ abs}\ abs$$

Example 3

$$\Sigma = f : \sigma \to \sigma \to \tau$$
$$\Gamma = x : \sigma$$

$$\Gamma \vdash f\,x\,x : \tau$$

---

[190]In Example 3, we have $f : \sigma \to \sigma \to \tau \in \Sigma$, and so f is a constant.

In Example 2, we have $f : \sigma \to \sigma \to \tau \in \Gamma$, and so f is a variable.

Looking at the different derivations of the type judgement $\Gamma \vdash f\,x\,x : \tau$ in Examples 2 and 3, you may find that they are very similar, and you may wonder: What is the point? Why do we distinguish between constants and variables?

In fact, one could simulate constants by variables. When setting up a type theory or programming language, there are choices to be made about whether there should be a distinction between variables and constants, and what it should look like. There is a famous epigram by Alan Perlis:

> One man's constant is another man's variable.

For our purposes, it is much clearer conceptually to make the distinction. For example, if we want to introduce the natural numbers in our $\lambda^{\to}$ language, then it is intuitive that there should be constants $1, 2, \ldots$ denoting the numbers. If $1, 2, \ldots$

Example 3

$$\Sigma = f : \sigma \to \sigma \to \tau$$
$$\Gamma = x : \sigma$$

$$\cfrac{\cfrac{f : \sigma \to \sigma \to \tau \in \Sigma}{\Gamma \vdash f : \sigma \to \sigma \to \tau}\ assum \qquad \Gamma \vdash x : \sigma}{\cfrac{\Gamma \vdash f\,x : \sigma \to \tau \qquad\qquad \Gamma \vdash x : \sigma}{\Gamma \vdash f\,x\,x : \tau}\ app}\ app$$

Note that this time, f is a constant[190].

---

[190]In Example 3, we have $f : \sigma \to \sigma \to \tau \in \Sigma$, and so f is a constant.

In Example 2, we have $f : \sigma \to \sigma \to \tau \in \Gamma$, and so f is a variable.

Looking at the different derivations of the type judgement $\Gamma \vdash f\,x\,x : \tau$ in Examples 2 and 3, you may find that they are very similar, and you may wonder: What is the point? Why do we distinguish between constants and variables?

In fact, one could simulate constants by variables. When setting up a type theory or programming language, there are choices to be made about whether there should be a distinction between variables and constants, and what it should look like. There is a famous epigram by Alan Perlis:

> One man's constant is another man's variable.

For our purposes, it is much clearer conceptually to make the distinction. For example, if we want to introduce the natural numbers in our $\lambda^{\to}$ language, then it is intuitive that there should be constants $1, 2, \ldots$ denoting the numbers. If $1, 2, \ldots$

Example 3

$$\Sigma \;=\; f : \sigma \to \sigma \to \tau$$
$$\Gamma \;=\; x : \sigma$$

$$
\dfrac{
\dfrac{\Gamma \vdash f : \sigma \to \sigma \to \tau \qquad \Gamma \vdash x : \sigma}{\Gamma \vdash f\,x : \sigma \to \tau}\; app
\qquad \Gamma \vdash x : \sigma
}{\Gamma \vdash f\,x\,x : \tau}\; app
$$

Note that this time, f is a constant[190].

We will often suppress applications of *assum*.

---

[190]In Example 3, we have $f : \sigma \to \sigma \to \tau \in \Sigma$, and so f is a constant.

In Example 2, we have $f : \sigma \to \sigma \to \tau \in \Gamma$, and so f is a variable.

Looking at the different derivations of the type judgement $\Gamma \vdash f\,x\,x : \tau$ in Examples 2 and 3, you may find that they are very similar, and you may wonder: What is the point? Why do we distinguish between constants and variables?

In fact, one could simulate constants by variables. When setting up a type theory or programming language, there are choices to be made about whether there should be a distinction between variables and constants, and what it should look like. There is a famous epigram by Alan Perlis:

> One man's constant is another man's variable.

For our purposes, it is much clearer conceptually to make the distinction. For example, if we want to introduce the natural numbers in our $\lambda^{\to}$ language, then it is intuitive that there should be constants $1, 2, \dots$ denoting the numbers. If $1, 2, \dots$

Type Assignment and $\alpha\beta\eta$-Conversion

Type construction:

- Type construction[191] is decidable.

---

were variables, then we could write strange expressions like $\lambda 2^{\mathbb{N}\to\mathbb{N}}.\, y$, so we could use 2 as a variable of type $\mathbb{N} \to \mathbb{N}$.

[191]Type construction is the problem of given a Σ, Γ and e, finding a τ such that Σ, $\Gamma \vdash e : \tau$.

Sometimes one also considers the problem where Γ is unknown and must also be constructed.

[192]$\alpha\beta\eta$-conversion is defined as for $\lambda^{\to}$. Given two (extended) λ-terms e and e', it is decidable whether $e =_{\alpha\beta\eta} e'$.

Type Assignment and $\alpha\beta\eta$-Conversion

Type construction:

- Type construction[191] is decidable.

- There is a practically useful implementation for type-construction (Hindley-Milner algorithm $\mathcal{W}$ [Mil78, NN99]).

  Term congruence[192] ($e =_{\alpha\beta\eta} e'$?) is decidable.

---

were variables, then we could write strange expressions like $\lambda 2^{\mathbb{N}\to\mathbb{N}}. y$, so we could use 2 as a variable of type $\mathbb{N} \to \mathbb{N}$.

[191]Type construction is the problem of given a Σ, Γ and e, finding a τ such that Σ, $\Gamma \vdash e : \tau$.

Sometimes one also considers the problem where Γ is unknown and must also be constructed.

[192]$\alpha\beta\eta$-conversion is defined as for $\lambda^{\to}$. Given two (extended) λ-terms e and e', it is decidable whether $e =_{\alpha\beta\eta} e'$.

8.3 Polymorphism and Type Classes

We will now look at the typed λ-calculus extended by polymorphism and type classes.

As we will see later, this is the universal representation for object logics in Isabelle.

Polymorphism: Intuition

In functional programming, the function $append$ for concatenating two lists works the same way on integer lists and on character lists: $append$ is polymorphic[193].

Type language must be generalized to include type variables (denoted by $\alpha, \beta \ldots$) and type constructors.

Example: $append$ has type $\alpha\ list \rightarrow \alpha\ list \rightarrow \alpha\ list$, and by type instantiation, it can also have type, say, $int\ list \rightarrow int\ list \rightarrow int\ list$.

---

[193]In functional programming, you will come across functions that operate uniformly on many different types. For example, a function $append$ for concatenating two lists works the same way on integer lists and on character lists. Such functions are called polymorphic.

More precisely, this kind of polymorphism, where a function does exactly the same thing regardless of the type instance, is called parametric polymorphism, as opposed to ad-hoc polymorphism.

In a type system with polymorphism, the notion of base type (which is just a type constant, i.e., one symbol) is generalized to a type constructor with an arity ≥ 0. A type constructor of arity n applied to n types is then a type. For example, there might be a type constructor $list$ of arity 1, and int of arity 0. Then, $int\ list$ is a type.

Note that application of a type constructor to a type is written in postfix notation, unlike any notation for function application we have seen. However, other conventions exist,

Polymorphism: Two Syntaxes

- Syntax for polymorphic types ($\mathcal{B}$ a set of type constructors[194] including $\rightarrow$), $T \in \mathcal{B}$, α is a type variable)

$$\tau ::= \alpha \mid (\tau, \dots, \tau)\, T$$

---

even within Isabelle.

See [Pau96, Tho95b, Tho99] for details on the polymorphic type systems of functional programming languages.

[194]As before, we define a type language, i.e., a language consisting of types, and a particular type language is characterized by a giving a certain set of symbols $\mathcal{B}$. But unlike before, $\mathcal{B}$ is now a set of type constructors. Each type constructor has an arity associated with it just like a function in first-order logic. The intention is that a type constructor may be applied to types.

Following the conventions of ML [Pau96], we write types in postfix notation, something we have not seen before. I.e., the type constructor comes after the arguments it is applied to.

It makes perfect sense to view the function construction arrow $\rightarrow$ as type constructor, however written infix rather than postfix.

So the $\mathcal{B}$ is some fixed set "defined by the user", but it should definitely always include $\rightarrow$.

Polymorphism: Two Syntaxes

- Syntax for polymorphic types ($\mathcal{B}$ a set of type construc-
  tors[194] including $\rightarrow$), $T \in \mathcal{B}$, α is a type variable)

$$\tau ::= \alpha \mid (\tau, \ldots, \tau)\, T$$

  Examples: $\mathbb{N}$, $\mathbb{N} \rightarrow \mathbb{N}$, $\alpha\, list$, $\mathbb{N}\, list$, $(\mathbb{N}, bool)\, pair$.

---

even within Isabelle.

See [Pau96, Tho95b, Tho99] for details on the polymorphic type systems of functional programming languages.

[194]As before, we define a type language, i.e., a language consisting of types, and a particular type language is characterized by a giving a certain set of symbols $\mathcal{B}$. But unlike before, $\mathcal{B}$ is now a set of type constructors. Each type constructor has an arity associated with it just like a function in first-order logic. The intention is that a type constructor may be applied to types.

Following the conventions of ML [Pau96], we write types in postfix notation, something we have not seen before. I.e., the type constructor comes after the arguments it is applied to.

It makes perfect sense to view the function construction arrow $\rightarrow$ as type constructor, however written infix rather than postfix.

So the $\mathcal{B}$ is some fixed set "defined by the user", but it should definitely always include $\rightarrow$.

Polymorphism: Two Syntaxes

- Syntax for polymorphic types ($\mathcal{B}$ a set of type constructors[194] including $\rightarrow$), $T \in \mathcal{B}$, α is a type variable)

$$\tau ::= \alpha \mid (\tau, \ldots, \tau)\, T$$

  Examples: $\mathbb{N}$, $\mathbb{N} \rightarrow \mathbb{N}$, $\alpha\, list$, $\mathbb{N}\, list$, $(\mathbb{N}, bool)\, pair$.

- Syntax for (raw) terms as before:

$$e ::= x \mid c \mid (ee) \mid (\lambda x^{\tau}.\, e)$$

$(x \in Var, \; c \in Const)$

---

even within Isabelle.

  See [Pau96, Tho95b, Tho99] for details on the polymorphic type systems of functional programming languages.

  [194]As before, we define a type language, i.e., a language consisting of types, and a particular type language is characterized by a giving a certain set of symbols $\mathcal{B}$. But unlike before, $\mathcal{B}$ is now a set of type constructors. Each type constructor has an arity associated with it just like a function in first-order logic. The intention is that a type constructor may be applied to types.

  Following the conventions of ML [Pau96], we write types in postfix notation, something we have not seen before. I.e., the type constructor comes after the arguments it is applied to.

  It makes perfect sense to view the function construction arrow $\rightarrow$ as type constructor, however written infix rather than postfix.

  So the $\mathcal{B}$ is some fixed set "defined by the user", but it should definitely always include $\rightarrow$.

Polymorphic Type Assignment Calculus

Type substitutions (denoted Θ) defined in analogy to substitutions in FOL[195]. Apart from application of Θ in rule *assum*, type assignment is as for $\lambda^{\rightarrow}$:

$$\frac{c : \tau \in \Sigma}{\Gamma \vdash c : \tau\Theta} \; assum^{*} \qquad \Gamma, x : \tau, \Delta \vdash x : \tau \quad hyp$$

$$\frac{\Gamma \vdash e : \sigma \rightarrow \tau \quad \Gamma \vdash e' : \sigma}{\Gamma \vdash ee' : \tau} \; app \qquad \frac{\Gamma, x : \sigma \vdash e : \tau}{\Gamma \vdash \lambda x^{\sigma}.\, e \; : \sigma \rightarrow \tau} \; abs$$

*: Θ is any type substitution.

---

[195]A type substitution replaces a type variable by a type, just like in first-order logic, a substitution replaces a variable by a term.

Type Classes: Intuition

Type classes[196] are a way of "making ad-hoc polymorphism[197]

---

[196]Type classes are a way of "making ad-hoc polymorphism less ad-hoc" [HHPW96, WB89].

Type classes are used to group together types with certain properties, in particular, types for which certain symbols are defined.

For example, for some types, a symbol $\leq$ (which is a binary infix predicate) may exist and for some it may not, and we could have a type class ord containing all types for which it exists.

Suppose you want to sort a list of elements (smaller elements should come before bigger elements). This is only defined for elements of a type for which the symbol $\leq$ exists.

Note that while a symbol such as $\leq$ may have a similar meaning for different types (for example, integers and reals), one cannot say that it means exactly the same thing regardless of the type of the argument to which it is applied. In fact, $\leq$ has to be defined separately for each type in ord.

This is in contrast to parametric poymorphism, but also

less ad-hoc"[HHPW96, WB89].

Type classes are used to group together types with certain properties, in particular, types for which certain symbols are defined.

We only sketch the formalization here, and refer to [HHPW96, Nip93, NP93] for details.

---

somewhat different from ad-hoc polymorphism: The types of the symbols must not be declared separately. E.g., one has to declare only once that $\leq$ is of type $(a :: ord, \alpha)$.

[197]Ad-hoc polymorphism, also called overloading, refers to functions that do different (although usually similar) things on different types. For example, a function $\leq$ may be defined as 'a' $\leq$ 'b' ... on characters and $1 \leq 2$ ... on integers. In this case, the symbol $\leq$ must be declared and defined separately for each type.

This is in contrast to parametric pomorphism, but also somewhat different from type classes.

Type classes are a way of "making ad-hoc polymorphism less ad-hoc"[HHPW96, WB89].

Type Classes in Isabelle

- Syntactic classes[198] (similarly as in Haskell): E.g., declare that there exists a class ord which is a subclass of class $term$, and that for any $\tau :: ord$, the constant $\leq$ is defined and has type $\tau \rightarrow \tau \rightarrow bool$. Isabelle has syntax for this.

---

[198]A syntactic class is a class of types for which certain symbols are declared to exist. Isabelle has a syntax for such declarations. E.g., the declaration

```
sort ord < term
const <= : ['a::ord, 'a] => bool
```

may form part of an Isabelle theory file. It declares a type class ord which is a subclass (that's what the $<$ means; in mathematical notation it will be written $\prec$) of a class $term$, meaning that any type in ord is also in $term$. We will write the "class judgement" $ord \prec term$. The class $term$ must be defined elsewhere.

The second line declares a symbol <=. Such a declaration is preceded by the keyword const. The notation $\alpha :: ord$ stands for a type variable constrained to be in class ord. So <= is declared to be of type $[\alpha :: ord, \alpha] \Rightarrow bool$, meaning that it takes two arguments of a type in the class ord and returns a term of type $bool$. The symbol $\Rightarrow (=>)$ is the function type arrow in Isabelle. Note that the second occurrence of α is

- Axiomatic classes[199]: Declare (axiomatize) that certain theorems should hold for a $\tau :: \kappa$ where κ is a type class. E.g., axiomatize that $\leq$ is reflexive by an (Isabelle) theorem $"x \leq x"$. Isabelle has syntax for this.

written without $:: ord$. This is because it is enough to state the class constraint once.

Note also that $[\alpha :: ord, \alpha] => bool$ is in fact just another way of writing $\alpha :: ord => \alpha => bool$, similarly as for goals.

Haskell [HHPW96] has type classes but ML [Pau96] hasn't.

[199]In addition to declaring the syntax of a type class, one can axiomatize the semantics of the symbols. Again, Isabelle has a syntax for such declarations. E.g., the declaration

```
 axclass order < ord
    order_refl: ''x <= x ''
    order_trans: ''[| x <= y; y <= z |] ==> x <= z''
    ...
```

may form part of an Isabelle theory file. It declares an axiomatic type class $order$ which is a subclass of ord defined above.

The next two lines are the axioms. Here, order_refl and order_trans are the names of the axioms. Recall that $\implies$ is the implication symbol in Isabelle (that is to say, the metalevel

To use a class, we can declare members[200] of it, e.g., $\mathbb{N}$ is a member of ord.

implication).

Whenever an Isabelle theory declares that a type is a member of such a class, it must prove those axioms.

The rationale of having axiomatic classes is that it allows for proofs that hold in different but similar mathematical structures to be done only once. So for example, all theorems that hold for dense orders can be proven for all dense orders with one single proof.

[200]One also speaks of a type being an instance of a type class, but this is slightly confusing, since we also say that a type can be an instance of another type, e.g., $\mathbb{N} \to \mathbb{N}$ is an instance of α, since $\alpha[\alpha \leftarrow (\mathbb{N} \to \mathbb{N})] = \mathbb{N} \to \mathbb{N}$. So it is better to speak of a member of a type class.

Isabelle provides a syntax for declaring that a type is a member of a type class, e.g.

```
instance nat :: ord
```

declares that type nat is a member of class ord.

If the class κ is a syntactic class, such a declaration must

Syntax: Classes, Types, and Terms

Based on

- a set of type classes[201], say $\mathcal{K} = \{ord, order, lattice, \ldots\}$,

come with a definition of the symbols that are declared to exist for κ.

In addition, if κ is an axiomatic class, such a declaration must come with a proof of the axioms.

If a type τ is (by declaration) a member of class κ, we write the "class judgement" $\tau :: \kappa$.

[201]The set $\mathcal{K}$ we gave is incomplete and just exemplary.

So the set of type classes involved in an Isabelle theory is a finite set of names (written lower-case), typically including ord, $order$, and $lattice$.

We have seen some Isabelle syntax for declaring the type classes previously.

In grammars and elsewhere, κ is the letter we use for "type class".

- a set of type constructors[202], say $\mathcal{B} = \{bool, \_ \rightarrow \_$[203]$, ind, \_\ list, \_\ set\ .$

---

[202]As before, the set $\mathcal{B}$ we gave is is incomplete (there are "...") and just exemplary. We might call $\mathcal{B}$ a type signature.
Note also that an _ is used to denote the arity of a type constructor.

- _ $list$ means that $list$ is unary type constructor;
- _ $\rightarrow$ _ means that $\rightarrow$ is a binary infix type constructor.

The notation using _ is slightly abusive since the _ is not actually part of the type constructor. _ $list$ is not a type constructor; $list$ is a type constructor.

So the set of type constructors involved in an Isabelle theory is a finite set of names (written lower-case) with each having an arity associated, typically including $bool$, $\rightarrow$, and $list$. Note however that $bool$ is fundamental (since object level predicates are modeled as functions taking terms to a Boolean), and so is $\rightarrow$, the constructor of the function space between two types.

In grammars and elsewhere, T is the letter we use for "type

- a set of constants $Const$ and a set of variables Var,

we define

- Polymorphic types[204]: $\tau ::= \alpha \mid \alpha :: \kappa \mid (\tau, \dots, \tau)\, T$

- Raw terms (as before): $e ::= x \mid c \mid (ee) \mid (\lambda x^\tau . e)$

(α is type variable, $T \in \mathcal{B}$, $\kappa \in \mathcal{K}$, $x \in Var$, $c \in Const$)

---

constructor".

[203]In $\lambda^\rightarrow$, types were built from base types using a "special symbol" $\rightarrow$.

When we generalize $\lambda^\rightarrow$ to a λ-calculus with polymorphism, this "special symbol" becomes a type constructor. However, the syntax is still special, and it is interpreted in a particular way.

[204] $\tau ::= \alpha \mid \alpha :: \kappa \mid (\tau, \dots, \tau)\, T$ (α is type variable)

is a grammar defining what polymorphic types are (syntactically). As before, τ is the non-terminal we use for (now: polymorphic) types.

This grammar is not exemplary but generic, and it deserves a closer look.

A type variable is a variable that stands for a type, as opposed to a term. We have not given a grammar for type variables, but assume that there is a countable set of type variables disjoint from the set of term variables. We use α as the non-terminal for a type variable (abusing notation, we

Type Assignment Calculus with Type Classes

Assume some syntax for declaring $\tau :: \kappa$ and $\kappa \prec \kappa'$. In addition introduce the rule

$$\frac{\tau :: \kappa \quad \kappa \prec \kappa'}{\tau :: \kappa'} \; subclass$$

Type assignment rules as before, but type substitution Θ in

$$\frac{c : \tau \in \Sigma}{\Gamma \vdash c : \tau\Theta} \; assum$$

must respect class constraints: for each $\alpha :: \kappa$ occurring in τ where $\alpha\Theta = \sigma$, judgement $\sigma :: \kappa$ must hold.

often also use α to denote an actual type variable).

First, note that a type variable may be followed by a class constraint $:: \kappa$ (recall that κ is the non-terminal for type classes). However, a type variable is not necessarily followed by such a constraint, for example if the type variable already occurs elsewhere and is constrained in that place. We have already seen this.

Moreover, a polymorphic type is obtained by preceding a type constructor with a tuple of types. The arity of the tuple must be equal to the declared arity of the type constructor.

It is not shown here that for some special type constructors, such as $\rightarrow$, the argument may also be written infix.

Example

Suppose that by virtue of declarations, we have $\mathbb{N} :: order$, $order \prec ord$, and $\leq: \alpha :: ord \rightarrow \alpha \rightarrow bool \in \Sigma$. Derive

$$\frac{\mathbb{N} :: order \quad order \prec ord}{\mathbb{N} :: ord} \; \textit{subclass}$$

and then $(\Theta = [\alpha \leftarrow \mathbb{N}])$

$$\frac{(\leq: (\alpha :: ord) \rightarrow \alpha \rightarrow bool) \in \Sigma}{\vdash \leq: \mathbb{N} \rightarrow \mathbb{N} \rightarrow bool} \; \textit{assum}$$

which respects the class constraint since the judgement $\mathbb{N} :: ord$ was derived above.

8.4 Higher-Order Unification

The λ-calculus is "the" metalogic. Hence we now (sometimes) call its variables "metavariables" for emphasis and we precede them with "?". E.g. they can stand for object-level formulae. More details later.

8.4 Higher-Order Unification

The λ-calculus is "the" metalogic. Hence we now (sometimes) call its variables "metavariables" for emphasis and we precede them with "?". E.g. they can stand for object-level formulae. More details later.

Two issues concerning metavariables are:

- suitable renamings[205] of metavariables;

- unification[206] before rule application.

---

[205]Whenever a rule is applied, the metavariables occurring in it must be renamed to fresh variables to ensure that no metavariable in the rule has been used in the proof before.

The notion fresh is often casually used in logic, and it means: this variable has never been used before. To be more precise, one should say: never been used before in the relevant context.

[206]The mechanism to instantiate metavariables as needed is called (higher-order) unification. Unification is the process of finding a substitution that makes two terms equal.

We will now see more formally what it is and later also where it is used.

What Is Higher-Order Unification?

Unification of terms e, e': find substitution θ for metavariables such that $e\theta =_{\alpha\beta\eta} e'\theta$.

   Examples[207]:

$$
\begin{aligned}
?X + ?Y &=_{\alpha\beta\eta} \; x + x \\
?P(x) &=_{\alpha\beta\eta} \; x + x \\
f(?X\,x) &=_{\alpha\beta\eta} \; ?Y\,x \\
?F(?G\,x) &=_{\alpha\beta\eta} \; f(g(x))
\end{aligned}
$$

---

[207]

  A solution for $?X + ?Y =_{\alpha\beta\eta} \; x + x$ is $[?X \leftarrow x, ?Y \leftarrow x]$.

  A solution for $?P(x) =_{\alpha\beta\eta} x + x$ is $[?P \leftarrow (\lambda y.y + y)]$.

  A solution for $f(?X x) =_{\alpha\beta\eta} ?Y\,x$ is $[?X \leftarrow (\lambda z.z), ?Y \leftarrow (\lambda z.f\,z)]$.

  Three solutions for $?F(?G\,x) =_{\alpha\beta\eta} f(g(x))$ are

$$
\begin{aligned}
&[?F \leftarrow f, \; ?G \leftarrow g], \\
&[?F \leftarrow (\lambda x.f(g\,x)), \; ?G \leftarrow (\lambda x.x)], \\
&[?F \leftarrow (\lambda x.x), \; ?G \leftarrow (\lambda x.f(g\,x))],
\end{aligned}
$$

What Is Higher-Order Unification?

Unification of terms e, e': find substitution θ for metavariables such that $e\theta =_{\alpha\beta\eta} e'\theta$.

Examples[207]:

$$
\begin{aligned}
?X + ?Y &=_{\alpha\beta\eta} \ x + x \\
?P(x) &=_{\alpha\beta\eta} \ x + x \\
f(?X\,x) &=_{\alpha\beta\eta} \ ?Y\,x \\
?F(?G\,x) &=_{\alpha\beta\eta} \ f(g(x))
\end{aligned}
$$

Why higher-order? Metavariables may be instantiated to functions, e.g. $[?P \leftarrow \lambda y.y + y]$.

₂₀₇

A solution for $?X + ?Y =_{\alpha\beta\eta} \ x + x$ is $[?X \leftarrow x, ?Y \leftarrow x]$.

A solution for $?P(x) =_{\alpha\beta\eta} x + x$ is $[?P \leftarrow (\lambda y.y + y)]$.

A solution for $f(?X\,x) =_{\alpha\beta\eta} ?Y\,x$ is $[?X \leftarrow (\lambda z.z), ?Y \leftarrow (\lambda z.f\ z)]$.

Three solutions for $?F(?G\,x) =_{\alpha\beta\eta} f(g(x))$ are

$$
\begin{aligned}
&[?F \leftarrow f,\ ?G \leftarrow g], \\
&[?F \leftarrow (\lambda x.f(g\,x)),\ ?G \leftarrow (\lambda x.x)], \\
&[?F \leftarrow (\lambda x.x),\ ?G \leftarrow (\lambda x.f(g\,x))],
\end{aligned}
$$

Higher-Order Unification: Facts

- Unification modulo[208] $\alpha\beta$ (HO-unification) is semi-decidable (in Isabelle: incomplete).

- Unification modulo $\alpha\beta\eta$ is undecidable (in Isabelle: incomplete).

---

[208]Unification of terms e, e' modulo $\alpha\beta$ means finding a substitution θ for metavariables such that $\theta(e) =_{\alpha\beta} \theta(e')$.

Likewise, unification of terms e, e' modulo $\alpha\beta\eta$ means finding a substitution σ for metavariables such that $\sigma(e) =_{\alpha\beta\eta} \sigma(e')$.

Higher-Order Unification: Facts

- Unification modulo[208] $\alpha\beta$ (HO-unification) is semi-decidable (in Isabelle: incomplete).

- Unification modulo $\alpha\beta\eta$ is undecidable (in Isabelle: incomplete).

- HO-unification is well-behaved for most practical cases.

- Important fragments (like HO-patterns) are decidable.

- HO-unification has possibly infinitely many solutions.

  We will look at some of these issues again later.

---

[208]Unification of terms e, e' modulo $\alpha\beta$ means finding a substitution θ for metavariables such that $\theta(e) =_{\alpha\beta} \theta(e')$.

Likewise, unification of terms e, e' modulo $\alpha\beta\eta$ means finding a substitution σ for metavariables such that $\sigma(e) =_{\alpha\beta\eta} \sigma(e')$.

8.5 Summary on λ-Calculus

λ-calculus is a formalism for writing functions.

β-reduction is the notion of "computing" in λ-calculus.

λ-calculus is Turing-complete.

$\lambda^{\rightarrow}$ restricts syntax to "meaningful" λ-terms.

Extension of typed λ-calculus used to represent syntax of object logics. λ-terms[209] stand for object terms/formulae, possibly containing "distinguished occurrences" of (object) variables. This will be explained thoroughly next lecture.

HO-unification important in constructing proofs.

---

[209]So just like first-order logic, the λ-calculus has a syntactic category called terms. Bit the word "term" has a different meaning for the λ-calculus than for first-order logic, and so one can say λ-term for emphasis.

Note that at this stage, we have no syntactic category called "formula" for the λ-calculus.

9 Encoding Syntax

Metatheory: Motivation

Previously, we have seen the (polymorphically) typed λ-calculus (with type classes).

Now, we will see how the typed λ-calculus can be used as a metalanguage for representing[210] the syntax of an object logic, e.g. first-order logic.

---

[210]In the following, we will distinguish between the object logic and the metalogic. We have already seen this kind of distinction before.

The object logic, or user-defined theory if you like, has a syntax and has a notion of proof. Both must be represented in the metalogic. This is what this lecture and a later lecture are about.

[211]

$\phi \in Prop$ *iff* $\ulcorner \phi \urcorner \in o$ means: The object level formula ϕ is a well-formed (according to the syntactic rules of the object logic) proposition if and only if its encoding in the metalogic, written $\ulcorner \phi \urcorner$, has type o.

Metatheory: Motivation

Previously, we have seen the (polymorphically) typed λ-calculus (with type classes).

Now, we will see how the typed λ-calculus can be used as a metalanguage for representing[210] the syntax of an object logic, e.g. first-order logic.

Idea: An object-level proposition is a meta-level term. Metalogic type o for propositions.

The terms of type o encode object level propositions: $\phi \in Prop$ iff $\ulcorner \phi \urcorner : o$[211].

---

[210]In the following, we will distinguish between the object logic and the metalogic. We have already seen this kind of distinction before.

The object logic, or user-defined theory if you like, has a syntax and has a notion of proof. Both must be represented in the metalogic. This is what this lecture and a later lecture are about.

[211]

$\phi \in Prop$ iff $\ulcorner \phi \urcorner \in o$ means: The object level formula ϕ is a well-formed (according to the syntactic rules of the object logic) proposition if and only if its encoding in the metalogic, written $\ulcorner \phi \urcorner$, has type o.

Metatheory: Motivation

Previously, we have seen the (polymorphically) typed λ-calculus (with type classes).

Now, we will see how the typed λ-calculus can be used as a metalanguage for representing[210] the syntax of an object logic, e.g. first-order logic.

Idea: An object-level proposition is a meta-level term. Metalogic type o for propositions.

The terms of type o encode object level propositions: $\phi \in Prop$ iff $\ulcorner \phi \urcorner : o$[211].

Later: How do we represent the proofs/provability?

---

[210]In the following, we will distinguish between the object logic and the metalogic. We have already seen this kind of distinction before.

The object logic, or user-defined theory if you like, has a syntax and has a notion of proof. Both must be represented in the metalogic. This is what this lecture and a later lecture are about.

[211]

$\phi \in Prop$ iff $\ulcorner \phi \urcorner \in o$ means: The object level formula ϕ is a well-formed (according to the syntactic rules of the object logic) proposition if and only if its encoding in the metalogic, written $\ulcorner \phi \urcorner$, has type o.

Why Have a Metalogic?

Why should we have a meta- or framework logic rather than implementing provers for each object logic individually?

$+$ Implement 'core'[212] only once

$+$ Shared support for automation[213]

$+$ Conceptual framework[214] for exploring what a logic is

But

$+/-$ Metalayer[215] between user and logic

$-$ Makes assumptions[216] about structure of logic

9.1 $\lambda^{\rightarrow}$: Review

[212] By the core we mean the syntax and proof rules of the metalogic. These should be simple, so that one can be reasonably confident that the implementation is correct.

[213] There are some general techniques involved in automating the search for a proof that work for various object logics. It is therefore useful to implement these techniques on a higher level, rather than considering each object logic individually.

[214] By implementing various object logics within the same metalogic, we can compare the object logics in a more formal way.

[215] Having a logic and a metalogic can be very mind-boggling. We already experienced that when working with Isabelle, it is sometimes confusing to know whether we are at the level of a particular theory, or at the level of general Isabelle syntax, or at the level of ML, the programming language that Isabelle is implemented in.

[216] Designing a metalogic is a bold endeavor.

How are we supposed to know that the metalogic is expressive enough to encode any object logic someone might

$\lambda^{\rightarrow}$ is sufficient for presentation here (no polymorphism, type classes).

- Syntax for types ($\mathcal{B}$ a set of base types, $T \in \mathcal{B}$)

$$\tau ::= T \mid \tau \rightarrow \tau$$

  Examples: $\mathbb{N}$, $\mathbb{N} \rightarrow \mathbb{N}$, $(\mathbb{N} \rightarrow \mathbb{N}) \rightarrow \mathbb{N}$, $\mathbb{N} \rightarrow \mathbb{N} \rightarrow \mathbb{N}$

- Syntax for terms: λ-calculus augmented with types

$$e ::= x \mid c \mid (ee) \mid (\lambda x^{\tau}. e)$$

$(x \in Var,\ c \in Const)$

---

invent?

There is probably no general satisfactory answer to this question.

In fact, we make assumptions that object logics are of a certain kind.

This is related to the nature of implication. Roughly speaking, we assume logics and proof systems for which the deduction theorem holds, i.e., for which $A \vdash B$ (B is derivable under assumption A) holds if and only if $\vdash A \rightarrow B$ ($A \rightarrow B$ is derivable without any assumption).

There are logics (modal, relevance logics) for which the theorem does not hold [BM00].

Type Assignment

- Signature $\Sigma ::= \langle \rangle \mid \Sigma, c : \tau.$

- Context $\Gamma ::= \langle \rangle \mid \Gamma, x : \tau.$

- Type assignment rules

$$\frac{c : \tau \in \Sigma}{\Gamma \vdash c : \tau} \; \textit{assum} \qquad \Gamma, x : \tau, \Delta \vdash x : \tau \quad \textit{hyp}$$

$$\frac{\Gamma \vdash e : \sigma \to \tau \quad \Gamma \vdash e' : \sigma}{\Gamma \vdash ee' : \tau} \; \textit{app} \qquad \frac{\Gamma, x : \sigma \vdash e : \tau}{\Gamma \vdash \lambda x^\sigma . e \; : \sigma \to \tau} \; \textit{abs}$$

9.2 Representing Syntax of Propositional Logic

Let the following fragment of propositional logic[217], which we call $Prop$, be our object logic:

$$P \ ::= \ x \ | \ \neg P \ | \ P \wedge P \ | \ P \rightarrow P$$

9.2 Representing Syntax of Propositional Logic

Let the following fragment of propositional logic[217], which we call $Prop$, be our object logic:

$$P \;::=\; x \;\mid\; \neg P \;\mid\; P \wedge P \;\mid\; P \to P$$

Let $\lambda^{\to}$[218] be our metalogic. Declare

- $\mathcal{B} = \{o\}$.

- Signature assigns types to constants[219]:

$$\Sigma = \langle not : \qquad , and : \qquad\qquad , imp : \qquad\qquad \rangle$$

9.2 Representing Syntax of Propositional Logic

Let the following fragment of propositional logic[217], which we call $Prop$, be our object logic:

$$P \;::=\; x \;\mid\; \neg P \;\mid\; P \wedge P \;\mid\; P \to P$$

Let $\lambda^{\to}$[218] be our metalogic. Declare

- $\mathcal{B} = \{o\}$.

- Signature assigns types to constants[219]:

$$\Sigma = \langle not : o \to o, and : o \to o \to o, imp : o \to o \to o \rangle$$

---

[217]We consider here the fragment of propositional logic containing the logical symbols $\neg, \wedge, \to$, and we call it $Prop$. We chose this small fragment because it is sufficient for our purposes, namely to demonstrate how encoding syntax in $\lambda^{\to}$ works. It would be trivial to adapt everything in the sequel to include $\vee$ or $\bot$.

[218]We will now show an approach to representing the syntax an object logic in $\lambda^{\to}$. This approach is called first-order syntax. For the moment, it is better not to worry about why the approach is called that. This will become clear later.

[219]Now the object/meta distinction starts becoming mind-boggling!

We declare

$$\Sigma = \langle not : o \to o, and : o \to o \to o, imp : o \to o \to o \rangle,$$

and so on the level of our metalogic $\lambda^{\to}$, not, and, and imp are constants. However, these constants represent the logical symbols of the object logic.

Note the types of the constants:

- Context assigns types to variables[220].

---

not has type $o \to o$, so it takes a proposition and returns a proposition.

and and *imp* have type $o \to o \to o$, so each takes two propositions and returns a proposition.

[220]We identify metalevel variables and object level propositional variables. Hence Γ should contain expressions of the form $a : o$, where a is a $\lambda^{\to}$ variable, representing a propositional variable. Note that under this agreement, Γ should not contain expressions like, e.g., $a : o \to o$.

Digression: Programming Languages

$\lambda^{\rightarrow}$ is the theory underlying typed functional programming. Our declaration of $\mathcal{B}$ and Σ on the previous slide corresponds to the declaration of an algebraic datatype in a functional programming language [Pau96]:

```
datatype Prop =
      VarInject of Variable  |   not of Prop
    | and of Prop * Prop     |   imp of Prop * Prop
```

Example of First-Order Syntax

$a : o \vdash imp\ (not\ a)\ a : o$[221]

---

[221]$a : o \vdash imp\ (not\ a)\ a : o$ is a judgement in $\lambda^{\rightarrow}$, which may or may not be provable.

If we set up everything correctly and if $a : o \vdash imp\ (not\ a)\ a : o$ is provable, then the judgement represents the fact $\neg a \rightarrow a$ is a proposition.

In this sense, we could then say that derivability in $\lambda^{\rightarrow}$ captures the syntax of $Prop$, i.e., it can distinguish a legal proposition from a "non-proposition".

Note that this has nothing to do with the question of whether it is a true proposition! So far, we are only talking about the representation of syntax.

Example of First-Order Syntax

$a : o \vdash imp\,(not\,a)\,a : o$[221]

$$\dfrac{a : o \vdash imp : o \to o \to o \qquad \dfrac{a : o \vdash not : o \to o \quad a : o \vdash a : o}{a : o \vdash not\,a : o}}{a : o \vdash imp\,(not\,a) : o \to o} \qquad a : o \vdash a : o}{a : o \vdash imp\,(not\,a)\,a : o}$$

Applications of *hyp* and *assum* suppressed. Otherwise always rule *app*.

---

[221]$a : o \vdash imp\,(not\,a)\,a : o$ is a judgement in $\lambda^{\to}$, which may or may not be provable.

If we set up everything correctly and if $a : o \vdash imp\,(not\,a)\,a : o$ is provable, then the judgement represents the fact $\neg a \to a$ is a proposition.

In this sense, we could then say that derivability in $\lambda^{\to}$ captures the syntax of $Prop$, i.e., it can distinguish a legal proposition from a "non-proposition".

Note that this has nothing to do with the question of whether it is a true proposition! So far, we are only talking about the representation of syntax.

Non-example of First-Order Syntax

$a : o \vdash not\ (imp\ a)\ a : o^{222}$

$^{222}a : o \vdash not\ (imp\ a)\ a : o$ is a judgement in $\lambda^{\rightarrow}$ which may or may not be provable.

If we set up everything correctly and if $a : o \vdash not\ (imp\ a)\ a : o$ is provable, then the judgement represents the fact that $(\rightarrow a)\neg a$ is a proposition.

However, you may observe that $(\rightarrow a)\neg a$ is gibberish. In fact, there is no formal sense whatsoever in saying that $not\ (imp\ a)\ a$ corresponds to $(\rightarrow a)\neg a$.

We will see that $a : o \vdash not\ (imp\ a)\ a : o$ isn't provable, and this reflects the fact that there is no proposition represented by $not\ (imp\ a)\ a$.

^{223}Generally, it is difficult to prove that a proof of a given judgement within a given proof system does not exist, since there are infinitely many possible proofs and it is not obvious to predict how big an existing proof might be.

However, under certain conditions, there are techniques for simplifying proofs. In fact, there may be normal form proofs, i.e., proofs simplified as much as possible. One can then

Non-example of First-Order Syntax

$$a : o \vdash not \; (imp \; a) \; a : o^{222}$$

$$\cfrac{a : o \vdash not : o \to o \qquad \cfrac{a : o \vdash imp : o \to o \to o \qquad a : o \vdash a : o}{a : o \vdash imp \; a : o \to o}}{???}$$

---

[222] $a : o \vdash not \; (imp \; a) \; a : o$ is a judgement in $\lambda^{\to}$ which may or may not be provable.

If we set up everything correctly and if $a : o \vdash not \; (imp \; a) \; a : o$ is provable, then the judgement represents the fact that $(\to a) \neg a$ is a proposition.

However, you may observe that $(\to \; a) \neg a$ is gibberish. In fact, there is no formal sense whatsoever in saying that $not \; (imp \; a) \; a$ corresponds to $(\to a) \neg a$.

We will see that $a : o \vdash not \; (imp \; a) \; a : o$ isn't provable, and this reflects the fact that there is no proposition represented by $not \; (imp \; a) \; a$.

[223] Generally, it is difficult to prove that a proof of a given judgement within a given proof system does not exist, since there are infinitely many possible proofs and it is not obvious to predict how big an existing proof might be.

However, under certain conditions, there are techniques for simplifying proofs. In fact, there may be normal form proofs, i.e., proofs simplified as much as possible. One can then

Non-example of First-Order Syntax

$a : o \vdash not \; (imp \; a) \; a : o$[222]

$$
\cfrac{a : o \vdash not : o \to o \qquad \cfrac{a : o \vdash imp : o \to o \to o \quad a : o \vdash a : o}{a : o \vdash imp \; a : o \to o}}{???}
$$

No proof possible! (Requires analysis of normal forms[223].)

---

[222] $a : o \vdash not \; (imp \; a) \; a : o$ is a judgement in $\lambda^{\to}$ which may or may not be provable.

If we set up everything correctly and if $a : o \vdash not \; (imp \; a) \; a : o$ is provable, then the judgement represents the fact that $(\to a)\neg a$ is a proposition.

However, you may observe that $(\to \; a)\neg a$ is gibberish. In fact, there is no formal sense whatsoever in saying that $not \; (imp \; a) \; a$ corresponds to $(\to a)\neg a$.

We will see that $a : o \vdash not \; (imp \; a) \; a : o$ isn't provable, and this reflects the fact that there is no proposition represented by $not \; (imp \; a) \; a$.

[223] Generally, it is difficult to prove that a proof of a given judgement within a given proof system does not exist, since there are infinitely many possible proofs and it is not obvious to predict how big an existing proof might be.

However, under certain conditions, there are techniques for simplifying proofs. In fact, there may be normal form proofs, i.e., proofs simplified as much as possible. One can then

Bijection between $Prop$ and o

We desire bijection[224] $\ulcorner \cdot \urcorner : Prop \rightarrow o$ that is

- adequate: each proposition in $Prop$ can be represented by a $\lambda^{\rightarrow}$-term of type o:

$$\text{If } P \in Prop \text{ then } \Gamma \vdash \ulcorner P \urcorner : o$$

---

argue: if a proof of a certain judgement exists, it must be no bigger than a certain size. By searching through all proofs smaller than this size, one can prove that no proof exists.

In this lecture, we do not go into the details of this topic [GLT89, Pra65].

[224]In general mathematical terminology, a bijection between A and B is a mapping $f : A \rightarrow B$ such that for all $a, a' \in A$, where $a \neq a'$, we have $f(a) \neq f(a')$, and for each $b \in B$, there exists an $a \in A$ such that $f(a) = b$.

For a bijection f, the inverse f^{-1} is always defined, and we have $f(f^{-1}(b)) = b$ for all $b \in B$ and $f^{-1}(f(a)) = a$ for all $a \in A$.

Bijection between $Prop$ and o

We desire bijection[224] $\ulcorner \cdot \urcorner : Prop \to o$ that is

- adequate: each proposition in $Prop$ can be represented by a $\lambda^{\to}$-term of type o:

$$\text{If } P \in Prop \text{ then } \Gamma \vdash \ulcorner P \urcorner : o$$

- faithful: each $\lambda^{\to}$ term of type o represents a proposition in $Prop$:

$$\text{If } \Gamma \vdash t : o \text{ then } \ulcorner t \urcorner^{-1} \in Prop$$

---

argue: if a proof of a certain judgement exists, it must be no bigger than a certain size. By searching through all proofs smaller than this size, one can prove that no proof exists.

In this lecture, we do not go into the details of this topic [GLT89, Pra65].

[224]In general mathematical terminology, a bijection between A and B is a mapping $f : A \to B$ such that for all $a, a' \in A$, where $a \neq a'$, we have $f(a) \neq f(a')$, and for each $b \in B$, there exists an $a \in A$ such that $f(a) = b$.

For a bijection f, the inverse f^{-1} is always defined, and we have $f(f^{-1}(b)) = b$ for all $b \in B$ and $f^{-1}(f(a)) = a$ for all $a \in A$.

Adequacy of Bijection

Example: $(\neg a) \to b \in Prop$ therefore $imp \ (not \ a) \ b : o$

---

[225]If $P \in Prop$, and if for each propositional variable x in P, we have $x : o \in \Gamma$, then $\Gamma \vdash \ulcorner P \urcorner : o$.

Proof: By structural induction on $Prop$.

Base case: P is a propositional variable.
Then $\ulcorner P \urcorner = P$, and so if $P : o \in \Gamma$, then we have $\Gamma \vdash \ulcorner P \urcorner : o$ by rule *hyp*.

Induction step: Suppose the claim holds for $P \in Prop$ and $Q \in Prop$.

Consider the propositional formula $\neg P$. We have $\ulcorner \neg P \urcorner = not \ \ulcorner P \urcorner$. Assume that for each propositional variable x in P, we have $x : o \in \Gamma$. By the induction hypothesis, $\Gamma \vdash \ulcorner P \urcorner : o$. Moreover $\Gamma \vdash not : o \to o$ by rule *assum*, and so $\Gamma \vdash not \ \ulcorner P \urcorner : o$ by rule *app*.

Now consider the propositional formula $P \wedge Q$. We have $\ulcorner P \wedge Q \urcorner = and \ \ulcorner P \urcorner \ \ulcorner Q \urcorner$. Assume that for each propositional variable x in P or Q, we have $x : o \in \Gamma$. By the induction hypothesis, $\Gamma \vdash \ulcorner P \urcorner : o$ and $\Gamma \vdash \ulcorner Q \urcorner : o$. Moreover $\Gamma \vdash and : o \to o \to o$ by rule *assum*, and so

Adequacy of Bijection

Example: $(\neg a) \to b \in Prop$ therefore $imp\ (not\ a)\ b : o$

Formalize mapping $\ulcorner \cdot \urcorner$:

$$
\begin{aligned}
\ulcorner x \urcorner &= x && \text{for } x \text{ a variable} \\
\ulcorner \neg P \urcorner &= not\ \ulcorner P \urcorner \\
\ulcorner P \wedge Q \urcorner &= and\ \ulcorner P \urcorner\ \ulcorner Q \urcorner \\
\ulcorner P \to Q \urcorner &= imp\ \ulcorner P \urcorner\ \ulcorner Q \urcorner
\end{aligned}
$$

---

[225]If $P \in Prop$, and if for each propositional variable x in P, we have $x : o \in \Gamma$, then $\Gamma \vdash \ulcorner P \urcorner : o$.

Proof: By structural induction on $Prop$.

Base case: P is a propositional variable. Then $\ulcorner P \urcorner = P$, and so if $P : o \in \Gamma$, then we have $\Gamma \vdash \ulcorner P \urcorner : o$ by rule *hyp*.

Induction step: Suppose the claim holds for $P \in Prop$ and $Q \in Prop$.

Consider the propositional formula $\neg P$. We have $\ulcorner \neg P \urcorner = not\ \ulcorner P \urcorner$. Assume that for each propositional variable x in P, we have $x : o \in \Gamma$. By the induction hypothesis, $\Gamma \vdash \ulcorner P \urcorner : o$. Moreover $\Gamma \vdash not : o \to o$ by rule *assum*, and so $\Gamma \vdash not\ \ulcorner P \urcorner : o$ by rule *app*.

Now consider the propositional formula $P \wedge Q$. We have $\ulcorner P \wedge Q \urcorner = and\ \ulcorner P \urcorner\ \ulcorner Q \urcorner$. Assume that for each propositional variable x in P or Q, we have $x : o \in \Gamma$. By the induction hypothesis, $\Gamma \vdash \ulcorner P \urcorner : o$ and $\Gamma \vdash \ulcorner Q \urcorner : o$. Moreover $\Gamma \vdash and : o \to o \to o$ by rule *assum*, and so

Adequacy of Bijection

Example: $(\neg a) \to b \in Prop$ therefore $imp\ (not\ a)\ b : o$

Formalize mapping $\ulcorner \cdot \urcorner$:

$$
\begin{aligned}
\ulcorner x \urcorner &= x && \text{for } x \text{ a variable} \\
\ulcorner \neg P \urcorner &= not\ \ulcorner P \urcorner \\
\ulcorner P \wedge Q \urcorner &= and\ \ulcorner P \urcorner \ulcorner Q \urcorner \\
\ulcorner P \to Q \urcorner &= imp\ \ulcorner P \urcorner \ulcorner Q \urcorner
\end{aligned}
$$

Formal statement accounts for variables:

If $P \in Prop$, and if for each propositional variable x in P, we have $x : o \in \Gamma$, then $\Gamma \vdash \ulcorner P \urcorner : o$.

---

[225]If $P \in Prop$, and if for each propositional variable x in P, we have $x : o \in \Gamma$, then $\Gamma \vdash \ulcorner P \urcorner : o$.

Proof: By structural induction on $Prop$.

Base case: P is a propositional variable. Then $\ulcorner P \urcorner = P$, and so if $P : o \in \Gamma$, then we have $\Gamma \vdash \ulcorner P \urcorner : o$ by rule *hyp*.

Induction step: Suppose the claim holds for $P \in Prop$ and $Q \in Prop$.

Consider the propositional formula $\neg P$. We have $\ulcorner \neg P \urcorner = not\ \ulcorner P \urcorner$. Assume that for each propositional variable x in P, we have $x : o \in \Gamma$. By the induction hypothesis, $\Gamma \vdash \ulcorner P \urcorner : o$. Moreover $\Gamma \vdash not : o \to o$ by rule *assum*, and so $\Gamma \vdash not\ \ulcorner P \urcorner : o$ by rule *app*.

Now consider the propositional formula $P \wedge Q$. We have $\ulcorner P \wedge Q \urcorner = and\ \ulcorner P \urcorner \ulcorner Q \urcorner$. Assume that for each propositional variable x in P or Q, we have $x : o \in \Gamma$. By the induction hypothesis, $\Gamma \vdash \ulcorner P \urcorner : o$ and $\Gamma \vdash \ulcorner Q \urcorner : o$. Moreover $\Gamma \vdash and : o \to o \to o$ by rule *assum*, and so

Adequacy of Bijection

Example: $(\neg a) \to b \in Prop$ therefore $imp\ (not\ a)\ b : o$

Formalize mapping $\ulcorner \cdot \urcorner$:

$$
\begin{aligned}
\ulcorner x \urcorner &= x & \text{for } x \text{ a variable} \\
\ulcorner \neg P \urcorner &= not\ \ulcorner P \urcorner \\
\ulcorner P \wedge Q \urcorner &= and\ \ulcorner P \urcorner \ulcorner Q \urcorner \\
\ulcorner P \to Q \urcorner &= imp\ \ulcorner P \urcorner \ulcorner Q \urcorner
\end{aligned}
$$

Formal statement accounts for variables:

If $P \in Prop$, and if for each propositional variable x in P, we have $x : o \in \Gamma$, then $\Gamma \vdash \ulcorner P \urcorner : o$. Proof by induction[225].

---

[225]If $P \in Prop$, and if for each propositional variable x in P, we have $x : o \in \Gamma$, then $\Gamma \vdash \ulcorner P \urcorner : o$.

Proof: By structural induction on $Prop$.

Base case: P is a propositional variable. Then $\ulcorner P \urcorner = P$, and so if $P : o \in \Gamma$, then we have $\Gamma \vdash \ulcorner P \urcorner : o$ by rule *hyp*.

Induction step: Suppose the claim holds for $P \in Prop$ and $Q \in Prop$.

Consider the propositional formula $\neg P$. We have $\ulcorner \neg P \urcorner = not\ \ulcorner P \urcorner$. Assume that for each propositional variable x in P, we have $x : o \in \Gamma$. By the induction hypothesis, $\Gamma \vdash \ulcorner P \urcorner : o$. Moreover $\Gamma \vdash not : o \to o$ by rule *assum*, and so $\Gamma \vdash not\ \ulcorner P \urcorner : o$ by rule *app*.

Now consider the propositional formula $P \wedge Q$. We have $\ulcorner P \wedge Q \urcorner = and\ \ulcorner P \urcorner \ulcorner Q \urcorner$. Assume that for each propositional variable x in P or Q, we have $x : o \in \Gamma$. By the induction hypothesis, $\Gamma \vdash \ulcorner P \urcorner : o$ and $\Gamma \vdash \ulcorner Q \urcorner : o$. Moreover $\Gamma \vdash and : o \to o \to o$ by rule *assum*, and so

Faithfulness of Bijection

Define $\ulcorner \cdot \urcorner^{-1}$

$$
\begin{aligned}
\ulcorner x \urcorner^{-1} &= x &&\text{for } x \text{ a variable} \\
\ulcorner not\ P \urcorner^{-1} &= \neg \ulcorner P \urcorner^{-1} \\
\ulcorner and\ P\ Q \urcorner^{-1} &= \ulcorner P \urcorner^{-1} \wedge \ulcorner Q \urcorner^{-1} \\
\ulcorner imp\ P\ Q \urcorner^{-1} &= \ulcorner P \urcorner^{-1} \rightarrow \ulcorner Q \urcorner^{-1}
\end{aligned}
$$

$\Gamma \vdash and\ \ulcorner P \urcorner\ \ulcorner Q \urcorner : o$ by two applications of rule *app*.

The case $P \rightarrow Q$ is completely analogous.

[226]By the definition of $Prop$ and the definition of $\ulcorner \cdot \urcorner$, it is clear that $\ulcorner P \urcorner$ is defined for all $P \in Prop$. It is very easy to show by induction on $Prop$ that $\ulcorner \ulcorner P \urcorner \urcorner^{-1} = P$.

Here is an example of a proof by induction on $Prop$.

Obviously, everything we say here depends on the particular fragment of propositional logic, but in an inessential way. It would be trivial to adapt to other fragments.

Faithfulness of Bijection

Define $\ulcorner \cdot \urcorner^{-1}$

$$\ulcorner x \urcorner^{-1} \;=\; x \qquad\qquad\qquad \text{for } x \text{ a variable}$$
$$\ulcorner not\ P \urcorner^{-1} \;=\; \neg \ulcorner P \urcorner^{-1}$$
$$\ulcorner and\ P\ Q \urcorner^{-1} \;=\; \ulcorner P \urcorner^{-1} \wedge \ulcorner Q \urcorner^{-1}$$
$$\ulcorner imp\ P\ Q \urcorner^{-1} \;=\; \ulcorner P \urcorner^{-1} \rightarrow \ulcorner Q \urcorner^{-1}$$

For bijection, should have $\ulcorner \ulcorner P \urcorner \urcorner^{-1} = P$ and $\ulcorner \ulcorner t \urcorner^{-1} \urcorner = t$.
Former is trivial[226], but what about latter?

$\Gamma \vdash and\ \ulcorner P \urcorner\ \ulcorner Q \urcorner : o$ by two applications of rule *app*.
    The case $P \rightarrow Q$ is completely analogous.

[226]By the definition of $Prop$ and the definition of $\ulcorner \cdot \urcorner$, it is clear that $\ulcorner P \urcorner$ is defined for all $P \in Prop$. It is very easy to show by induction on $Prop$ that $\ulcorner \ulcorner P \urcorner \urcorner^{-1} = P$.
    Here is an example of a proof by induction on $Prop$.
    Obviously, everything we say here depends on the particular fragment of propositional logic, but in an inessential way. It would be trivial to adapt to other fragments.

$\ulcorner t \urcorner^{-1}$ Is not Total

Example: For $t = not\,((\lambda x^o.\,x)a)$, we have $a : o \vdash t : o$

$$
\cfrac{a : o \vdash not : o \to o \qquad \cfrac{\cfrac{\cfrac{a : o, x : o \vdash x : o}{a : o \vdash \lambda x^o.\,x : o \to o}\;abs \qquad a : o \vdash a : o}{a : o \vdash (\lambda x^o.\,x)\,a : o}\;app}{}}{a : o \vdash not\,((\lambda x^o.\,x)\,a) : o}\;app
$$

But $\ulcorner t \urcorner^{-1}$ is undefined!

Normal Forms

If $t : o$, then there exists a t' such that $t =_{\beta\eta} t'$, where $t' : o$ and t' is in canonical ($\beta\eta$-long) normal[227] form, e.g.

$$
\begin{aligned}
not\,((\lambda x^o.\,x)\,a) &=_{\beta\eta} & not\,a \\
not &=_{\beta\eta} & \lambda x^o.\,not\,x \\
imp\,(not\,((\lambda x^o.\,x)\,a)) &=_{\beta\eta} & \lambda x^o.\,imp\,(not\,a)\,x
\end{aligned}
$$

---

[227]

A canonical $\beta\eta$-long normal form of a λ-term is obtained by applying first β-reduction as long as possible, and then computing the maximal η-expansion.

You may wonder: Why is there such a thing as a maximal η-expansion? Can't I expand a λ-term to $\lambda x_1 \ldots x_n.\,M\,x_1 \ldots x_n$ for arbitrary n? In the untyped λ-calculus, this is indeed the case. But in the typed λ-calculus, the answer is no! Consider this example:

not can be expanded to $\lambda x.\,not\,x$ since not is of function type: it has type $o \to o$. Therefore, $not\,x$ can be assigned a type, which is an intermediate step in typing $\lambda x.\,not\,x$:

$$
\cfrac{\cfrac{\Gamma, x : o \vdash not : o \to o \quad \Gamma, x : o \vdash x : o}{\Gamma, x : o \vdash not\,x : o}\ app}{\Gamma \vdash \lambda x.\,not\,x : o \to o}\ abs
$$

But we cannot, say, expand not to $\lambda xy.\,not\,x\,y$ since it is impossible to assign a type to $not\,x\,y$.

Bijection Theorem

The encoding $\ulcorner \cdot \urcorner$ is a bijection between propositional formulae with variables in Γ[228] and canonical terms t', where $\Gamma \vdash t' : o$.

Effectively, when a term of type $\tau_1 \rightarrow \tau_n \rightarrow \tau$ is η-expanded, it will have the form $\lambda x_1 x_2 \ldots x_n.e$.

Normal forms are unique.

[228]Saying that a propositional formula has variables in Γ is an abuse of terminology, i.e., it isn't exactly true, but it is trusted that the reader can guess the exact formulation.

What we mean is: a propositional formula such that for each propositional variable x occurring in the formula, we have $x : o \in \Gamma$.

[229]What this picture says is that if the left hand side is a fragment from a proof tree, deriving the judgement $\vdash (\lambda x^\sigma. e)e' : \tau$, then there exists a proof of the judgement $\vdash e[x \leftarrow e'] : \tau$.

Be aware however that our argument here is very sketchy. We do not go into the details in this course.

[230]Simply writing $t : o$ is again a bit sloppy. We should write: $\Gamma \vdash t : o$ for some Γ containing only expressions of the form $x : o$, where x is a propositional variable in $Prop$.

Bijection Theorem

The encoding $\ulcorner \cdot \urcorner$ is a bijection between propositional formulae with variables in Γ[228] and canonical terms t', where $\Gamma \vdash t' : o$.

Proof: Based on normalization

$$\cfrac{\cfrac{x : \sigma \vdash e : \tau}{\vdash \lambda x^\sigma . e \; : \sigma \to \tau} \; abs \qquad \vdash e' : \sigma}{\vdash (\lambda x^\sigma . e)e' \; : \tau} \; app \qquad\qquad \Rightarrow^{229} \qquad \vdash e[x \leftarrow e'] : \tau$$

Effectively, when a term of type $\tau_1 \to \tau_n \to \tau$ is η-expanded, it will have the form $\lambda x_1 x_2 \ldots x_n.e$.

Normal forms are unique.

[228]Saying that a propositional formula has variables in Γ is an abuse of terminology, i.e., it isn't exactly true, but it is trusted that the reader can guess the exact formulation.

What we mean is: a propositional formula such that for each propositional variable x occurring in the formula, we have $x : o \in \Gamma$.

[229]What this picture says is that if the left hand side is a fragment from a proof tree, deriving the judgement $\vdash (\lambda x^\sigma . e)e' : \tau$, then there exists a proof of the judgement $\vdash e[x \leftarrow e'] : \tau$.

Be aware however that our argument here is very sketchy. We do not go into the details in this course.

[230]Simply writing $t : o$ is again a bit sloppy. We should write: $\Gamma \vdash t : o$ for some Γ containing only expressions of the form $x : o$, where x is a propositional variable in $Prop$.

Bijection Theorem

The encoding $\ulcorner \cdot \urcorner$ is a bijection between propositional formulae with variables in Γ[228] and canonical terms t', where $\Gamma \vdash t' : o$.

   Proof: Based on normalization

$$\dfrac{\dfrac{x : \sigma \vdash e : \tau}{\vdash \lambda x^\sigma . e \; : \sigma \to \tau} \; abs \quad \vdash e' : \sigma}{\vdash (\lambda x^\sigma . e) e' \; : \tau} \; app \qquad \Rightarrow^{229} \quad \vdash e[x \leftarrow e'] : \tau$$

Corollary: If $t : o$[230] then $t =_{\beta\eta} t'$ and $\ulcorner t' \urcorner^{-1} \in Prop$ for some canonical t'.

   Effectively, when a term of type $\tau_1 \to \tau_n \to \tau$ is η-expanded, it will have the form $\lambda x_1 x_2 \ldots x_n . e$.

   Normal forms are unique.

   [228]Saying that a propositional formula has variables in Γ is an abuse of terminology, i.e., it isn't exactly true, but it is trusted that the reader can guess the exact formulation.

   What we mean is: a propositional formula such that for each propositional variable x occurring in the formula, we have $x : o \in \Gamma$.

   [229]What this picture says is that if the left hand side is a fragment from a proof tree, deriving the judgement $\vdash (\lambda x^\sigma . e) e' : \tau$, then there exists a proof of the judgement $\vdash e[x \leftarrow e'] : \tau$.

   Be aware however that our argument here is very sketchy. We do not go into the details in this course.

   [230]Simply writing $t : o$ is again a bit sloppy. We should write: $\Gamma \vdash t : o$ for some Γ containing only expressions of the form $x : o$, where x is a propositional variable in $Prop$.

9.3 Representing Syntax of First-Order Logic

In $Prop$, we only have the syntactic category of formulae (propositions), represented in $\lambda^{\rightarrow}$ by the type o.

9.3 Representing Syntax of First-Order Logic

In $Prop$, we only have the syntactic category of formulae (propositions), represented in $\lambda^{\rightarrow}$ by the type o.

In first-order[231] logic, we also have the syntactic category of terms. For representation in $\lambda^{\rightarrow}$, we now introduce type i, so $\mathcal{B} = \{i, o\}$.

9.3 Representing Syntax of First-Order Logic

In $Prop$, we only have the syntactic category of formulae (propositions), represented in $\lambda^\rightarrow$ by the type o.

In first-order[231] logic, we also have the syntactic category of terms. For representation in $\lambda^\rightarrow$, we now introduce type i, so $\mathcal{B} = \{i, o\}$.

Just like $\Gamma \vdash a : o$ means that a represents a proposition, $\Gamma \vdash t : i$ means that t represents a term.

---

[231]In the previous section, we have seen how we can use first-order syntax (of $\lambda^\rightarrow$) to represent the syntax of an object logic, then $Prop$. We haven't really understood yet why we speak of first-order syntax, but note that the notion "first-order" refers to $\lambda^\rightarrow$, i.e., the metalevel.

We will now consider first-order logic as object language. So we will now attempt to represent the syntax of first-order logic (the object language) using first-order $\lambda^\rightarrow$ syntax (the metalanguage). To avoid confusion, it is best to imagine that it is a mere coincidence that both the object and the metalanguage are described as "first-order". Of course there are reasons why both languages are called like that, but it is best to understand this separately for both levels. We will come back to this.

Example: First-Order Arithmetic (FOA)

Following fragment of FOA is our object level language[232]:

$$\text{Terms} \quad T \ ::= \ x \ | \ 0 \ | \ s^{233} \, T \ | \ T + T \ | \ T \times T$$
$$\text{Formulae} \ F \ ::= \ T = T \ | \ \neg F \ | \ F \wedge F \ | \ F \rightarrow F$$

Example: First-Order Arithmetic (FOA)

Following fragment of FOA is our object level language[232]:

$$\text{Terms} \quad T \quad ::= \quad x \mid 0 \mid s^{233}\, T \mid T + T \mid T \times T$$
$$\text{Formulae} \quad F \quad ::= \quad T = T \mid \neg F \mid F \wedge F \mid F \rightarrow F$$

In $\lambda^{\rightarrow}$ (on metalevel), define signature $\Sigma = \Sigma_{\mathcal{F}}^{234} \cup \Sigma_{\mathcal{P}} \cup \Sigma_{\mathcal{C}}$:

$$\Sigma_{\mathcal{F}} = \langle zero : i,\ succ : i \rightarrow i,\ plus : i \rightarrow i \rightarrow i,$$
$$times : i \rightarrow i \rightarrow i \rangle$$
$$\Sigma_{\mathcal{P}} = \langle eq : i \rightarrow i \rightarrow o \rangle$$
$$\Sigma_{\mathcal{C}} = \langle not : o \rightarrow o,\ and : o \rightarrow o \rightarrow o,\ imp : o \rightarrow o \rightarrow o \rangle$$

---

[232]With this grammar, we specify a certain language of a fragment (since quantifiers, $\vee$, and $\bot$ are missing) of first-order logic.

Alternatively, we could say that $\mathcal{F} = \{0, s, +, \times\}$ and $\mathcal{P} = \{=\}$. However, the way we defined first-order logic, the language thus obtained would also include quantifiers, $\vee$, and $\bot$. For the moment we want to restrict ourselves to the fragment given by the grammar for FOA.

[233]s is a unary prefix function, so s applied to T is written $s\, T$.

[234]We have defined

$$\Sigma_{\mathcal{F}} = \langle zero : i,\ succ : i \rightarrow i,\ plus : i \rightarrow i \rightarrow i,\ times : i \rightarrow i \rightarrow i \rangle$$
$$\Sigma_{\mathcal{P}} = \langle eq : i \rightarrow i \rightarrow o \rangle$$

Example: First-Order Arithmetic (FOA)

Following fragment of FOA is our object level language[232]:

$$\text{Terms} \quad T \quad ::= \quad x \mid 0 \mid s^{233}\, T \mid T + T \mid T \times T$$
$$\text{Formulae} \quad F \quad ::= \quad T = T \mid \neg F \mid F \wedge F \mid F \to F$$

In $\lambda^{\to}$ (on metalevel), define signature $\Sigma = \Sigma_{\mathcal{F}}^{234} \cup \Sigma_{\mathcal{P}} \cup \Sigma_{\mathcal{C}}$:

$$\Sigma_{\mathcal{F}} = \langle zero : i,\ succ : i \to i,\ plus : i \to i \to i,$$
$$times : i \to i \to i \rangle$$
$$\Sigma_{\mathcal{P}} = \langle eq : i \to i \to o \rangle$$
$$\Sigma_{\mathcal{C}} = \langle not : o \to o,\ and : o \to o \to o,\ imp : o \to o \to o \rangle$$

---

[232]With this grammar, we specify a certain language of a fragment (since quantifiers, $\vee$, and $\perp$ are missing) of first-order logic.

Alternatively, we could say that $\mathcal{F} = \{0, s, +, \times\}$ and $\mathcal{P} = \{=\}$. However, the way we defined first-order logic, the language thus obtained would also include quantifiers, $\vee$, and $\perp$. For the moment we want to restrict ourselves to the fragment given by the grammar for FOA.

[233]s is a unary prefix function, so s applied to T is written $s\, T$.

[234]We have defined

$$\Sigma_{\mathcal{F}} = \langle zero : i,\ succ : i \to i,\ plus : i \to i \to i,\ times : i \to i \to i \rangle$$
$$\Sigma_{\mathcal{P}} = \langle eq : i \to i \to o \rangle$$

$zero : i$ means: viewed on the object level, 0 is a term. $plus : i \to i \to i$ means: viewed on the object level, $plus$ is a function that takes two terms and returns a term. $eq : i \to i \to o$ means: viewed on the object level, $=$ is a predicate that takes two terms and returns a proposition.

Example: $\ulcorner x + s\,0 \urcorner^{235} =$

On the metalevel (level of $\lambda^{\rightarrow}$), $zero$, $plus$ and eq are constants. Note that we could also formalize them as variables.

Recall that we encoded the non-logical symbols of an object logic as constants. It would however be possible to set up the encoding in such a way that the non-logical symbols are encoded as variables, so we would have a context $\Gamma_{\mathcal{F}} \cup \Gamma_{\mathcal{P}}$ and instead of our $\Sigma_{\mathcal{F}} \cup \Sigma_{\mathcal{P}}$. This is in line with Perlis' epigram. We will sometimes take this approach in the exercises as the encoding of $\lambda^{\rightarrow}$ in Isabelle makes it more straightforward to play around with different Γ's than with different Σ's.

[235]We extend the definition of $\ulcorner \cdot \urcorner$ as follows:

$$\begin{aligned}
\ulcorner x \urcorner &= x \\
\ulcorner 0 \urcorner &= zero \\
\ulcorner s\,t \urcorner &= succ\,\ulcorner t \urcorner \\
\ulcorner r + t \urcorner &= plus\,\ulcorner r \urcorner \ulcorner t \urcorner \\
\ulcorner r \times t \urcorner &= times\,\ulcorner r \urcorner \ulcorner t \urcorner
\end{aligned}$$

Example: $\ulcorner x + s\,0 \urcorner^{235} = plus\ x\ (succ\ zero)$.

On the metalevel (level of $\lambda^\rightarrow$), $zero$, $plus$ and eq are constants. Note that we could also formalize them as variables.

Recall that we encoded the non-logical symbols of an object logic as constants. It would however be possible to set up the encoding in such a way that the non-logical symbols are encoded as variables, so we would have a context $\Gamma_\mathcal{F} \cup \Gamma_\mathcal{P}$ and instead of our $\Sigma_\mathcal{F} \cup \Sigma_\mathcal{P}$. This is in line with Perlis' epigram. We will sometimes take this approach in the exercises as the encoding of $\lambda^\rightarrow$ in Isabelle makes it more straightforward to play around with different Γ's than with different Σ's.

[235]We extend the definition of $\ulcorner \cdot \urcorner$ as follows:

$$\ulcorner x \urcorner \;=\; x$$
$$\ulcorner 0 \urcorner \;=\; zero$$
$$\ulcorner s\ t \urcorner \;=\; succ\ \ulcorner t \urcorner$$
$$\ulcorner r + t \urcorner \;=\; plus\ \ulcorner r \urcorner\ \ulcorner t \urcorner$$
$$\ulcorner r \times t \urcorner \;=\; times\ \ulcorner r \urcorner\ \ulcorner t \urcorner$$

Encoding FOL in General

In general, to encode some first-order language, we must define $\Sigma_{\mathcal{F}}$ and $\Sigma_{\mathcal{P}}$ so that for each n-ary $f \in \mathcal{F}$, $p \in \mathcal{P}$

$$f_{enc} : \underbrace{i \to \ldots \to i}_{n \text{ times}} \to i \ \in \ \Sigma_{\mathcal{F}},$$

$$p_{enc} : \underbrace{i \to \ldots \to i}_{n \text{ times}} \to o \ \in \ \Sigma_{\mathcal{P}},$$

and then $\ulcorner f(t_1, \ldots t_n) \urcorner = f_{enc} \ulcorner t_1 \urcorner \ldots \ulcorner t_n \urcorner$ and $\ulcorner p(t_1, \ldots t_n) \urcorner = p_{enc} \ulcorner t_1 \urcorner \ldots \ulcorner t_n \urcorner$.

Abusing notation, we might skip the subscript enc.

Note that here, on the object level, x is a first-order variable (a variable is a term), and hence on the metalevel, it has type i.

Quantifiers in First-Order Syntax

Along the same lines, one might suggest

$$all : var \to o \to o, \qquad \text{so} \quad \ulcorner \forall x.\, P \urcorner = all\ x\ \ulcorner P \urcorner$$

But this approach has some problems:

---

[236]In first-order logic, variables are not a syntactic category of their own, but rather they are a "sub-category" of terms. Therefore one should expect that var should be a "subtype" of i, that is to say, every term of type var is automatically also of type i. However, there is no such notion in $\lambda^{\to}$.

[237]There is a notion of substitution in $\lambda^{\to}$, hence on the metalevel. But all is just a constant like any other on the level of $\lambda^{\to}$, and hence $(and\ (p\ x)(all\ x\ (q\ x)))[x \leftarrow a] = (and\ (p\ a)(all\ a\ (q\ a)))$, and not $(and\ (p\ a)(all\ x\ (q\ x)))$ as one should expect.

That is to say, the standard operation of substitution, which exists on the metalevel, is of no use for implementing substitution on the object level. Instead, substitution on the object level must be "programmed explicitly".

Note that the following question arises: on the $\lambda^{\to}$ level, should the terms of type var be variables or constants?

One could imagine that they are variables. This means that

Quantifiers in First-Order Syntax

Along the same lines, one might suggest

$$all : var \to o \to o, \qquad \text{so} \quad \ulcorner \forall x.\, P \urcorner = all\ x\ \ulcorner P \urcorner$$

But this approach has some problems:

- Variables are also terms, so "$var \subseteq i$"[236]? No subtyping!

---

[236]In first-order logic, variables are not a syntactic category of their own, but rather they are a "sub-category" of terms. Therefore one should expect that var should be a "subtype" of i, that is to say, every term of type var is automatically also of type i. However, there is no such notion in $\lambda^\to$.

[237]There is a notion of substitution in $\lambda^\to$, hence on the metalevel. But all is just a constant like any other on the level of $\lambda^\to$, and hence $(and\ (p\ x)(all\ x\ (q\ x)))[x \leftarrow a] = (and\ (p\ a)(all\ a\ (q\ a)))$, and not $(and\ (p\ a)(all\ x\ (q\ x)))$ as one should expect.

That is to say, the standard operation of substitution, which exists on the metalevel, is of no use for implementing substitution on the object level. Instead, substitution on the object level must be "programmed explicitly".

Note that the following question arises: on the $\lambda^\to$ level, should the terms of type var be variables or constants?

One could imagine that they are variables. This means that

Quantifiers in First-Order Syntax

Along the same lines, one might suggest

$$all : var \to o \to o, \qquad \text{so} \quad \ulcorner \forall x.\, P \urcorner = all\ x\ \ulcorner P \urcorner$$

But this approach has some problems:

- Variables are also terms, so "$var \subseteq i$" [236]? No subtyping!

- all is not a binding operator in $\lambda^{\to}$. E.g., $(p(x) \wedge \forall x.\, q(x))[x \leftarrow a]$ cannot be modeled[237] as $(and\ (p\ x)(all\ x\ (q\ x)))[x \leftarrow a]$.

---

[236]In first-order logic, variables are not a syntactic category of their own, but rather they are a "sub-category" of terms. Therefore one should expect that var should be a "subtype" of i, that is to say, every term of type var is automatically also of type i. However, there is no such notion in $\lambda^{\to}$.

[237]There is a notion of substitution in $\lambda^{\to}$, hence on the metalevel. But all is just a constant like any other on the level of $\lambda^{\to}$, and hence $(and\ (p\ x)(all\ x\ (q\ x)))[x \leftarrow a] = (and\ (p\ a)(all\ a\ (q\ a)))$, and not $(and\ (p\ a)(all\ x\ (q\ x)))$ as one should expect.

That is to say, the standard operation of substitution, which exists on the metalevel, is of no use for implementing substitution on the object level. Instead, substitution on the object level must be "programmed explicitly".

Note that the following question arises: on the $\lambda^{\to}$ level, should the terms of type var be variables or constants?

One could imagine that they are variables. This means that

9.4 Higher-Order Abstract Syntax (HOAS)

Example, full FOA: $F ::= \ldots \forall x.\, A \mid \exists x.\, A$ $\Sigma = \Sigma_{\mathcal{F}} \cup \Sigma_{\mathcal{P}} \cup \Sigma_{\mathcal{C}} \cup \Sigma_{\mathcal{Q}}$:

$$\Sigma_{\mathcal{Q}} = \langle all : (i \to o^{238}) \to o,\ exists : (i \to o) \to o \rangle$$

9.4 Higher-Order Abstract Syntax (HOAS)

Example, full FOA: $F ::= \ldots \forall x.\, A \mid \exists x.\, A$ $\Sigma = \Sigma_{\mathcal{F}} \cup \Sigma_{\mathcal{P}} \cup \Sigma_{\mathcal{C}} \cup \Sigma_{\mathcal{Q}}$:

$$\Sigma_{\mathcal{Q}} = \langle all : (i \to o^{238}) \to o,\ exists : (i \to o) \to o \rangle$$

Extend the definition of $\ulcorner . \urcorner$:

$$
\begin{aligned}
\ulcorner \forall x.\, P \urcorner &= all\ (\lambda x^i.\, \ulcorner P \urcorner) \\
\ulcorner \exists x.\, P \urcorner &= exists\ (\lambda x^i.\, \ulcorner P \urcorner)
\end{aligned}
$$

---

the signature Σ would not contain any constants of type var or $\ldots \to var$. The only terms of type var would be variables. In this case, a $\lambda^{\to}$ term like $(and\ (p\ x)(all\ x\ (q\ x)))$ could only be typed in a context Γ containing $x : var$.

Alternatively, one could imagine that they are constants. The signature signature Σ would contain expressions of the form $x : var$, where x would be a $\lambda^{\to}$ constant. One thing that isn't nice about this approach is that Σ cannot be an infinite sequence, and so we would have to fix a finite set of variables that can be represented in $\lambda^{\to}$.

In either case, the operation of substitution on the metalevel is of no use for implementing substitution on the object level.

[238]Some intuition: a proposition is represented by a term of type o. Now a term of type $i \to o$ represents a proposition where some positions are marked in a special way. For example, in $\lambda x^i.\, eq\ x\ x$, the positions where x occurs are marked in a special way, by virtue of the fact that the λ in front of the expression binds the x. This "marking" allows us to "insert"

Adequacy and faithfulness as before[239].

---

other terms in place of x. We will see this soon.

 all is a constant which can be applied to a term of type $i \rightarrow o$.

[239]Terms and formulae are represented by (canonical) members of i and o. The principle is similar as for $Prop$.

Examples

$$\ulcorner \forall x. \, x = x \urcorner \qquad\qquad = \, all(\lambda x^i. \, eq \, x \, x)$$

$$\ulcorner \forall x. \, \exists y. \, \neg(x + x = y) \urcorner \; =$$
$$all(\lambda x^i. \, exists(\lambda y^i. \, not \, (eq \, (plus \, x \, x) \, y)))$$

Examples

$$\ulcorner \forall x.\, x = x \urcorner \qquad\qquad = \; all(\lambda x^i.\, eq\, x\, x)$$
$$\ulcorner \forall x.\, \exists y.\, \neg(x + x = y) \urcorner \; =$$
$$all(\lambda x^i.\, exists(\lambda y^i.\, not\, (eq\, (plus\, x\, x)\, y)))$$

Example derivation (all but one steps use rule *app*):

$$
\dfrac{
\vdash all : (i \to o) \to o
\qquad
\dfrac{
\dfrac{
\dfrac{
x : i \vdash eq : i \to i \to o \quad x : i \vdash x : i
}{
x : i \vdash eq\, x : i \to o
}
\quad x : i \vdash x : i
}{
\dfrac{x : i \vdash eq\, x\, x : o}{\vdash \lambda x^i.\, eq\, x\, x : i \to o}\; abs
}
}{
\vdash all(\lambda x^i.\, eq\, x\, x) : o
}
$$

221

Order

Order of a type: For type τ written $\tau_1 \to \ldots \to \tau_n$, right associated, $\tau_n \in \mathcal{B}$:

- $Ord(\tau) = 0$ if $\tau \in \mathcal{B}$, i.e., if $n = 1$;

- $Ord(\tau) = 1 + max(Ord(\tau_i))$,

---

[240]A term of first-order type is a function taking (an arbitrary number of) arguments all of which must be of base type.

A term of second-order type is a function taking (an arbitrary number of) arguments some of which may be functions (of first order type).

A term of third-order type is a function taking (an arbitrary number of) arguments some of which may be functions, which again take functions (of first order type) as arguments.

. . .

Obviously, it would be wrong to think of the order as "number of arrows in a type". Instead, one can think of order as the "nesting depth of arrows in a type".

Sometimes, the notion "second-order" is used in the context of type theories for quite a different concept, but we will avoid that other use here.

Order

Order of a type: For type τ written $\tau_1 \rightarrow \ldots \rightarrow \tau_n$, right associated, $\tau_n \in \mathcal{B}$:

- $Ord(\tau) = 0$ if $\tau \in \mathcal{B}$, i.e., if $n = 1$;

- $Ord(\tau) = 1 + max(Ord(\tau_i))$,

  Intuition: "functions as arguments"[240].
  A type of order 1 is first-order, of order 2 second-order etc.
  A type of order > 1 is called higher order.

---

[240]A term of first-order type is a function taking (an arbitrary number of) arguments all of which must be of base type.

A term of second-order type is a function taking (an arbitrary number of) arguments some of which may be functions (of first order type).

A term of third-order type is a function taking (an arbitrary number of) arguments some of which may be functions, which again take functions (of first order type) as arguments.

. . .

Obviously, it would be wrong to think of the order as "number of arrows in a type". Instead, one can think of order as the "nesting depth of arrows in a type".

Sometimes, the notion "second-order" is used in the context of type theories for quite a different concept, but we will avoid that other use here.

Why "Higher Order"?

Constants representing propositional operators (logical symbols) or non-logical symbols are first-order (hence first-order syntax):

$$and : o \rightarrow o \rightarrow o$$

Why "Higher Order"?

Constants representing propositional operators (logical symbols) or non-logical symbols are first-order (hence first-order syntax):

$$and : o \rightarrow o \rightarrow o$$

Variable binding operators are higher-order (hence higher-order syntax):

$$all : (i \rightarrow o) \rightarrow o$$

Exercise: Summation Operator

What is the order of the summation operator $\sum$?

Exercise: Summation Operator

What is the order of the summation operator $\sum$?

$$sum : i \rightarrow i \rightarrow (i \rightarrow i) \rightarrow i$$

$$\ulcorner \sum_{x=0}^{n} (x+2) \urcorner =$$

Exercise: Summation Operator

What is the order of the summation operator $\sum$?

$$sum : i \rightarrow i \rightarrow (i \rightarrow i) \rightarrow i$$

$$\ulcorner \sum_{x=0}^{n} (x+2) \urcorner = sum\ zero\ n\ (\lambda x^i.\ plus\ x\ (succ\ succ\ zero))$$

So the order is 2.

Why "Abstract"?

HOAS looks quite different from the concrete object level syntax and hence "abstracts" from this object level syntax.

More specifically, different object level binding operators are represented by a combination of a constant (all, $exists$) and the generic λ-operator.

Thanks to this technique, standard operations on syntax need no special encoding, but are supported implicitly by $\lambda^{\rightarrow}$.

We will now see this.

Binding

Binding on the object level and metalevel coincide.

So in $\forall x.\, P$, all occurrences of x in P are bound, and likewise, in $all(\lambda x^i.\, \ulcorner P \urcorner)$, all occurrences of x in $\ulcorner P \urcorner$ are bound.

This provides support for substitution.

Substitution

Recall rules for $\forall$:

$$\frac{\forall x.\, P(x)}{P(t)} \; \forall\text{-}E$$

Substitution

Recall rules for $\forall$:

$$\frac{\forall x.\, P(x)}{P(t)} \;\forall\text{-}E \qquad\qquad \rightsquigarrow \qquad \frac{all\ P}{P(t)} \;\forall\text{-}E$$

Substitution

Recall rules for $\forall$:

$$\dfrac{\forall x.\, P(x)}{P(t)} \;\forall\text{-}E \qquad\qquad \rightsquigarrow \qquad \dfrac{all\ P}{P(t)} \;\forall\text{-}E$$

$$\dfrac{\forall x.\, x = x}{x = x[x \leftarrow 0]} \;\forall\text{-}E$$

Now apply substitution...

Substitution

Recall rules for $\forall$:

$$\frac{\forall x.\, P(x)}{P(t)} \; \text{$\forall$-E} \qquad \rightsquigarrow \qquad \frac{all\ P}{P(t)} \; \text{$\forall$-E}$$

$$\frac{\forall x.\, x = x}{0 = 0} \; \text{$\forall$-E}$$

Now apply substitution...

Substitution

Recall rules for $\forall$:

$$\frac{\forall x.\,P(x)}{P(t)}\;\forall\text{-}E \qquad \rightsquigarrow \qquad \frac{all\ P}{P(t)}\;\forall\text{-}E$$

$$\frac{\forall x.\,x = x}{0 = 0}\;\forall\text{-}E \quad\rightsquigarrow\quad \frac{all\ (\lambda x^i.\,eq\ x\ x)}{(\lambda x^i.\,eq\ x\ x)\ zero}\;\forall\text{-}E$$

Now apply substitution...
Now apply β-reduction...

Substitution

Recall rules for $\forall$:

$$\frac{\forall x.\,P(x)}{P(t)}\ \forall\text{-}E \qquad \rightsquigarrow \qquad \frac{all\ P}{P(t)}\ \forall\text{-}E$$

$$\frac{\forall x.\,x = x}{0 = 0}\ \forall\text{-}E \quad \rightsquigarrow \quad \frac{all\ (\lambda x^{i}.\,eq\ x\ x)}{eq\ zero\ zero}\ \forall\text{-}E$$

Now apply substitution. . .
Now apply β-reduction. . .
We now understand "marked positions in a formula".

Equivalence under Bound Variable Renaming

On the object level, formulae are equivalent under renaming of bound variables:

$$(\forall x.\, P \leftrightarrow \forall y.\, P[x \leftarrow y])$$

Equivalence under Bound Variable Renaming

On the object level, formulae are equivalent under renaming of bound variables:

$$(\forall x.\, P \leftrightarrow \forall y.\, P[x \leftarrow y])$$

Likewise, on the metalevel, formulae obtained by bound variable renaming are α-equivalent:

$$all(\lambda x^i.\, P) =_\alpha all(\lambda y^i.\, P[x \leftarrow y])$$

9.5 Summary of Encoding Syntax

Object Language Metalanguage

Syntactic category

Term, *Prop*

9.5 Summary of Encoding Syntax

Object Language	Metalanguage
Syntactic category	Type declaration $\{i, o\} = \mathcal{B}$
$Term$, $Prop$	
Variable x	

9.5 Summary of Encoding Syntax

Object Language	Metalanguage
Syntactic category $Term$, $Prop$	Type declaration $\{i, o\} = \mathcal{B}$
Variable x	Variable[241] x
Non-logical symb. $+$	

9.5 Summary of Encoding Syntax

Object Language	Metalanguage
Syntactic category *Term*, *Prop*	Type declaration $\{i, o\} = \mathcal{B}$
Variable x	Variable[241] x
Non-logical symb. $+$	1st-order constant $plus : i \rightarrow i \rightarrow i$
Logical symbol $\wedge$	

9.5 Summary of Encoding Syntax

Object Language	Metalanguage
Syntactic category $Term$, $Prop$	Type declaration $\{i, o\} = \mathcal{B}$
Variable x	Variable[241] x
Non-logical symb. $+$	1st-order constant $plus : i \to i \to i$
Logical symbol $\wedge$	1st-order constant $and : o \to o \to o$
Binding operator $\forall$	

9.5 Summary of Encoding Syntax

Object Language	Metalanguage
Syntactic category $Term$, $Prop$	Type declaration $\{i, o\} = \mathcal{B}$
Variable x	Variable[241] x
Non-logical symb. $+$	1st-order constant $plus : i \rightarrow i \rightarrow i$
Logical symbol $\wedge$	1st-order constant $and : o \rightarrow o \rightarrow o$
Binding operator $\forall$	2nd-order const. $all : (i \rightarrow o) \rightarrow o$
Meaningful expr. $a \wedge b \in Prop$	

9.5 Summary of Encoding Syntax

Object Language	Metalanguage
Syntactic category $Term$, $Prop$	Type declaration $\{i, o\} = \mathcal{B}$
Variable x	Variable[241] x
Non-logical symb. $+$	1st-order constant $plus : i \to i \to i$
Logical symbol $\wedge$	1st-order constant $and : o \to o \to o$
Binding operator $\forall$	2nd-order const. $all : (i \to o) \to o$
Meaningful expr. $a \wedge b \in Prop$	Member of type $(and\, a\, b) : o$

[241]Although propositional variables and first-order variables are quite different concepts, for both the representation in $\lambda^{\to}$ used $\lambda^{\to}$ variables. Technically however, there is a difference between the representations of propositional variables and first-order variables. In particular, propositional variables are represented as $\lambda^{\to}$ variables of type o, and first-order variables are represented as $\lambda^{\to}$ variables of type i.

10 Resolution

Three Sections on Deduction Techniques

After encoding syntax, the next topic in the theory is encoding proofs.

But before, we look at some more practical issues:

- Resolution

- Proof search

- Term rewriting

We will explain many techniques relevant for Isabelle, but not in extreme detail and rigor. We want to understand better how Isabelle works, but not provide a formal proof that she works correctly, or be able to rebuild her.

Resolution

Resolution is the basic mechanism for transforming proof states in Isabelle in order to construct a proof.

It involves unifying a certain part of the current goal (state) with a certain part of a rule, and replacing that part of the current goal.

We have already explained this in the labs and you have been working with it all the time, but now we want to understand it more thoroughly (in the next lecture, we will look at it more abstractly).

We look at several variants of resolution.

Resolution (`rtac`, as in Prolog[242])

$$\frac{\phi_1 \quad \cdots \quad \phi_i \quad \cdots \quad \phi_n}{\psi}$$

$\phi_1, \ldots, \phi_n$ are current subgoals and ψ is original goal. Isabelle displays

```
Level ... (n subgoals)
```
ψ

  1. ϕ_1

    $\vdots$

  n. ϕ_n

---

[242]Prolog is a logic programming language [Apt97].

The computation mechanism of Prolog is resolution of a current goal (corresponding to our $\phi_1, \ldots, \phi_n$) with a Horn clause (corresponding to our $[\![\alpha_1; \ldots; \alpha_m]\!] \Longrightarrow \beta$).

Resolution (`rtac`, as in Prolog[242])

$$\frac{\alpha_1 \ldots \alpha_m}{\beta}$$

$$\frac{\phi_1 \quad \cdots \quad \phi_i \quad \cdots \quad \phi_n}{\psi}$$

$\phi_1, \ldots, \phi_n$ are current subgoals and ψ is original goal. Isabelle displays

```
Level ... (n subgoals)
```
ψ

  1. ϕ_1

   ⋮

  n. ϕ_n

$[\alpha_1; \ldots; \alpha_m] \Longrightarrow \beta$ is rule.

---

[242]Prolog is a logic programming language [Apt97].

The computation mechanism of Prolog is resolution of a current goal (corresponding to our $\phi_1, \ldots, \phi_n$) with a Horn clause (corresponding to our $[\alpha_1; \ldots; \alpha_m] \Longrightarrow \beta$).

Resolution (`rtac`, as in Prolog[242])

$$\frac{\alpha_1 \ldots \alpha_m}{\beta}$$

$$\phi_1 \quad \cdots \quad \phi_i \quad \cdots \quad \phi_n$$
$$\psi$$

Simple scenario where ϕ_i has no premises[243]. Now β must be unifiable with selected sub-goal ϕ_i.

---

[242]Prolog is a logic programming language [Apt97].

The computation mechanism of Prolog is resolution of a current goal (corresponding to our $\phi_1, \ldots, \phi_n$) with a Horn clause (corresponding to our $[\alpha_1; \ldots; \alpha_m] \Longrightarrow \beta$).

Resolution (`rtac`, as in Prolog[242])

$$\frac{\alpha'_1 \cdots \alpha'_m}{\textcircled{β'}}$$

$$\frac{\phi'_1 \quad \cdots \quad \textcircled{ϕ'_i} \quad \cdots \quad \phi'_n}{\psi'}$$

Simple scenario where ϕ_i has no premises[243]. Now β must be unifiable with selected subgoal ϕ_i.

We apply the unifier ($'$[244])

---

[242]Prolog is a logic programming language [Apt97].

The computation mechanism of Prolog is resolution of a current goal (corresponding to our $\phi_1, \ldots, \phi_n$) with a Horn clause (corresponding to our $[\![\alpha_1; \ldots; \alpha_m]\!] \Longrightarrow \beta$).

Resolution (`rtac`, as in Prolog[242])

$$\frac{\phi_1' \cdots \alpha_1' \cdots \alpha_m' \cdots \phi_n'}{\psi'}$$

Simple scenario where ϕ_i has no premises[243]. Now β must be unifiable with selected subgoal ϕ_i.

We apply the unifier ($'$[244])

We replace ϕ_i' by the premises of the rule.

---

[242]Prolog is a logic programming language [Apt97].

The computation mechanism of Prolog is resolution of a current goal (corresponding to our $\phi_1, \ldots, \phi_n$) with a Horn clause (corresponding to our $[\![\alpha_1; \ldots; \alpha_m]\!] \implies \beta$).

Resolution (with Lifting over Parameters)

$$\frac{\phi_1 \quad \cdots \quad \bigwedge x.\phi_i \quad \cdots \quad \phi_n}{\psi}$$

Now suppose the i'th (selected) subgoal is preceded by $\bigwedge$ (metalevel universal quantifier[245]).

Resolution (with Lifting over Parameters)

$$\frac{\alpha_1 \quad \cdots \quad \alpha_m}{\beta}$$

$$\frac{\phi_1 \quad \cdots \quad \bigwedge x.\phi_i \quad \cdots \quad \phi_n}{\psi}$$

Rule

Resolution (with Lifting over Parameters)

$$\frac{\bigwedge x.\alpha_1[x] \cdots \bigwedge x.\alpha_m[x]}{\bigwedge x.\beta[x]}$$

$$\frac{\phi_1 \quad \cdots \quad \bigwedge x.\phi_i \quad \cdots \quad \phi_n}{\psi}$$

Rule is lifted[246] over x: Apply $[?X \leftarrow ?X(x)]$.

Resolution (with Lifting over Parameters)

$$\bigwedge x . \alpha_1[x] \cdots \bigwedge x . \alpha_m[x]$$
$$\bigwedge x . \beta[x]$$

$$\phi_1 \quad \cdots \quad \bigwedge x . \phi_i \quad \cdots \quad \phi_n$$
$$\psi$$

Rule is lifted[246] over x: Apply $[?X \leftarrow ?X(x)]$.

As before, β must be unifiable with ϕ_i;

Resolution (with Lifting over Parameters)

$$\bigwedge x.\alpha'_1[x] \cdots \bigwedge x.\alpha'_m[x]$$
$$\overline{\bigwedge x.\beta'[x]}$$

$$\phi'_1 \quad \cdots \quad \bigwedge x.\phi'_i \quad \cdots \quad \phi'_n$$
$$\overline{\qquad\qquad\qquad\qquad}$$
$$\psi'$$

Rule is lifted[246] over x: Apply $[?X \leftarrow ?X(x)]$.

As before, β must be unifiable with ϕ_i; apply the unifier.

Resolution (with Lifting over Parameters)

$$\frac{\phi'_1 \cdots \bigwedge x.\alpha'_1[x] \cdots \bigwedge x.\alpha'_m[x] \cdots \phi'_n}{\psi'}$$

Rule is lifted[246] over x: Apply $[?X \leftarrow ?X(x)]$.

As before, β must be unifiable with ϕ_i; apply the unifier.

We replace ϕ'_i by the premises of the rule. $\alpha'_1, \ldots, \alpha'_m$ are preceded by $\bigwedge x$.

Resolution (with Lifting over Assumptions)

$$[\phi_{i1} \cdots \phi_{ik_i}]$$
$$\vdots$$

$$\phi_1 \qquad \cdots \qquad \phi_i \qquad \cdots \qquad \phi_n$$

$$\psi$$

Now, suppose the i'th (selected) subgoal has assumptions $\phi_{i1}, \ldots, \phi_{ik_i}$.

Resolution (with Lifting over Assumptions)

$$\frac{\alpha_1 \quad \cdots \quad \alpha_m}{\beta}$$

$$\frac{\phi_1 \quad \cdots \quad \overset{\overset{\textstyle [\phi_{i1}\cdots\phi_{ik_i}]}{\vdots}}{\phi_i} \quad \cdots \quad \phi_n}{\psi}$$

As before, we have a rule. In general β is not[247] unifiable with the i'th subgoal, even assuming that β is unifiable with ϕ_i.

Resolution (with Lifting over Assumptions)

$$
\cfrac{
\begin{array}{c}
[\phi_{i1}\cdots\phi_{ik_i}] \quad [\phi_{i1}\cdots\phi_{ik_i}] \\
\vdots \quad \cdots \quad \vdots \\
\alpha_1 \quad \cdots \quad \alpha_m
\end{array}
}{
\begin{array}{c}
[\phi_{i1}\cdots\phi_{ik_i}] \\
\vdots \\
\beta
\end{array}
}
$$

$$
\cfrac{
\begin{array}{ccccccc}
 & & [\phi_{i1}\cdots\phi_{ik_i}] & & & & \\
 & & \vdots & & & & \\
\phi_1 & \cdots & \phi_i & \cdots & \phi_n & & \beta
\end{array}
}{\psi}
$$

Rule must be lifted over assumptions[248]. No unification so far!

Resolution (with Lifting over Assumptions)
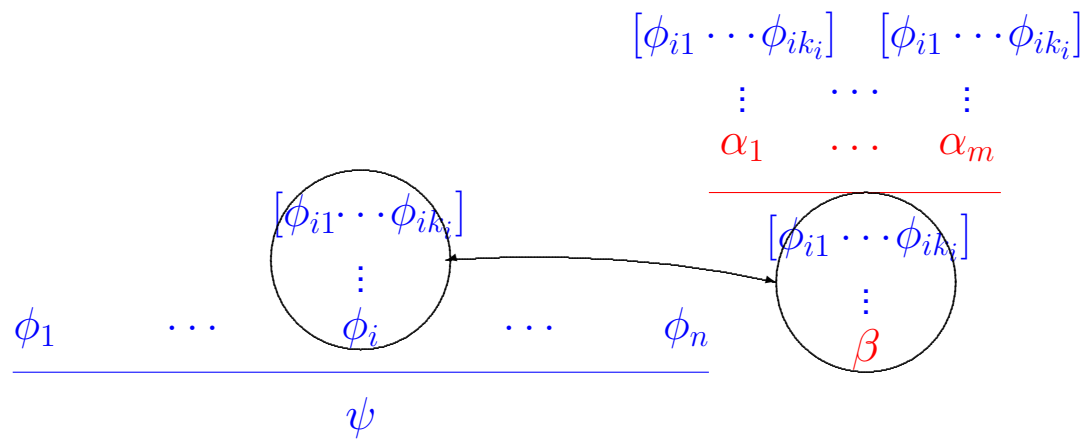
$$[\phi_{i1}\cdots\phi_{ik_i}] \quad [\phi_{i1}\cdots\phi_{ik_i}]$$

$$\vdots \quad \cdots \quad \vdots$$

$$\alpha_1 \quad \cdots \quad \alpha_m$$

$$\phi_1 \quad \cdots \quad \phi_i \quad \cdots \quad \phi_n$$

$$\psi$$

Now, subgoal and rule conclusion (below the bar) are unifiable[249].

235

Resolution (with Lifting over Assumptions)

$$
\cfrac{
\begin{array}{ccc}
[\phi_{i1}\cdots\phi_{ik_i}] & & [\phi_{i1}\cdots\phi_{ik_i}] \\
\vdots & \cdots & \vdots \\
\alpha_1 & \cdots & \alpha_m
\end{array}
}{[\phi_{i1}\cdots\phi_{ik_i}]}
$$

$$
\cfrac{
\begin{array}{cccccc}
 & & [\phi_{i1}\cdots\phi_{ik_i}] & & & \\
 & & \vdots & & & [\phi_{i1}\cdots\phi_{ik_i}] \\
 & & & & & \vdots \\
\phi_1 & \cdots & \fbox{ϕ_i} & \cdots & \phi_n & \fbox{β}
\end{array}
}{\psi}
$$
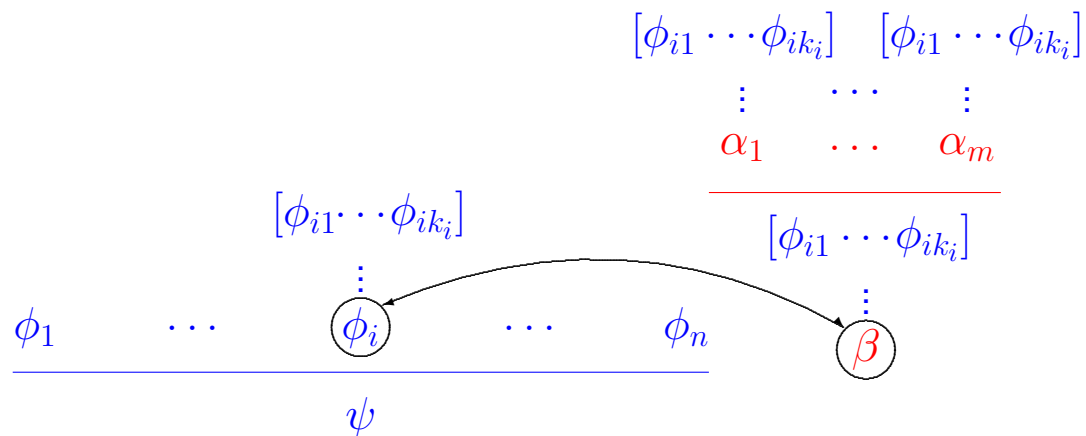
Now, subgoal and rule conclusion (below the bar) are unifiable[249].
Non-trivially[250], β must be unifiable with ϕ_i.

Resolution (with Lifting over Assumptions)

$$\cfrac{\cfrac{[\phi'_{i1}\cdots\phi'_{ik_i}]}{\vdots}}{\alpha'_1} \cdots \cfrac{\cfrac{[\phi'_{i1}\cdots\phi'_{ik_i}]}{\vdots}}{\alpha'_m}$$

$$\cfrac{[\phi'_{i1}\cdots\phi'_{ik_i}]}{\vdots}{\beta'}$$

$$\cfrac{\phi'_1 \qquad \cdots \qquad \cfrac{[\phi'_{i1}\cdots\phi'_{ik_i}]}{\vdots}{\phi'_i} \qquad \cdots \qquad \phi'_n}{\psi'}$$

We apply the unifier.

Resolution (with Lifting over Assumptions)

$$\frac{\begin{array}{ccc} [\phi'_{i1} \cdots \phi'_{ik_i}] & [\phi'_{i1} \cdots \phi'_{ik_i}] \\ \vdots & \cdots & \vdots \\ \phi'_1 \cdots \phi'_{i-1} \ \alpha'_1 & \cdots & \alpha'_m \ \phi'_{i+1} \cdots \phi'_n \end{array}}{\psi'}$$

We replace the subgoal.

Rule Premises Containing $\implies$

$$[\phi'_{i1} \cdots \phi'_{ik_i}]$$

$$\vdots$$

$$\phi'_1 \cdots \qquad \alpha'_j \qquad \cdots \phi'_n$$

$$\psi'$$

What if some α'_j has the form $[\gamma_1; \ldots; \gamma_l] \implies \delta$?

---

251 Generally, Isabelle makes no distinction between

$$[\psi_1; \ldots; \psi_n] \implies [\mu_1; \ldots; \mu_k] \implies \phi$$

and

$$[\psi_1; \ldots; \psi_n; \mu_1; \ldots; \mu_k] \implies \phi$$

and displays the second form. Semantically, this corresponds to the equivalence of $A_1 \wedge \ldots \wedge A_n \to B$ and $A_1 \to \ldots \to A_n \to B$.

We have seen this in the exercises.

Rule Premises Containing $\Longrightarrow$

$$[\phi'_{i1} \cdots \phi'_{ik_i}]$$
$$\vdots$$
$$\phi'_1 \cdots \quad [\![\gamma_1; \ldots; \gamma_l]\!] \Longrightarrow \delta \quad \cdots \phi'_n$$
$$\rule{8cm}{0.4pt}$$
$$\psi'$$

What if some α'_j has the form $[\![\gamma_1; \ldots; \gamma_l]\!] \Longrightarrow \delta$?
Is this what we get?

---

[251] Generally, Isabelle makes no distinction between

$$[\![\psi_1; \ldots; \psi_n]\!] \Longrightarrow [\![\mu_1; \ldots; \mu_k]\!] \Longrightarrow \phi$$

and

$$[\![\psi_1; \ldots; \psi_n; \mu_1; \ldots; \mu_k]\!] \Longrightarrow \phi$$

and displays the second form. Semantically, this corresponds to the equivalence of $A_1 \wedge \ldots \wedge A_n \to B$ and $A_1 \to \ldots \to A_n \to B$.

We have seen this in the exercises.

Rule Premises Containing $\Longrightarrow$

$$[\phi'_{i1} \cdots \phi'_{ik_i}; \gamma'_1 \cdots \gamma'_l]$$

$$\vdots$$

$$\phi'_1 \cdots \qquad \delta' \qquad \cdots \phi'_n$$

$$\psi'$$

What if some α'_j has the form $[\![\gamma_1; \ldots; \gamma_l]\!] \Longrightarrow \delta$?

Is this what we get?

Well, we write $:$ for $\Longrightarrow$, and use $A \Longrightarrow B \Longrightarrow C \equiv [A; B] \Longrightarrow C$[251].

---

[251]Generally, Isabelle makes no distinction between

$$[\![\psi_1; \ldots; \psi_n]\!] \Longrightarrow [\![\mu_1; \ldots; \mu_k]\!] \Longrightarrow \phi$$

and

$$[\![\psi_1; \ldots; \psi_n; \mu_1; \ldots; \mu_k]\!] \Longrightarrow \phi$$

and displays the second form. Semantically, this corresponds to the equivalence of $A_1 \wedge \ldots \wedge A_n \to B$ and $A_1 \to \ldots \to A_n \to B$.

We have seen this in the exercises.

Elimination-Resolution

$$\frac{\alpha_1 \cdots \alpha_m}{\beta}$$

$$[\phi_{i1} \cdots \quad \phi_{il} \quad \cdots \phi_{ik_i}]$$
$$\vdots$$
$$\frac{\phi_1 \qquad \cdots \qquad \phi_i \qquad \cdots \qquad \phi_n}{\psi}$$

Same scenario as before[252]

---

[252]So the scenario looks as for resolution with lifting over assumptions. However, this time we do not show the lifting over assumptions in our animation.

[253]Elimination-resolution is used to eliminate a connective in the premises.

For example, if the current goal is

$$[A \wedge B]$$
$$\vdots$$
$$\frac{B}{A \wedge B \to B}$$

and the rule is

$$[P; Q]$$
$$\vdots$$
$$\frac{P \wedge Q \qquad R}{R} \wedge\text{-}E$$

Elimination-Resolution

Same scenario as before[252], but now β must be unifiable with ϕ_i, and α_1 must be unifiable with ϕ_{il}, for some l.

---

[252]So the scenario looks as for resolution with lifting over assumptions. However, this time we do not show the lifting over assumptions in our animation.

[253]Elimination-resolution is used to eliminate a connective in the premises.

For example, if the current goal is

$$[A \wedge B]$$
$$\vdots$$
$$\frac{B}{A \wedge B \rightarrow B}$$

and the rule is

$$[P; Q]$$
$$\vdots$$
$$\frac{P \wedge Q \qquad R}{R} \wedge\text{-}E$$

Elimination-Resolution

$$\cfrac{}{[\phi'_{i,1}\cdots \boxed{\phi'_{il}} \cdots \phi'_{ik_i}] \qquad \boxed{\alpha'_1}\cdots\alpha'_m}$$

$$\cfrac{\phi'_1 \qquad \cdots \qquad \boxed{\phi'_i} \qquad \cdots \qquad \phi'_n}{\psi'}$$
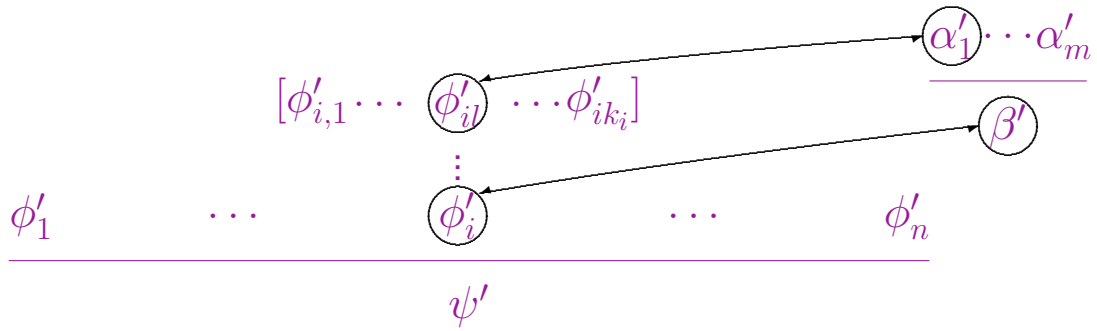
Same scenario as before[252], but now β must be unifiable with ϕ_i, and α_1 must be unifiable with ϕ_{il}, for some l.

Apply the unifier.

---

[252]So the scenario looks as for resolution with lifting over assumptions. However, this time we do not show the lifting over assumptions in our animation.

[253]Elimination-resolution is used to eliminate a connective in the premises.

For example, if the current goal is

$$\cfrac{\begin{array}{c}[A \wedge B]\\ \vdots \\ B\end{array}}{A \wedge B \rightarrow B}$$

and the rule is

$$\cfrac{P \wedge Q \qquad \begin{array}{c}[P;Q]\\ \vdots\\ R\end{array}}{R} \wedge\text{-}E$$

Elimination-Resolution

$$[\phi'_{i1} \cdots \phi'_{i,l-1}, \phi'_{i,l+1} \cdots \phi'_{ik_i}] \qquad [\phi'_{i1} \cdots \phi'_{i,l-1}, \phi'_{i,l+1} \cdots \phi'_{ik_i}]$$
$$\vdots \qquad\qquad\qquad\qquad \vdots$$
$$\underline{\phi'_1 \ \cdots \ \phi'_{i-1} \ \alpha'_2 \qquad\qquad \cdots \qquad\qquad \alpha'_m \ \phi'_{i+1} \cdots \phi'_n}$$
$$\psi'$$

Same scenario as before[252], but now β must be unifiable with ϕ_i, and α_1 must be unifiable with ϕ_{il}, for some l.

Apply the unifier.

We replace ϕ'_i by the premises of the rule except the first[253].
$\alpha'_2, \ldots, \alpha'_m$ inherit the assumptions of ϕ'_i, except ϕ'_{il}.

[252]So the scenario looks as for resolution with lifting over assumptions. However, this time we do not show the lifting over assumptions in our animation.

[253]Elimination-resolution is used to eliminate a connective in the premises.

For example, if the current goal is

$$[A \wedge B]$$
$$\vdots$$
$$\underline{B}$$
$$A \wedge B \to B$$

and the rule is

$$[P; Q]$$
$$\vdots$$
$$\frac{P \wedge Q \qquad R}{R} \ \wedge\text{-}E$$

Destruct-Resolution

$$\frac{\alpha}{\beta}$$

$$[\phi_{i1} \cdots \phi_{il} \cdots \phi_{ik_i}]$$
$$\vdots$$
$$\frac{\phi_1 \quad \cdots \quad \phi_i \quad \cdots \quad \phi_n}{\psi}$$

Simple rule

10.1 Summary on Resolution

- Build proof resembling sequent style notation;

- technically: replace goals with rule premises, or goal premises with rule conclusions;

then the result of elimination resolution is
$$[A; B]$$
$$\vdots$$
$$\frac{B}{A \wedge B \rightarrow B}$$

Effectively, the interplay between elimination rules and elimination-resolution is such that one "does not throw any information away". Before we had the assumption $A \wedge B$. This was replaced by the components A and B as separate assumptions.

[254]Destruct-resolution is used to eliminate a connective in the premises. The difference compared to elimination-resolution can be seen in the following example. Unlike elimination-resolution, destruct-resolution "throws information away".

Destruct-Resolution

$$[\phi_{i1} \cdots \phi_{il} \cdots \phi_{ik_i}] \quad \dfrac{\alpha}{\beta}$$

$$\vdots$$

$$\phi_1 \quad \cdots \quad \phi_i \quad \cdots \quad \phi_n$$

$$\psi$$

Simple rule, and α must be unifiable with ϕ_{il}, for some l.

10.1 Summary on Resolution

- Build proof resembling sequent style notation;

- technically: replace goals with rule premises, or goal premises with rule conclusions;

then the result of elimination resolution is

$$[A; B]$$
$$\vdots$$
$$\dfrac{B}{A \wedge B \to B}$$

Effectively, the interplay between elimination rules and elimination-resolution is such that one "does not throw any information away". Before we had the assumption $A \wedge B$. This was replaced by the components A and B as separate assumptions.

[254]Destruct-resolution is used to eliminate a connective in the premises. The difference compared to elimination-resolution can be seen in the following example. Unlike elimination-resolution, destruct-resolution "throws information away".

Destruct-Resolution

$$[\phi'_{i1} \cdots \boxed{\phi'_{il}} \cdots \phi'_{ik_i}] \quad \dfrac{\alpha'}{\beta'}$$

$$\vdots$$

$$\phi'_1 \quad \cdots \quad \phi'_i \quad \cdots \quad \phi'_n$$

$$\overline{\qquad\qquad\qquad\qquad\qquad}$$

$$\psi'$$

Simple rule, and α must be unifiable with ϕ_{il}, for some l. We apply the unifier.

10.1 Summary on Resolution

- Build proof resembling sequent style notation;

- technically: replace goals with rule premises, or goal premises with rule conclusions;

then the result of elimination resolution is

$$[A; B]$$
$$\vdots$$
$$\dfrac{B}{A \wedge B \to B}$$

Effectively, the interplay between elimination rules and elimination-resolution is such that one "does not throw any information away". Before we had the assumption $A \wedge B$. This was replaced by the components A and B as separate assumptions.

[254]Destruct-resolution is used to eliminate a connective in the premises. The difference compared to elimination-resolution can be seen in the following example. Unlike elimination-resolution, destruct-resolution "throws information away".

Destruct-Resolution

$$[\phi'_{i1} \cdots \beta' \cdots \phi'_{ik_i}]$$
$$\vdots$$
$$\phi'_1 \quad \cdots \quad \phi'_i \quad \cdots \quad \phi'_n$$
$$\overline{\qquad\qquad\qquad\qquad\qquad}$$
$$\psi'$$

Simple rule, and α must be unifiable with ϕ_{il}, for some l.
We apply the unifier.
We replace premise[254] ϕ'_{il} with the conclusion of the rule.

10.1 Summary on Resolution

- Build proof resembling sequent style notation;

- technically: replace goals with rule premises, or goal premises with rule conclusions;

then the result of elimination resolution is
$$[A; B]$$
$$\vdots$$
$$B$$
$$\overline{A \wedge B \rightarrow B}$$

Effectively, the interplay between elimination rules and elimination-resolution is such that one "does not throw any information away". Before we had the assumption $A \wedge B$. This was replaced by the components A and B as separate assumptions.

[254]Destruct-resolution is used to eliminate a connective in the premises. The difference compared to elimination-resolution can be seen in the following example. Unlike elimination-resolution, destruct-resolution "throws information away".

- metavariables and unification to obtain appropriate instance of rule, delay commitments;

- lifting over parameters and assumptions;

- various techniques to manipulate premises or conclusions, as convenient: `rtac`, `etac`, `dtac`.

For example, if the current goal is

$$\frac{\begin{array}{c} [A \wedge B] \\ \vdots \\ B \end{array}}{A \wedge B \to B}$$

and the rule is

$$\frac{P \wedge Q}{Q} \; \texttt{conjunct2}$$

then the result of destruct-resolution is

$$\frac{\begin{array}{c} [B] \\ \vdots \\ B \end{array}}{A \wedge B \to B}$$

If we had instead used rule

$$\frac{P \wedge Q}{P} \; \texttt{conjunct2}$$

the result would have been

$$
\begin{array}{c}
[A] \\
\vdots \\
B \\
\hline
A \wedge B \rightarrow B
\end{array}
$$

and we would be stuck. We accidentally "threw away" the assumption B.

Outline of this Part

- Proof search (à la tableau proving) and backtracking

- Classifying rules

- Proof procedures

11.1 Proof Search and Backtracking

- Need for more automation[255]

- Some aspects in proof construction are highly non-deterministic:

    – unification: which unifier to choose?

---

[255]We have seen in the exercises that doing a proof step by step is very tedious and often involves difficult guessing or alternatively, backtracking. We cannot hope to prove anything about realistic systems if proving simple theorems is so tedious.

Efficiency considerations are important for automation. The non-determinacy in proof search obviously leads to inefficiencies as many possibilities have to be explored.

- resolution: where[256] to apply a rule (which 'subgoal')?
- which rule to apply?

• How to organize proof-search technically[257]?

---

[256]We have seen in the exercises (and also in the lecture) that one can choose the subgoal to which one wants to apply a rule.

[257]We have seen in the previous lecture that resolution transforms a proof state into a new proof state. But how does one organize all those potential proof states in order to find proofs?

Organizing Proof Search Conceptually

Organize proof search as a tree[258] of theorems[259] (`thm`'s).

A sketch of an exemplary proof search:

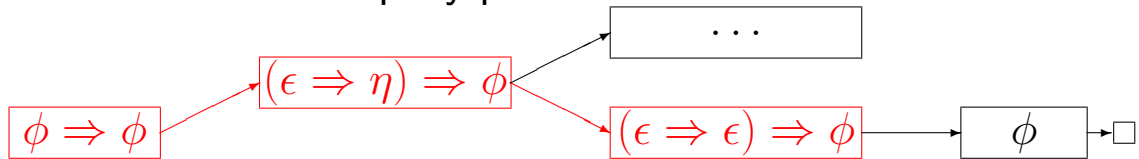[258]We have seen in the previous lecture that resolution transforms a proof state into a new proof state. Since in general, a proof state has several successor states (states that can be obtained by one resolution step), conceptually one obtains a tree where the children of a state are the successors.

[259]Technically, a proof state is an Isabelle theorem, (`thm`), i.e. something which Isabelle regards as true.

Organizing Proof Search Conceptually

Organize proof search as a tree[258] of theorems[259] (`thm`'s).

A sketch of an exemplary proof search:

$$\boxed{\phi \Rightarrow \phi}$$

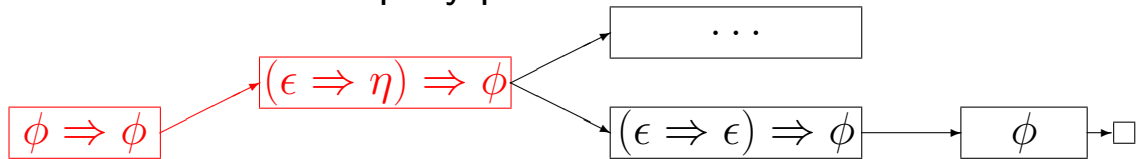`Goal` ϕ will create the proof state $\phi \Longrightarrow \phi$[260].

---

[258]We have seen in the previous lecture that resolution transforms a proof state into a new proof state. Since in general, a proof state has several successor states (states that can be obtained by one resolution step), conceptually one obtains a tree where the children of a state are the successors.

[259]Technically, a proof state is an Isabelle theorem, (`thm`), i.e. something which Isabelle regards as true.

Organizing Proof Search Conceptually

Organize proof search as a tree[258] of theorems[259] (`thm`'s).

A sketch of an exemplary proof search:
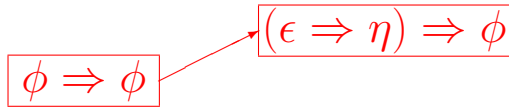
One tactic step (by . . .) transforms into proof state $(\epsilon \Rightarrow \eta) \Rightarrow \phi$.

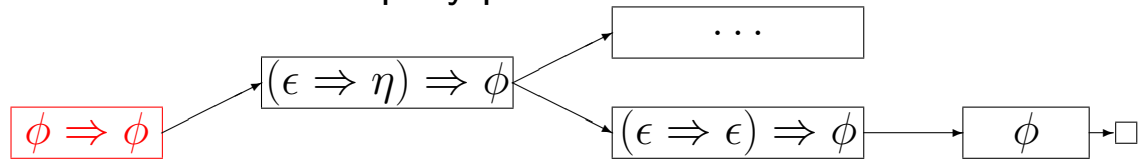[258]We have seen in the previous lecture that resolution transforms a proof state into a new proof state. Since in general, a proof state has several successor states (states that can be obtained by one resolution step), conceptually one obtains a tree where the children of a state are the successors.

[259]Technically, a proof state is an Isabelle theorem, (`thm`), i.e. something which Isabelle regards as true.

Organizing Proof Search Conceptually

Organize proof search as a tree[258] of theorems[259] (`thm`'s).

A sketch of an exemplary proof search:
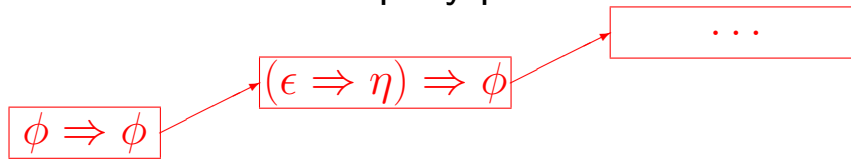
Next tactic step yields dead-end (no tactic application possible).

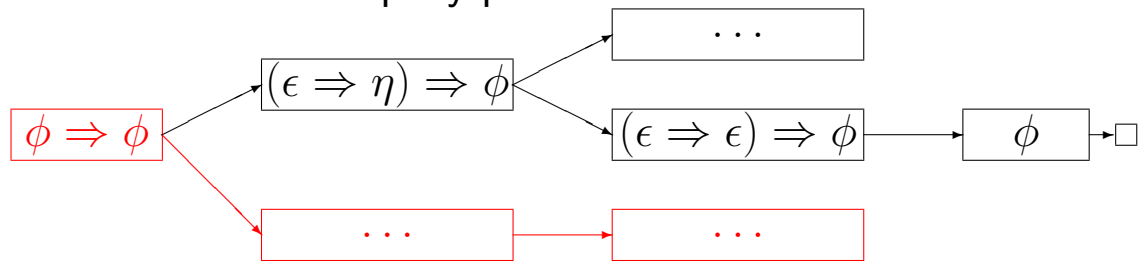[258]We have seen in the previous lecture that resolution transforms a proof state into a new proof state. Since in general, a proof state has several successor states (states that can be obtained by one resolution step), conceptually one obtains a tree where the children of a state are the successors.

[259]Technically, a proof state is an Isabelle theorem, (`thm`), i.e. something which Isabelle regards as true.

Organizing Proof Search Conceptually

Organize proof search as a tree[258] of theorems[259] (`thm`'s).

A sketch of an exemplary proof search:
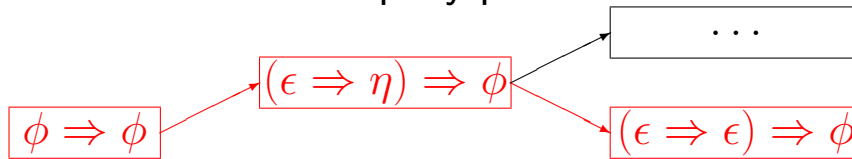
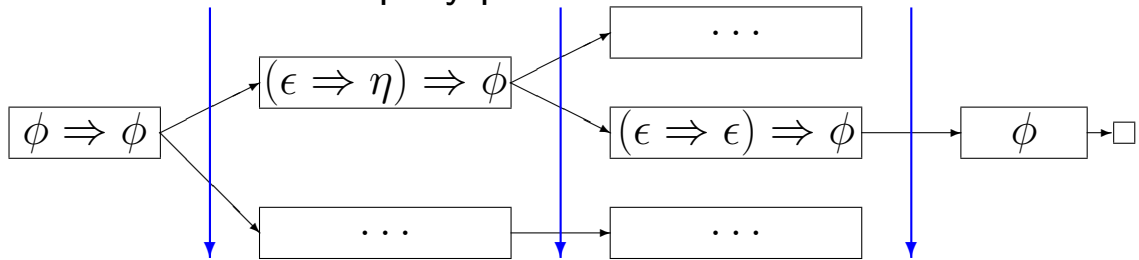`back();` tries an alternative successor of $(\epsilon \Rightarrow \eta) \Rightarrow \phi$.

---

[258]We have seen in the previous lecture that resolution transforms a proof state into a new proof state. Since in general, a proof state has several successor states (states that can be obtained by one resolution step), conceptually one obtains a tree where the children of a state are the successors.

[259]Technically, a proof state is an Isabelle theorem, (`thm`), i.e. something which Isabelle regards as true.

Organizing Proof Search Conceptually

Organize proof search as a tree[258] of theorems[259] (`thm`'s).

A sketch of an exemplary proof search:
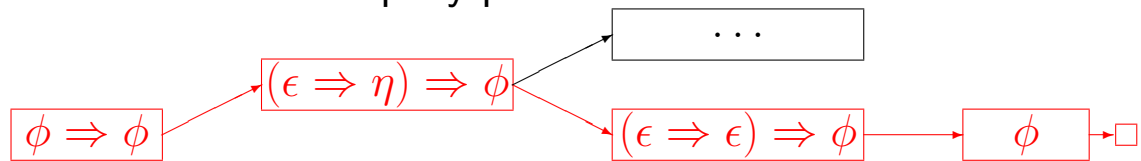
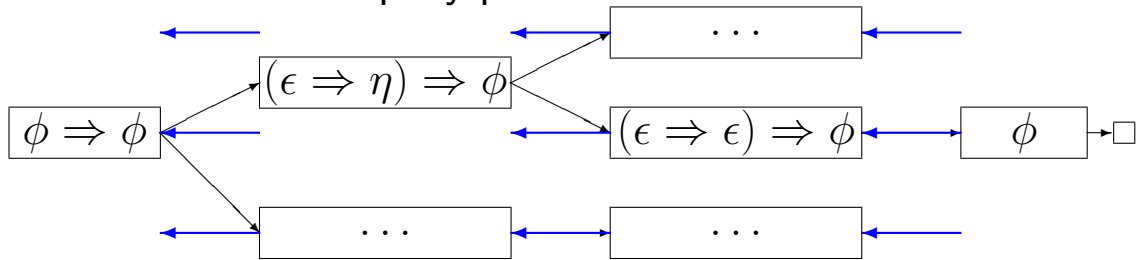Now $(\epsilon \Rightarrow \epsilon) \Rightarrow \phi$ is solvable using `atac`[261]. `qed`.

---

[258] We have seen in the previous lecture that resolution transforms a proof state into a new proof state. Since in general, a proof state has several successor states (states that can be obtained by one resolution step), conceptually one obtains a tree where the children of a state are the successors.

[259] Technically, a proof state is an Isabelle theorem, (`thm`), i.e. something which Isabelle regards as true.

Organizing Proof Search Conceptually

Organize proof search as a tree[258] of theorems[259] (`thm`'s).

A sketch of an exemplary proof search:

Use `undo();` three times to go to previous proof states.

---

[258]We have seen in the previous lecture that resolution transforms a proof state into a new proof state. Since in general, a proof state has several successor states (states that can be obtained by one resolution step), conceptually one obtains a tree where the children of a state are the successors.

[259]Technically, a proof state is an Isabelle theorem, (`thm`), i.e. something which Isabelle regards as true.

Organizing Proof Search Conceptually

Organize proof search as a tree[258] of theorems[259] (`thm`'s).

A sketch of an exemplary proof search:

Use `undo();` three times to go to previous proof states.

---

[258]We have seen in the previous lecture that resolution transforms a proof state into a new proof state. Since in general, a proof state has several successor states (states that can be obtained by one resolution step), conceptually one obtains a tree where the children of a state are the successors.

[259]Technically, a proof state is an Isabelle theorem, (`thm`), i.e. something which Isabelle regards as true.

Organizing Proof Search Conceptually

Organize proof search as a tree[258] of theorems[259] (`thm`'s).

A sketch of an exemplary proof search:

Use `undo();` three times to go to previous proof states.

---

[258]We have seen in the previous lecture that resolution transforms a proof state into a new proof state. Since in general, a proof state has several successor states (states that can be obtained by one resolution step), conceptually one obtains a tree where the children of a state are the successors.

[259]Technically, a proof state is an Isabelle theorem, (`thm`), i.e. something which Isabelle regards as true.

Organizing Proof Search Conceptually

Organize proof search as a tree[258] of theorems[259] (`thm`'s).

A sketch of an exemplary proof search:

Use `back();` to try alternative successor. . . .

---

[258]We have seen in the previous lecture that resolution transforms a proof state into a new proof state. Since in general, a proof state has several successor states (states that can be obtained by one resolution step), conceptually one obtains a tree where the children of a state are the successors.

[259]Technically, a proof state is an Isabelle theorem, (`thm`), i.e. something which Isabelle regards as true.

Organizing Proof Search Conceptually

Organize proof search as a tree[258] of theorems[259] (`thm`'s).
A sketch of an exemplary proof search:

Summary:
`back();` to try alternative successors[262] ($\Rightarrow$ different unifiers).

---

[258]We have seen in the previous lecture that resolution transforms a proof state into a new proof state. Since in general, a proof state has several successor states (states that can be obtained by one resolution step), conceptually one obtains a tree where the children of a state are the successors.
[259]Technically, a proof state is an Isabelle theorem, (`thm`), i.e. something which Isabelle regards as true.

Organizing Proof Search Conceptually

Organize proof search as a tree[258] of theorems[259] (`thm`'s).

A sketch of an exemplary proof search:

Summary:

$\texttt{back}()$; to try alternative successors[262] ($\Rightarrow$ different unifiers).

$\texttt{undo}()$; to go to previous proof state.

---

[258]We have seen in the previous lecture that resolution transforms a proof state into a new proof state. Since in general, a proof state has several successor states (states that can be obtained by one resolution step), conceptually one obtains a tree where the children of a state are the successors.

[259]Technically, a proof state is an Isabelle theorem, (`thm`), i.e. something which Isabelle regards as true.

Problems

The search space of proof search can be thought of as such a tree, but it cannot be implemented like this straightaway:

- Branching of the tree infinite in general (HO-unification).

- Explicit tree representation[263] expensive in time and space.

As an aside[264], it is also possible to understand proof search more abstractly. But we are interested in the operational aspects.

---

[263]Obviously, an infinite tree cannot be represented explicitly. But even if the tree is finite, it is generally expensive to represent it explicitly. In particular, the tree may contain many failing branches and only few successful ones, which begs the question if representing the unsuccessful branches cannot be avoided somehow.

[264]The explicit tree representation is not very abstract in that each node has a defined order of the children (first successor, second successor, ...). This order is an artefact of the order in which unifiers are enumerated by the unification algorithm used. It is inessential for the proofs that are contained in the tree.

As a more abstract understanding of proof search, one can organize proof search as a relation on theorems (thm's)

$$prooftrees = \mathcal{P}(\texttt{thm} \times \texttt{thm})$$

More precisely, one can look at a fragment of a tree of theorems as before, e.g.:

Organizing Proof Search Operationally
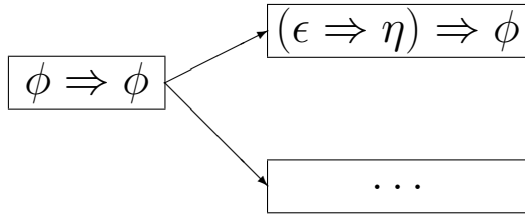
One could say that each tactic application (with a particular rule) gives rise to a relations on theorems. That is to say, ϕ and ϕ' are in the relation if ϕ' is a successor proof state of ϕ.

This is abstract in that there is no order among the successors of a proof state.

Also, one does not represent a tree explicitly.

Advantage: we have an abstract algebra.

- $PT_1 \circ PT_2$: sequential composition ("then").

  Given two relations between `thm`'s, PT_1 and PT_2, we define composition $PT_1 \circ PT_2$ as the relation

  $\{(\phi, \psi) \mid \text{there is } \eta \text{ such that } (\phi, \eta) \in PT_1 \text{ and } (\eta, \psi) \in PT_2\}$

- $PT_1 \cup PT_2$: alternative of proof attempts ("or")

The union of two relations is defined as usual for sets. If PT_1 and PT_2 each model the application of a particular tactic, then $PT_1 \cup PT_2$ models the application of "first tactic or second tactic".

- PT^* : reflexive transitive closure ("repeat ")

- $(\phi \Rightarrow \phi, \phi) \in PT^* \quad \equiv \quad$ "there is a proof for ϕ"

  PT^* is inductively defined as the smallest set where

  - $(\phi, \phi) \in PT^*$ for all ϕ;
  - if $(\phi, \eta) \in PT$ and $(\eta, \psi) \in PT^*$ then $(\phi, \psi) \in PT^*$.

  So if PT models the application of a particular tactic, then PT^* models the application of that tactic arbitrarily many times.

However, this also has some disadvantages:

- Union $\cup$ is difficult to implement (needs comparison with all previous results since one wants to avoid duplicates).

Organize proof search as a function on theorems[265] (`thm`'s)

$$\text{type } \texttt{tactic} = \texttt{thm} \rightarrow \texttt{thm seq}$$

where `seq`[266] is the type constructor for infinite lists.

This allows us to have tacticals[267]:

- `THEN`

- `ORELSE`

- `REPEAT`

- `INTLEAVE, BREADTHFIRST, DEPTHFIRST, ...`

- More operational, strategic interpretations of union $\cup$ are desirable (try this — then that, interleave attempts in PT_1 with attempts in PT_2, and so forth).

[265] This way of understanding and origanizing proof search is not so abstract, but rather operational. Instead of saying that ϕ and ϕ' are in a relation, one says that ϕ' is in the sequence returned by the tactic applied to ϕ. There is an order among the successors of a proof state.

One still does not represent a tree explicitly, although conceptually, proof search is about exploring this tree.

[266] For any type τ, the type τ `seq` (recall the notation) is the type of (possibly) infinite lists of elements of type τ. This is of course an abstract datatype. There should be functions to return the head and the tail of such an infinite list.

An abstract datatype is a type whose terms cannot be represented explicitly and accessed directly, but only via certain functions for that type.

[267]

11.2 Classifying Rules

How to Organize Proof Rules?

Observation: Some rules can always be applied blindly in backward reasoning, e.g. $\rightarrow$-*I* or $\wedge$-*I*. Others are problematic, e.g. $\wedge$-*EL* or $\wedge$-*ER* (you do not know which to apply to get rid of a $\wedge$ in the premises).

11.2 Classifying Rules

How to Organize Proof Rules?

Observation: Some rules can always be applied blindly in backward reasoning, e.g. →-*I* or ∧-*I*. Others are problematic, e.g. ∧-*EL* or ∧-*ER* (you do not know which to apply to get rid of a ∧ in the premises).

But proof rules can be tailored to be applied blindly.

In the following we will explain this using sequent style notation.

- `THEN`

- `ORELSE`

- `REPEAT`

- `INTLEAVE, BREADTHFIRST, DEPTHFIRST, ...`

are called tacticals.

Tacticals are operations on tactics. They play an important role in automating proofs in Isabelle. The most basic tacticals are `THEN` and `ORELSE`. Both of those tacticals are of type `tactic * tactic → tactic` and are written infix: tac_1 `THEN` tac_2 applies tac_1 and then tac_2, while tac_1 `ORELSE` tac_2 applies tac_1 if possible and otherwise applies tac_2 [Pau05, Ch. 4].

Review: Sequent Notation

$$\Gamma \vdash A \quad (\text{where } A \in \Gamma) \qquad \frac{\Gamma \vdash B}{A, \Gamma \vdash B} \; weaken$$

$$\frac{\Gamma \vdash A \quad \Gamma \vdash B}{\Gamma \vdash A \wedge B} \; \wedge\text{-}I \qquad \frac{\Gamma \vdash A \wedge B}{\Gamma \vdash A} \; \wedge\text{-}EL \qquad \frac{\Gamma \vdash A \wedge B}{\Gamma \vdash B} \; \wedge\text{-}ER$$

$$\frac{A, \Gamma \vdash B}{\Gamma \vdash A \rightarrow B} \; \rightarrow\text{-}I \qquad \frac{\Gamma \vdash A \rightarrow B \quad \Gamma \vdash A}{\Gamma \vdash B} \; \rightarrow\text{-}E$$

$$\frac{\Gamma \vdash A}{\Gamma \vdash A \vee B} \; \vee\text{-}IL \qquad \frac{\Gamma \vdash B}{\Gamma \vdash A \vee B} \; \vee\text{-}IR$$

$$\frac{\Gamma \vdash A \vee B \quad \Gamma, A \vdash C \quad \Gamma, B \vdash C}{\Gamma \vdash C} \; \vee\text{-}E$$

Example: $\wedge$-E'

In the sequent calculus[268], one writes $\wedge$-E[269] as:

$$\frac{A, B, \Gamma \vdash C}{A \wedge B, \Gamma \vdash C} \; \wedge\text{-}E'$$

This mimics[270] the effect of using $\wedge$-E (conjE of Isabelle), here written in sequent style notation,

$$\frac{\Gamma \vdash A \wedge B \quad A, B, \Gamma \vdash C}{\Gamma \vdash C} \; \wedge\text{-}E$$

in combination with etac. We now derive $\wedge$-E'.

[268] Tableau proving is a derivation system [Fit96].

It turns out that the language of tableaux is equivalent to the sequent calculus (recall our use of sequent style notation) [Pau97]. The techniques Isabelle uses for automating proofs can thereby be understood as tableau proving [Pau97].

[269] In Isabelle notation, it looks as follows:

$$[\![P \& Q; \; [\![P; \; Q]\!] \Longrightarrow R]\!] \Longrightarrow R$$

(see IFOL_lemmas.ML).

[270] That is to say, $\wedge$-E' behaves for the sequent notation as conjE+etac behaves for Isabelle.

Deriving $\wedge$-E' from $\wedge$-E

$$\frac{\Gamma \vdash A \wedge B \quad A, B, \qquad \Gamma \vdash C}{\Gamma \vdash C} \wedge\text{-}E$$

Deriving $\wedge\text{-}E'$ from $\wedge\text{-}E$

$$\frac{A \wedge B, \Gamma \vdash A \wedge B \quad A, B, A \wedge B, \Gamma \vdash C}{A \wedge B, \Gamma \vdash C} \wedge\text{-}E$$

If we replace Γ with $A \wedge B, \Gamma$ (just instantiation),

Deriving $\wedge$-E' from $\wedge$-E

$$\frac{A, B, A \wedge B, \Gamma \vdash C}{A \wedge B, \Gamma \vdash C} \wedge\text{-}E$$

If we replace Γ with $A \wedge B, \Gamma$ (just instantiation), then one part holds by the assumption rule,

Deriving $\wedge$-E' from $\wedge$-E

$$\cfrac{\cfrac{A, B, \Gamma \vdash C}{A, B, A \wedge B, \Gamma \vdash C} \; \textit{weaken}}{A \wedge B, \Gamma \vdash C} \; \wedge\text{-}E$$

If we replace Γ with $A \wedge B, \Gamma$ (just instantiation), then one part holds by the assumption rule, and we can apply weakening.

Deriving $\wedge\text{-}E'$ from $\wedge\text{-}E$

$$\frac{A, B, \Gamma \vdash C}{A \wedge B, \Gamma \vdash C} \, \wedge\text{-}E'$$

A Proof by Blind Rule Application

$$\vdash (\rho \wedge \phi) \rightarrow \psi \rightarrow \phi$$

The topmost connective is $\rightarrow$, which asks for $\rightarrow$-*I*.

---

[271]See now that we first derived the rule $\wedge$-E', which is a rule that can be used blindly to decompose a conjunction in the assumptions. This was not something ad-hoc to prove this particular formula. The rule $\wedge$-E' should be used generally instead of $\wedge$-*EL* or $\wedge$-*EL*, because it has the advantage that it can be applied blindly.

The essential point about being able to apply a rule blindly is that the application does not throw any information away. This is indeed the case for $\wedge$-E'. We remove the assumption $\phi \wedge \psi$, but we get the two conjuncts ϕ and ψ as assumptions instead.

The rule $\wedge$-E' mimics the effect of using $\wedge$-*E* in combination with `etac`, which you can see by looking again at the exercises on `etac`.

A Proof by Blind Rule Application

$$\frac{\rho \wedge \phi \vdash \psi \to \phi}{\vdash (\rho \wedge \phi) \to \psi \to \phi} \to\text{-}I$$

The topmost connective is $\to$, which asks for $\to$-*I*.
Again $\to$-*I*.

---

[271]See now that we first derived the rule $\wedge$-E', which is a rule that can be used blindly to decompose a conjunction in the assumptions. This was not something ad-hoc to prove this particular formula. The rule $\wedge$-E' should be used generally instead of $\wedge$-*EL* or $\wedge$-*EL*, because it has the advantage that it can be applied blindly.

The essential point about being able to apply a rule blindly is that the application does not throw any information away. This is indeed the case for $\wedge$-E'. We remove the assumption $\phi \wedge \psi$, but we get the two conjuncts ϕ and ψ as assumptions instead.

The rule $\wedge$-E' mimics the effect of using $\wedge$-E in combination with `etac`, which you can see by looking again at the exercises on `etac`.

A Proof by Blind Rule Application

$$\frac{\dfrac{\rho \wedge \phi, \psi \vdash \phi}{\rho \wedge \phi \vdash \psi \to \phi} \;{\to}\text{-}I}{\vdash (\rho \wedge \phi) \to \psi \to \phi} \;{\to}\text{-}I$$

The topmost connective is $\to$, which asks for $\to$-*I*.
Again $\to$-*I*.
To decompose[271] the assumption $\rho \wedge \phi$, use $\wedge$-*E′*.

---

[271]See now that we first derived the rule $\wedge$-*E′*, which is a rule that can be used blindly to decompose a conjunction in the assumptions. This was not something ad-hoc to prove this particular formula. The rule $\wedge$-*E′* should be used generally instead of $\wedge$-*EL* or $\wedge$-*EL*, because it has the advantage that it can be applied blindly.

The essential point about being able to apply a rule blindly is that the application does not throw any information away. This is indeed the case for $\wedge$-*E′*. We remove the assumption $\phi \wedge \psi$, but we get the two conjuncts ϕ and ψ as assumptions instead.

The rule $\wedge$-*E′* mimics the effect of using $\wedge$-*E* in combination with `etac`, which you can see by looking again at the exercises on `etac`.

A Proof by Blind Rule Application

$$\dfrac{\dfrac{\dfrac{\rho, \phi, \psi \vdash \phi}{\rho \wedge \phi, \psi \vdash \phi} \wedge\text{-}E'}{\rho \wedge \phi \vdash \psi \to \phi} \to\text{-}I}{\vdash (\rho \wedge \phi) \to \psi \to \phi} \to\text{-}I$$

The topmost connective is $\to$, which asks for $\to$-I.

Again $\to$-I.

To decompose[271] the assumption $\rho \wedge \phi$, use $\wedge$-E'.

The proof can now be completed by the assumption rule.

---

[271]See now that we first derived the rule $\wedge$-E', which is a rule that can be used blindly to decompose a conjunction in the assumptions. This was not something ad-hoc to prove this particular formula. The rule $\wedge$-E' should be used generally instead of $\wedge$-EL or $\wedge$-EL, because it has the advantage that it can be applied blindly.

The essential point about being able to apply a rule blindly is that the application does not throw any information away. This is indeed the case for $\wedge$-E'. We remove the assumption $\phi \wedge \psi$, but we get the two conjuncts ϕ and ψ as assumptions instead.

The rule $\wedge$-E' mimics the effect of using $\wedge$-E in combination with `etac`, which you can see by looking again at the exercises on `etac`.

Safe and Unsafe Rules

Combined tactics rely on classification of rules, maintained in Isabelle data structure `claset`[272], and accessed by functions[273] of type `claset * thm list → claset`.

Class:	To add use function:
Safe introduction rules	`addSIs`
Safe elimination rules	`addSEs`
Unsafe introduction rules	`addIs`
Unsafe elimination rules	`addEs`

---

[272]`claset` is an abstract datatype. Overloading notation, `claset` is also an ML unit function which will return a term of that datatype when applied to $()$, namely, the current classifier set.

A classifier set determines which rules are safe and unsafe introduction, respectively elimination rules. The current classifier set is a classifier set used by default in certain tactics.

The current classifier set can be accessed via special functions for that purpose.

[273]The functions `addSIs`, `addSEs`, `addIs`, `addEs` are all of type `claset * thm list → claset`. They add rules to the current classifier set. For example, `addSIs` adds a rule as safe introduction rule.

Adapting Rules for Automated Proof Search

As seen for $\wedge$-*E*, rules must be suitably adapted in order to be useful in automated proof search. Another example:

$$\frac{}{\vdash (\alpha \to \beta) \vee (\beta \to \alpha)} \; \vee\text{-}swap^{274}$$

Neither $\vee$-*IL* nor $\vee$-*IR* would work here. Uses classical logic.

Adapting Rules for Automated Proof Search

As seen for $\wedge$-*E*, rules must be suitably adapted in order to be useful in automated proof search. Another example:

$$\cfrac{\cfrac{}{\neg(\alpha \to \beta) \vdash \beta \to \alpha}\ \to\text{-}I}{\vdash (\alpha \to \beta) \vee (\beta \to \alpha)}\ \vee\text{-}swap^{274}$$

Adapting Rules for Automated Proof Search

As seen for $\wedge$-E, rules must be suitably adapted in order to be useful in automated proof search. Another example:

$$\cfrac{\cfrac{\cfrac{}{\neg(\alpha \rightarrow \beta), \beta \vdash \alpha} \ \rightarrow\text{-}swapE^{275}}{\neg(\alpha \rightarrow \beta) \vdash \beta \rightarrow \alpha} \ \rightarrow\text{-}I}{\vdash (\alpha \rightarrow \beta) \vee (\beta \rightarrow \alpha)} \ \vee\text{-}swap^{274}$$

Adapting Rules for Automated Proof Search

As seen for $\wedge$-*E*, rules must be suitably adapted in order to be useful in automated proof search. Another example:

$$\cfrac{\cfrac{\cfrac{\neg\alpha, \alpha, \beta \vdash \beta}{\neg(\alpha \to \beta), \beta \vdash \alpha} \to\text{-}swapE^{275}}{\neg(\alpha \to \beta) \vdash \beta \to \alpha} \to\text{-}I}{\vdash (\alpha \to \beta) \vee (\beta \to \alpha)} \vee\text{-}swap^{274}$$

---

[274] The rule $\vee$-*swap* is

$$\frac{\neg A, \Gamma \vdash B}{\Gamma \vdash A \vee B} \vee\text{-}swap$$

To derive it you need classical reasoning, as the rule exploits the equivalence of $A \to B$ and $\neg A \vee B$.

  This is a derived rule which is explicitly contained in the Isabelle classifier set as the clasical introduction rule for $\vee$. It is called `disjCI` (check out `FOL_lemmas1.ML`)!

[275] The rule $\to$-*swapE* is

$$\frac{A, \neg C, \Gamma \vdash B}{\neg(A \to B), \Gamma \vdash C} \to\text{-}swapE$$

To derive it you need classical reasoning, as the rule exploits the equivalence of $\neg(A \to B)$ and $A \wedge \neg B$.

  This is a standard technique in Isabelle, based on swapping. For dealing with negated formulas in the premises of the current subgoal, introduction rules are combined with `swap` using `etac`.

Principle: Emulate sequent calculus[276] with derived rules.

Generally, we have a formula $\neg(A \circ B)$ in the premises, where $\circ$ is some binary connective. Swapping will put $(A \circ B)$ in the conclusion and put the old conclusion into the premises after negating it. Afterwards, an introduction rule for $\circ$ will be used [Pau05, Section 11.2].

[276] The sequent calculus works with expressions of the form $A_1, \ldots, A_n \vdash B_1, \ldots, B_m$ which should be interpreted as: under the assumptions $A_1, \ldots, A_n$, at least one of $B_1, \ldots, B_m$ can be proven. So as a formula, this would be $A_1 \wedge \ldots \wedge A_n \rightarrow B_1 \vee \ldots \vee B_m$.

In Isabelle (and the proof trees we have seen, e.g,. in this lecture), we only have sequents with one formula to the right of the $\vdash$. We have said that we use sequent notation.

The important point to note here is that in the sequent calculus, one can shift a formula from left to right or vice versa, but one has to negate it, or more precisely, turn A into $\neg A$ and $\neg A$ into A. This is called swapping and is an important technique for combined tactics.

Handling Quantifiers

Can derive[277] $\forall\text{-}E'$ ($\equiv$ `allE`[278]) using $\forall\text{-}E$ ($\equiv$ `spec`):

$$\begin{array}{c} \quad [A(x) \qquad\qquad] \\ \quad \vdots \\ \dfrac{\forall x.A(x) \qquad\quad B}{B} \;\;\forall\text{-}E' \end{array}$$

The sequent calculus inherently relies on classical reasoning [Pau05, Ch. 11].

[277]You should do it in Isabelle. The rule is:

$$[\![\text{ALL } x.\ P(x);\ P(x) \Longrightarrow R]\!] \Longrightarrow R$$

[278]As you may have noticed earlier, there is a confusion between the names of proof rules as we present them for the theory and the names used in Isabelle. For example, rule $\rightarrow\text{-}E$ is called `mp` in Isabelle. This confusion concerns elimination rules.

There is however a good reason for these choices. In traditional presentations of logic, one sets up the simplest possible elimination rules for the connectives which naturally arise from the meaning of those connectives. This is what we have done as well. However, as we see in this lecture, these rules cannot be applied blindly and are thus not very suitable for automa-

Handling Quantifiers

Can derive[277] $\forall$-E' ($\equiv$ allE[278]) using $\forall$-E ($\equiv$ spec):

$$\cfrac{\forall x.A(x) \qquad \begin{array}{c} [A(x), \textcolor{red}{\forall x.A(x)}] \\ \vdots \\ B \end{array}}{B} \ \forall\text{-}dupE$$

The sequent calculus inherently relies on classical reasoning [Pau05, Ch. 11].

[277]You should do it in Isabelle. The rule is:

$$[\text{ALL } x. \ P(x); \ P(x) \implies R] \implies R$$

[278]As you may have noticed earlier, there is a confusion between the names of proof rules as we present them for the theory and the names used in Isabelle. For example, rule $\rightarrow$-E is called mp in Isabelle. This confusion concerns elimination rules.

There is however a good reason for these choices. In traditional presentations of logic, one sets up the simplest possible elimination rules for the connectives which naturally arise from the meaning of those connectives. This is what we have done as well. However, as we see in this lecture, these rules cannot be applied blindly and are thus not very suitable for automa-

256

tion. Therefore, combined tactics in Isabelle use derived rules such as ∧-*E* (called `conjE` in Isabelle).

Since this is of such central importance for Isabelle, one prefers to have the obvious names `conjE`, `allE` etc. for the rules that are actually used in "advanced" applications of Isabelle.

What is the difference[279] to $\exists\text{-}E$[280]?

What is the difference[279] to $\exists$-E[280]?

Problem: $\forall x. A(x)$ may still be needed.

What is the difference[279] to ∃-E[280]?

Problem: $\forall x.A(x)$ may still be needed.

---

[279]The difference between

$$\frac{\exists x.A(x) \qquad \begin{array}{c} [A(x)] \\ \vdots \\ B \end{array}}{B} \; \exists\text{-}E$$

and

$$\frac{\forall x.A(x) \qquad \begin{array}{c} [A(x)] \\ \vdots \\ B \end{array}}{B} \; \forall\text{-}E'$$

is that the first rule has a side condition: x must not occur free in any assumption on which B depends. See also what this means in terms of Isabelle.

[280]The rule

$$\frac{\exists x.A(x) \qquad \begin{array}{c} [A(x)] \\ \vdots \\ B \end{array}}{B} \; \exists\text{-}E$$

was derived previously (but in Isabelle, it is a basic rule in `IFOL.ML`). It is

$$[\![\mathtt{ALL}\ x.\ P(x);\ !!x.\ P(x) \Longrightarrow R]\!] \Longrightarrow R$$

Note that the rule `allE` ($\forall\text{-}E'$) is

$$[\![\mathtt{ALL}\ x.\ P(x);\ P(x) \Longrightarrow R]\!] \Longrightarrow R$$

The difference is that the former rule contains a metalevel universal quantifier. In terms of paper-and-pencil proofs, $\exists\text{-}E$ has the side condition that x must not occur free in any assumption on which B (see tree!) depends. There is no such side condition for $\forall\text{-}E'$.

Principle: Introduce duplicating[281] rules. Turns search infinite[282]!
Check out `allE` and `all_dupE` in `IFOL_lemmas.ML`[283]!

---

[281]You should recall that elimination rules are used in combination with `etac`. Using `allE` will eliminate the quantifier.

You should try a proof of the formula $(\forall x.P(x)) \rightarrow (P(a) \wedge P(b))$ in Isabelle to convince yourself that this is a problem since the quantified formula $\forall x.P(x)$ is needed twice as an assumption, with two different instantiations of x.

The duplicating rule $\forall$-*dupE* has the effect that the universally quantified formula will still remain as an assumption.

[282]Given only the rules so far (in combination with the appropriate tactics, `rtac and etac`, and swapping), excluding $\forall$-*dupE*, the proof search would be finite.

The rule $\forall$-*dupE* is responsible for making the proof search infinite. This can be no surprise however, as first-order logic is undecidable [And02], and so there can be no automatic procedure for proving all true first-order formulas.

[283]These files should be contained in your Isabelle distribution. Or, if you only have an Isabelle executable, you can find the sources here:

11.3 Proof Procedures (Simplified)

Tactics in Isabelle are performed in order[284]:

1. REPEAT (rtac $safe\_I\_rules$ ORELSE etac $safe\_E\_rules$)

2. canonize: propagate "$x = t$" throughout subgoal

3. rtac $unsafe\_I\_rules$ ORELSE etac $unsafe\_E\_rules$

4. atac

---

http://isabelle.in.tum.de/library/

[284]Tactics in Isabelle are performed in order:

1. REPEAT (rtac $safe\_I\_rules$ ORELSE etac $safe\_E\_rules$);

2. canonize: propagate "$x = t$" ...throughout subgoal;

3. rtac $unsafe\_I\_rules$ ORELSE etac $unsafe\_E\_rules$;

4. atac.

One elementary proof step consists of trying a safe introduction rule with rtac, or, if that is not possible, a safe elimination rule with etac. This will be repeated as long as possible.

Then in the current subgoal, any assumption of the form $x = t$ (where x is a metavariable) will be propagated throughout the subgoal, i.e., all occurrences of x wil be replaced by t.

Then Isabelle will try one application of an unsafe introduction rule with rtac, or, if that is not possible, an unsafe elimination rule with etac.

Combined Proof Search Tactics

- `step_tac : claset → int → tactic`
  (just safe steps)

- `fast_tac : claset → int → tactic`
  (safe and unsafe steps in depth-first stategy)

- `best_tac : claset → int → tactic`
  (safe and unsafe steps in breadth-first stategy)

- `slow_tac : claset → int → tactic`
  (like `fast_tac`, but with backtracking `atac`'s)

- `blast_tac : claset → int → tactic`
  (like `fast_tac`, but often more powerful)

Finally, she will use `atac`. Note that `atac` is unsafe. In general, there are several premises in a subgoal and `atac` may unify the conclusion of the subgoal with the wrong premise.

11.4 Summary on Automated Proof Search

- Proof search can be organized as a tree of theorems.

- Calculi can be set up to facilitate proof search (although this must be done by specialists).

- Combined with search strategies, powerful automatic procedures arise. Can prove well-known hard problems such as $((\exists y.\forall x.J(y,x)\vee\neg J(x,x)) \to \neg(\forall x.\exists y.\forall z.J(z,y)\vee \neg J(z,x))$

- Unfortunately, failure is difficult to interpret[285].

---

[285] `fast_tac`, `blast_tac` just tell you that the tactic failed, but not why. And it would be difficult to do that, since backtracking means that all attempts failed. This can have several reasons: a rule is missing, a rule has been classified wrongly, the search strategy was not adequate for the problem, enumeration of unifiers in a bad order. Or a combination thereof. Or it might be that too many unsafe steps are needed, since `fast_tac` limits their number.

12 Term Rewriting

Outline of this Part

- Higher-order rewriting

- Extensions: Ordered, pattern, congruence, splitting rewriting

- Organizing simplification rules

In this context, a term is a λ-term, since we use the λ-calculus to encode object logics.

12.1 Higher-Order Rewriting

Motivation:

- Simplification is a very important part of deduction, e.g.:

$$0 + (x + 0)^{286} = x$$

$$[a, b, d] @ [a, b]^{287} = [a, b, d, a, b]$$

- Based on rewrite rules as in functional programming[288]:

$$x + 0 = x, \qquad\qquad 0 + x = x$$

$$[] @ X = X, \qquad (x :: X) @ Y = x :: (X @ Y)$$

---

[286]Simplifying $0 + (x + 0)$ to x is something you have learned in school. It is justified by the usual semantics of arithmetic expressions. Here, however, we want to see more formally how such simplification works, rather than why it is justified.

[287]Lists are a common datatype in functional programming. $[a, b, d, a, b]$ is a list. Actually, this notation is syntactic sugar for $a :: (b :: (d :: (a :: (b :: []))))$. Here, $[]$ is the empty list and $::$ is a term constructor taking an alement and a list and returning a list. @ stands for list concatenation.

Intuitively, it is clear that $[a, b, d]$ concatenated with $[a, b]$ yields $[a, b, d, a, b]$.

Term constructor is usual terminology in functional programming. In first-order logic, we would speak of a function symbol. In the λ-calculus, we would speak of a (special kind of) constant (this will become clear later).

[288]For example, the lines

$$[] @ X = X$$
$$(x :: X) @ Y = x :: (X @ Y)$$

Term Rewriting: Foundation

- Recall: An equational theory consists of rules

$$\frac{}{x = x} \ \textit{refl} \qquad \frac{x = y}{y = x} \ \textit{sym} \qquad \frac{x = y \quad y = z}{x = z} \ \textit{trans}$$

$$\frac{x = y \quad P(x)}{P(y)} \ \textit{subst}$$

- plus additional (possibly conditional) rules of the form
  $\phi_1 = \psi_1, \ldots, \phi_n = \psi_n \Rightarrow \phi = \psi$.

The additional rules can be interpreted as rewrite rules[289], i.e. they are applied from left to right.

define the list concatenation function @.

---

[289]An equational theory is a formalism based on equational rules of the form $\phi_1 = \psi_1, \ldots, \phi_n = \psi_n \Longrightarrow \phi = \psi$.

A term rewriting system (to be defined shortly) is another formalism, based of rewrite rules. They also have the form $\phi_1 = \psi_1, \ldots, \phi_n = \psi_n \Longrightarrow \phi = \psi$, but they have a different flavor in that $=$ must be interpreted as a directed symbol. One could also write $\rightsquigarrow$ instead of $=$ to emphasize this.

Incomplete Decision Procedure for Goal $e = e'$

To decide if $e = e'$ in an equational theory:

1. stop if the goal is solved, i.e., $e \equiv e'$ (syntactical equality)

---

[290] Given two terms s and t, a unifier is a substitution θ such that $s\theta = t\theta$. A match is a substitution which only instantiates one of s or t, so $s\theta = t$ or $s = t\theta$ (one should usually clarify in the given context which of the terms is instantiated).

[291] This means that the procedure is called recusively for the conditions of the rewrite rule.

[292] The procedure defines a term rewriting system [BN98, Klo93].

Equational theories, term rewriting systems, propositional logic, first-order logic, different versions of the λ-calculus — with all those different formalisms playing a role here, we must agree on some terminology. In particular, the words term, function, predicate, constant and variable are used somewhat differently in the different formalisms.

Our point of reference for the terminology is the λ-calculus as it is built into Isabelle for representing object logics. In particular:

- A term is a λ-term; object-level formulae (including equa-

Incomplete Decision Procedure for Goal $e = e'$

To decide if $e = e'$ in an equational theory:

1. stop if the goal is solved, i.e., $e \equiv e'$ (syntactical equality)

2. make a rewrite step:

   (a) pick a subterm t in $e(t)$

---

[290]Given two terms s and t, a unifier is a substitution θ such that $s\theta = t\theta$. A match is a substitution which only instantiates one of s or t, so $s\theta = t$ or $s = t\theta$ (one should usually clarify in the given context which of the terms is instantiated).
[291]This means that the procedure is called recusively for the conditions of the rewrite rule.
[292]The procedure defines a term rewriting system [BN98, Klo93].

Equational theories, term rewriting systems, propositional logic, first-order logic, different versions of the λ-calculus — with all those different formalisms playing a role here, we must agree on some terminology. In particular, the words term, function, predicate, constant and variable are used somewhat differently in the different formalisms.

Our point of reference for the terminology is the λ-calculus as it is built into Isabelle for representing object logics. In particular:

• A term is a λ-term; object-level formulae (including equa-

Incomplete Decision Procedure for Goal $e = e'$

To decide if $e = e'$ in an equational theory:

1. stop if the goal is solved, i.e., $e \equiv e'$ (syntactical equality)

2. make a rewrite step:

   (a) pick a subterm t in $e(t)$

   (b) for a rewrite rule $\phi = \psi$, match[290] (unify) ϕ against t, i.e., find θ such that $\phi\theta = t$

---

[290]Given two terms s and t, a unifier is a substitution θ such that $s\theta = t\theta$. A match is a substitution which only instantiates one of s or t, so $s\theta = t$ or $s = t\theta$ (one should usually clarify in the given context which of the terms is instantiated).
[291]This means that the procedure is called recusively for the conditions of the rewrite rule.
[292]The procedure defines a term rewriting system [BN98, Klo93].

Equational theories, term rewriting systems, propositional logic, first-order logic, different versions of the λ-calculus — with all those different formalisms playing a role here, we must agree on some terminology. In particular, the words term, function, predicate, constant and variable are used somewhat differently in the different formalisms.

Our point of reference for the terminology is the λ-calculus as it is built into Isabelle for representing object logics. In particular:

- A term is a λ-term; object-level formulae (including equa-

Incomplete Decision Procedure for Goal $e = e'$

To decide if $e = e'$ in an equational theory:

1. stop if the goal is solved, i.e., $e \equiv e'$ (syntactical equality)

2. make a rewrite step:

   (a) pick a subterm t in $e(t)$

   (b) for a rewrite rule $\phi = \psi$,
   match[290] (unify) ϕ against t, i.e., find θ such that $\phi\theta = t$

   (d) replace $e(t)$ by $e(\psi\theta)$

---

[290]Given two terms s and t, a unifier is a substitution θ such that $s\theta = t\theta$. A match is a substitution which only instantiates one of s or t, so $s\theta = t$ or $s = t\theta$ (one should usually clarify in the given context which of the terms is instantiated).
[291]This means that the procedure is called recusively for the conditions of the rewrite rule.
[292]The procedure defines a term rewriting system [BN98, Klo93].

Equational theories, term rewriting systems, propositional logic, first-order logic, different versions of the λ-calculus — with all those different formalisms playing a role here, we must agree on some terminology. In particular, the words term, function, predicate, constant and variable are used somewhat differently in the different formalisms.

Our point of reference for the terminology is the λ-calculus as it is built into Isabelle for representing object logics. In particular:

- A term is a λ-term; object-level formulae (including equa-

Incomplete Decision Procedure for Goal $e = e'$

To decide if $e = e'$ in an equational theory:

1. **stop** if the goal is **solved**, i.e., $e \equiv e'$ (syntactical equality)

2. make a rewrite step:

   (a) pick a subterm t in $e(t)$

   (b) for a rewrite rule $\phi = \psi$,
       match[290] (unify) ϕ against t, i.e., find θ such that $\phi\theta = t$

   (d) replace $e(t)$ by $e(\psi\theta)$

3. **goto 1**

---

[290]Given two terms s and t, a **unifier** is a substitution θ such that $s\theta = t\theta$. A **match** is a substitution which only instantiates one of s or t, so $s\theta = t$ or $s = t\theta$ (one should usually clarify in the given context which of the terms is instantiated).
[291]This means that the procedure is called recusively for the conditions of the rewrite rule.
[292]The procedure defines a **term rewriting system** [BN98, Klo93].

Equational theories, term rewriting systems, propositional logic, first-order logic, different versions of the λ-calculus — with all those different formalisms playing a role here, we must agree on some terminology. In particular, the words **term**, **function**, **predicate**, **constant** and **variable** are used somewhat differently in the different formalisms.

Our point of reference for the terminology is the λ-calculus as it is built into Isabelle for representing object logics. In particular:

- A **term** is a λ-term; object-level formulae (including equa-

Incomplete Decision Procedure for Goal $e = e'$

To decide if $e = e'$ in an equational theory:

1. stop if the goal is solved, i.e., $e \equiv e'$ (syntactical equality)

2. make a rewrite step:

   (a) pick a subterm t in $e(t)$ (resp. $e'(t)$)

   (b) for a rewrite rule $\phi = \psi$, match[290] (unify) ϕ against t, i.e., find θ such that $\phi\theta = t$

   (d) replace $e(t)$ by $e(\psi\theta)$ (resp. $e'(t)$ by $e'(\psi\theta)$)

3. goto 1

---

[290]Given two terms s and t, a unifier is a substitution θ such that $s\theta = t\theta$. A match is a substitution which only instantiates one of s or t, so $s\theta = t$ or $s = t\theta$ (one should usually clarify in the given context which of the terms is instantiated).

[291]This means that the procedure is called recusively for the conditions of the rewrite rule.

[292]The procedure defines a term rewriting system [BN98, Klo93].

Equational theories, term rewriting systems, propositional logic, first-order logic, different versions of the λ-calculus — with all those different formalisms playing a role here, we must agree on some terminology. In particular, the words term, function, predicate, constant and variable are used somewhat differently in the different formalisms.

Our point of reference for the terminology is the λ-calculus as it is built into Isabelle for representing object logics. In particular:

- A term is a λ-term; object-level formulae (including equa-

Incomplete Decision Procedure for Goal $e = e'$

To decide if $e = e'$ in an equational theory:

1. stop if the goal is solved, i.e., $e \equiv e'$ (syntactical equality)

2. make a rewrite step:

   (a) pick a subterm t in $e(t)$ (resp. $e'(t)$)

   (b) for a rewrite rule $\phi_1 = \psi_1, \ldots, \phi_n = \psi_n \implies \phi = \psi$, match[290] (unify) ϕ against t, i.e., find θ such that $\phi\theta = t$

   (c) solve[291] $(\phi_1 = \psi_1, \ldots, \phi_n = \psi_n)\theta$

   (d) replace $e(t)$ by $e(\psi\theta)$ (resp. $e'(t)$ by $e'(\psi\theta)$)

3. goto 1

---

[290] Given two terms s and t, a unifier is a substitution θ such that $s\theta = t\theta$. A match is a substitution which only instantiates one of s or t, so $s\theta = t$ or $s = t\theta$ (one should usually clarify in the given context which of the terms is instantiated).

[291] This means that the procedure is called recusively for the conditions of the rewrite rule.

[292] The procedure defines a term rewriting system [BN98, Klo93].

Equational theories, term rewriting systems, propositional logic, first-order logic, different versions of the λ-calculus — with all those different formalisms playing a role here, we must agree on some terminology. In particular, the words term, function, predicate, constant and variable are used somewhat differently in the different formalisms.

Our point of reference for the terminology is the λ-calculus as it is built into Isabelle for representing object logics. In particular:

- A term is a λ-term; object-level formulae (including equa-

Incomplete Decision Procedure for Goal $e = e'$

To decide if $e = e'$ in an equational theory:

1. stop if the goal is solved, i.e., $e \equiv e'$ (syntactical equality)

2. make a rewrite step:

   (a) pick a subterm t in $e(t)$ (resp. $e'(t)$)

   (b) for a rewrite rule $\phi_1 = \psi_1, \ldots, \phi_n = \psi_n \implies \phi = \psi$, match[290] (unify) ϕ against t, i.e., find θ such that $\phi\theta = t$

   (c) solve[291] $(\phi_1 = \psi_1, \ldots, \phi_n = \psi_n)\theta$

   (d) replace $e(t)$ by $e(\psi\theta)$ (resp. $e'(t)$ by $e'(\psi\theta)$)

3. goto 1

This procedure + the rules define a term rewriting system[292].

---

[290]Given two terms s and t, a unifier is a substitution θ such that $s\theta = t\theta$. A match is a substitution which only instantiates one of s or t, so $s\theta = t$ or $s = t\theta$ (one should usually clarify in the given context which of the terms is instantiated).

[291]This means that the procedure is called recusively for the conditions of the rewrite rule.

[292]The procedure defines a term rewriting system [BN98, Klo93].

Equational theories, term rewriting systems, propositional logic, first-order logic, different versions of the λ-calculus — with all those different formalisms playing a role here, we must agree on some terminology. In particular, the words term, function, predicate, constant and variable are used somewhat differently in the different formalisms.

Our point of reference for the terminology is the λ-calculus as it is built into Isabelle for representing object logics. In particular:

- A term is a λ-term; object-level formulae (including equa-

Term Rewriting is Non-Trivial

- There are two major problems: this decision procedure may fail because:

    - it diverges (the rules are not terminating), e.g. $x+y = y + x$ or $x = y \implies x = y$;

Term Rewriting is Non-Trivial

- There are two major problems: this decision procedure may fail because:

  - it diverges (the rules are not terminating), e.g. $x+y = y+x$ or $x = y \implies x = y$;
  - rewriting does not yield a unique normal form (the

  tions) as well as object-level terms are all represented as λ-terms, and so for example, when we rewrite an equation, we rewrite a term.

- One could say that a function is any λ-term of functional type, i.e., of type containing at least one $\rightarrow$. Apart from that, there may be function symbols in some object logic. On the metalevel (and hence also for the purpose of term rewriting), these would be constants.

- There may be predicate symbols in some object logic. On the metalevel (and hence also for the purpose of term rewriting), these would be constants.

- A constant is a λ-term consisting of just one symbol from a set $Const$. Constants of the λ-calculus may be used to represent connectives, quantifiers, functions, predicates or any other symbols that an object logic may contain.

- The notion of variable is that of the metalevel, and so we

usually mean "variables including metavariables".

Nevertheless, some confusion may arise wherever we use the terminology from the point of view of an object logic.

See the following example:

The following is an example rewrite sequence, using the rules for lists. The picked subterm which is being replaced is underlined in each step:

$$
\begin{aligned}
\underline{(a :: (b :: (d :: []))) \; @ \; (a :: (b :: []))} &= [a, b, d, a, b] \; \rightsquigarrow \\
a :: \underline{((b :: (d :: [])) \; @ \; (a :: (b :: [])))} &= [a, b, d, a, b] \; \rightsquigarrow \\
a :: (b :: \underline{((d :: []) \; @ \; (a :: (b :: [])))}) &= [a, b, d, a, b] \; \rightsquigarrow \\
a :: (b :: (d :: \underline{([] \; @ \; (a :: (b :: [])))})) &= [a, b, d, a, b] \; \rightsquigarrow \\
a :: (b :: (d :: (a :: (b :: [])))) &= [a, b, d, a, b] \; \rightsquigarrow
\end{aligned}
$$

Note the we are done now, as the right-hand side is identical to the left-hand side, modulo the use of syntactic sugar.

Note that generally, a term rewriting sequence rewrites arbitrary terms. Here we only rewrite equations. From the point of view of term rewriting, an equation is just a special case of

rules are not confluent), e.g. rules $a = b$, $a = c$[293].

---

a term.

One could also imagine that object-level function and predicate symbols are represented as variables, as is done in LF. Recall Perlis' epigram.

[293]For a rewriting system consisting of rules $a = b$, $a = c$, one cannot rewrite $b = c$ to prove the equality, although it holds:

$$\dfrac{\dfrac{a = b}{b = a}\ \textit{sym} \qquad a = c}{b = c}\ \textit{trans}$$

rules are not confluent), e.g. rules $a = b$, $a = c$[293].

- Providing criteria for terminating and confluent rule sets is an active research area (see [BN98, Klo93], RTA, ...).

---

a term.

One could also imagine that object-level function and predicate symbols are represented as variables, as is done in LF. Recall Perlis' epigram.

[293]For a rewriting system consisting of rules $a = b$, $a = c$, one cannot rewrite $b = c$ to prove the equality, although it holds:

$$\frac{\dfrac{a = b}{b = a}\ sym \qquad a = c}{b = c}\ trans$$

12.2 Extensions of Rewriting

- Symmetric rules are problematic, e.g. ACI:[294]

$$
\begin{aligned}
(x + y) + z &= x + (y + z) \quad \text{(A)} \\
x + y &= y + x \qquad\quad \text{(C)} \\
x + x &= x \qquad\qquad \text{(I)}
\end{aligned}
$$

12.2 Extensions of Rewriting

- Symmetric rules are problematic, e.g. ACI:[294]

$$(x + y) + z = x + (y + z) \quad \text{(A)}$$
$$x + y = y + x \qquad\qquad \text{(C)}$$
$$x + x = x \qquad\qquad\quad\, \text{(I)}$$

- Idea: apply only if replaced term gets smaller w.r.t. some term ordering. In example, if $(y + x)\theta$ is smaller than $(x + y)\theta$.

- Ordered rewriting solves rewriting modulo ACI[295], using derived rules (exercise).

---

[294]ACI stands for associative, commutative and idempotent. In

$$(x + y) + z = x + (y + z) \quad \text{(A)}$$
$$x + y = y + x \qquad\qquad \text{(C)}$$
$$x + x = x \qquad\qquad\quad\, \text{(I)}$$

the constant $+$ is written infix.

[295]Consider an equational theory consisting only of those rules (apart from refl, sym, trans, subst). Apart from that, the language may contain arbitrary other constant symbols. For such a language, it is possible to give a term ordering that will assign more weight to the same term on the left-hand-side of a $+$ than on the right-hand side. We can base such a term ordering on a norm[296]. For example, the inductive definition of a norm $|\_|$ might include the line:

$$|s + t| := 2|s| + |t|$$

This means that if $|s| > |t|$, then $|s + t| = 2|s| + |t| > 2|t| + |s| = |t + s|$.

This has two effects:

Extension: HO-Pattern Rewriting

Rules such as $F(G\,c) = \ldots$[297] lead to highly ambiguous matching and hence inefficiency.

Solution: restrict l.h.s. of a rule to higher-order patterns.

– Applications of (A) or (I) always decrease the weight of a term (provided the weight of s is > 0):
$$|(s + t) + r| = 2|s + t| + |r| = 4|s| + 2|t| + |r| \;>\; 2|s| + 2|t| + |r| = 2|s| + |t + r| = |s + (t + r)|.$$

– Applications of (C) are only possible if the left-hand side is heavier than the right-hand side.

We haven't worked out here how the norm should be defined for the other symbols of the language. This would have to depend on that language.

The notation $|\_|$ (the argument is between the bars) is used in standard mathematics for the absolute value of a number and is standard for norms as well.

[297] For higher-order rewriting, it is very problematic to have rules containing terms of the form $F(G\,c)$ on the left-hand side, where F and G are free variables and c is a constant or bound variable. The reason can be seen in an example: Suppose you want to rewrite the term $f(g(h(i\,c)))$ where f,

Extension: HO-Pattern Rewriting

Rules such as $F(G\,c) = \ldots$[297] lead to highly ambiguous matching and hence inefficiency.

Solution: restrict l.h.s. of a rule to higher-order patterns.

A term t is a HO-pattern if

- it is in β-normal form; and

- any free F in t occurs in a subterm $F\,x_1 \ldots x_n$ where the x_i are η-equivalent to distinct bound variables.

- Applications of (A) or (I) always decrease the weight of a term (provided the weight of s is > 0):
$$|(s+t)+r| = 2|s+t| + |r| = 4|s| + 2|t| + |r| \;>\; 2|s| + 2|t| + |r| = 2|s| + |t+r| = |s+(t+r)|.$$

- Applications of (C) are only possible if the left-hand side is heavier than the right-hand side.

We haven't worked out here how the norm should be defined for the other symbols of the language. This would have to depend on that language.

The notation $|_-|$ (the argument is between the bars) is used in standard mathematics for the absolute value of a number and is standard for norms as well.

[297]For higher-order rewriting, it is very problematic to have rules containing terms of the form $F(G\,c)$ on the left-hand side, where F and G are free variables and c is a constant or bound variable. The reason can be seen in an example: Suppose you want to rewrite the term $f(g(h(i\,c)))$ where f,

Extension: HO-Pattern Rewriting

Rules such as $F(G\,c) = \ldots$[297] lead to highly ambiguous matching and hence inefficiency.

Solution: restrict l.h.s. of a rule to higher-order patterns.
A term t is a HO-pattern if

- it is in β-normal form; and

- any free F in t occurs in a subterm $F\,x_1 \ldots x_n$ where the x_i are η-equivalent to distinct bound variables.

Matching (unification) is decidable, unitary ('unique') and efficient algorithms exist.

- Applications of (A) or (I) always decrease the weight of a term (provided the weight of s is > 0):
$$|(s + t) + r| = 2|s + t| + |r| = 4|s| + 2|t| + |r| >$$
$$2|s| + 2|t| + |r| = 2|s| + |t + r| = |s + (t + r)|.$$

- Applications of (C) are only possible if the left-hand side is heavier than the right-hand side.

We haven't worked out here how the norm should be defined for the other symbols of the language. This would have to depend on that language.

The notation $|\_|$ (the argument is between the bars) is used in standard mathematics for the absolute value of a number and is standard for norms as well.

[297] For higher-order rewriting, it is very problematic to have rules containing terms of the form $F(G\,c)$ on the left-hand side, where F and G are free variables and c is a constant or bound variable. The reason can be seen in an example: Suppose you want to rewrite the term $f(g(h(i\,c)))$ where f,

HO-Pattern Rewriting (Cont.)

A rule $\ldots \Rightarrow \phi = \psi$ is a HO-pattern rule if:

- ϕ is a HO-pattern;

- all free variables in ψ occur also in ϕ; and

- ϕ is constant-head, i.e. of the form $\lambda x_1..x_m.c\,p_1 \ldots p_n$ (where c is a constant, $m \geq 0$, $n \geq 0$).

---

g, h, i are all constants. There are four unifiers of $F\,(G\,c)$ and $f(g(h(i\,c)))$:

$$[F \leftarrow f,\ G \leftarrow (\lambda x.g(h(i\,x)))],$$
$$[F \leftarrow (\lambda x.f(g\,x)),\ G \leftarrow (\lambda x.h(i\,x))],$$
$$[(F \leftarrow \lambda x.f(g(h\,x))),\ G \leftarrow (\lambda x.i\,x)],$$
$$[(F \leftarrow \lambda x.f(g(h(i\,x)))),\ G \leftarrow (\lambda x.x)].$$

This ambiguity makes such TRSs very inefficient.
[298]Further examples:

- $(\exists x.Px \vee Qx) = (\exists x.Px) \vee (\exists x.Qx)$

- $(\exists x.P \to Qx) = P \to (\exists x.Qx)$

- $(\exists x.Px \to Q) = (\forall x.Px) \to Q$

In these examples, you may assume that first-order logic is our object logic.

On the metalevel, and hence also for the sake of term rewriting, $\forall, \exists$ are constants.

294

HO-Pattern Rewriting (Cont.)

A rule $\ldots \Rightarrow \phi = \psi$ is a HO-pattern rule if:

- ϕ is a HO-pattern;

- all free variables in ψ occur also in ϕ; and

- ϕ is constant-head, i.e. of the form $\lambda x_1..x_m.c\, p_1 \ldots p_n$ (where c is a constant, $m \geq 0$, $n \geq 0$).

  Example:[298] $(\forall x.Px \wedge Qx) = (\forall x.Px) \wedge (\forall x.Qx)$

  Result: HO-pattern allows for very effective quantifier reasoning.

---

g, h, i are all constants. There are four unifiers of $F\,(G\,c)$ and $f(g(h(i\,c)))$:

$$[F \leftarrow f,\ G \leftarrow (\lambda x.g(h(i\,x)))],$$
$$[F \leftarrow (\lambda x.f(g\,x)),\ G \leftarrow (\lambda x.h(i\,x))],$$
$$[(F \leftarrow \lambda x.f(g(h\,x))),\ G \leftarrow (\lambda x.i\,x)],$$
$$[(F \leftarrow \lambda x.f(g(h(i\,x)))),\ G \leftarrow (\lambda x.x)].$$

This ambiguity makes such TRSs very inefficient.

[298]Further examples:

- $(\exists x.Px \vee Qx) = (\exists x.Px) \vee (\exists x.Qx)$

- $(\exists x.P \rightarrow Qx) = P \rightarrow (\exists x.Qx)$

- $(\exists x.Px \rightarrow Q) = (\forall x.Px) \rightarrow Q$

In these examples, you may assume that first-order logic is our object logic.

On the metalevel, and hence also for the sake of term rewriting, $\forall, \exists$ are constants.

Extensions Related to $\mathtt{if-then-else}$

The `if-then-else` construct will play an important role later. It asks for special rewrite rules.

In the notation $(\forall x.Px \land Qx)$, the symbols P and Q are metavariables (as far as term rewriting is concerned, simply think: variables).

Actually, $(\forall x.Px \land Qx)$ mixes object and metalevel syntax in a way which is typical for Isabelle: $(\forall x.Px \land Qx)$ is a "pretty-printed" version of `ALL` $(P \,\&\, Q)$.

You may want to look at a theory file (say, `IFOL.thy`) to get a flavor of this. The principle was explained thoroughly before.

Extension: Congruence Rewriting

Problem :
$$\texttt{if } A \texttt{ then } P \texttt{ else } Q \;=\; \texttt{if } A \texttt{ then } P' \texttt{ else } Q$$
$$\text{where } P = P' \text{ under condition } A$$

is not a rule[299].

Solution in Isabelle: explicitly admit this extra class of rules (congruence rewriting)
$$[\![A \Longrightarrow P = P']\!] \Longrightarrow$$
$$\texttt{if } A \texttt{ then } P \texttt{ else } Q \;=\; \texttt{if } A \texttt{ then } P' \texttt{ else } Q$$

---

[299]Rewrite rules have the form $\phi_1 = \psi_1, \ldots, \phi_n = \psi_n \Longrightarrow \phi = \psi$ (several equations imply one equation). It is not possible that any of the equations $\phi_1 = \psi_1, \ldots, \phi_n = \psi_n$ again depend on some condition, as in
$$\texttt{if } A \texttt{ then } P \texttt{ else } Q \;=\; \texttt{if } A \texttt{ then } P' \texttt{ else } Q$$
$$\text{where } P = P' \text{ under condition } A$$

Extension: Splitting Rewriting

Problem:

$$P(\texttt{if } A \texttt{ then } x \texttt{ else } y) = \texttt{if } A \texttt{ then } (P\,x) \texttt{ else } (P\,y)$$

is not a HO-pattern rule (since it is not constant-head).

Solution in Isabelle: explicitely admit this extra class of rules (case splitting).

12.3 Organizing Simplification Rules

- Standard (HO-pattern conditional ordered rewrite) rules;

- congruence rules;

- splitting rules.

Isabelle data structure: simpset[300]. Some operations[301]:

- addsimps : simpset * thm list $\rightarrow$ simpset

- delsimps : simpset * thm list $\rightarrow$ simpset

- addcongs : simpset * thm list $\rightarrow$ simpset

- addsplits : simpset * thm list $\rightarrow$ simpset

Commutativity can be added without losing termination.

---

[300]The simpset is an abstract datatype and at the same time an ML unit function for returning the current simplifier set. This is in analogy to the classifier set.

[301]These function manipulate the simplifier set, in analogy to the classifier set.

How to Apply the Simplifier?

Several versions of the simplifier:

- `simp_tac : simpset → int → tactic`

- `asm_simp_tac : simpset → int → tactic`
  (includes assumptions into `simpset`)

- `asm_full_simp_tac : simpset → int → tactic`
  (rewrites assumptions, and includes them into `simpset`)

Using global[302] simplifier sets: `Simp_tac`, `Asm_simp_tac`, `Asm_full_simp_tac`.

---

[302]`Simp_tac`, `Asm_simp_tac`, `Asm_full_simp_tac` work like their lower-case counterparts but use the current (global) simplifier set and hence do not take a simplifier set as first argument (e.g., `Simp_tac` has type `int → tactic`)

There are analogous capitalized versions for the tactics of the classical reasoner.

12.4 Summary on Term Rewriting

Simplifier is a powerful proof tool for

- conditional equational formulas

- ACI-rewriting

- quantifier reasoning

- congruence rewriting

- automatic proofs by case splitting

Fortunately, failure is quite easy to interpret[303].

---

[303]When you use `simp_tac`, usually you can just look at the term that you get to understand which simplification has not worked although you think that it should have worked.

12.5 Summary on Last Three Sections

- Although Isabelle is an interactive proof construction, it is a flexible environment with powerful automated proof procedures.

- For classical logic and set theory, tableau-like procedures like `blast_tac` and `fast_tac` decide many tautologies.

- For equational theories (datatypes, evaluating functional programs, but also higher-order logic) `simp_tac` decides many tautologies (and is fairly easy to control).

13 Isabelle's Metalogic

Representing Syntax and Proofs

- Previously, we have seen how the (polymorphically) typed λ-calculus can be used to represent the syntax of an object logic.

- Today, we will extend the λ-calculus to a logic (with formulae and inference rules): Isabelle's metalogic, which goes under the names of $\mathcal{M}$, Pure[304], HOL.

This lecture is based on Paulson's work [Pau89]. It is maybe the most challenging lecture of this course.

---

[304]In Isabelle jargon, the metalogic is called Pure.

In this course, we will avoid calling the Isabelle metalogic HOL, although you may find such uses in the literature.

In the literature and in Isabelle formalizations, we find various definitions of higher-order logic (HOL) that differ more or less substantially.

But the important point to remember here is this: The Isabelle metalogic $\mathcal{M}$ we study here is not identical to the logic we will study during the entire second half of this course. And the most important difference between $\mathcal{M}$ and HOL is not in the logics themselves, but in the way we use them:

$\mathcal{M}$ is a (the) metalogic!

HOL is an object logic!

What Is Formality anyway?

- Ultimately, logic and formal reasoning have to resort to natural language. Proofs of, say, the soundness of a derivation system employ the usual mathematical rigor, but that's all. Imagine this for the situation that we just want to do reasoning[305] in propositional logic and nothing else.

- We will now introduce a logic $\mathcal{M}$. Its proof system is small!

---

[305]We would formalize the language and the proof system as we did in the first lecture. Any proofs of soundness and completeness or other meta-properties should be rigorous, but they still resort to natural language.

Proof Techniques = Meta-Theorems

- When constructing proofs, there are

  - aspects that are specific to certain logics and its logical symbols: the proof rules;

  - aspects that reflect general principles of proof building: making and discharging assumptions, substitution, side conditions, resolution.

  It seems that the latter must be justified by complicated (and thus error-prone) explanations in natural language.

Proof Techniques = Meta-Theorems

- When constructing proofs, there are

  - aspects that are specific to certain logics and its logical symbols: the proof rules;

  - aspects that reflect general principles of proof building: making and discharging assumptions, substitution, side conditions, resolution.

  It seems that the latter must be justified by complicated (and thus error-prone) explanations in natural language.

- Using a metalogic such as $\mathcal{M}$ has two benefits:

  - Shared implementational support for the "general principles";

– to a wide extent, the "general principles" are formally derived in $\mathcal{M}$. This gives a high degree of confidence.

13.1 The Logic $\mathcal{M}$

We first introduce $\mathcal{M}$ just like any other logic, without considering its special role as metalogic. Nonetheless, we use the qualification "meta" to avoid confusion later.

Some variations are possible (mainly: polymorphism/type classes or not), but those are not so important for us.

– to a wide extent, the "general principles" are formally
derived in $\mathcal{M}$. This gives a high degree of confidence.

13.1 The Logic $\mathcal{M}$

We first introduce $\mathcal{M}$ just like any other logic, without con-
sidering its special role as metalogic. Nonetheless, we use the
qualification "meta" to avoid confusion later.

Some variations are possible (mainly: polymorphism/type
classes or not), but those are not so important for us.

$\mathcal{M}$ will be based on $\lambda^{\rightarrow}$. Would you call $\lambda^{\rightarrow}$ a logic?

– to a wide extent, the "general principles" are formally derived in $\mathcal{M}$. This gives a high degree of confidence.

13.1 The Logic $\mathcal{M}$

We first introduce $\mathcal{M}$ just like any other logic, without considering its special role as metalogic. Nonetheless, we use the qualification "meta" to avoid confusion later.

Some variations are possible (mainly: polymorphism/type classes or not), but those are not so important for us.

$\mathcal{M}$ will be based on $\lambda^{\rightarrow}$. Would you call $\lambda^{\rightarrow}$ a logic?

So far, $\lambda^{\rightarrow}$ is not a logic (no connectives, no formulae). We will now extend it to a logic.

Logic Based on $\lambda^{\rightarrow}$

Assume some $\mathcal{B}$ where $bool \in \mathcal{B}$, and some[306] signature Σ where

- $\Rightarrow\colon bool \rightarrow bool \rightarrow bool \in \Sigma$,

- $\equiv_\sigma\colon \sigma \rightarrow \sigma \rightarrow bool \in \Sigma$ for all types σ, and

- $\bigwedge_\sigma : (\sigma \rightarrow bool) \rightarrow bool \in \Sigma$ for all types σ.

We usually omit type subscripts[307] and write $\equiv$, $\bigwedge$.

$\Rightarrow$, $\equiv$, and $\bigwedge$[308] are the logical symbols of $\mathcal{M}$. $\Rightarrow$ and $\equiv$ are written infix.

Terms of type $bool$ are called

---

[306]Σ contains $\Rightarrow$, $\equiv$ and $\bigwedge$, but in addition, Σ may specify other symbols.

[307]Alternatively, we could define that

- $\equiv_\alpha\colon \alpha \rightarrow \alpha \rightarrow bool \in \Sigma$, and

- $\bigwedge_\alpha : (\alpha \rightarrow bool) \rightarrow bool \in \Sigma$,

where α is a type variable.

[308]$\Rightarrow$ is called meta-implication, $\equiv$ is called meta-equality, and $\bigwedge$ is called meta-universal-quantification.

Logic Based on $\lambda^{\rightarrow}$

Assume some $\mathcal{B}$ where $bool \in \mathcal{B}$, and some[306] signature Σ where

- $\Rightarrow: bool \rightarrow bool \rightarrow bool \in \Sigma$,

- $\equiv_\sigma: \sigma \rightarrow \sigma \rightarrow bool \in \Sigma$ for all types σ, and

- $\bigwedge_\sigma : (\sigma \rightarrow bool) \rightarrow bool \in \Sigma$ for all types σ.

We usually omit type subscripts[307] and write $\equiv$, $\bigwedge$.

$\Rightarrow$, $\equiv$, and $\bigwedge$[308] are the logical symbols of $\mathcal{M}$. $\Rightarrow$ and $\equiv$ are written infix.

Terms of type $bool$ are called (meta-)formulae: types generalize syntactic categories.

---

[306]Σ contains $\Rightarrow$, $\equiv$ and $\bigwedge$, but in addition, Σ may specify other symbols.

[307]Alternatively, we could define that

- $\equiv_\alpha: \alpha \rightarrow \alpha \rightarrow bool \in \Sigma$, and

- $\bigwedge_\alpha : (\alpha \rightarrow bool) \rightarrow bool \in \Sigma$,

where α is a type variable.

[308]$\Rightarrow$ is called meta-implication, $\equiv$ is called meta-equality, and $\bigwedge$ is called meta-universal-quantification.

Folding Assumptions

Lists of (meta-)formulae are denoted by Φ, Ψ, Ω. If Φ is the list $[\phi_1, \ldots, \phi_n]$, then

$$[\phi_1, \ldots, \phi_n] \Rightarrow \psi, \ \text{i.e.}$$
$$\Phi \Rightarrow \psi$$

abbreviates the meta-formula $\phi_1 \Rightarrow \ldots \Rightarrow \phi_n \Rightarrow \psi$.

You have seen this in the exercises.

Note that $[\phi_1, \ldots, \phi_n]$ on its own is not a term in $\mathcal{M}$!

Proof System for $\mathcal{M}$

The proof system will be presented in the style of natural deduction.

This is as formal as we get (for the metalogic): derivation trees in natural deduction style are authoritative.

The judgements[309], just like for natural deduction proofs in propositional logic or first-order logic, are formulae, i.e., terms of type $bool$. This is in contrast to derivability judgements or type judgements.

---

[309]We define our proof system for $\mathcal{M}$ using natural deduction.

The judgements are formulae, i.e., term of type $bool$. This means that a node ϕ in a derivation tree, as in

$$\frac{\cdots}{\phi} \cdots$$

must be a term of type $bool$. It cannot be a derivability judgement or type judgement or a term of type, say $bool \rightarrow bool$.

Rules for $\Rightarrow$

$$\frac{\begin{array}{c}[\phi]\\ \vdots\\ \psi\end{array}}{\phi \Rightarrow \psi} \Rightarrow\text{-}I \qquad \frac{\phi \Rightarrow \psi \quad \phi}{\psi} \Rightarrow\text{-}E$$

Just like rules for $\rightarrow$!

Rules for $\Rightarrow$

$$\frac{\begin{array}{c}[\phi]\\ \vdots\\ \psi\end{array}}{\phi \Rightarrow \psi} \Rightarrow\text{-}I \qquad \frac{\phi \Rightarrow \psi \quad \phi}{\psi} \Rightarrow\text{-}E$$

Just like rules for $\rightarrow$!

For layout reasons we sometimes use

$$\frac{\phi \quad \phi \Rightarrow \psi}{\psi} \Rightarrow\text{-}E$$

Rules for $\bigwedge$

Meta-universal-quantification is formalized in the style of higher-order abstract syntax; may write $\bigwedge x.\phi$ as syntactic sugar for $\bigwedge(\lambda x.\phi)$.

Note: quantification over terms of arbitrary type!

Rules for $\bigwedge$

Meta-universal-quantification is formalized in the style of higher-order abstract syntax; may write $\bigwedge x.\phi$ as syntactic sugar for $\bigwedge(\lambda x.\phi)$.

Note: quantification over terms of arbitrary type!

Rules:

$$\frac{\phi}{\bigwedge x.\phi} \bigwedge\text{-}I^* \qquad \frac{\bigwedge x.\phi}{\phi[x \leftarrow b]} \bigwedge\text{-}E$$

Side (eigenvariable) condition $*$: x is not free in any assumption on which ϕ depends.

Just like rules for $\forall$.

Rules for ≡: Equivalence Relation

$$\frac{}{a \equiv a} \; \equiv\text{-}refl \qquad\qquad \frac{a \equiv b}{b \equiv a} \; \equiv\text{-}symm$$

$$\frac{a \equiv b \quad b \equiv c}{a \equiv c} \; \equiv\text{-}trans$$

Just like rules for $=$.

Rules for $\equiv$: λ (i.e., α, β, η) Conversions

$$\frac{}{(\lambda x.a) \equiv (\lambda y.a[x \leftarrow y])} \; \alpha^{*} \qquad \frac{}{(\lambda x.a)b \equiv (a[x \leftarrow b])} \; \beta$$

$$\frac{}{(\lambda x.f\,x) \equiv f} \; \eta^{**}$$

Side condition $*$: y is not free in a.

Side condition $**$: x is not free in f.

Just like rules for $=_{\alpha,\beta,\eta}$.

η is equivalent to extensionality[310].

---

[310]Extensionality is the rule

$$\frac{f\,x \equiv g\,x}{f \equiv g}$$

where the side condition is that x must not be free in f or g or any assumption on which the proof of $f\,x \equiv g\,x$ depend. It is equivalent to the η-axiom [HS90, pages 72-74].

Recall that we have used the notion of extensionality before, for sets. The idea is the same here.

Rules for $\equiv$: Abstraction, Combination

$$\frac{a \equiv b}{(\lambda x.a) \equiv (\lambda x.b)} \equiv\text{-}abstr^* \qquad \frac{f \equiv g \quad a \equiv b}{f\ a \equiv g\ b} \equiv\text{-}comb$$

Side (eigenvariable) condition $*$: x is not free in any assumption on which $a \equiv b$ depends.

As defined for $\rightarrow_\beta$ before, $\equiv$ is propagated into contexts.

Conversion is built into the proof system!

Recall that $e \equiv e'$ is decidable in $\lambda^\rightarrow$ ($\equiv$-rules so far).

However, $e \equiv e'$ is not decidable in $\mathcal{M}$ (see next slide).

Rules for ≡: Introduction and Elimination

$$
\begin{array}{cc}
[\phi] & [\psi] \\
\vdots & \vdots \\
\psi & \phi \\
\end{array}
$$

$$
\dfrac{\psi \quad \phi}{\phi \equiv \psi} \; \equiv\text{-}I
\qquad
\dfrac{\phi \equiv \psi \quad \phi}{\psi} \; \equiv\text{-}E
$$

What is the type of ϕ and ψ here?

Rules for ≡: Introduction and Elimination

$$\frac{\begin{array}{cc} [\phi] & [\psi] \\ \vdots & \vdots \\ \psi & \phi \end{array}}{\phi \equiv \psi} \text{ ≡-I} \qquad \frac{\phi \equiv \psi \quad \phi}{\psi} \text{ ≡-E}$$

What is the type of ϕ and ψ here? ϕ and ψ are formulae, hence $bool$.

What object-level connective does ≡ correspond to?

Rules for $\equiv$: Introduction and Elimination

$$
\begin{array}{cc}
[\phi] & [\psi] \\
\vdots & \vdots \\
\psi & \phi \\
\hline
\phi \equiv \psi
\end{array} \equiv\text{-}I
\qquad
\dfrac{\phi \equiv \psi \quad \phi}{\psi} \equiv\text{-}E
$$

What is the type of ϕ and ψ here? ϕ and ψ are formulae, hence $bool$.

What object-level connective does $\equiv$ correspond to? $\leftrightarrow$.

13.2 Encoding Syntax and Provability

We use FOL and its subset propositional logic (which we call here $Prop$) as exemplary object logic.

We already know how to encode syntax.

We will now see how to encode proof rules and mimic proofs of the object logic.

To encode a particular object logic L, we have to extend $\mathcal{M}$ by extending the type language, the term language (the signature) and the proof rules. The thus extended logic will be called $\mathcal{M}_L$.

Encoding Syntax: Review

As before, $i, o \in \mathcal{B}$. Previously:

$$\Sigma \supseteq \ \langle not : o \rightarrow o, and : o \rightarrow o \rightarrow o, imp : o \rightarrow o \rightarrow o,$$
$$all : (i \rightarrow o) \rightarrow o, exists : (i \rightarrow o) \rightarrow o \rangle$$

Two types[311] for truth values: o and $bool$.

We now need a more concise (sweeter) syntax or things will become hopelessly unreadable.

But this is also quite demanding: you should always be able to "unsugar" the syntax.

---

[311]So we have truth values in the metalogic (type $bool$) and in the object logic (type o). To distinguish them clearly there are two different types for them.

Encoding Syntax Readably

$$\Sigma \supseteq \langle \perp : o,$$
$$\wedge, \vee, \to \,^{312} : o \to o \to o,$$
$$\forall, \exists : (i \to o) \to o,$$
$$true : o \to bool \rangle.$$

$\to$ is both a constant declared in Σ and the function type arrow.

$\wedge, \vee, \to$ will be written infix, and we may write $\forall x.\phi$ for $\forall(\lambda x.\phi)$, and likewise for $\exists$.

$true\ A^{313}$ is usually written $[A]$.

---

[312]We write

$$\langle \perp : o,$$
$$\wedge, \vee, \to : o \to o \to o,$$
$$\forall, \exists : (i \to o) \to o,$$
$$true : o \to bool \rangle$$

as shorthand for

$$\langle \perp : o,$$
$$\wedge : o \to o \to o,$$
$$\vee : o \to o \to o,$$
$$\to : o \to o \to o,$$
$$\forall : (i \to o) \to o,$$
$$\exists : (i \to o) \to o$$
$$true : o \to bool \rangle$$

[313]So we have truth values in the metalogic (type $bool$) and in the object logic (type o).

Paulson [Pau89] says: "the meta-formula $[A]$ abbreviates $true\ A$ and means that A is true". More precisely, we can say that $[A]$ is a meta-formula that may or may not be derivable

Encoding the Rules

The rules of the object logic are encoded as axioms of the metalogic. These axioms are added to the proof system of $\mathcal{M}$ (to obtain $\mathcal{M}_L$).

To avoid confusion, we will use distinctive terminology:

There is a meta-rule called $\Rightarrow$-E.

There is a similar object rule that we call the $\rightarrow$-E rule.

It is encoded as a meta-axiom that we call the $\rightarrow$-E axiom.

---

in $\mathcal{M}_L$, and that this should reflect derivability of A in L.

In the file `IFOL.thy` in your Isabelle distribution, you find

```
Trueprop    :: "o => prop"
```

`Trueprop` corresponds to $true$.

Encoding of the Rules of Propositional Logic

$$\bigwedge AB.[A] \Rightarrow ([B] \Rightarrow [A \wedge B]) \qquad (\wedge\text{-}I)$$

$$\bigwedge AB.[A \wedge B] \Rightarrow [A] \qquad (\wedge\text{-}EL)$$

$$\bigwedge AB.[A \wedge B] \Rightarrow [B] \qquad (\wedge\text{-}ER)$$

$$\bigwedge AB.[A] \Rightarrow [A \vee B] \qquad (\vee\text{-}IL)$$

$$\bigwedge AB.[B] \Rightarrow [A \vee B] \qquad (\vee\text{-}IR)$$

$$\bigwedge ABC.[A \vee B] \Rightarrow \\ ([A] \Rightarrow [C]) \Rightarrow ([B] \Rightarrow [C]) \Rightarrow [C] \qquad (\vee\text{-}E)$$

$$\bigwedge AB.([A] \Rightarrow [B]) \Rightarrow [A \to B] \qquad (\to\text{-}I)$$

$$\bigwedge AB.[A \to B] \Rightarrow [A] \Rightarrow [B] \qquad (\to\text{-}E)$$

$$\bigwedge A.[\bot] \Rightarrow [A] \qquad (\bot\text{-}E)$$

Faithful Metalogics

For any object logic L, we define:

- $\mathcal{M}_L$ is sound for L if, for every proof of $[\![B]\!]$ from assumptions $[\![A_1]\!], \ldots, [\![A_m]\!]$ in $\mathcal{M}_L$, there is a proof of B from assumptions $A_1, \ldots, A_m$ in L.

- $\mathcal{M}_L$ is complete for L if, for every proof of B from assumptions $A_1, \ldots, A_m$ in L, there is a proof of $[\![B]\!]$ from assumptions $[\![A_1]\!], \ldots, [\![A_m]\!]$ in $\mathcal{M}_L$.

- $\mathcal{M}_L$ is faithful for L if $\mathcal{M}_L$ is sound and complete for L.

Using concepts of Prawitz [Pra65, Pra71], one can show by structural induction that $\mathcal{M}_{Prop}$ is faithful for $Prop$.

An Example Proof

$$\frac{}{\bigwedge AB.[\![A \wedge B]\!] \Rightarrow [\![A]\!]} \wedge\text{-}EL$$

An Example Proof

$$\frac{\overline{\bigwedge AB.[\![A \wedge B]\!]}}{\begin{array}{c}\Rightarrow [\![A]\!]\end{array}} \wedge\text{-}EL$$

$$\frac{}{\begin{array}{c}\bigwedge B.[\![P \wedge B]\!] \\ \Rightarrow [\![P]\!]\end{array}} \bigwedge\text{-}E$$

An Example Proof

$$\dfrac{\dfrac{\dfrac{}{\bigwedge AB.[\![A \wedge B]\!] \Rightarrow [\![A]\!]} \;\wedge\text{-}EL}{\bigwedge B.[\![P \wedge B]\!] \Rightarrow [\![P]\!]} \;\bigwedge\text{-}E}{[\![P \wedge Q]\!] \Rightarrow [\![P]\!]} \;\bigwedge\text{-}E$$

An Example Proof

$$\cfrac{\cfrac{\cfrac{\overline{\bigwedge AB.[A \wedge B] \Rightarrow [\![A]\!]} \; \wedge\text{-}EL}{\bigwedge B.[P \wedge B] \Rightarrow [\![P]\!]} \; \bigwedge\text{-}E}{[\![P \wedge Q]\!] \Rightarrow [\![P]\!]} \; \bigwedge\text{-}E \qquad [\![P \wedge Q]\!]}{[\![P]\!]} \; \Rightarrow\text{-}E$$

An Example Proof

$$
\cfrac{
  \cfrac{
    \cfrac{
      \cfrac{
        \overline{\bigwedge AB.[A \wedge B] \Rightarrow [A]}\ \wedge\text{-}EL
      }{\bigwedge B.[P \wedge B] \Rightarrow [P]}\ \bigwedge\text{-}E
    }{[P \wedge Q] \Rightarrow [P]}\ \bigwedge\text{-}E
  }{
    [[P \wedge Q]]^1 \qquad [P \wedge Q] \Rightarrow [P]
  }
}{
  \cfrac{[P]}{[P \wedge Q] \Rightarrow [P]}\ \Rightarrow\text{-}I^1
}\ \Rightarrow\text{-}E
$$

An Example Proof

$$\frac{\overline{\bigwedge AB.([A] \Rightarrow [B])} \quad \rightarrow\text{-}I}{\frac{\Rightarrow [A \to B]}{\frac{\bigwedge B.([P \wedge Q] \Rightarrow [B])}{\Rightarrow [P \wedge Q \to B]} \quad \bigwedge\text{-}E}{\frac{([P \wedge Q] \Rightarrow [P])}{\Rightarrow [P \wedge Q \to P]}} \quad \bigwedge\text{-}E}$$

$$\frac{\overline{\bigwedge AB.[A \wedge B]} \quad \wedge\text{-}EL}{\frac{\Rightarrow [A]}{\frac{\bigwedge B.[P \wedge B]}{\Rightarrow [P]}} \quad \bigwedge\text{-}E}$$

$$\frac{[[P \wedge Q]]^1 \quad \overline{[P \wedge Q] \Rightarrow [P]} \quad \bigwedge\text{-}E}{\frac{[P]}{[P \wedge Q] \Rightarrow [P]} \quad \Rightarrow\text{-}I^1} \quad \Rightarrow\text{-}E$$

An Example Proof

$$
\cfrac{
  \cfrac{
    \cfrac{
      \overline{\bigwedge AB.([\![A]\!] \Rightarrow [\![B]\!]) \Rightarrow [\![A \to B]\!]} \;\; \text{→-}I
    }{
      \bigwedge B.([\![P \wedge Q]\!] \Rightarrow [\![B]\!]) \Rightarrow [\![P \wedge Q \to B]\!]
    } \;\; \text{\Large$\wedge$-}E
  }{
    ([\![P \wedge Q]\!] \Rightarrow [\![P]\!]) \Rightarrow [\![P \wedge Q \to P]\!]
  } \;\; \text{\Large$\wedge$-}E
  \qquad
  \cfrac{
    [[P \wedge Q]]^1 \quad
    \cfrac{
      \cfrac{
        \cfrac{
          \overline{\bigwedge AB.[\![A \wedge B]\!] \Rightarrow [\![A]\!]} \;\; \text{$\wedge$-}EL
        }{
          \bigwedge B.[\![P \wedge B]\!] \Rightarrow [\![P]\!]
        } \;\; \text{\Large$\wedge$-}E
      }{
        [\![P \wedge Q]\!] \Rightarrow [\![P]\!]
      } \;\; \text{\Large$\wedge$-}E
    }
    {\;}
  }{
    \cfrac{[\![P]\!]}{[\![P \wedge Q]\!] \Rightarrow [\![P]\!]} \;\; \text{⇒-}I^1
  } \;\; \text{⇒-}E
}{
  [\![P \wedge Q \to P]\!]
} \;\; \text{⇒-}E
$$

302

An Example Proof

$$\cfrac{\cfrac{\cfrac{\overline{\bigwedge AB.([A] \Rightarrow [B])}}{\Rightarrow [A \to B]} \;\text{→-I}}{\cfrac{\bigwedge B.([P \land Q] \Rightarrow [B])}{\Rightarrow [P \land Q \to B]} \;\text{$\bigwedge$-E}}}{\cfrac{([P \land Q] \Rightarrow [P])}{\Rightarrow [P \land Q \to P]} \;\text{$\bigwedge$-E}} \qquad \cfrac{[[P \land Q]]^1 \quad \cfrac{\cfrac{\cfrac{\overline{\bigwedge AB.[A \land B]}}{\Rightarrow [A]} \;\text{$\wedge$-EL}}{\cfrac{\bigwedge B.[P \land B]}{\Rightarrow [P]} \;\text{$\bigwedge$-E}}}{[P \land Q] \Rightarrow [P]} \;\text{$\bigwedge$-E}}{\cfrac{[P]}{[P \land Q] \Rightarrow [P]} \;\text{⇒-I}^1} \;\text{⇒-E}$$

$$\cfrac{}{[P \land Q \to P]} \;\text{⇒-E}$$

Example Proof Simplified

$$
\cfrac{
  \cfrac{
    \cfrac{
      \overline{\textstyle\bigwedge AB.([A] \Rightarrow [B]) \Rightarrow [A \to B]} \;\; \to\text{-}I
    }{
      \textstyle\bigwedge B.([P \land Q] \Rightarrow [B]) \Rightarrow [P \land Q \to B]
    } \;\; \bigwedge\text{-}E
  }{
    ([P \land Q] \Rightarrow [P]) \Rightarrow [P \land Q \to P]
  } \;\; \bigwedge\text{-}E
  \qquad
  \cfrac{
    \cfrac{
      \cfrac{
        \overline{\textstyle\bigwedge AB.[A \land B] \Rightarrow [A]} \;\; \land\text{-}EL
      }{
        \textstyle\bigwedge B.[P \land B] \Rightarrow [P]
      } \;\; \bigwedge\text{-}E
    }{
      [P \land Q] \Rightarrow [P]
    } \;\; \bigwedge\text{-}E
  }{}
}{
  [P \land Q \to P]
} \;\; \Rightarrow\text{-}E
$$

Remarks about Example Proof

- $\wedge$-*EL* and $\rightarrow$-*E* are not object rules but meta-axioms!

- The first, more complicated proof corresponds to the construction one would use to show that $\mathcal{M}_{Prop}$ is complete for $Prop$.

- Proof fragments of the form

$$\cfrac{\cfrac{\phi \Rightarrow \psi \quad [\phi]}{\psi} \Rightarrow\text{-}E}{\phi \Rightarrow \psi} \Rightarrow\text{-}I$$

can be collapsed into $\phi \Rightarrow \psi$: proof normalization.

13.3 Reasoning with Resolution

In Isabelle, we mainly use backwards reasoning: we construct a proof tree starting from the root working to the leaves.

13.3 Reasoning with Resolution

In Isabelle, we mainly use backwards reasoning: we construct a proof tree starting from the root working to the leaves.

On the meta-level, this proof is in fact a forwards proof: working from the leaves to the root.

This is achieved by starting the proof of ψ with the trivial meta-theorem $\psi \Rightarrow \psi$[314] and using a technique called resolution.

---

[314]We have seen this before as a proof in propositional logic.

$$\frac{[\psi]^1}{\psi \to \psi} \Rightarrow\text{-}I^1$$

The Resolution Rule

For any formulae $\psi_1, \ldots, \psi_n, \psi, \phi_1, \ldots, \phi_m, \phi$ where $FV(\phi_1, \ldots, \phi_m, \phi) \subseteq$ $\{x_1, \ldots, x_k\}$, and $\phi\theta \equiv \psi_i$ for some $i \in \{1, \ldots, n\}$, resolution is the following rule:

$$\frac{\bigwedge x_1 \ldots x_k.[\phi_1, \ldots, \phi_m] \Rightarrow \phi \quad [\psi_1, \ldots, \psi_n] \Rightarrow \psi}{[\psi_1, \ldots, \psi_{i-1}, \phi_1\theta, \ldots, \phi_m\theta, \psi_{i+1}, \ldots, \psi_n] \Rightarrow \psi} \; res$$

Intuition: $\bigwedge x_1 \ldots x_k.[\phi_1, \ldots, \phi_m] \Rightarrow \phi$ is a meta-axiom such as $\wedge$-*EL*, $[\psi_1, \ldots, \psi_n] \Rightarrow \psi$ is the current goal (proof state).

Compare to phrasing using $\vee$[315]!

We will now derive the rule.

---

[315]You may have seen the following formulation of the resolution rule:

$$\frac{A_1 \vee \ldots \vee A_n \quad B_1 \vee \ldots \vee B_m}{(A_1 \vee \ldots \vee A_{i-1}, A_{i+1} \vee \ldots \vee A_n \vee B_1 \vee \ldots \vee B_{j-i}, B_{j+1} \vee \ldots \vee B_m)\theta}$$

where either $A_i\theta = \neg B_j\theta$ or $\neg A_i\theta = B_j\theta$.

You can see the correspondence to the rule given here by recalling that in first-order logic, $\phi_1 \to \ldots \to \phi_m \to \phi$ is equivalent to $\phi_1 \wedge \ldots \wedge \phi_m \to \phi$, which is in turn equivalent to $\neg\phi_1 \vee \ldots \vee \neg\phi_m \vee \phi$.

You may still be wondering though why in the rule *res*, we only allow instantiation of $[\phi_1, \ldots, \phi_m] \Rightarrow \phi$. This restriction will in fact be lifted later.

Resolution as Derived Meta-Rule

$$\bigwedge x_1 \ldots x_k.$$
$$[\phi_1, \ldots, \phi_m] \Rightarrow \phi$$

---

[316]Recall that $\phi\theta \equiv \psi_i$.

Resolution as Derived Meta-Rule

$$\frac{\begin{array}{c} \bigwedge x_1 \ldots x_k.\\ [\phi_1, \ldots, \phi_m] \Rightarrow \phi \end{array}}{[\phi_1\theta, \ldots, \phi_m\theta] \Rightarrow \phi\theta} \; \bigwedge\text{-}E$$

---

[316]Recall that $\phi\theta \equiv \psi_i$.

Resolution as Derived Meta-Rule

$$
\phi_1\theta \quad \ldots \quad \phi_m\theta \qquad \cfrac{\begin{array}{c}\bigwedge x_1\ldots x_k.\\ [\phi_1,\ldots,\phi_m]\Rightarrow\phi\end{array}}{[\phi_1\theta,\ldots,\phi_m\theta]\Rightarrow\phi\theta}\ \bigwedge\text{-}E
$$

---

[316]Recall that $\phi\theta \equiv \psi_i$.

307

Resolution as Derived Meta-Rule

$$
\cfrac{\phi_1\theta \quad \ldots \quad \phi_m\theta \quad \cfrac{\cfrac{\bigwedge x_1 \ldots x_k.\ [\phi_1, \ldots, \phi_m] \Rightarrow \phi}{[\phi_1\theta, \ldots, \phi_m\theta] \Rightarrow \phi\theta}\ \text{$\bigwedge$-}E}{}}{\phi\theta}\ \Rightarrow\text{-}E
$$

---

[316] Recall that $\phi\theta \equiv \psi_i$.

Resolution as Derived Meta-Rule

$$
\dfrac{\phi_1\theta \quad \ldots \quad \phi_m\theta \quad \dfrac{\dfrac{\bigwedge x_1\ldots x_k.\; [\phi_1,\ldots,\phi_m] \Rightarrow \phi}{[\phi_1\theta,\ldots,\phi_m\theta] \Rightarrow \phi\theta}\; \wedge\text{-}E}{}}{\phi\theta}\; \Rightarrow\text{-}E
$$

$$
\begin{array}{c} [\psi_1,\ldots,\psi_n] \\ \Rightarrow \psi \end{array}
$$

---

[316] Recall that $\phi\theta \equiv \psi_i$.

Resolution as Derived Meta-Rule

$$
\cfrac{\phi_1\theta \quad \dots \quad \phi_m\theta \qquad \cfrac{\bigwedge x_1\dots x_k.\,[\phi_1,\dots,\phi_m]\Rightarrow\phi}{[\phi_1\theta,\dots,\phi_m\theta]\Rightarrow\phi\theta}\ \bigwedge\text{-}E}{\phi\theta}\ \Rightarrow\text{-}E
\qquad
\psi_1 \quad \dots \ \psi_{i-1}
\qquad
\begin{array}{c}[\psi_1,\dots,\psi_n]\\ \Rightarrow\psi\end{array}
$$

---

[316] Recall that $\phi\theta \equiv \psi_i$.

Resolution as Derived Meta-Rule

$$\dfrac{\phi_1\theta \quad \dots \quad \phi_m\theta \qquad \dfrac{\bigwedge x_1\dots x_k.\ [\phi_1,\dots,\phi_m]\Rightarrow\phi}{[\phi_1\theta,\dots,\phi_m\theta]\Rightarrow\phi\theta}\ \bigwedge\text{-}E}{\phi\theta}\ \Rightarrow\text{-}E \qquad \dfrac{\psi_1 \quad \dots \ \psi_{i-1} \qquad \begin{matrix}[\psi_1,\dots,\psi_n]\\ \Rightarrow\psi\end{matrix}}{[\psi_i,\dots,\psi_n]\Rightarrow\psi}\ \Rightarrow\text{-}E$$

---

[316]Recall that $\phi\theta \equiv \psi_i$.

Resolution as Derived Meta-Rule

$$\cfrac{\phi_1\theta \quad \ldots \quad \phi_m\theta \quad \cfrac{\cfrac{\bigwedge x_1 \ldots x_k.}{[\phi_1, \ldots, \phi_m] \Rightarrow \phi}}{[\phi_1\theta, \ldots, \phi_m\theta] \Rightarrow \phi\theta} \wedge\text{-}E}{\phi\theta} \Rightarrow\text{-}E \quad \cfrac{\psi_1 \quad \ldots \quad \psi_{i-1} \quad \cfrac{[\psi_1, \ldots, \psi_n] \Rightarrow \psi}{[\psi_i, \ldots, \psi_n] \Rightarrow \psi} \Rightarrow\text{-}E}{[\psi_{i+1}, \ldots, \psi_n] \Rightarrow \psi} \Rightarrow\text{-}E^{316}$$

---

[316]Recall that $\phi\theta \equiv \psi_i$.

Resolution as Derived Meta-Rule

$$\cfrac{[\phi_1\theta]^2 \ldots [\phi_m\theta]^2 \quad \cfrac{\bigwedge x_1 \ldots x_k.\ [\phi_1, \ldots, \phi_m] \Rightarrow \phi}{[\phi_1\theta, \ldots, \phi_m\theta] \Rightarrow \phi\theta} \bigwedge\text{-}E}{\cfrac{\phi\theta \qquad \qquad \qquad \psi_1 \quad \ldots \quad \psi_{i-1} \qquad \cfrac{[\psi_1, \ldots, \psi_n] \Rightarrow \psi}{[\psi_i, \ldots, \psi_n] \Rightarrow \psi} \Rightarrow\text{-}E}{[\psi_{i+1}, \ldots, \psi_n] \Rightarrow \psi} \Rightarrow\text{-}E^{316}} \Rightarrow\text{-}E}{[\phi_1\theta, \ldots, \phi_m\theta, \psi_{i+1}, \ldots, \psi_n] \Rightarrow \psi} \Rightarrow\text{-}I^2$$

---

[316] Recall that $\phi\theta \equiv \psi_i$.

Resolution as Derived Meta-Rule

$$\cfrac{[\phi_1\theta]^2\dots[\phi_m\theta]^2 \quad \cfrac{\bigwedge x_1\dots x_k.\;[\phi_1,\dots,\phi_m]\Rightarrow\phi}{[\phi_1\theta,\dots,\phi_m\theta]\Rightarrow\phi\theta}\;\bigwedge\text{-}E}{\cfrac{\phi\theta \qquad \cfrac{[\psi_1]^1\dots[\psi_{i-1}]^1 \quad \cfrac{[\psi_1,\dots,\psi_n]\Rightarrow\psi}{[\psi_i,\dots,\psi_n]\Rightarrow\psi}\;\Rightarrow\text{-}E}{[\psi_{i+1},\dots,\psi_n]\Rightarrow\psi}\;\Rightarrow\text{-}E^{316}}{\cfrac{[\phi_1\theta,\dots,\phi_m\theta,\psi_{i+1},\dots,\psi_n]\Rightarrow\psi}{[\psi_1,\dots,\psi_{i-1},\phi_1\theta,\dots,\phi_m\theta,\psi_{i+1},\dots,\psi_n]\Rightarrow\psi}\;\Rightarrow\text{-}I^1}\;\Rightarrow\text{-}I^2}\;\Rightarrow\text{-}E}$$

---

[316] Recall that $\phi\theta \equiv \psi_i$.

Deriving Resolution: Remarks

- We collapsed iterated applications of rules (denoted by double horizontal line).

Deriving Resolution: Remarks

- We collapsed iterated applications of rules (denoted by double horizontal line).

- This is not just a matter of simplicity. The derivation is schematic not just in the sense that the Greek letters could stand for arbitrary formulae; we don't even know how many formulae are involved (k, m, n, i could be any natural numbers).

- But for any concrete $\psi_1, \ldots, \psi_n, \psi, \phi_1, \ldots, \phi_m, \phi$, you could do the formal derivation in $\mathcal{M}$.

Dropping Outer Quantifiers

We adopt the convention that outer quantifiers in meta-formulae are dropped. E.g. $[A] \Rightarrow [B] \Rightarrow [A \wedge B]$ instead of $\bigwedge AB.[A] \Rightarrow [B] \Rightarrow [A \wedge B]$.

In addition: use renaming for freshness[317].

Then we can write the resolution rule as follows:

$$\frac{[\phi_1, \ldots, \phi_m] \Rightarrow \phi \quad [\psi_1, \ldots, \psi_n] \Rightarrow \psi}{[\psi_1, \ldots, \psi_{i-1}, \phi_1\theta, \ldots, \phi_m\theta, \psi_{i+1}, \ldots, \psi_n] \Rightarrow \psi} \; res$$

where $\phi\theta \equiv \psi_i$.

We will now work with this schematic form.

---

[317] The schematic form of the resolution rule is:

$$\frac{[\phi_1, \ldots, \phi_m] \Rightarrow \phi \quad [\psi_1, \ldots, \psi_n] \Rightarrow \psi}{[\psi_1, \ldots, \psi_{i-1}, \phi_1\theta, \ldots, \phi_m\theta, \psi_{i+1}, \ldots, \psi_n] \Rightarrow \psi} \; res$$

where $\phi\theta \equiv \psi$.

We will work with this schematic form, but remember: if necessary, you could construct an actual derivation in $\mathcal{M}$.

In this schematic form, it is always assumed that the free variables in $[\phi_1, \ldots, \phi_m] \Rightarrow \phi$ are fresh, i.e. $FV([\phi_1, \ldots, \phi_m] \Rightarrow \phi) \cap FV([\psi_1, \ldots, \psi_n] \Rightarrow \psi) = \emptyset$.

This assumption may be justified considering the formal derivation of the resolution rule. Suppose that the free variables in $[\phi_1, \ldots, \phi_m] \Rightarrow \phi$ are not all fresh, and consider $\bigwedge x_1' \ldots x_k'.[\phi_1', \ldots, \phi_m'] \Rightarrow \phi'$, obtained from $\bigwedge x_1 \ldots x_k.[\phi_1, \ldots, \phi_m] \Rightarrow \phi$ by replacing each x_i with x_i', where the x_i' are fresh.

It is easy to see that in the formal derivation of the resolution

Proof of $A \wedge B \to C \to A \wedge C$ **(1)**

Let's prove $A \wedge B \to (C \to A \wedge C)$ by resolution. We start by resolution with →-*I*:

$$\frac{([A_1] \Rightarrow [B_1]) \qquad [A \wedge B \to (C \to A \wedge C)]}{([A \wedge B] \Rightarrow [C \to A \wedge C])} \text{ res}$$
$$\Rightarrow [A_1 \to B_1] \qquad \Rightarrow [A \wedge B \to (C \to A \wedge C)]$$
$$\Rightarrow [A \wedge B \to (C \to A \wedge C)]$$

rule, one can replace

$$\frac{\bigwedge x_1 \ldots x_k.[\phi_1, \ldots, \phi_m] \Rightarrow \phi}{[\phi_1\theta, \ldots, \phi_m\theta] \Rightarrow \phi\theta} \text{ } \bigwedge\text{-}E$$

with

$$\frac{\bigwedge x'_1 \ldots x'_k.[\phi'_1, \ldots, \phi'_m] \Rightarrow \phi'}{[\phi_1\theta, \ldots, \phi_m\theta] \Rightarrow \phi\theta} \text{ } \bigwedge\text{-}E$$

Therefore we can assume without loss of generality that the free variables in $[\phi_1, \ldots, \phi_m] \Rightarrow \phi$ are fresh.

The next question is: why do we want fresh variables? Maybe this is clear intuitively: A rule is always meant to be schematic and the choice of variables names in a rule should be irrelevant. More concretely, one may say that if one does not rename the variables in a rule and hence there is some variable, say A, that occurs in the current subgoal, then resolution may lead to a subgoal containing occurrences of A originating from the goal and others originating from the rule, and these are inadvertently identified, leading to a proof state

What to do next[318]?

---

that is more instantiated than it should be.

[318]On the one hand, we want to resolve

$$([A \wedge B] \Rightarrow [C \rightarrow A \wedge C]) \Rightarrow [A \wedge B \rightarrow (C \rightarrow A \wedge C)],$$

i.e., we have to match $([A \wedge B] \Rightarrow [C \rightarrow A \wedge C])$ against the conclusion of some meta-axiom.

On the other hand, think what Isabelle would display in this situation. The (only) subgoal would be

$$1. \ A \wedge B \Rightarrow C \rightarrow A \wedge C,$$

so we have to show $C \rightarrow A \wedge C$ (using assumption $A \wedge B$). So you should look at $C \rightarrow A \wedge C$ to guess which meta-axiom should be used now.

[319]In our current situation, Isabelle would display:

$$\text{Level } 1(1 \text{ subgoal})$$
$$A \wedge B \rightarrow (C \rightarrow A \wedge C)$$
$$1. \ A \wedge B \Longrightarrow C \rightarrow A \wedge C$$

From your experience with Isabelle, it is clear that since the

What to do next[318]? Again resolution with →-*I*.

Problem: the conclusion of →-*I* is not unifiable[319] with $[A \wedge B] \Rightarrow [C \rightarrow A \wedge C]$.

---

that is more instantiated than it should be.

[318]On the one hand, we want to resolve

$$([A \wedge B] \Rightarrow [C \rightarrow A \wedge C]) \Rightarrow [A \wedge B \rightarrow (C \rightarrow A \wedge C)],$$

i.e., we have to match $([A \wedge B] \Rightarrow [C \rightarrow A \wedge C])$ against the conclusion of some meta-axiom.

On the other hand, think what Isabelle would display in this situation. The (only) subgoal would be

$$1.\ A \wedge B \Rightarrow C \rightarrow A \wedge C,$$

so we have to show $C \rightarrow A \wedge C$ (using assumption $A \wedge B$). So you should look at $C \rightarrow A \wedge C$ to guess which meta-axiom should be used now.

[319]In our current situation, Isabelle would display:

$$\texttt{Level } 1(1 \texttt{ subgoal})$$
$$A \wedge B \rightarrow (C \rightarrow A \wedge C)$$
$$1.\ A \wedge B \Longrightarrow C \rightarrow A \wedge C$$

From your experience with Isabelle, it is clear that since the

Lifting over Assumptions

The rule for lifting an object rule (meta-axiom) $[\phi_1, \ldots, \phi_m] \Rightarrow \phi$ over a list of assumptions Ψ is

$$\frac{[\phi_1, \ldots, \phi_m] \Rightarrow \phi}{[\Psi \Rightarrow \phi_1, \ldots, \Psi \Rightarrow \phi_m] \Rightarrow (\Psi \Rightarrow \phi)} \ \textit{a-lift}$$

We will now derive it for one assumption, so $\Psi = [\psi]$.

---

top-level symbol in $C \rightarrow A \wedge C$ is $\rightarrow$, you would use $\rightarrow$-*I*.

But look at the resolution rule again. We would take a fresh instance of $\rightarrow$-*I*, say $([A_2] \Rightarrow [B_2]) \Rightarrow [A_2 \rightarrow B_2]$. The problem is that $[A_2 \rightarrow B_2]$ is not unifiable with $[A \wedge B] \Rightarrow [C \rightarrow A \wedge C]$, and so *res* is not applicable.

Deriving Assumption Lifting for one Assumption

$$[\phi_1, \ldots, \phi_m] \Rightarrow \phi \qquad \dfrac{\psi \Rightarrow \phi_1 \qquad \psi}{\phi_1} \Rightarrow\text{-}E \ldots \dfrac{\psi \Rightarrow \phi_m \qquad \psi}{\phi_m} \Rightarrow\text{-}E$$

$$\psi \Rightarrow \phi$$

Deriving Assumption Lifting for **one** Assumption

$$[\phi_1, \ldots, \phi_m] \Rightarrow \phi \quad \cfrac{\cfrac{\psi \Rightarrow \phi_1 \qquad \psi}{\phi_1} \Rightarrow\text{-}E \quad \ldots \quad \cfrac{\psi \Rightarrow \phi_m \qquad \psi}{\phi_m} \Rightarrow\text{-}E}{\phi} \Rightarrow\text{-}E$$

Deriving Assumption Lifting for one Assumption

$$\cfrac{[\phi_1, \ldots, \phi_m] \Rightarrow \phi \qquad \cfrac{\cfrac{\psi \Rightarrow \phi_1 \qquad [\psi]^2}{\phi_1} \Rightarrow\text{-}E \ldots \cfrac{\psi \Rightarrow \phi_m \qquad [\psi]^2}{\phi_m} \Rightarrow\text{-}E}{\phi} \Rightarrow\text{-}E}{\cfrac{\phi}{\psi \Rightarrow \phi} \Rightarrow\text{-}I^2}$$

Deriving Assumption Lifting for **one** Assumption

$$\cfrac{[\phi_1, \ldots, \phi_m] \Rightarrow \phi \qquad \cfrac{\cfrac{[\psi \Rightarrow \phi_1]^1 \quad [\psi]^2}{\phi_1} \Rightarrow\text{-}E \ \ldots \ \cfrac{[\psi \Rightarrow \phi_m]^1 \quad [\psi]^2}{\phi_m} \Rightarrow\text{-}E}{\cfrac{\cfrac{\phi}{\psi \Rightarrow \phi} \Rightarrow\text{-}I^2}{} } \Rightarrow\text{-}E}{[\psi \Rightarrow \phi_1, \ldots, \psi \Rightarrow \phi_m] \Rightarrow (\psi \Rightarrow \phi)} \Rightarrow\text{-}I^1$$

This process can be repeated for any number of assumptions to get the general rule.

Proof of $A \wedge B \to (C \to A \wedge C)$ **(2)**

We do resolution using the $\to$-*I* rule[320] lifted over $A \wedge B$:

$$\frac{([\![A \wedge B]\!] \Rightarrow ([\![A_2]\!] \Rightarrow [\![B_2]\!])) \quad \overset{\vdots}{([\![A \wedge B]\!] \Rightarrow [\![C \to A \wedge C]\!])}}{\Rightarrow ([\![A \wedge B]\!] \Rightarrow [\![A_2 \to B_2]\!]) \quad \Rightarrow [\![A \wedge B \to (C \to A \wedge C)]\!]} \; res}{\begin{array}{c} ([\![A \wedge B]\!] \Rightarrow [\![C]\!] \Rightarrow [\![A \wedge C]\!]) \\ \Rightarrow [\![A \wedge B \to (C \to A \wedge C)]\!] \end{array}}$$

320

$$([\![A \wedge B]\!] \Rightarrow ([\![A_2]\!] \Rightarrow [\![B_2]\!])) \Rightarrow ([\![A \wedge B]\!] \Rightarrow [\![A_2 \to B_2]\!])$$

is the $\to$-*I*-rule (meta-axiom) lifted over the assumption $A \wedge B$.

Proof of $A \wedge B \to (C \to A \wedge C)$ **(2)**

We do resolution using the $\to$-*I* rule[320] lifted over $A \wedge B$:

$$
\dfrac{
\begin{array}{cc}
\begin{array}{l}
(\llbracket A \wedge B \rrbracket \Rightarrow (\llbracket A_2 \rrbracket \Rightarrow \llbracket B_2 \rrbracket)) \\
\Rightarrow (\llbracket A \wedge B \rrbracket \Rightarrow \llbracket A_2 \to B_2 \rrbracket)
\end{array}
&
\dfrac{\vdots}{
\begin{array}{l}
(\llbracket A \wedge B \rrbracket \Rightarrow \llbracket C \to A \wedge C \rrbracket) \\
\Rightarrow \llbracket A \wedge B \to (C \to A \wedge C) \rrbracket
\end{array}
}
\end{array}
}{
\begin{array}{c}
(\llbracket A \wedge B \rrbracket \Rightarrow \llbracket C \rrbracket \Rightarrow \llbracket A \wedge C \rrbracket) \\
\Rightarrow \llbracket A \wedge B \to (C \to A \wedge C) \rrbracket
\end{array}
} \; res
$$

Before we proceed, we introduce the abbreviations

$$ \omega = \llbracket A \wedge B \to (C \to A \wedge C) \rrbracket, \;\; \Omega = [\llbracket A \wedge B \rrbracket, \llbracket C \rrbracket] $$

---

[320]

$$ (\llbracket A \wedge B \rrbracket \Rightarrow (\llbracket A_2 \rrbracket \Rightarrow \llbracket B_2 \rrbracket)) \Rightarrow (\llbracket A \wedge B \rrbracket \Rightarrow \llbracket A_2 \to B_2 \rrbracket) $$

is the $\to$-*I*-rule (meta-axiom) lifted over the assumption $A \wedge B$.

Proof of $A \wedge B \to (C \to A \wedge C)$ **(2)**

We do resolution using the $\to$-*I* rule[320] lifted over $A \wedge B$:

$$
\frac{
\begin{array}{cc}
\begin{array}{l}
([\![A \wedge B]\!] \Rightarrow ([\![A_2]\!] \Rightarrow [\![B_2]\!])) \\
\Rightarrow ([\![A \wedge B]\!] \Rightarrow [\![A_2 \to B_2]\!])
\end{array}
&
\begin{array}{l}
\vdots \\
\overline{([\![A \wedge B]\!] \Rightarrow [\![C \to A \wedge C]\!])} \\
\Rightarrow \omega
\end{array}
\end{array}
}{
\begin{array}{c}
(\qquad\qquad \Omega \quad \Rightarrow [\![A \wedge C]\!]) \\
\Rightarrow \omega
\end{array}
} \; res
$$

Before we proceed, we introduce the abbreviations

$$\omega = [\![A \wedge B \to (C \to A \wedge C)]\!], \; \Omega = [[\![A \wedge B]\!], [\![C]\!]]$$

---

[320]

$$([\![A \wedge B]\!] \Rightarrow ([\![A_2]\!] \Rightarrow [\![B_2]\!])) \Rightarrow ([\![A \wedge B]\!] \Rightarrow [\![A_2 \to B_2]\!])$$

is the $\to$-*I*-rule (meta-axiom) lifted over the assumption $A \wedge B$.

Proof of $A \wedge B \rightarrow (C \rightarrow A \wedge C)$ **(3)**

We do resolution using the $\wedge$-*I* rule[321] lifted over Ω:

$$\dfrac{\begin{array}{c}(\Omega \Rightarrow [\![A_3]\!]) \Rightarrow (\Omega \Rightarrow [\![B_3]\!]) \\ \Rightarrow (\Omega \Rightarrow [\![A_3 \wedge B_3]\!])\end{array} \quad \dfrac{\vdots}{(\Omega \Rightarrow [\![A \wedge C]\!]) \Rightarrow \omega}}{(\Omega \Rightarrow [\![A]\!]) \Rightarrow (\Omega \Rightarrow [\![C]\!]) \Rightarrow \omega} \; res$$

At this point, Isabelle would display $\Omega \Rightarrow [\![A]\!]$ and $\Omega \Rightarrow [\![C]\!]$ as two subgoals.

The next step is to solve $\Omega \Rightarrow [\![C]\!]$ by assumption, but this must be formalized.

---

[321]
$$(\Omega \Rightarrow [\![A_3]\!]) \Rightarrow (\Omega \Rightarrow [\![B_3]\!]) \Rightarrow (\Omega \Rightarrow [\![A_3 \wedge B_3]\!])$$

is the $\wedge$-*I*-rule (meta-axiom) lifted over the assumption list Ω. Recall that Ω was an abbreviation for $[\![A \wedge B]\!], [\![C]\!]$, but this is obviously irrelevant for the process of lifting.

The Assumption Axiom

The assumption axiom is: for any $i \in \{1, \ldots, m\}$

$$\frac{}{[\phi_1, \ldots, \phi_m] \Rightarrow \phi_i} \; assum$$

It has a simple (schematic[322]) derivation:

$$\frac{\dfrac{\dfrac{[\phi_i]^1}{[\phi_{i+1}, \ldots, \phi_m] \Rightarrow \phi_i} \Rightarrow\text{-}I}{[\phi_i, \ldots, \phi_m] \Rightarrow \phi_i} \Rightarrow\text{-}I^1}{[\phi_1, \ldots, \phi_m] \Rightarrow \phi_i} \Rightarrow\text{-}I^{323}$$

---

[322]The assumption axiom

$$\frac{}{[\phi_1, \ldots, \phi_m] \Rightarrow \phi_i} \; assum$$

is schematic is two senses:

- the Greek letters could stand for arbitrary formulae;

- just like for resolution rule, we don't even know how many formulae are involved (m, i could be any natural numbers).

However, one could also write the axiom as

$$\frac{}{[A_1, \ldots, A_m] \Rightarrow A_i} \; assum$$

where the A's are variables (of type $bool$) and instantiate it later when it is used in some resolution step.

[323]Recall here that the rule $\Rightarrow$-I, just like $\rightarrow$-I, allows you to discharge zero or more assumptions. In the present derivation, we discharge the assumption ϕ_i at some point but we do not

Proof of $A \wedge B \to (C \to A \wedge C)$ (4)

We do resolution using the assumption axiom:

$$\frac{\Omega \Rightarrow [\![C]\!] \quad \overline{[\Omega \Rightarrow [\![A]\!], \Omega \Rightarrow [\![C]\!]] \Rightarrow \omega}^{\vdots}}{(\Omega \Rightarrow [\![A]\!]) \Rightarrow \omega} \; res$$

We used the correct instance of the assumption axiom. Alternatively[324], we could have use the more generic $[A_4, B_4] \Rightarrow B_4$.

What to do next? (Recall that $\Omega = [\![A \wedge B]\!], [\![C]\!]$.)

---

discharge any other assumptions.

[324]As explained previously, we could use a more generic variant of the assumption axiom, in that we have variables in it that may become instantiated upon resolution. As in previous proof steps we assume that these variables are suitably renamed; for this purpose we index them by 4.

Note however that the variant is still specific in the sense that $m = 2$. Like in meta-axioms used before, we use letters from the beginning of the alphabet, so the variant of the assumption axiom that we use is $[A_4, B_4] \Rightarrow B_4$. The proof fragment would then look as follows:

$$\frac{[A_4, B_4] \Rightarrow B_4 \quad \overline{[\Omega \Rightarrow [\![A]\!], \Omega \Rightarrow [\![C]\!]] \Rightarrow \omega}^{\vdots}}{(\Omega \Rightarrow [\![A]\!]) \Rightarrow \omega} \; res$$

where $\theta = \{A_4 \leftarrow [\![A \wedge B]\!], B_4 \leftarrow [\![C]\!]\}$.

317

Proof of $A \wedge B \to (C \to A \wedge C)$ **(4)**

We do resolution using the assumption axiom:

$$\cfrac{\Omega \Rightarrow \llbracket C \rrbracket \quad \cfrac{\vdots}{\llbracket \Omega \Rightarrow \llbracket A \rrbracket, \Omega \Rightarrow \llbracket C \rrbracket \rrbracket \Rightarrow \omega}}{(\Omega \Rightarrow \llbracket A \rrbracket) \Rightarrow \omega} \; res$$

We used the correct instance of the assumption axiom. Alternatively[324],
we could have use the more generic $[A_4, B_4] \Rightarrow B_4$.

What to do next? (Recall that $\Omega = \llbracket \llbracket A \wedge B \rrbracket, \llbracket C \rrbracket \rrbracket$.) Resolution
with $\wedge$-*EL*.

discharge any other assumptions.

[324] As explained previously, we could use a more generic vari-
ant of the assumption axiom, in that we have variables in it
that may become instantiated upon resolution. As in previ-
ous proof steps we assume that these variables are suitably
renamed; for this purpose we index them by 4.

Note however that the variant is still specific in the sense
that $m = 2$. Like in meta-axioms used before, we use letters
from the beginning of the alphabet, so the variant of the
assumption axiom that we use is $[A_4, B_4] \Rightarrow B_4$. The proof
fragment would then look as follows:

$$\cfrac{[A_4, B_4] \Rightarrow B_4 \quad \cfrac{\vdots}{\llbracket \Omega \Rightarrow \llbracket A \rrbracket, \Omega \Rightarrow \llbracket C \rrbracket \rrbracket \Rightarrow \omega}}{(\Omega \Rightarrow \llbracket A \rrbracket) \Rightarrow \omega} \; res$$

where $\theta = \{ A_4 \leftarrow \llbracket A \wedge B \rrbracket, B_4 \leftarrow \llbracket C \rrbracket \}$.

Proof of $A \wedge B \rightarrow (C \rightarrow A \wedge C)$ **(5)**

Magically, we guess the right instance of $\wedge$-*EL* and lift it over Ω:

$$\frac{(\Omega \Rightarrow \llbracket A \wedge B \rrbracket) \Rightarrow (\Omega \Rightarrow \llbracket A \rrbracket) \quad \overset{\vdots}{(\Omega \Rightarrow \llbracket A \rrbracket) \Rightarrow \omega}}{(\Omega \Rightarrow \llbracket A \wedge B \rrbracket) \Rightarrow \omega} \ \textit{res}$$

What to do next? (Recall that $\Omega = [\llbracket A \wedge B \rrbracket, \llbracket C \rrbracket]$.)

Proof of $A \wedge B \to (C \to A \wedge C)$ **(5)**

Magically, we guess the right instance of $\wedge$-*EL* and lift it over Ω:

$$\frac{(\Omega \Rightarrow \llbracket A \wedge B \rrbracket) \Rightarrow (\Omega \Rightarrow \llbracket A \rrbracket) \quad \overset{\vdots}{(\Omega \Rightarrow \llbracket A \rrbracket) \Rightarrow \omega}}{(\Omega \Rightarrow \llbracket A \wedge B \rrbracket) \Rightarrow \omega} \; res$$

What to do next? (Recall that $\Omega = \llbracket \llbracket A \wedge B \rrbracket, \llbracket C \rrbracket \rrbracket$.) Prove the subgoal by assumption.

Proof of $A \wedge B \to (C \to A \wedge C)$ **(6)**

We do resolution using the assumption axiom:

$$\cfrac{\Omega \Rightarrow [\![A \wedge B]\!] \quad \cfrac{\vdots}{(\Omega \Rightarrow [\![A \wedge B]\!]) \Rightarrow \omega}}{\omega} \; res$$

Recall that $\omega = [\![A \wedge B \to (C \to A \wedge C)]\!]$. Done!

Getting Rid of the Magic

In one step, we had to guess the right instance of $\wedge$-*EL*. This is not practical.

Solutions:

- Generalize the resolution rule to allow for instantiation of the current proof state and not just of meta-axioms.

- Derive

$$\bigwedge ABC.[[A \wedge B], ([[A], [B]] \Rightarrow [C])] \Rightarrow [C]$$

which encodes the $\wedge$-*E* object rule.

The Whole Proof at a Glance

Compare proof in $\mathcal{M}_{Prop}$ with corresponding proof in $Prop$:

$$
\cfrac{\text{\small $\wedge$-EL} \quad \cfrac{a. \quad \cfrac{\text{\small $\wedge$-I} \quad \cfrac{\text{\small $\rightarrow$-I} \quad \cfrac{\text{\small $\rightarrow$-I} \quad \omega \Rightarrow \omega}{\ldots \Rightarrow \omega}}{\ldots \Rightarrow \omega}}{\ldots \Rightarrow \omega}}{\ldots \Rightarrow \omega}}{\omega}
$$

$$
\cfrac{\cfrac{\cfrac{\cfrac{[A \wedge B]^1}{A} \wedge\text{-EL} \quad [C]^2}{A \wedge C} \wedge\text{-I}}{C \to A \wedge C} \rightarrow\text{-}I^2}{A \wedge B \to (C \to A \wedge C)} \rightarrow\text{-}I^1
$$

---

[325]Intuitively, as far as the order in which the object rules, resp. meta-axioms, are applied, the proof in $\mathcal{M}_{Prop}$ is the proof in $Prop$ turned upside-down.

However, this may seem suspicious for two reasons:

- In derivation trees, the direction of implication (forgetting about whether it is meta- or object implication) is "downwards": whatever is above implies whatever is below. So it seems strange that this order should be reversed just because we go from the object to the meta-level.

- In general, a derivation tree in the object level is a proper tree, i.e., there are nodes where it branches. So what sense does it make to "turn it upside-down"? The result would not be any tree at all.

These points will now be addressed.

The Whole Proof at a Glance

Compare proof in $\mathcal{M}_{Prop}$ with corresponding proof in $Prop$:

$$
\cfrac{
  \cfrac{
    \cfrac{
      \text{$a.$ } \cfrac{
        \text{$\wedge$-}I \; \cfrac{
          \text{$\rightarrow$-}I \; \cfrac{
            \text{$\rightarrow$-}I \quad \omega \Rightarrow \omega
          }{\ldots \Rightarrow \omega}
        }{\ldots \Rightarrow \omega}
      }{\ldots \Rightarrow \omega}
    }{\ldots \Rightarrow \omega}
  }{\text{$a.$ } \cfrac{\wedge\text{-}EL}{\ldots \Rightarrow \omega}}
}{\omega}
\qquad
\cfrac{
  \cfrac{
    \cfrac{
      \cfrac{[A \wedge B]^1}{A} \wedge\text{-}EL \qquad [C]^2
    }{A \wedge C} \wedge\text{-}I
  }{C \rightarrow A \wedge C} \rightarrow\text{-}I^2
}{A \wedge B \rightarrow (C \rightarrow A \wedge C)} \rightarrow\text{-}I^1
$$

"The meta-level proof is the object level proof upside-down[325]."

---

[325]Intuitively, as far as the order in which the object rules, resp. meta-axioms, are applied, the proof in $\mathcal{M}_{Prop}$ is the proof in $Prop$ turned upside-down.

However, this may seem suspicious for two reasons:

- In derivation trees, the direction of implication (forgetting about whether it is meta- or object implication) is "downwards": whatever is above implies whatever is below. So it seems strange that this order should be reversed just because we go from the object to the meta-level.

- In general, a derivation tree in the object level is a proper tree, i.e., there are nodes where it branches. So what sense does it make to "turn it upside-down"? The result would not be any tree at all.

These points will now be addressed.

Direction of the Implication

Is the direction of the implication reversed just because we go from the object to the meta-level?

Direction of the Implication

Is the direction of the implication reversed just because we go from the object to the meta-level?

No! The direction is reversed because we start from the trivial meta-theorem $\omega \Rightarrow \omega$, and the resolution steps modify the left-hand side of this meta-theorem.

How Can One Turn a Tree Upside-Down?

A proper tree has nodes where it branches. Also, in Isabelle proofs, we frequently have to prove several subgoals. So how is this branching reflected in the meta-proof?

---

[326]If one pictures the object level proof and how it is modeled in $\mathcal{M}_{Prop}$, one intutive way of thinking of it is as follows: Each rule application in the object level proof must also be performed at the meta-level. Now, starting at the root of the object level proof, we may do any rule application that is the child of a rule application we have done previously. Take for example the following object level proof:

$$\cfrac{\cfrac{[A \wedge (B \wedge C)]^1}{A} \wedge\text{-}EL^3 \qquad \cfrac{\cfrac{\cfrac{[A \wedge (B \wedge C)]^1}{B \wedge C} \wedge\text{-}ER^5}{C} \wedge\text{-}ER^4}{}}{\cfrac{A \wedge C}{A \wedge (B \wedge C) \to A \wedge C} \to\text{-}I^1} \wedge\text{-}I^2$$

Then in the meta-proof, the meta-axioms might be applied in the following orders:

$\to\text{-}I^1$, $\wedge\text{-}I^2$, $\wedge\text{-}ER^4$, $\wedge\text{-}ER^5$, $\wedge\text{-}EL^3$, or

$\to\text{-}I^1$, $\wedge\text{-}I^2$, $\wedge\text{-}EL^3$, $\wedge\text{-}ER^4$, $\wedge\text{-}ER^5$, or

$\to\text{-}I^1$, $\wedge\text{-}I^2$, $\wedge\text{-}ER^4$, $\wedge\text{-}EL^3$, $\wedge\text{-}ER^5$.

How Can One Turn a Tree Upside-Down?

A proper tree has nodes where it branches. Also, in Isabelle proofs, we frequently have to prove several subgoals. So how is this branching reflected in the meta-proof?

A meta-formula of the form $\psi_1 \Rightarrow \ldots \Rightarrow \psi_n \Rightarrow \psi$ corresponds to a branching point in the object level proof. It means that there are subgoals $\psi_1, \ldots, \psi_n$. But in the derivation tree in $\mathcal{M}_{Prop}$, there is no branching.

---

[326]If one pictures the object level proof and how it is modeled in $\mathcal{M}_{Prop}$, one intuitive way of thinking of it is as follows: Each rule application in the object level proof must also be performed at the meta-level. Now, starting at the root of the object level proof, we may do any rule application that is the child of a rule application we have done previously. Take for example the following object level proof:

$$\cfrac{\cfrac{[A \wedge (B \wedge C)]^1}{A} \wedge\text{-}EL^3 \qquad \cfrac{\cfrac{\cfrac{[A \wedge (B \wedge C)]^1}{B \wedge C} \wedge\text{-}ER^5}{C} \wedge\text{-}ER^4}{}}{\cfrac{A \wedge C}{A \wedge (B \wedge C) \to A \wedge C} \to\text{-}I^1} \wedge\text{-}I^2$$

Then in the meta-proof, the meta-axioms might be applied in the following orders:

$\to\text{-}I^1$, $\wedge\text{-}I^2$, $\wedge\text{-}ER^4$, $\wedge\text{-}ER^5$, $\wedge\text{-}EL^3$, or

$\to\text{-}I^1$, $\wedge\text{-}I^2$, $\wedge\text{-}EL^3$, $\wedge\text{-}ER^4$, $\wedge\text{-}ER^5$, or

$\to\text{-}I^1$, $\wedge\text{-}I^2$, $\wedge\text{-}ER^4$, $\wedge\text{-}EL^3$, $\wedge\text{-}ER^5$.

How Can One Turn a Tree Upside-Down?

A proper tree has nodes where it branches. Also, in Isabelle proofs, we frequently have to prove several subgoals. So how is this branching reflected in the meta-proof?

A meta-formula of the form $\psi_1 \Rightarrow \ldots \Rightarrow \psi_n \Rightarrow \psi$ corresponds to a branching point in the object level proof. It means that there are subgoals $\psi_1, \ldots, \psi_n$. But in the derivation tree in $\mathcal{M}_{Prop}$, there is no branching.

In the construction of a meta-proof (just like in Isabelle), one is always free to choose which subgoal to solve next. Interleaving[326] is possible.

---

[326]If one pictures the object level proof and how it is modeled in $\mathcal{M}_{Prop}$, one intutive way of thinking of it is as follows: Each rule application in the object level proof must also be performed at the meta-level. Now, starting at the root of the object level proof, we may do any rule application that is the child of a rule application we have done previously. Take for example the following object level proof:

$$\cfrac{\cfrac{[A \wedge (B \wedge C)]^1}{A} \wedge\text{-}EL^3 \qquad \cfrac{\cfrac{\cfrac{[A \wedge (B \wedge C)]^1}{B \wedge C} \wedge\text{-}ER^5}{C} \wedge\text{-}ER^4}{A \wedge C} \wedge\text{-}I^2}{A \wedge (B \wedge C) \to A \wedge C} \to\text{-}I^1$$

Then in the meta-proof, the meta-axioms might be applied in the following orders:

$\to\text{-}I^1$, $\wedge\text{-}I^2$, $\wedge\text{-}ER^4$, $\wedge\text{-}ER^5$, $\wedge\text{-}EL^3$, or

$\to\text{-}I^1$, $\wedge\text{-}I^2$, $\wedge\text{-}EL^3$, $\wedge\text{-}ER^4$, $\wedge\text{-}ER^5$, or

$\to\text{-}I^1$, $\wedge\text{-}I^2$, $\wedge\text{-}ER^4$, $\wedge\text{-}EL^3$, $\wedge\text{-}ER^5$.

13.4 Quantification

We have to add the following meta-axioms to obtain $\mathcal{M}_{\text{FOL}}$:

$$\bigwedge F.(\bigwedge x.[\![F\,x]\!]) \Rightarrow [\![\forall x.F\,x]\!] \qquad\qquad (\forall\text{-}I)$$
$$\bigwedge Fy.[\![\forall x.F\,x]\!] \Rightarrow [\![F\,y]\!] \qquad\qquad (\forall\text{-}E)$$
$$\bigwedge Fy.[\![F\,y]\!] \Rightarrow [\![\exists x.F\,x]\!] \qquad\qquad (\exists\text{-}I)$$
$$\bigwedge FB.[\![\exists x.F\,x]\!] \Rightarrow (\bigwedge x.[\![F\,x]\!] \Rightarrow [\![B]\!]) \Rightarrow [\![B]\!] \quad (\exists\text{-}E)$$

Similarly as for $Prop$, one can show that $\mathcal{M}_{\text{FOL}}$ is faithful for FOL.

Side condition checking is shifted to the meta-level.

We now consider resolution proofs for FOL.

But this is not new to you: In Isabelle, you are always free to choose the subgoal that you want to work on next, and so you can interleave the proofs of the different subgoals.

Proof of $(\forall z.G\,z) \rightarrow (\forall z.G\,z \vee H\,z)$ **(1)**

$$\frac{([\![A_1]\!] \Rightarrow [\![B_1]\!]) \qquad [\![(\forall z.G\,z) \rightarrow (\forall z.G\,z \vee H\,z)]\!]}{([\![\forall z.G\,z]\!] \Rightarrow [\![\forall z.G\,z \vee H\,z]\!])} \;res$$
$$\Rightarrow [\![A_1 \rightarrow B_1]\!] \qquad \Rightarrow [\![(\forall z.G\,z) \rightarrow (\forall z.G\,z \vee H\,z)]\!]$$
$$\Rightarrow [\![(\forall z.G\,z) \rightarrow (\forall z.G\,z \vee H\,z)]\!]$$

What to do next?

Proof of $(\forall z.G\,z) \rightarrow (\forall z.G\,z \vee H\,z)$ **(1)**

$$
\frac{
\begin{array}{cc}
\begin{array}{l}
([A_1] \Rightarrow [B_1]) \\
\Rightarrow [A_1 \rightarrow B_1]
\end{array}
&
\begin{array}{l}
[\![(\forall z.G\,z) \rightarrow (\forall z.G\,z \vee H\,z)]\!] \\
\Rightarrow [\![(\forall z.G\,z) \rightarrow (\forall z.G\,z \vee H\,z)]\!]
\end{array}
\end{array}
}{
\begin{array}{l}
([\![\forall z.G\,z]\!] \Rightarrow [\![\forall z.G\,z \vee H\,z]\!]) \\
\Rightarrow [\![(\forall z.G\,z) \rightarrow (\forall z.G\,z \vee H\,z)]\!]
\end{array}
} \; res
$$

What to do next? Resolution with $\forall$-*I* lifted over assumption $[\![\forall z.G\,z]\!]$.

Proof of $(\forall z.G\,z) \to (\forall z.G\,z \vee H\,z)$ **(2)**

$$
\frac{
\begin{array}{c}
(\llbracket \forall z.G\,z \rrbracket \Rightarrow (\bigwedge x.\llbracket F_1\,x \rrbracket)) \\
\Rightarrow (\llbracket \forall z.G\,z \rrbracket \Rightarrow \llbracket \forall x.F_1\,x \rrbracket)
\end{array}
\qquad
\frac{\vdots}{
\begin{array}{c}
(\llbracket \forall z.G\,z \rrbracket \Rightarrow \llbracket \forall z.G\,z \vee H\,z \rrbracket) \\
\Rightarrow \llbracket (\forall z.G\,z) \to (\forall z.G\,z \vee H\,z) \rrbracket
\end{array}
}
}{
\begin{array}{c}
(\llbracket \forall z.G\,z \rrbracket \Rightarrow (\bigwedge z.\llbracket G\,z \vee H\,z \rrbracket)) \\
\Rightarrow \llbracket (\forall z.G\,z) \to (\forall z.G\,z \vee H\,z) \rrbracket
\end{array}
}\; \textit{res}
$$

The substitution θ is $[F_1 \leftarrow \lambda w.G\,w \vee H\,w]$.

We suppress conversion, assuming terms are in normal form.
What to do next?

Proof of $(\forall z.G\,z) \to (\forall z.G\,z \vee H\,z)$ **(2)**

$$
\frac{
\begin{array}{cc}
\begin{array}{l}
(\llbracket \forall z.G\,z \rrbracket \Rightarrow (\bigwedge x.\llbracket F_1\,x \rrbracket)) \\
\Rightarrow (\llbracket \forall z.G\,z \rrbracket \Rightarrow \llbracket \forall x.F_1\,x \rrbracket)
\end{array}
&
\dfrac{\vdots}{
\begin{array}{l}
(\llbracket \forall z.G\,z \rrbracket \Rightarrow \llbracket \forall z.G\,z \vee H\,z \rrbracket) \\
\Rightarrow \llbracket (\forall z.G\,z) \to (\forall z.G\,z \vee H\,z) \rrbracket
\end{array}
}
\end{array}
}{
\begin{array}{l}
(\llbracket \forall z.G\,z \rrbracket \Rightarrow (\bigwedge z.\llbracket G\,z \vee H\,z \rrbracket)) \\
\Rightarrow \llbracket (\forall z.G\,z) \to (\forall z.G\,z \vee H\,z) \rrbracket
\end{array}
}\ res
$$

The substitution θ is $[F_1 \leftarrow \lambda w.G\,w \vee H\,w]$.

We suppress conversion, assuming terms are in normal form.

What to do next? Resolution with $\vee$-*IL* after lifting over assumption. Problem: the conclusion of $\vee$-*IL* is not unifiable with $\bigwedge z.\llbracket G\,z \vee H\,z \rrbracket)$.

Lifting over Parameters

Lifting over parameters seems easier to explain if outer $\bigwedge$'s are not dropped. The rule for lifting a meta-axiom $\bigwedge y_1 \ldots y_k.[\phi_1, \ldots, \phi_m] \Rightarrow \phi$ over a parameter z is

$$\frac{\bigwedge y_1 \ldots y_k.[\phi_1, \ldots, \phi_m] \Rightarrow \phi}{\bigwedge f_1 \ldots f_k.[\bigwedge z.\phi'_1, \ldots, \bigwedge z.\phi'_m] \Rightarrow (\bigwedge z.\phi')} \text{ p-lift}$$

where $'$ stands for application of the substitution $[y_1 \leftarrow f_1\, z, \ldots, y_k \leftarrow f_k\, z]$.

We will now derive it.

Deriving Parameter Lifting for **one** Parameter

$'$ stands for application of $[y_1 \leftarrow f_1(z), \ldots, y_k \leftarrow f_k(z)]$.

$$\cfrac{\cfrac{\bigwedge y_1 \ldots y_k.[\phi_1, \ldots, \phi_m] \Rightarrow \phi}{[\phi'_1, \ldots, \phi'_m] \Rightarrow \phi'} \text{$\bigwedge$-E} \quad \cfrac{\bigwedge z.\phi'_1}{\phi'_1} \text{$\bigwedge$-E} \ldots \cfrac{\bigwedge z.\phi'_m}{\phi'_m} \text{$\bigwedge$-E}}{\cfrac{\phi'}{\bigwedge z.\phi'} \text{$\bigwedge$-I}} \text{$\Rightarrow$-E}$$

Deriving Parameter Lifting for one Parameter

$'$ stands for application of $[y_1 \leftarrow f_1(z), \ldots, y_k \leftarrow f_k(z)]$.

$$\dfrac{\dfrac{\bigwedge y_1 \ldots y_k.[\phi_1, \ldots, \phi_m] \Rightarrow \phi}{[\phi'_1, \ldots, \phi'_m] \Rightarrow \phi'} \bigwedge\text{-}E \quad \dfrac{[\bigwedge z.\phi'_1]^1}{\phi'_1} \bigwedge\text{-}E \ldots \dfrac{[\bigwedge z.\phi'_m]^1}{\phi'_m} \bigwedge\text{-}E}{\dfrac{\dfrac{\dfrac{\phi'}{\bigwedge z.\phi'} \bigwedge\text{-}I}{[\bigwedge z.\phi'_1, \ldots, \bigwedge z.\phi'_m] \Rightarrow \bigwedge z.\phi'} \Rightarrow\text{-}I^1}{\bigwedge f_1 \ldots f_k.[\bigwedge z.\phi'_1, \ldots, \bigwedge z.\phi'_m] \Rightarrow \bigwedge z.\phi'} \bigwedge\text{-}I} \Rightarrow\text{-}E}$$

Deriving Parameter Lifting for **one** Parameter

$'$ stands for application of $[y_1 \leftarrow f_1(z), \ldots, y_k \leftarrow f_k(z)]$.

$$
\cfrac{
  \cfrac{
    \cfrac{\bigwedge y_1 \ldots y_k.[\phi_1, \ldots, \phi_m] \Rightarrow \phi}{[\phi_1', \ldots, \phi_m'] \Rightarrow \phi'} \wedge\text{-}E
    \quad
    \cfrac{[\bigwedge z.\phi_1']^1}{\phi_1'} \wedge\text{-}E
    \ldots
    \cfrac{[\bigwedge z.\phi_m']^1}{\phi_m'} \wedge\text{-}E
  }{
    \cfrac{\phi'}{\bigwedge z.\phi'} \wedge\text{-}I
  } \Rightarrow\text{-}E
}{
  \cfrac{[\bigwedge z.\phi_1', \ldots, \bigwedge z.\phi_m'] \Rightarrow \bigwedge z.\phi'}{[\bigwedge z.\phi_1', \ldots, \bigwedge z.\phi_m'] \Rightarrow \bigwedge z.\phi'} \wedge\text{-}I
} \Rightarrow\text{-}I^1
$$

After parameter lifting, we drop outer quantifiers again.

328

Lifting $\vee$-*IL*

Lifting $\bigwedge AB.[A] \Rightarrow [A \vee B]$ ($\vee$-*IL*) over z gives

$$\bigwedge G_2 H_2.(\bigwedge z.[G_2\, z]) \Rightarrow (\bigwedge z.[G_2\, z \vee H_2\, z]).$$

Lifting $\vee$-*IL*

Lifting $\bigwedge AB.[A] \Rightarrow [A \vee B]$ ($\vee$-*IL*) over z gives

$$\left(\bigwedge z.[G_2\, z] \right) \Rightarrow \left(\bigwedge z.[G_2\, z \vee H_2\, z] \right).$$

We drop outer quantifiers and lift over assumption $[\![\forall z.G\, z]\!]$ to obtain

$$([\![\forall z.G\, z]\!] \Rightarrow \bigwedge z.[G_2\, z]) \Rightarrow$$
$$([\![\forall z.G\, z]\!] \Rightarrow \bigwedge z.[G_2\, z \vee H_2\, z])$$

This rule will be applied in the next step.

Proof of $(\forall z.G\,z) \to (\forall z.G\,z \vee H\,z)$ **(3)**

$$
\cfrac{
\begin{array}{cc}
\begin{array}{l}
([\![\forall z.G\,z]\!] \Rightarrow \bigwedge z.[\![G_2\,z]\!]) \Rightarrow \\
([\![\forall z.G\,z]\!] \Rightarrow \bigwedge z.[\![G_2\,z \vee H_2\,z]\!])
\end{array}
&
\cfrac{\vdots}{
\begin{array}{l}
([\![\forall z.G\,z]\!] \Rightarrow (\bigwedge z.[\![G\,z \vee H\,z]\!])) \\
\Rightarrow [\![(\forall z.G\,z) \to (\forall z.G\,z \vee H\,z)]\!]
\end{array}}
\end{array}
}{
\begin{array}{c}
([\![\forall z.G\,z]\!] \Rightarrow \bigwedge z.[\![G\,z]\!]) \Rightarrow \\
[\![(\forall z.G\,z) \to (\forall z.G\,z \vee H\,z)]\!]
\end{array}
} \; res
$$

What to do next?

Proof of $(\forall z.G\,z) \rightarrow (\forall z.G\,z \vee H\,z)$ (3)

$$\frac{\begin{array}{cc} & \vdots \\ ([\![\forall z.G\,z]\!] \Rightarrow \bigwedge z.[\![G_2\,z]\!]) \Rightarrow & \overline{([\![\forall z.G\,z]\!] \Rightarrow (\bigwedge z.[\![G\,z \vee H\,z]\!]))} \\ ([\![\forall z.G\,z]\!] \Rightarrow \bigwedge z.[\![G_2\,z \vee H_2\,z]\!]) & \Rightarrow [\![(\forall z.G\,z) \rightarrow (\forall z.G\,z \vee H\,z)]\!] \end{array}}{\begin{array}{c} ([\![\forall z.G\,z]\!] \Rightarrow \bigwedge z.[\![G\,z]\!]) \Rightarrow \\ [\![(\forall z.G\,z) \rightarrow (\forall z.G\,z \vee H\,z)]\!] \end{array}}\ res$$

What to do next? Resolution with ∀-*E* lifted over z. However, this cannot be guessed from looking at $\bigwedge z.[\![G\,z]\!]$, but rather from looking at premise $[\![\forall z.G\,z]\!]$.

Lifting of $\forall$-E over z

Lifting $\bigwedge F y. [\![\forall x. F\ x]\!] \Rightarrow [\![F\ y]\!]$ ($\forall$-E) over parameter z gives

$$\bigwedge G_3 f_3. (\bigwedge z. [\![\forall x.(G_3\ z)x]\!]) \Rightarrow (\bigwedge z. [\![G_3\ z(f_3\ z)]\!]).$$

We drop outer quantifiers and lift over assumption $[\![\forall z. G\ z]\!]$ to obtain

$$([\![\forall z. G\ z]\!] \Rightarrow \bigwedge z. [\![\forall x.(G_3\ z)x]\!]) \Rightarrow$$
$$([\![\forall z. G\ z]\!] \Rightarrow \bigwedge z. [\![G_3\ z(f_3\ z)]\!])$$

This rule will be applied in the next step.

Proof of $(\forall z.G\,z) \rightarrow (\forall z.G\,z \vee H\,z)$ **(4)**

$$
\dfrac{
\begin{array}{cc}
\begin{array}{c}
(\llbracket \forall z.G\,z \rrbracket \Rightarrow \bigwedge z.\llbracket \forall x.(G_3\,z)x \rrbracket) \Rightarrow \\
(\llbracket \forall z.G\,z \rrbracket \Rightarrow \bigwedge z.\llbracket G_3\,z(f_3\,z) \rrbracket)
\end{array}
&
\dfrac{\vdots}{
\begin{array}{c}
(\llbracket \forall z.G\,z \rrbracket \Rightarrow \bigwedge z.\llbracket G\,z \rrbracket) \Rightarrow \\
\llbracket (\forall z.G\,z) \rightarrow (\forall z.G\,z \vee H\,z) \rrbracket
\end{array}
}
\end{array}
}{
\begin{array}{c}
(\llbracket \forall z.G\,z \rrbracket \Rightarrow \bigwedge z.\llbracket \forall x.G\,x \rrbracket) \Rightarrow \\
\llbracket (\forall z.G\,z) \rightarrow (\forall z.G\,z \vee H\,z) \rrbracket
\end{array}
} \; res
$$

The substitution θ is $[f_3 \leftarrow \lambda w.w,\ G_3 \leftarrow \lambda vw.G\,w]$.

We suppress conversion, assuming terms are in normal form.

What to do next?

Proof of $(\forall z.G\,z) \to (\forall z.G\,z \vee H\,z)$ **(4)**

$$
\begin{array}{c}
\vdots \\
\hline
\end{array}
$$

$$
\cfrac{
\begin{array}{cc}
\begin{array}{c}
([\![\forall z.G\,z]\!] \Rightarrow \bigwedge z.[\![\forall x.(G_3\,z)x]\!]) \Rightarrow \\
([\![\forall z.G\,z]\!] \Rightarrow \bigwedge z.[\![G_3\,z(f_3\,z)]\!])
\end{array}
&
\begin{array}{c}
([\![\forall z.G\,z]\!] \Rightarrow \bigwedge z.[\![G\,z]\!]) \Rightarrow \\
[\![(\forall z.G\,z) \to (\forall z.G\,z \vee H\,z)]\!]
\end{array}
\end{array}
}{
\begin{array}{c}
([\![\forall z.G\,z]\!] \Rightarrow \bigwedge z.[\![\forall x.G\,x]\!]) \Rightarrow \\
[\![(\forall z.G\,z) \to (\forall z.G\,z \vee H\,z)]\!]
\end{array}
}\;\textit{res}
$$

The substitution θ is $[f_3 \leftarrow \lambda w.w,\ G_3 \leftarrow \lambda vw.G\,w]$.

　　We suppress conversion, assuming terms are in normal form.

　　What to do next? Since $z \notin FV(\forall x.G\,x)$, we can use a modified assumption axiom.

Modified Assumption Axiom

$$\frac{}{[\phi_1, \ldots, \phi_m] \Rightarrow \bigwedge z.\phi_i} \; assum \qquad \text{where } z \notin FV(\phi_i).$$

It has the following derivation:

$$\frac{\dfrac{\dfrac{[\phi_i]^1}{\bigwedge z.\phi_i} \bigwedge\text{-}I}{[\phi_{i+1}, \ldots, \phi_m] \Rightarrow \bigwedge z.\phi_i} \Rightarrow\text{-}I}{\dfrac{[\phi_i, \ldots, \phi_m] \Rightarrow \bigwedge z.\phi_i}{[\phi_1, \ldots, \phi_m] \Rightarrow \bigwedge z.\phi_i} \Rightarrow\text{-}I} \Rightarrow\text{-}I^1$$

Instance of Modified Assumption Axiom

In the next step, we will use the instance

$$[\![\forall z.G\,z]\!] \Rightarrow \bigwedge z.[\![\forall x.G\,x]\!]$$

of

$$[\phi_1, \ldots, \phi_m] \Rightarrow \bigwedge z.\phi_i.$$

We identified $\forall z.G\,z$ and $\forall x.G\,x$ by conversion.

Proof of $(\forall z.G\,z) \to (\forall z.G\,z \vee H\,z)$ **(5)**

$$
\cfrac{
\cfrac{\vdots}{[\![\forall z.G\,z]\!] \Rightarrow \bigwedge z.[\![\forall x.G\,x]\!]}
\quad
\cfrac{\vdots}{([\![\forall z.G\,z]\!] \Rightarrow \bigwedge z.[\![\forall x.G\,x]\!]) \Rightarrow [\![(\forall z.G\,z) \to (\forall z.G\,z \vee H\,z)]\!]}
}{[\![(\forall z.G\,z) \to (\forall z.G\,z \vee H\,z)]\!]}\ \textit{res}
$$

Done!

Remark on Step 2

Recall Step 2:

$$
\frac{([\![\forall z.G\,z]\!] \Rightarrow (\bigwedge x.[\![F_1\,x]\!])) \quad \frac{\vdots}{([\![\forall z.G\,z]\!] \Rightarrow [\![\forall z.G\,z \vee H\,z]\!])}}{([\![\forall z.G\,z]\!] \Rightarrow (\bigwedge z.[\![G\,z \vee H\,z]\!]))} \; res
$$

$$
\Rightarrow ([\![\forall z.G\,z]\!] \Rightarrow [\![\forall x.F_1\,x]\!]) \qquad \Rightarrow [\![(\forall z.G\,z) \to (\forall z.G\,z \vee H\,z)]\!]
$$

$$
\Rightarrow [\![(\forall z.G\,z) \to (\forall z.G\,z \vee H\,z)]\!]
$$

One could have obtained $\bigwedge z.([\![\forall z.G\,z]\!] \Rightarrow ([\![G\,z \vee H\,z]\!]))$ instead of $([\![\forall z.G\,z]\!] \Rightarrow (\bigwedge z.[\![G\,z \vee H\,z]\!]))$ by lifting $\forall$-I in a different way[327]. There will be an exercise.

---

[327]In our proof, we lifted $\forall$-I over assumption $[\![\forall z.G\,z]\!]$ as follows:

$$
([\![\forall z.G\,z]\!] \Rightarrow (\bigwedge x.[\![F_1\,x]\!])) \Rightarrow ([\![\forall z.G\,z]\!] \Rightarrow [\![\forall x.F_1\,x]\!])
$$

It would have been possible to derive (formally, in $\mathcal{M}$) the following rule instead:

$$
(\bigwedge x.[\![\forall z.G\,z]\!] \Rightarrow [\![F_1\,x]\!]) \Rightarrow ([\![\forall z.G\,z]\!] \Rightarrow [\![\forall x.F_1\,x]\!])
$$

This is essentially so since $z \notin FV[\![\forall z.G\,z]\!]$. If we had done it like that, step 2 would have looked as follows

$$
\frac{(\bigwedge x.[\![\forall z.G\,z]\!] \Rightarrow [\![F_1\,x]\!]) \quad \frac{\vdots}{([\![\forall z.G\,z]\!] \Rightarrow [\![\forall z.G\,z \vee H\,z]\!])}}{(\bigwedge z.[\![\forall z.G\,z]\!] \Rightarrow [\![G\,z \vee H\,z]\!])} \; res
$$

$$
\Rightarrow ([\![\forall z.G\,z]\!] \Rightarrow [\![\forall x.F_1\,x]\!]) \qquad \Rightarrow [\![(\forall z.G\,z) \to (\forall z.G\,z \vee H\,z)]\!]
$$

$$
\Rightarrow [\![(\forall z.G\,z) \to (\forall z.G\,z \vee H\,z)]\!]
$$

The rest of the proof would then have looked slightly differently due to the different scope of the $\bigwedge$. For example,

Checking Side Conditions

To demonstrate how side conditions are checked, we show a proof attempt that fails due to a side condition.

Take $\exists u. \forall w.\ w = u$ in FOL with equality, so assume we have a meta-axiom for reflexivity:

$$\bigwedge z. [\![z = z]\!] \quad (\textit{refl})$$

---

it would have been necessary to lift $\vee$-*IL* over assumptions before lifting it over parameters.

In fact, if we denote a vector of variables by overlining, then we can derive the following rule for lifting over assumptions:

$$\frac{[(\bigwedge \bar{x}_1.\phi_1), \dots, (\bigwedge \bar{x}_m.\phi_m)] \Rightarrow \phi}{[(\bigwedge \bar{x}_1.\Psi \Rightarrow \phi_1), \dots, (\bigwedge \bar{x}_1.\Psi \Rightarrow \phi_m)] \Rightarrow (\Psi \Rightarrow \phi)}$$

where $\bar{x}_1, \dots \bar{x}_m \notin FV(\Psi)$. Compare this to rule *a-lift*. Using the more complicated rule, where the assumption list Ψ is pulled into the scope of $\bigwedge$'s surrounding each rule premise ϕ_i, would probably have made the presentation here somewhat more complicated. On the other hand, this is indeed what happens in Isabelle (try to do the proof of $(\forall z. G\, z) \rightarrow (\forall z. G\, z \vee H\, z)$ in Isabelle).

Failed Proof Attempt of $\exists u. \forall w.\, w = u$

$$
\dfrac{
  (\bigwedge x.[F_2\, x]) \Rightarrow [\forall x.F_2\, x]
  \qquad
  \dfrac{
    [F_1\, y_1] \Rightarrow [\exists x.F_1\, x]
    \qquad
    [\exists u.\forall w.\, w = u] \Rightarrow [\exists u.\forall w.\, w = u]
  }{
    [\forall w.\, w = y_1] \Rightarrow [\exists u.\forall w.\, w = u]
  }\ res
}{
  (\bigwedge x.[x = y_1]) \Rightarrow [\exists y.\forall x.\, x = y]
}\ res
$$

Substitution? $[F_1 \leftarrow \qquad\qquad , F_2 \leftarrow \qquad\qquad].$

---

[328]Note that lifting *refl*

$$\bigwedge z.[z = z]$$

over x gives

$$\bigwedge g_3. \bigwedge x.[g_3\, x = g_3\, x].$$

Here the variable z in *refl* was replaced by the variable g_3 that depends on x. However, we drop the outer quantification $\bigwedge g_3$. In this particular case, $\bigwedge x$ is also an outer quantification, but we keep it, since obtaining this quantification was the very purpose of lifting (recall that lifting is done to achieve unifiability).

[329]Recall that $\bigwedge x.\phi$ is syntactic sugar for $\bigwedge x.(\lambda x.\phi)$.

So we have to unify $\lambda x.[x = y_1]$ and $\lambda x.[g_3\, x = g_3\, x]$.

It turns out that this task can be decomposed into having to unify $\lambda x.x$ and $\lambda x.g_3\, x$ on the one hand, and $\lambda x.y_1$ and $\lambda x.g_3\, x$ on the other hand. Unification of $\lambda x.x$ and $\lambda x.g_3\, x$ forces g_3 to be $\lambda x.x$, so we are left with having to unify $\lambda x.y_1$ and $\lambda x.x$. But these terms are not unifiable!

Failed Proof Attempt of $\exists u.\forall w.\, w = u$

$$
\cfrac{
(\bigwedge x.[F_2\, x]) \Rightarrow [\forall x.F_2\, x]
\qquad
\cfrac{
\begin{array}{l}
[F_1\, y_1] \Rightarrow \\
[\exists x.F_1\, x]
\end{array}
\quad
\begin{array}{l}
[\exists u.\forall w.\, w = u] \Rightarrow \\
[\exists u.\forall w.\, w = u]
\end{array}
}{
[\forall w.\, w = y_1] \Rightarrow [\exists u.\forall w.\, w = u]
}\ res
}{
(\bigwedge x.[x = y_1]) \Rightarrow [\exists y.\forall x.\, x = y]
}\ res
$$

Substitution? $[F_1 \leftarrow \lambda v.\forall w.\, w = v, F_2 \leftarrow \lambda v.\, v = y_1]$.

What to do next?

---

[328]Note that lifting *refl*

$$
\bigwedge z.[z = z]
$$

over x gives

$$
\bigwedge g_3.\, \bigwedge x.[g_3\, x = g_3\, x].
$$

Here the variable z in *refl* was replaced by the variable g_3 that depends on x. However, we drop the outer quantification $\bigwedge g_3$. In this particular case, $\bigwedge x$ is also an outer quantification, but we keep it, since obtaining this quantification was the very purpose of lifting (recall that lifting is done to achieve unifiability).

[329]Recall that $\bigwedge x.\phi$ is syntactic sugar for $\bigwedge x.(\lambda x.\phi)$.

So we have to unify $\lambda x.[x = y_1]$ and $\lambda x.[g_3\, x = g_3\, x]$.

It turns out that this task can be decomposed into having to unify $\lambda x.x$ and $\lambda x.g_3\, x$ on the one hand, and $\lambda x.y_1$ and $\lambda x.g_3\, x$ on the other hand. Unification of $\lambda x.x$ and $\lambda x.g_3\, x$ forces g_3 to be $\lambda x.x$, so we are left with having to unify $\lambda x.y_1$ and $\lambda x.x$. But these terms are not unifiable!

338

Failed Proof Attempt of $\exists u.\forall w.\, w = u$

$$
\cfrac{
(\bigwedge x.\llbracket F_2\, x\rrbracket) \Rightarrow \llbracket \forall x.F_2\, x\rrbracket
\qquad
\cfrac{
\llbracket F_1\, y_1\rrbracket \Rightarrow \llbracket \exists x.F_1\, x\rrbracket
\qquad
\llbracket \exists u.\forall w.\, w = u\rrbracket \Rightarrow \llbracket \exists u.\forall w.\, w = u\rrbracket
}{
\llbracket \forall w.\, w = y_1\rrbracket \Rightarrow \llbracket \exists u.\forall w.\, w = u\rrbracket
}\ res
}{
(\bigwedge x.\llbracket x = y_1\rrbracket) \Rightarrow \llbracket \exists y.\forall x.\, x = y\rrbracket
}\ res
$$

Substitution? $[F_1 \leftarrow \lambda v.\forall w.\, w = v,\ F_2 \leftarrow \lambda v.\, v = y_1]$.

What to do next? Resolution with *refl* lifted over parameter x: $\bigwedge x.\llbracket g_3\, x = g_3\, x\rrbracket$[328]. But $\bigwedge x.\llbracket x = y_1\rrbracket$ and $\bigwedge x.\llbracket g_3\, x = g_3\, x\rrbracket$ are not unifiable[329]. Proof fails!

---

[328]Note that lifting *refl*

$$\bigwedge z.\llbracket z = z\rrbracket$$

over x gives

$$\bigwedge g_3.\,\bigwedge x.\llbracket g_3\, x = g_3\, x\rrbracket.$$

Here the variable z in *refl* was replaced by the variable g_3 that depends on x. However, we drop the outer quantification $\bigwedge g_3$. In this particular case, $\bigwedge x$ is also an outer quantification, but we keep it, since obtaining this quantification was the very purpose of lifting (recall that lifting is done to achieve unifiability).

[329]Recall that $\bigwedge x.\phi$ is syntactic sugar for $\bigwedge x.(\lambda x.\phi)$.

So we have to unify $\lambda x.\llbracket x = y_1\rrbracket$ and $\lambda x.\llbracket g_3\, x = g_3\, x\rrbracket$.

It turns out that this task can be decomposed into having to unify $\lambda x.x$ and $\lambda x.g_3\, x$ on the one hand, and $\lambda x.y_1$ and $\lambda x.g_3\, x$ on the other hand. Unification of $\lambda x.x$ and $\lambda x.g_3\, x$ forces g_3 to be $\lambda x.x$, so we are left with having to unify $\lambda x.y_1$ and $\lambda x.x$. But these terms are not unifiable!

13.5 Free Variables in Goals

The resolution rule can be generalized to allow for instantiation of variables in goals:

$$\frac{[\phi_1, \ldots, \phi_m] \Rightarrow \phi \quad [\psi_1, \ldots, \psi_n] \Rightarrow \psi}{([\psi_1, \ldots, \psi_{i-1}, \phi_1, \ldots, \phi_m, \psi_{i+1}, \ldots, \psi_n] \Rightarrow \psi)\theta} \; res$$

where $\phi\theta \equiv \psi_i\theta$.

But then we must distinguish the status of the free variables. Denote the universal closure[330] of ψ by $\bigwedge \_.\psi$. Then
. . .

This was just a semi-formal argument that $\bigwedge x.[x = y_1]$ and $\bigwedge x.[g_3\, x = g_3\, x]$ are not unifiable, but it gives you the idea.

[330]The universal closure of a meta-formula ψ is the formula $\bigwedge x_1 \ldots x_n.\psi$ where $FV(\psi) = \{x_1 \ldots x_n\}$.

As might be expected, the same concept is also used for FOL formulae where it is defined in analogy using $\forall$ instead of $\bigwedge$.

Instantiation of the Initial Goal

Previously, when we proved ψ we in fact proved $\bigwedge\_.\psi$.

$$
\begin{array}{c}
\dfrac{\psi \Rightarrow \psi}{} \\
\vdots \qquad\qquad \vdots \\
\hline
\psi
\end{array}
$$

---

[331]Suppose we want to prove $((A \to B) \to A) \to A$. If we allow for instantiation of the free variables A and B, we could easily end up proving $((A \to A) \to A) \to A$. This is probably not what we want. In fact the proof has little to do with the proof of $((A \to B) \to A) \to A$ that is schematic in A and B.

In terms of $\mathcal{M}_{Prop}$, we want to prove $\bigwedge AB.[\![((A \to B) \to A) \to A]\!]$

Recall that $((A \to B) \to A) \to A$ is Peirce's law.

[332]The more free variables in the goal we allow Isabelle to instantiate, the more unifiers there are. This may increase the search space to the extent of making it impossible to find a proof.

Instantiation of the Initial Goal

Previously, when we proved ψ we in fact proved $\bigwedge \_.\psi$.

$$\cfrac{\cfrac{\bigwedge \_.\psi \Rightarrow \psi}{\psi \Rightarrow \psi} \; \bigwedge\text{-}E}{\cfrac{\vdots \qquad \vdots}{\cfrac{\psi}{\bigwedge \_.\psi} \; \bigwedge\text{-}I}}$$

---

[331]Suppose we want to prove $((A \to B) \to A) \to A$. If we allow for instantiation of the free variables A and B, we could easily end up proving $((A \to A) \to A) \to A$. This is probably not what we want. In fact the proof has little to do with the proof of $((A \to B) \to A) \to A$ that is schematic in A and B.

In terms of $\mathcal{M}_{Prop}$, we want to prove $\bigwedge AB.[\![((A \to B) \to A) \to A]\!]$

Recall that $((A \to B) \to A) \to A$ is Peirce's law.

[332]The more free variables in the goal we allow Isabelle to instantiate, the more unifiers there are. This may increase the search space to the extent of making it impossible to find a proof.

Instantiation of the Initial Goal

Previously, when we proved ψ we in fact proved $\bigwedge\_.\psi$.

Now, allowing for instantiation of ψ, we in fact prove $\bigwedge\_.\psi\theta$.

$$\frac{\dfrac{\bigwedge\_.\psi \Rightarrow \psi}{\psi \Rightarrow \psi}\; \bigwedge\text{-}E}{\vdots \qquad \vdots}$$

$$\frac{\dfrac{\psi\theta}{\bigwedge\_.\psi\theta}\; \bigwedge\text{-}I}{}$$

---

[331]Suppose we want to prove $((A \to B) \to A) \to A$. If we allow for instantiation of the free variables A and B, we could easily end up proving $((A \to A) \to A) \to A$. This is probably not what we want. In fact the proof has little to do with the proof of $((A \to B) \to A) \to A$ that is schematic in A and B.

In terms of $\mathcal{M}_{Prop}$, we want to prove $\bigwedge AB.[\![((A \to B) \to A) \to A]\!]$

Recall that $((A \to B) \to A) \to A$ is Peirce's law.

[332]The more free variables in the goal we allow Isabelle to instantiate, the more unifiers there are. This may increase the search space to the extent of making it impossible to find a proof.

Instantiation of the Initial Goal

Previously, when we proved ψ we in fact proved $\bigwedge_-.\psi$.

Now, allowing for instantiation of ψ, we in fact prove $\bigwedge_-.\psi\theta$.

$$\dfrac{\dfrac{\bigwedge_-.\psi \Rightarrow \psi}{\psi \Rightarrow \psi}\;\bigwedge\text{-}E}{\vdots \qquad \vdots}$$

$$\dfrac{\dfrac{\psi\theta}{\bigwedge_-.\psi\theta}\;\bigwedge\text{-}I}{}$$

This may not be what we want[331].

Problem: more unifiers, hence bigger search space[332].

---

[331]Suppose we want to prove $((A \to B) \to A) \to A$. If we allow for instantiation of the free variables A and B, we could easily end up proving $((A \to A) \to A) \to A$. This is probably not what we want. In fact the proof has little to do with the proof of $((A \to B) \to A) \to A$ that is schematic in A and B.

In terms of $\mathcal{M}_{Prop}$, we want to prove $\bigwedge AB.[\![((A \to B) \to A) \to A]\!]$

Recall that $((A \to B) \to A) \to A$ is Peirce's law.

[332]The more free variables in the goal we allow Isabelle to instantiate, the more unifiers there are. This may increase the search space to the extent of making it impossible to find a proof.

Two Kinds of Free Variables

In Isabelle, control over instantiation is given by having two kinds of free variables:

- ordinary variables must not become instatiated;

- metavariables (unknowns, schematic variables) may become instantiated.

In goals we can have both kinds, in rules we have metavariables. Try it out in Isabelle![333]

Once a theorem is proven, any free variables will be made metavariables[334], and the reading is as for rules: The theorem is implicitly universally quantified over the free variables.

---

[333]To understand the difference, try proving $A \wedge B \rightarrow P$ and $A \wedge B \rightarrow ?P$ in Isabelle. The first won't succeed while the second may succeed in various ways.

[334]Prove $A \wedge B \rightarrow ?P$ in Isabelle and save (qed) it as a theorem and then have a look at the theorem.

13.6 Conclusion on Isabelle's Metalogic

The logic $\mathcal{M}$ and its proof system are small.

What makes $\mathcal{M}$ powerful enough to encode a large variety of object logics?

- The λ-calculus is very powerful for expressing syntax and syntactic manipulations ($\rightarrow$ substitution). $\mathcal{M}$ must be extended by appropriate signature for an object logic.

- Rules of the object logic can be encoded and added to $\mathcal{M}$[335] as axioms.

---

[335]In some course on propositional logic, you may have learned that the connective $\rightarrow$ is not really necessary since $A \rightarrow B$ is equivalent to $\neg A \vee B$. Likewise, we considered $\neg A$ as syntactic sugar for $A \rightarrow \perp$.

Therefore, when we introduce a logic $\mathcal{M}$ that is so extremely simple as far as the number of logical symbols is concerned (just $\Rightarrow$, $\equiv$, $\bigwedge$), one might think that the idea is that all the other logical symbols one usually needs are just syntactic sugar. This is not the case!

To encode propositional logic or FOL in $\mathcal{M}$, we must add their rules as axioms.

Later, we will be working with a logic just slightly richer than $\mathcal{M}$ but still quite simple, and there the idea is indeed that all the other logical symbols one usually needs are just syntactic sugar.

Conclusion (2)

General principles of proof building (e.g. resolution, proving by assumption, side condition checking) are not something that must be justified by complicated (and thus error-prone) explanations in natural language — they are formal derivations in the metalogic.

This has two big advantages: shared support and high degree of confidence.

14 HOL: Foundations

14.1 Overview

HOL is expressive foundation[336] for

- **Mathematics**: analysis, algebra, . . .

- **Computer science**: program correctness, hardware verification, . . .

14 HOL: Foundations

14.1 Overview

HOL is expressive foundation[336] for

- Mathematics: analysis, algebra, . . .

- Computer science: program correctness, hardware verification, . . .

HOL is very similar to $\mathcal{M}$, but it "is" an object logic[337]!

- HOL is classical[338].

- Still[339] important: modeling of problems/domains (now within HOL).

- Still important: deriving relevant reasoning principles.

---

[336]Theorem proving in higher-order logic is an active research area with some impressive applications.

[337]The differences between $\mathcal{M}$ and HOL are subtle and the matter is further complicated by the fact that there are some variations in the way in which the Isabelle metalogic $\mathcal{M}$ on the one hand and the object logic HOL on the other hand are presented.

But what matters for us here is that HOL is an object logic, i.e., it is one of the object logic that can be represented by $\mathcal{M}$, just like propositional logic or first-order logic. That is to say, we use HOL as object logic.

[338]Recall the distinction between classical and intuitionistic logics. There is a particular rule in HOL from which the rule of the excluded middle can be derived. This is in contrast to constructive (intuitionistic) logics.

[339]We have previously looked at metatheory, i.e., how can one logic be represented/modeled in a metalogic.

In particular, we have seen how general reasoning principles

Isabelle/HOL vs. Alternatives

We will use Isabelle/HOL[340].

- Could forgo the use of a metalogic[341] and employ alternatives, e.g., HOL system or PVS, or constructive provers[342] such as Coq, Nuprl.

- Choice depends on culture and application.

---

can be derived in the metalogic.

We now set aside the issue of metalogics, but there is still an issue of modeling one system within another: how do we model problems/domains within HOL? How do we derive reasoning principles?

[340]We use Isabelle/HOL, and this means that HOL is an object logic represented by the metalogic $\mathcal{M}$.

[341]There are theorem proving systems that have no metalogic, but rather have a particular logic hard-wired into them, e.g. a HOL system or PVS.

[342]Constructive provers are based on intuitionistic logic. The rationale is that one has to give evidence for any statement. Coq and Nuprl are examples of such systems.

Safety through Strength

Safety[343] via conservative (definitional) extensions:

- Small kernel of constants and rules;

- extend theory with new constants and types defined using existing ones;

- derive properties/theorems.

Contrast with:

- Weak logics (e.g., propositional logic): can't define much;

- axiomatic extensions[344]: can lead to inconsistency.

Bertrand Russel once likened the advantages of postulation over definition to the advantages of theft over honest toil!

---

[343]The principle is simple: the smaller a system is, the easier it is to check that it is correct, and the more confident one can be about it.

We have seen this before when we argued for the use of metalogics. However, in that context, we still had to add further axioms to $\mathcal{M}$. Here this is not the case.

Safety through strength means: HOL is strong enough to model interesting systems without having to add further axioms – that's what makes it safe.

[344]What we attempt to do here has similarities to the process of representing an object logic in a metalogic. But an important difference must be noted.

We will see many extensions of the HOL kernel by constants (and types). The definitions of those constants and types involve axioms that must be added according to a strict discipline. Other than that, we will not add any axioms!

Set Theory as Alternative?

Set theory is the logician's choice as basis for modern mathematics.

- ZFC[345] (in Isabelle): impressive applications!

- Bernays-Gödel: used for resolution since finitely axiomatizable.

Set theories (both) distinguish between sets and classes.

- Consistency maintained as some collections are "too big" to be sets, e.g., class of all sets V is not a set.

- A class cannot belong to another class (let alone a set)!

---

[345]ZFC stands for Zermelo-Fränkel set theory with choice [Dev93, Ebb94].

Finally: We Choose HOL!

HOL developed by [Chu40, Hen50] and rediscovered by [And02, GM93].

- Rationale: one usually works with typed entities.

- Reasoning is then easier with support for types.

  HOL is classical logic based on $\lambda^{\rightarrow}$.

- Isabelle/HOL also supports "mod cons" [346] like polymorphism and type classes!

  HOL is weaker than ZF set theory, but for most applications this does not matter. If you prefer ML to Lisp, you will probably prefer HOL to ZF. (Paulson)

---

[346] "Mod cons" stands for "modern conveniences".

What Does Higher-Order Mean?

"Type" order[347]	Logic order	Example
0		$A \wedge B \to B \wedge A$
+ quantification		
1	1	$\forall x, y.\, R(x,y) \to R(y,x)$
+ quantification	2	$False \equiv \forall P.\, P$
		$P \wedge Q \equiv \forall R.\, (P \to Q \to R)$
2	3	
+ quantification	4	$\forall X.\, (X(R,S) \leftrightarrow (\forall x.\, R(x) \to S(x)))$
		$\to X(R', S')\ (\equiv subrel(R', S')^{348})$

---

[347]Recall the definition of an order on types.

Assume here, as we did in the lecture on representing syntax, that there is a type i of individuals and a type o for truth values. These objects are of 0th-order type.

A predicate symbol is a function taking individuals to a truth value, so it is an object of first-order type. A function symbol is a function taking individuals to an individual, so it is also an object of first-order type.

In the table, we classify logics by the order of the non-logical symbols (e.g., for first-order logic: variables, function symbols, predicate symbols).

A hierarchy of logics is obtained by the following alternation:

- admit an additional order for the non-logical symbols in the logic;

- admit quantification over symbols of that order.

If one only admits symbols of type o (propositional variables), one obtains propositional logic. At this stage, one also admits symbols of type i, although this is still degenerate since one cannot form any formulas with them (without any function or predicate symbols).

One can now admit quantification over individuals, although one can still not do much with that since one cannot form any formulas.

Consider the binary predicate $subrel$ which takes two unary relations as arguments. $subrel(R, S)$ is defined as true whenever R is a subrelation of S, i.e. when $\forall x.\, R(x) \rightarrow S(x)$.

Now instead of defining such a predicate and writing, say, a formula $subrel(R', S')$, one could abstract from that name and write

$$\forall X.\, (X(R, S) \leftrightarrow (\forall x.\, R(x) \rightarrow S(x))) \rightarrow X(R', S')$$

The subformula $X(R, S) \leftrightarrow (\forall x.\, R(x) \rightarrow S(x))$ is true if and only if X is indeed the predicate $subrel$ and so the entire formula is true if R' is indeed a subrelation of S'.

HOL = Union of All Finite Orders

ω-order logic, also called finite-type theory or higher-order logic (HOL), includes logics of all finite orders.

14.2 Syntax

Syntactically, HOL is a polymorphic (although not necessarily) variant of $\lambda^{\rightarrow}$ with certain default types and constants. Default constants can be called logical symbols.

Types (Review)

Given a set of type constructors, say $\mathcal{B}^{349} = \{bool, \_ \rightarrow \_, ind^{350}, \_ \times \_^{351}, \_ \; list, \_ \; set, \ldots\}$, polymorphic types are defined by $\tau ::= \alpha \mid (\tau, .., \tau) \; T$, where α is type variable.

- $bool$ is also called o in literature [Chu40, And02]. Confusingly, the truth value type in Isabelle/HOL (i.e., object-level) is called $bool$.

- $bool$ and $\rightarrow$ always present in HOL; ind will also play a special role; other type constructors may be defined.

- Note polymorphism[352]!

---

[349]As before, we use the letter $\mathcal{B}$ to denote a particular set of type constructors.

Note that this set is not hard-wired into HOL, but can be specified as part of a particular HOL language. One can therefore speak of $\mathcal{B}$ as a type signature.

$\mathcal{B}$ is some fixed set "defined by the user". In Isabelle, there is a syntax provided for this purpose.

However, some type constructors are always present.

[350]ind ("indefinite") is a type constructor which stands for a type with infinitely many members, a concept which is central in HOL, as we will see later.

[351]For any two types τ and σ, we write $\tau \times \sigma$ for the type of pairs where the first component is of type τ and the second component is of type σ.

The infix syntax is in analogy to $\rightarrow$.

The pair type is not in the core of HOL, but it can be defined in it.

[352]We have seen the generalization of $\lambda^{\rightarrow}$ to polymorphism.

Terms

Reminder: $e ::= x \mid c \mid (ee) \mid (\lambda x^{\tau^{353}}.e)$

Typing rules as in polymorphic λ-calculus, with Σ defining and typing constants.

Terms of type $bool$ are called (well-formed) formulae.

In HOL, Σ always includes:

$$
\begin{aligned}
True, False^{354} &: bool \\
= &: \alpha \to \alpha \to bool \quad \text{(polymorphic, or set}^{355}\text{)} \\
\to &: bool \to bool \to bool \\
\epsilon &: (\alpha \to bool) \to \alpha \quad \text{(in Isabelle: } \texttt{Eps} \text{ or } \texttt{SOME}^{356}\text{)}
\end{aligned}
$$

Note that in order to simplify the presentation, we neglect polymorphism in the section on semantics. In that section, τ and σ will be metavariables (used in the description of the formalism) ranging over types, rather than type variables of a polymorphic type system.

14.3 Semantics

Intuitively: many-sorted semantics + functions

- FOL: structure is domain and functions/relations.

$$\mathcal{A} = \langle \mathcal{D} \qquad , I_{\mathcal{A}} \rangle$$

14.3 Semantics

Intuitively: many-sorted semantics + functions

- FOL: structure is domain and functions/relations. Many-sorted FOL: domains are sort-indexed

$$\mathcal{A} = \langle \mathcal{D}_1, \ldots, \mathcal{D}_n, I_\mathcal{A} \rangle$$

14.3 Semantics

Intuitively: many-sorted semantics + functions

- FOL: structure is domain and functions/relations. Many-sorted FOL: domains are sort-indexed

$$\mathcal{A} = \langle \mathcal{D}_1, \ldots, \mathcal{D}_n, I_{\mathcal{A}} \rangle$$

- HOL extends idea: $\mathcal{D}$ indexed by (infinitely many) types.

- Complications due to polymorphism [GM93].

- We only give a monomorphic variant of semantics here!

Model Based on Universe of Sets $\mathcal{U}$

$\mathcal{U}$ is a collection of sets (domains), fulfilling closure conditions:

Inhab: Each $X \in \mathcal{U}$ is a nonempty set

Sub: If $X \in \mathcal{U}$ and $Y \neq \emptyset \subseteq X$, then $Y \in \mathcal{U}$

Prod: If $X, Y \in \mathcal{U}$ then $X \times Y \in \mathcal{U}$.

Pow: If $X \in \mathcal{U}$ then $\wp(X) = \{Y \mid Y \subseteq X\} \in \mathcal{U}$

Infty: $\mathcal{U}$ contains a distinguished infinite set[357] I

Choice: There is a function $ch \in \Pi_{X \in \mathcal{U}}.X$.

---

[357] The infinity axiom

$$\frac{}{\exists f^{(ind \to ind)}.injective\ f \wedge \neg surjective\ f}\ infty$$

says that there is a function from I to I (the postulated infinite set in $\mathcal{U}$) which is injective (any two different elements e, e' of I have different images under f) but not surjective (there exists an element of I which is not the image of any element).

Such a function can only exist if I is infinite, and in fact the axiom expresses the very essence of infinity, as we will see later.

Think of the natural numbers and the successor function as an example: for any two different natural numbers, the successors are different, and the number 0 is not the successor of any number.

Prod: Encoding $X \times Y$

$X \times Y$ is the Cartesian product, i.e., the set of pairs (x, y) such that $x \in X$ and $y \in Y$.

One can actually "encode" a tuple (x, y) without explicitly postulating the "existence of tuples"[358]. E.g.: $(x, y) \equiv \{\{x\}, \{x, y\}\}$.

---

[358]According to usual mathematical practice, one would argue that if two sets A and B are well-defined, then the set $A \times B$ of pairs (tuples) (a, b) where $a \in A$ and $b \in B$ is also well-defined.

That is, we assume that if one understands what a and b are, then one also understands what the pair (a, b) is. A pair is a "semantic object".

Ultimately, semantics can only be understood using one's intuition, and only be explained using natural language. One should try to base the semantics on a very small number of fundamental concepts.

Therefore, one might want to avoid having a concept "pair" ("tuple") explicitly, or put differently, one might want to reduce "pairs" to something even more fundamental. That's what is intended by the encoding $\{\{x\}, \{x, y\}\}$.

Note that this reduction step somehow makes the type discipline invisible, because x and y might be semantic objects "of different type".

Choice: Picking a Member

The function ch takes a set $X \in \mathcal{U}$ as argument and returns a member of X.

We hence write $ch \in \Pi_{X \in \mathcal{U}}.X$[359], i.e., ch is of dependent type.

Essentially, the constant ϵ will be interpreted as ch, but you will see the technical details later.

---

[359]When we write $ch \in \Pi_{X \in \mathcal{U}}.X$, i.e., ch is of dependent type, then this is a statement on the semantic level. The expression $\Pi_{X \in \mathcal{U}}.X$ is not part of the formal syntax of HOL (unlike in LF, a system we have not treated here), and its meaning is only described in plain English, by saying that ch takes a set $X \in \mathcal{U}$ as argument and returns a member of X.

Function Space in $\mathcal{U}$

Define set $X \to Y$ as (graph of) functions[360] from X to Y.

- For nonempty Y[361], this set is nonempty and is a subset of $\wp(X \times Y)$.

- From closure conditions: $X, Y \in \mathcal{U}$ then $X \to Y \in \mathcal{U}$.

---

[360]In any basic math course on algebra, we learn that a binary relation between X and Y is set of a pairs of tuples of the form (x, y) where $x \in X$ and $y \in Y$. One also calls such a set a graph since one can view pairs (x, y) as edges.

We also learn that a relation R is called a function from X to Y if for each $x \in X$, there exists exactly one $y \in Y$ such that $(x, y) \in R$. Provided that Y is nonempty, a function from X to Y always exists.

Thus the set of functions from X to Y, denoted $X \to Y$, is a nonempty subset of the set of relations on X and Y, i.e., $\wp(X \times Y)$. Since $X \to Y$ is nonempty, by **Prod** we have that $X \to Y \in \mathcal{U}$.

[361]It is crucial in the semantics that any type is inhabited, i.e., has an element. The function space $X \to Y$ is empty if X is empty.

This means that $\mathcal{D}_{\tau \to \sigma}$ would necessarily be empty if τ is empty. This in turn would lead to a very nasty behavior of β-reduction: for any types τ, σ and a constant c of type σ,

Distinguished Sets

From

Infty: $\mathcal{U}$ contains a distinguished infinite set I

Sub: If $X \in \mathcal{U}$ and $Y \neq \emptyset \subseteq X$, then $Y \in \mathcal{U}$

it follows that the following sets exist in $\mathcal{U}$:

Unit: A distinguished 1-element[362] set $\{1\}$

Bool: A distinguished 2-element set $\{T, F\}$.

---

we expect $(\lambda x^\tau . c)x = c$. But this wouldn't hold if we cannot give a semantics to $(\lambda x^\tau . c)$ since $\mathcal{D}_{\tau \to \sigma}$ is empty.

Therefore: inhabitation.

We will also see two more specific points where inhabitation is crucial, one concerning the ϵ-operator and one concerning sets.

[362]Of course, the conditions on $\mathcal{U}$ do not per se enforce the existence of sets containing the elements 1 or T or F. Just as well, one could say that they enforce the existence of sets containing elements ☕ or 🚲 or ⚽.

It is only because the name of a semantic element is ultimately irrelevant that we claim, without loss of generality, that there is a 1-element set $\{1\}$ and a 2-element set $\{T, F\}$. We say that these sets are distinguished because they play a special role in the setup of the semantics.

Frames

For semantics, we neglect polymorphism. τ and σ range over types.

A frame is a collection $\{\mathcal{D}_\tau\}_\tau$ of non-empty sets (domains) $\mathcal{D}_\tau \in \mathcal{U}$, one for each type τ, where:

- $\mathcal{D}_{bool} = \{T, F\}$;

- $\mathcal{D}_{\tau \to \sigma} \subseteq \mathcal{D}_\tau \to \mathcal{D}_\sigma$, i.e., some collection of functions from $\mathcal{D}_\tau$ to $\mathcal{D}_\sigma$.

- $\mathcal{D}_{ind} = I$.

Note: for fundamental reasons discussed later, one cannot simply define $\mathcal{D}_{\tau \to \sigma} = \mathcal{D}_\tau \to \mathcal{D}_\sigma$ at this stage.

Interpretations

An interpretation $\mathfrak{M} = \langle \{\mathcal{D}_\tau\}_\tau, \mathcal{J} \rangle$ is a frame $\{\mathcal{D}_\tau\}_\tau$ and a denotation function $\mathcal{J}$ mapping each constant of type τ to an element of $\mathcal{D}_\tau$ where:

- $\mathcal{J}(\mathit{True}) = T$ and $\mathcal{J}(\mathit{False}) = F$;

- $\mathcal{J}(=_{\tau \to \tau \to bool})^{363}$ is equality on $\mathcal{D}_\tau$;

- $\mathcal{J}(\to)$ is implication function over $\mathcal{D}_{bool}$. For $b, b' \in \{T, F\}$,

$$\mathcal{J}(\to)(b, b') = \begin{cases} F & \text{if } b = T \text{ and } b' = F \\ T & \text{otherwise} \end{cases}$$

---

[363] For $=$ and ϵ, we give type subscripts in the presentation of the semantics since we assume, conceptually, that there are infinitely many copies of those constants, one for each type. We do this to avoid explicit polymorphism in this presentation.

- $\mathcal{J}(\epsilon_{(\tau \to bool) \to \tau})$ is defined by (for $f \in (\mathcal{D}_\tau \to \mathcal{D}_{bool})$):

$$\mathcal{J}(\epsilon_{(\tau \to bool) \to \tau})(f)^{364} = \begin{cases} ch(f^{-1}(\{T\})) & \text{if } f^{-1}(\{T\}) \neq \emptyset \\ ch(\mathcal{D}_\tau) & \text{otherwise} \end{cases}$$

Note: If a frame $\{\mathcal{D}_\tau\}_\tau$ does not contain all of the functions used above, then $\{\mathcal{D}_\tau\}_\tau$ cannot belong to any interpretation.

---

[364]We have

$$\mathcal{J}(\epsilon_{(\tau \to bool) \to \tau})(f) = \begin{cases} ch(f^{-1}(\{T\})) & \text{if } f^{-1}(\{T\}) \neq \emptyset \\ ch(\mathcal{D}_\tau) & \text{otherwise} \end{cases}$$

ch is a (semantic) function which takes a nonempty set and returns an element from that set. f is a semantic function from $\mathcal{D}_\tau$ to $\mathcal{D}_{bool}$. However, f can be interpreted as set. This is done in all formality here: we write $f^{-1}(\{T\})$. One says that f is the characteristic function of the set $f^{-1}(\{T\})$.

Now the type of ϵ is $(\tau \to bool) \to \tau$ (for any τ), so ϵ expects a function as argument, which can be interpreted as a set as just stated. This set can be empty or nonempty. In case it is nonempty, an element is picked from the set non-deterministically. If the set is empty, an element from the type τ (which must be nonempty since each type is interpreted as nonempty set). Note the importance of inhabitation.

A Terminological Note

The terminology is slightly different from FOL:

In FOL, "$\langle \{\mathcal{D}_\tau\}_\tau, \mathcal{J} \rangle$" is called structure and "$\mathcal{J}$" is called interpretation.

In HOL, $\langle \{\mathcal{D}_\tau\}_\tau, \mathcal{J} \rangle$ is called interpretation and $\mathcal{J}$ is called denotation function.

The Value of Terms (Naive)

In analogy to FOL, given an interpretation $\mathfrak{M} = \langle \{\mathcal{D}_\tau\}_\tau, \mathcal{J} \rangle$ and a type-indexed collection of assignments[365] $A = \{A_\tau\}_\tau$, define $\mathcal{V}_A^{\mathfrak{M}}$ such that $\mathcal{V}_A^{\mathfrak{M}}(t_\rho) \in \mathcal{D}_\rho$ for all t, as follows:

1. $\mathcal{V}_A^{\mathfrak{M}}(x_\tau) = A(x_\tau)$;

2. $\mathcal{V}_A^{\mathfrak{M}}(c) = \mathcal{J}(c)$ for c a constant;

3. $\mathcal{V}_A^{\mathfrak{M}}(s_{\tau \to \sigma}{}^{[366]}t_\tau) = (\mathcal{V}_A^{\mathfrak{M}}(s))(\mathcal{V}_A^{\mathfrak{M}}(t))$, i.e., the value of the function $\mathcal{V}_A^{\mathfrak{M}}(s)$ at the argument $\mathcal{V}_A^{\mathfrak{M}}(t)$;

4. $\mathcal{V}_A^{\mathfrak{M}}(\lambda x^\tau. t_\sigma) = $ the function from $\mathcal{D}_\tau$ into $\mathcal{D}_\sigma$ whose value for each $e \in \mathcal{D}_\tau$ is $\mathcal{V}_{A[x \leftarrow e]}^{\mathfrak{M}}{}^{[367]}(t)$.

What is the problem?

---

[365]An assignment (previously called valuation) maps variables to elements of a domain.

A type-indexed collection of assignments is an assignment that respects the types: a variable of type τ will be assigned to a member of $\mathcal{D}_\tau$ [GM93]. Note that a variable has a type by virtue of a context Γ, which is suppressed in our presentation of models.

[366]In the presentation of models, we give type subscripts for the cases $\mathcal{V}_A^{\mathfrak{M}}(s_{\tau \to \sigma}t_\tau)$ and $\mathcal{V}_A^{\mathfrak{M}}(\lambda x^\tau. t_\sigma)$ to indicate the types of s and t in those definitions. Note that a term has a type in a certain context Γ, which is suppressed in our presentation of models. The semantics is only defined for well-formed terms, in particular, applications and abstractions having types of the indicated forms.

[367]$A[x \leftarrow e]$ denotes the assignment that is identical to A except that $A(x) = e$.

The Value of Terms (Naive)

In analogy to FOL, given an interpretation $\mathfrak{M} = \langle \{\mathcal{D}_\tau\}_\tau, \mathcal{J} \rangle$ and a type-indexed collection of assignments[365] $A = \{A_\tau\}_\tau$, define $\mathcal{V}_A^{\mathfrak{M}}$ such that $\mathcal{V}_A^{\mathfrak{M}}(t_\rho) \in \mathcal{D}_\rho$ for all t, as follows:

1. $\mathcal{V}_A^{\mathfrak{M}}(x_\tau) = A(x_\tau)$;

2. $\mathcal{V}_A^{\mathfrak{M}}(c) = \mathcal{J}(c)$ for c a constant;

3. $\mathcal{V}_A^{\mathfrak{M}}(s_{\tau \to \sigma}{}^{[366]}t_\tau) = (\mathcal{V}_A^{\mathfrak{M}}(s))(\mathcal{V}_A^{\mathfrak{M}}(t))$, i.e., the value of the function $\mathcal{V}_A^{\mathfrak{M}}(s)$ at the argument $\mathcal{V}_A^{\mathfrak{M}}(t)$;

4. $\mathcal{V}_A^{\mathfrak{M}}(\lambda x^\tau . t_\sigma) =$ the function from $\mathcal{D}_\tau$ into $\mathcal{D}_\sigma$ whose value for each $e \in \mathcal{D}_\tau$ is $\mathcal{V}_{A[x \leftarrow e]}^{\mathfrak{M}}{}^{[367]}(t)$.

What is the problem? Condition 4!

---

[365]An assignment (previously called valuation) maps variables to elements of a domain.

A type-indexed collection of assignments is an assignment that respects the types: a variable of type τ will be assigned to a member of $\mathcal{D}_\tau$ [GM93]. Note that a variable has a type by virtue of a context Γ, which is suppressed in our presentation of models.

[366]In the presentation of models, we give type subscripts for the cases $\mathcal{V}_A^{\mathfrak{M}}(s_{\tau \to \sigma} t_\tau)$ and $\mathcal{V}_A^{\mathfrak{M}}(\lambda x^\tau . t_\sigma)$ to indicate the types of s and t in those definitions. Note that a term has a type in a certain context Γ, which is suppressed in our presentation of models. The semantics is only defined for well-formed terms, in particular, applications and abstractions having types of the indicated forms.

[367]$A[x \leftarrow e]$ denotes the assignment that is identical to A except that $A(x) = e$.

Condition 4 Is Critical

For $\mathcal{V}_A^{\mathfrak{M}}$ to be well-defined, the function from $\mathcal{D}_\tau$ into $\mathcal{D}_\sigma$ in condition 4 must live

- in some domain of $\mathcal{U}$ (since it is required that $\mathcal{V}_A^{\mathfrak{M}}(t_\rho) \in \mathcal{D}_\rho$ for all t, and $\mathcal{D}_\rho \in \mathcal{U}$): this is guaranteed by closure conditions on $\mathcal{U}$;

---

[368] In condition 4, the semantics of $\lambda x^\tau . t_\sigma$ is defined unambiguously as a certain function. But in general, there is no guarantee that this function is actually in $\mathcal{D}_{\tau \to \sigma}$, and in this case, $\mathfrak{M} = \langle \{\mathcal{D}_\tau\}_\tau, \mathcal{J} \rangle$ would not be a model.
[369] General models must be distinguished from standard models, as well see later.

We sometimes omit the word "general" in general model.

Condition 4 Is Critical

For $\mathcal{V}_A^{\mathfrak{M}}$ to be well-defined, the function from $\mathcal{D}_\tau$ into $\mathcal{D}_\sigma$ in condition 4 must live

- in some domain of $\mathcal{U}$ (since it is required that $\mathcal{V}_A^{\mathfrak{M}}(t_\rho) \in \mathcal{D}_\rho$ for all t, and $\mathcal{D}_\rho \in \mathcal{U}$): this is guaranteed by closure conditions on $\mathcal{U}$;

- in a certain domain of $\mathcal{U}$, namely $\mathcal{D}_{\tau \to \sigma}$[368]; for this, $\mathcal{D}_{\tau \to \sigma}$ must be big enough.

If $\mathcal{V}_A^{\mathfrak{M}}$ is well-defined, we call $\mathfrak{M} = \langle \mathcal{D}_\tau, \mathcal{J} \rangle$ a (general)[369] model.

---

[368]In condition 4, the semantics of $\lambda x^\tau . t_\sigma$ is defined unambiguously as a certain function. But in general, there is no guarantee that this function is actually in $\mathcal{D}_{\tau \to \sigma}$, and in this case, $\mathfrak{M} = \langle \{\mathcal{D}_\tau\}_\tau, \mathcal{J} \rangle$ would not be a model.

[369]General models must be distinguished from standard models, as well see later.

We sometimes omit the word "general" in general model.

Models

Hence: Not all interpretations are general models, but we restrict our attention to the general models.

If $\mathcal{D}_{\tau \to \sigma}$ is the set of all functions from $\mathcal{D}_\tau$ to $\mathcal{D}_\sigma$, then it is certainly "big enough". In this case, we speak of a standard model. Important for completeness.

If $\mathfrak{M}$ is a general model and A an assignment, then $\mathcal{V}_A^{\mathfrak{M}}$ is uniquely determined.

$\mathcal{V}_A^{\mathfrak{M}}(t)$ is value of t in $\mathfrak{M}$ wrt. A.

Note that in contrast to first-order logic, "model" does not mean "an interpretation that makes a formula true".

Satisfiability and Validity

A formula (term of type $bool$) ϕ is satisfiable wrt. a model $\mathfrak{M}$ if there exists an assignment A such that $\mathcal{V}_A^{\mathfrak{M}}(\phi) = T$.

A formula ϕ is valid wrt. a model $\mathfrak{M}$ if for all assignments A, we have $\mathcal{V}_A^{\mathfrak{M}}(\phi) = T$.

A formula ϕ is valid in the general sense if it is valid in every general model.

A formula ϕ is valid in the standard sense if it is valid in every standard model.

Existence of Values

Closure conditions for general models guarantee every well-formed term has a value under every assignment, and this means that certain values must exist, e.g.,

- Closure under functions: since $\mathcal{V}_A^{\mathfrak{M}}(\lambda x^\tau. x)$ is defined, the identity function from $\mathcal{D}_\tau$ to $\mathcal{D}_\tau$ must always belong to $\mathcal{D}_{\tau \to \tau}$.

- Closure under application: if $\mathcal{D}_\mathbb{N}$ is natural numbers, and $\mathcal{D}_{\mathbb{N} \to \mathbb{N} \to \mathbb{N}}$ contains addition function p where $p\, x\, y = x + y$, then $\mathcal{D}_{\mathbb{N} \to \mathbb{N}}$ must contain k where $k\, x = 2x + 5$, since $k = \mathcal{V}_A^{\mathfrak{M}}(\lambda x_\mathbb{N}. f(f\, x\, x)\, y)$ where $A(f) = p$ and $A(y) = 5$.

14.4 Basic Rules

We now give the core calculus of HOL. Its rules can be stated using only the constants $=$, $\rightarrow$, and ϵ. However, there will be one rule, *tof* ("true or false"), which would be hard to read if we did that.

So we allow ourselves to "cheat"[370] and also use constants *True*, *False*, $\vee$ to write rule *tof*.

Later we will define those constants, i.e., regard them as syntactic sugar.

---

[370]Rule *tof* can be written as follows:

$$\frac{}{} \quad \textit{tof}$$

$$(\lambda\psi.\ (\phi = (\lambda x.x = \lambda x.x) \rightarrow \psi) \rightarrow$$
$$\quad\quad (\phi = ((\lambda\eta.\eta) = \lambda x.(\lambda x.x = \lambda x.x)) \rightarrow \psi) \rightarrow \psi) =$$
$$(\lambda x.(\lambda x.x = \lambda x.x))$$

Our notation for rule *tof* is thus based on the following definitions:

$$
\begin{aligned}
\textit{True} \ &= \ (\lambda x^{bool}.x = \lambda x.x) \\
\textit{False} \ &= \ \forall \phi^{bool}.\phi \\
\vee \ &= \ \lambda\phi\eta.\forall\psi.(\phi \rightarrow \psi) \rightarrow (\eta \rightarrow \psi) \rightarrow \psi
\end{aligned}
$$

Basic Rules in Sequent Notation

$$\frac{}{\Gamma \vdash \phi = \phi} \; refl \qquad\qquad \frac{\Gamma \vdash \phi = \eta \quad \Gamma \vdash P(\phi)}{\Gamma \vdash P(\eta)} \; subst$$

$$\frac{\Gamma \vdash \phi\, x = \eta\, x}{\Gamma \vdash \phi = \eta} \; ext^{*371} \qquad\qquad \frac{\Gamma, \phi \vdash \eta}{\Gamma \vdash \phi \to \eta} \; impl$$

$$\frac{\Gamma \vdash \phi \to \eta \quad \Gamma \vdash \phi}{\Gamma \vdash \eta} \; mp$$

$$\frac{}{\Gamma \vdash (\phi \to \eta) \to (\eta \to \phi) \to (\phi = \eta)} \; iff$$

$$\frac{}{\phi = True \lor \phi = False} \; tof \qquad \frac{\Gamma \vdash \phi x}{\Gamma \vdash \phi(\epsilon x.\phi x^{373})} \; selectI^{372}$$

---

[371]The rule

$$\frac{\Gamma \vdash \phi\, x = \eta\, x}{\Gamma \vdash \phi = \eta} \; ext$$

has the side condition that $x \notin FV(\Gamma)$.

   Phrased like

$$\frac{\phi\, x = \eta\, x}{\phi = \eta} \; ext$$

the rule has the side condition that x must not occur freely in the derivation of $\phi\, x = \eta\, x$.

   [372]You may wonder why there is no rule for eliminating ϵ. We will later see a rule derivation where an ϵ is effectively eliminated, and we will also see that this is done without requiring a rule explicitly for this purpose.

   Apart from that, the ϵ-operator is used in HOL as basis for defining $\exists$ and the if-then-else constructs. Once we have derived the appropriate rules for those, we will not explicitly encounter ϵ anymore.

   [373]For readability, we will frequently use a syntax that one is

Axiom of Infinity

There is one additional rule (axiom) that will give us the existence of infinite sets:

$$\frac{}{\exists f^{(ind \to ind)}.injective^{374} \ f \wedge \neg surjective \ f} \ \textit{infty}$$

Has special role. Interesting to look at HOL with or without infinity. Won't consider infinity today.

Note "cheating" (use of $\exists$).

These eight (nine) rules are the entire basis!

---

more used to than higher-order abstract syntax:

$\epsilon x.\phi x$ stands for $\epsilon(\phi)$.

$\forall x.\phi(x)$ stands for $\forall(\phi)$, and likewise for $\exists$.

We have done the same previously for $\mathcal{M}$.

Soundness and Completeness

Soundness is straightforward [And02, p. 240].

Completeness only follows w.r.t. general models, as opposed to standard models. Recall that a standard model is one where $\mathcal{D}_{\tau \to \sigma}$ is always the set of all functions from $\mathcal{D}_{\tau}$ to $\mathcal{D}_{\sigma}$.

There are formulas that are valid in all standard models, but not in all general models, and which cannot be proven in our calculus. Our calculus can prove the formulas that are true in all general models including non-standard ones. This reconciles HOL with Gödel's incompleteness theorem[375].

If we consider a version of HOL without infinity, then every model is a standard model[376] and so completeness holds.

---

[375]This is a standard trick when faced with the problem that a deductive system is not complete. One can either enlarge the set of axioms, or one can weaken the models by permitting more models. If we allow more models, then fewer theorems will be valid (i.e., hold in all models), and so fewer theorems will have to be provable in the derivation system.

Here, completeness is based on general models, and not standard models. This resolves the apparent contradiction with Gödel's incompleteness theorem: HOL with infinity contains I, hence the natural numbers, hence arithmetic ...). By Gödel's incompleteness theorem, there cannot be a consistent derivation system that can prove all valid theorems in the natural numbers.

A readable account on this problem can be found in [And02, ch. 7].

[376]We might consider a version of HOL without infinity, i.e., one where each domain is finite (note that $\mathcal{U}$ is still infinite, since there are infinitely many types, e.g., $bool$, $bool \to bool$,

14.5 Isabelle/HOL

We now look at a particular instance of HOL (given by defining certain types and constants) which essentially corresponds to the HOL theory of Isabelle[377].

We present language and rules[378] using "mathematical" syntax, but also comparing with Isabelle (concrete/HOAS) syntax.

We take polymorphism back on board.

$bool \rightarrow bool \rightarrow bool, \ldots$)).

One can see that every function in such a finite domain is representable as a λ-term, and so for any σ and τ, we must have $\mathcal{D}_{\tau \rightarrow \sigma} = \mathcal{D}_\tau \rightarrow \mathcal{D}_\sigma$.

For details consult [And02, §54].

[377]This file should be contained in your Isabelle distribution. Or, if you only have an Isabelle executable, you can find the sources here:

http://isabelle.in.tum.de/library/

Moreover, in HOL_lemmas.ML, you find all the derivations of the rules presented in this lecture.

However, the presentation of this lecture is still partly based on HOL.thy of Isabelle 98. E.g., the definition of Ex_def is now different from the one presented here.

Note also that here in the slides, we use a style of displaying Isabelle files which uses some symbols beyond the usual ASCII set.

[378]We will mix natural deduction (with discharging assump-

(Central Parts of the) Language

tions), natural deduction written in sequent style, and Isabelle syntax.

For a thorough account of this, consult [SH84].

Some general remarks about the correspondence: A rule

$$\frac{\psi}{\phi}$$

in ND notation corresponds to an Isabelle rule $\psi \implies \phi$.

A rule

$$\begin{array}{c} [\rho] \\ \vdots \\ \psi \\ \hline \phi \end{array}$$

is written as

$$\frac{\rho, \Gamma \vdash \psi}{\Gamma \vdash \phi}$$

in sequent style or

$$\frac{\rho \implies \psi}{\phi}$$

using the Isabelle meta-implication $\implies$.

A rule
$$\frac{\psi}{\phi(x)}$$

with side condition that x must not occur free in any undischarged assumption on which ψ depends is written as
$$\frac{\Gamma \vdash \psi}{\Gamma \vdash \phi(x)}$$

in sequent style, where the side condition reads: x must not occur free in Γ. Using the Isabelle meta-universal quantification, the rule is written
$$\frac{\bigwedge x.\psi}{\phi(x)}$$

We will switch between the various ways of writing the rules! This means in particular that we will use $\implies$ and $\bigwedge$ from Isabelle's metalogic.

$$\Sigma_0 =$$

$$
\begin{aligned}
\{ \quad & True, \ False^{379} && : bool, \\
& \neg_{\_}{}^{380} && : bool \rightarrow bool, \\
& {}_{\_}\wedge_{\_}, \ {}_{\_}\vee_{\_}, \ {}_{\_}\rightarrow_{\_} && : bool \rightarrow bool \rightarrow bool, \\
& \forall_{\_}, \ \exists_{\_} && : (\alpha \rightarrow bool) \rightarrow bool, \\
& \epsilon_{\_} && : (\alpha \rightarrow bool) \rightarrow \alpha, \\
& if\_then\_else_{\_} && : bool \rightarrow \alpha \rightarrow \alpha \rightarrow \alpha, \\
& {}_{\_} = {}_{\_} && : \alpha \rightarrow \alpha \rightarrow bool \}
\end{aligned}
$$

---

[379] For convenience (and to save space, we write $\ldots a : \tau,\ b : \tau \ldots$ as $\ldots a, b : \tau \ldots$ in a signature. This is of course syntactic sugar.

[380] We use a notation with _ to indicate the arity and fixity of constants, as this has been done for type constructors before.

The whole matter of arity of fixity is one of notational convenience. For example, as the type of $\wedge$ indicates, we should write $(\wedge \phi)\psi$ (Curryed notation), but we write $\phi \wedge \psi$ since it is more what we are used to.

Basic Rules in Isabelle Notation

```
refl:             "t = t"
subst:            "[| s = t; P(s) |] ==> P(t)"
ext:              "(!!x. (f x) = g x) ==>
                   (%x. f x) = (%x. g x)"
impI:             "(P ==> Q) ==> P-->Q"
mp:               "[| P-->Q;  P |] ==> Q"
iff:              "(P-->Q) --> (Q-->P) --> (P=Q)"
True_or_False: "(P=True) | (P=False)"
selectI:          "P (x) ==> P (@x. P x)"
```

See HOL.thy.

Basic Rules in Mixed Notation

$$\frac{}{\phi = \phi}\; \textit{refl} \qquad\qquad \frac{\phi = \eta \quad P(\phi)}{P(\eta)}\; \textit{subst}$$

$$\frac{\phi\, x = \eta\, x}{\phi = \eta}\; \textit{ext}^* \qquad\qquad \frac{\phi \implies \eta}{\phi \rightarrow \eta}\; \textit{impl}$$

$$\frac{\phi \rightarrow \eta \quad \phi}{\eta}\; \textit{mp}$$

$$\frac{}{(\phi \rightarrow \eta) \rightarrow (\eta \rightarrow \phi) \rightarrow (\phi = \eta)}\; \textit{iff}$$

$$\frac{}{\phi = \textit{True} \lor \phi = \textit{False}}\; \textit{tof} \qquad \frac{\phi x}{\phi(\epsilon x . \phi x)}\; \textit{selectI}$$

No more "Cheating": The Definitions

$$
\begin{aligned}
True^{381} &=^{382} (\lambda x^{bool}.x = \lambda x.x) \\
\forall^{383} &= \lambda\phi^{\alpha\to bool}.(\phi = \lambda x.\,True) \\
False^{384} &= \forall\phi^{bool^{385}}.\phi^{386} \\
\vee^{387} &= \lambda\phi\eta.\forall\psi.(\phi \to \psi) \to (\eta \to \psi) \to \psi \\
\wedge^{388} &= \lambda\phi\eta.\forall\psi.(\phi \to \eta \to \psi) \to \psi \\
\neg^{389} &= \lambda\phi.(\phi \to False) \\
\exists^{390} &= (\lambda\phi.\phi(\epsilon x.\phi x)) \\
If^{391} &= \lambda\phi^{bool}xy.\epsilon z.(\phi = True \to z = x)\wedge \\
&\quad\ (\phi = False \to z = y)
\end{aligned}
$$

---

[381]

$$
True = (\lambda x^{bool}.x = \lambda x.x)
$$

The term $\lambda x^{bool}.x = \lambda x.x$ evaluates to T, and so it is a suitable definition for the constant $True$.

Note that we give the type for x once. The right-hand side $\lambda x.x$ will thereby also be forced to be of type $bool \to bool$.

This is necessary for reasons that will become clear later.

Note that $(\lambda x^{bool}.x = \lambda x.x)$ is closed. Definitions must always be closed.

[382]It is a design choice if we want to add these definitions at the level of the object logic (HOL) or at the level of the $\mathcal{M}$. In the first case, we would use $=$ and have axioms such as

$$
True = (\lambda x^{bool}.x = \lambda x.x)
$$

In the second case, we would have meta-axioms

$$
True \equiv (\lambda x^{bool}.x = \lambda x.x)
$$

This would mean that we would regard $True$ merely as syntactic sugar. The second way corresponds to what is done in Isabelle, see HOL.thy. It is technically more convenient since rewriting is based on meta-level equalities.

Logically, it is not a big difference which way one chooses. We will have an exercise on this.

[383]

$$
\forall = \lambda\phi.(\phi \overset{382}{=} \lambda x.\,True)
$$

Note the use of HOAS here. $\forall$ should be a function that

$$If = \lambda\phi xy.\epsilon z.(\phi = True \to z = x) \wedge (\phi = False \to z = y)$$

The constant If stands for the if-then-else construct. Note first that $\epsilon z.(\phi = True \to z = x) \wedge (\phi = False \to z = y)$ is η-equivalent to $\epsilon z.(\lambda z.(\phi = True \to z = x) \wedge (\phi = False \to z = y)) z$, which is written $\epsilon(\lambda z.(\phi = True \to z = x) \wedge (\phi = False \to z = y))$ in the "real" HOL syntax, which uses the concept of HOAS.

The expression $\epsilon(\lambda z.(\phi = True \to z = x) \wedge (\phi = False \to z = y))$ picks a term from the set of terms z such that $(\phi = True \to z = x) \wedge (\phi = False \to z = y)$ holds. But this means that $z = x$ if $\phi = True$, or $z = y$ if $\phi = False$.

Since If should be a function which takes ϕ, x and y as arguments, we must abstract over those variables, giving $\lambda\phi xy.\epsilon z.(\phi = True \to z = x) \wedge (\phi = False \to z = y)$.

Note: Different Syntaxes

Mathematical	vs.	Isabelle, e.g.

$\neg\phi$ `Not Phi`

$\lambda x^{bool}.P$ $\%^{392}\texttt{x} :: {}^{393}\texttt{bool}.P$

HOAS	vs.	concrete, e.g.

$\forall\,(\lambda x^\tau.(\wedge p(x)\,q(x)))$ $\forall x^\tau.p(x)\wedge q(x)$

$\epsilon\,(P)$ $\epsilon x.P(x)$

We use all those forms as convenient. For displaying Isabelle files, we will sometimes use a style where some ASCII words (e.g. %) are replaced with mathematical symbols (e.g. λ).

---

[392] Note that the λ-binder of the object logic HOL is not distinguished from the λ-binder of Isabelle's metalogic $\mathcal{M}$. One could introduce an object level constant $lambda$, but one quickly sees that it would be an unnecessary overhead.

[393] As we have learned previously, λ-abstracted variables should have a type superscript, although this superscript is often omitted since the type can be inferred.

Since $\forall x.p(x)\wedge q(x)$ is the "concrete syntax" version of $\forall\,(\lambda x.(\wedge p(x)\,q(x)))$, it makes sense that we allow an optional superscript also for $\forall$-bound (and likewise for $\exists$-bound) variables.

In Isabelle the optional type annotation is written using :: instead of a superscript.

14.6 Conclusions on HOL

- HOL generalizes semantics of FOL:

  - $bool$ serves as type of propositions;

  - Syntax/semantics allows for higher-order functions.

- Logic is rather minimal: 8 or 9 rules, based on 3 constants, soundness straightforward.

- Logic complete (w.r.t. general models, but not standard models).

- Next lecture we will see how all well-known inference rules can be derived.

15 HOL: Deriving Rules

Outline

Last lecture: Introduction to HOL

- Basic syntax and semantics

- Basic eight (or nine) rules

- Definitions of $True$, $False$, $\wedge$, $\vee$, $\forall$ ...

Today:

- Deriving rules for the defined constants

- Outlook on the rest of this course

Reminder: Different Syntaxes

Mathematical   vs.   Isabelle, e.g.

$\neg\phi$ Not Phi

$\lambda x^{bool}.P$ %x :: bool. P

HOAS   vs.   concrete, e.g.

$\forall\,(\lambda x^\tau.(\wedge p(x)\,q(x)))$ $\forall x^\tau.p(x)\wedge q(x)$

$\epsilon\,(P)$ $\epsilon x.P(x)$

We use all those forms as convenient. For displaying Isabelle files, we will sometimes use a style where some ASCII words (e.g. %) are replaced with mathematical symbols (e.g. λ).

Reminder: Definitions

$$
\begin{aligned}
\textit{True} &= (\lambda x^{bool}.x = \lambda x.x) \\
\forall &= \lambda \phi^{\alpha \to bool}.(\phi = \lambda x.\textit{True}) \\
\textit{False} &= \forall \phi^{bool}.\phi \\
\vee &= \lambda \phi \eta. \forall \psi.(\phi \to \psi) \to (\eta \to \psi) \to \psi \\
\wedge &= \lambda \phi \eta. \forall \psi.(\phi \to \eta \to \psi) \to \psi \\
\neg &= \lambda \phi.(\phi \to \textit{False}) \\
\exists &= (\lambda \phi.\phi(\epsilon x.\phi x)) \\
\textit{If} &= \lambda \phi xy.\epsilon z.(\phi = \textit{True} \to z = x) \wedge \\
&\qquad (\phi = \textit{False} \to z = y)
\end{aligned}
$$

Derived Rules

The definitions can be understood either semantically (checking if each definition captures the usual meaning of that constant) or by their properties (= derived rules).

We now look at the constants in turn and derive rules for them. We will present derivations in natural deduction style.

We usually proceed as follows: first show a rule involving a constant, then replace the constant with its definition (if applicable), then show the derivation.

15.1 Equality

- Rule *sym*

$$\frac{s = t}{t = s} \; \textit{sym}$$

Derived Rules

The definitions can be understood either semantically (checking if each definition captures the usual meaning of that constant) or by their properties (= derived rules).

We now look at the constants in turn and derive rules for them. We will present derivations in natural deduction style.

We usually proceed as follows: first show a rule involving a constant, then replace the constant with its definition (if applicable), then show the derivation.

15.1 Equality

- Rule *sym* and ND derivation[394]

$$\frac{s = t \quad \dfrac{}{s = s} \; \text{refl}}{t = s} \; \text{subst}$$

---

[394]We present most of those proofs by giving a derivation tree for it, but sometimes, we also give an Isabelle proof script.

Note also the mix of syntaxes.

- Isabelle rule s=t ==> t=s. Proof script:

```
Goal "s=t ==> t=s";
by (etac subst 1);          (* P is %x.x=s *)
by (rtac refl 1);           (* s=s *)
qed "sym";
```

Equality: Transitivity and Congruences

- Rule *trans*

$$\frac{\begin{array}{c} r = s \\ s = t \end{array}}{r = t} \; trans$$

Equality: Transitivity and Congruences

- Rule *trans* and ND derivation

$$\frac{\dfrac{r = s}{s = r}\;sym \qquad s = t}{r = t}\;subst$$

Isabelle rule [| r=s; s=t |] ==> r=t

Equality: Transitivity and Congruences

- Rule *trans* and ND derivation

$$\dfrac{\dfrac{r = s}{s = r}\ sym \qquad s = t}{r = t}\ subst$$

  Isabelle rule `[| r=s; s=t |] ==> r=t`

- Congruences (only Isabelle forms):

  `(f::'a=>'b) = g ==> f(x)=g(x)` (*fun_cong*)

  `x=y ==> f(x)=f(y)` (*arg_cong*)

  Isabelle proofs using *subst* and *refl*.

Equality of Booleans (*iffI*)

Rule *iffI*

$$
\begin{array}{cc}
\begin{array}{c}
[P] \\
\vdots \\
Q
\end{array}
&
\begin{array}{c}
[Q] \\
\vdots \\
P
\end{array}
\end{array}
$$

$$
\frac{}{P = Q} \; iffI
$$

Equality of Booleans (*iffI*)

Rule *iffI* and ND derivation

$$
\cfrac{
\cfrac{(P \to Q) \to (Q \to P) \to (P = Q)
\quad iff \quad
\cfrac{\begin{array}{c}[P]\\\vdots\\Q\end{array}}{P \to Q}\ impl
}{(Q \to P) \to (P = Q)}\ mp
\qquad
\cfrac{\begin{array}{c}[Q]\\\vdots\\P\end{array}}{Q \to P}\ impl
}{P = Q}\ mp
$$

Isabelle rule [| P ==> Q; Q ==> P |] ==> P=Q.

Equality of Booleans ($iffD2$)

Rule $iffD2$

$$\frac{\begin{array}{c} P = Q \\ Q \end{array}}{P} \; iffD2$$

Equality of Booleans (*iffD2*)

Rule *iffD2* and ND derivation

$$\cfrac{\cfrac{P = Q}{Q = P} \; \text{sym} \qquad Q}{P} \; subst$$

Isabelle rule `[| P=Q; Q |] ==> P`.

15.2 *True*

$$True = ((\lambda x^{bool}.x) = (\lambda x.x))$$

- Rule *TrueI*

$$\frac{}{True} \; \textit{TrueI}$$

15.2 *True*

$$True = ((\lambda x^{bool}.x) = (\lambda x.x))$$

- Rule *TrueI*

$$\frac{}{(\lambda x.x) = (\lambda x.x)} \; \textit{TrueI}$$

15.2 *True*

$$True = ((\lambda x^{bool}.x) = (\lambda x.x))$$

- Rule *TrueI* and ND derivation

$$\frac{}{(\lambda x.x) = (\lambda x.x)} \; \textit{refl}$$

15.2 *True*

$$True = ((\lambda x^{bool}.x) = (\lambda x.x))$$

- Rule *TrueI* and ND derivation

$$\frac{}{(\lambda x.x) = (\lambda x.x)}\ \textit{refl}$$

- Rule *eqTrueE*

$$\frac{P = True}{P}\ \textit{eqTrueE}$$

15.2 *True*

$$True = ((\lambda x^{bool}.x) = (\lambda x.x))$$

- Rule *TrueI* and ND derivation

$$\frac{}{(\lambda x.x) = (\lambda x.x)} \; \textit{refl}$$

- Rule *eqTrueE* and ND derivation

$$\frac{P = True \quad \dfrac{}{True} \; \textit{TrueI}}{P} \; \textit{iffD2}$$

Isabelle rule `P=True ==> P`.

- Rule *eqTrueI*

$$\frac{P}{P = True} \ \textit{eqTrueI}$$

True **(Cont.)**

- Rule *eq TrueI* and ND derivation

$$\frac{\dfrac{}{\mathit{True}}\ \mathit{TrueI} \qquad P}{P = \mathit{True}}\ \mathit{iffI}$$

  Note that 0 assumptions were discharged.

  Isabelle rule `P  ==> P=True`.

15.3 Universal Quantification

$$\forall P = (P = (\lambda x.\, True))$$

- Rule *allI*

$$\frac{P(x)}{\forall P} \;\; \textit{allI}$$

15.3 Universal Quantification

$\forall P = (P = (\lambda x. \mathit{True}))$

- Rule *allI*

$$P(x)$$

$$\frac{}{P = \lambda x.\ \mathit{True}}\ \mathit{allI}$$

15.3 Universal Quantification

$\forall P = (P = (\lambda x. \mathit{True}))$

- Rule *allI* and ND derivation

$$\cfrac{\cfrac{P(x)}{P(x) = \mathit{True}} \ \textit{eqTrueI}}{P = \lambda x. \ \mathit{True}} \ \textit{ext}$$

Inherits the side condition of *ext*: x must not occur freely in the derivation of $P(x)$.

Isabelle rule `(!!x.  P(x)) ==> !x.  P(x).`

Example Illustrating Side Condition

$$\dfrac{\dfrac{[r(x)]^1}{r(x) \to r(x)} \to\text{-}I^1}{\forall x.\, r(x) \to r(x)} \text{ allI}$$

Why is this correct?

Example Illustrating Side Condition

$$\frac{\dfrac{[r(x)]^1}{r(x) \rightarrow r(x)} \rightarrow\text{-}I^1}{\forall x.\, r(x) \rightarrow r(x)} \; \textit{allI}$$

Why is this correct? Let's do it without using *allI* explicitly:

$$\frac{\dfrac{\dfrac{[r(x)]^2}{r(x) \rightarrow r(x)} \rightarrow\text{-}I^2}{(r(x) \rightarrow r(x)) = \textit{True}} \; \textit{eqTrueI}}{\lambda x.\,(r(x) \rightarrow r(x)) = \lambda x.\, \textit{True}} \; \textit{ext}$$

The side condition is respected.

Universal Quantification (Cont.)

- Rule *spec* (recall $\forall P$ means $\forall x. Px$)

$$\forall P$$

$$\frac{}{P(t)} \; spec$$

Universal Quantification (Cont.)

- Rule *spec* (recall $\forall P$ means $\forall x. Px$)

$$P = \lambda x.\,True$$

$$\frac{}{P(t)}\ \textit{spec}$$

Universal Quantification (Cont.)

- Rule *spec* (recall $\forall P$ means $\forall x.Px$) and ND derivation

$$\cfrac{\cfrac{P = \lambda x.\,True}{P(t) = True}\; \textit{fun\_cong}}{P(t)}\; \textit{eqTrueE}$$

Isabelle rule !x::'a.   P(x) ==> P(x).

Note: Need universal quantification to reason about $False$ (since $False = (\forall P.P)$).

15.4 *False*

$$False = (\forall P.P) \qquad (= \forall(\lambda P.P))$$

- FalseI:

15.4 *False*

$$False = (\forall P.P) \qquad \left(= \forall(\lambda P.P) \right)$$

- FalseI: No rule!

- Rule *FalseE*

$$\frac{False}{P} \; FalseE$$

15.4 *False*

$$False = (\forall P.P) \qquad \left(= \forall(\lambda P.P)\right)$$

- FalseI: No rule!

- Rule *FalseE*

$$\frac{\forall P.\, P}{P} \; FalseE$$

15.4 *False*

$$False = (\forall P.P) \qquad \left(= \forall(\lambda P.P) \right)$$

- FalseI: No rule!

- Rule *FalseE* and ND derivation

$$\frac{\forall P.\, P}{P}\ \textit{spec}$$

Isabelle rule `False ==> P`.

False **(Cont.)**

- Rule *False_neq_True*

$$False = True$$

$$\frac{\quad\quad}{P} \quad \textit{False\_neq\_True}$$

False **(Cont.)**

- Rule *False_neq_True* and ND derivation

$$\frac{\dfrac{False = True}{False} \; eqTrueE}{P} \; FalseE$$

  Isabelle rule `False=True ==> P`.

- Similar:

$$\frac{True = False}{P} \; \textit{True\_neq\_False}$$

15.5 Negation

$\neg P = P \rightarrow False$

- Rule *notI*

$$\frac{\begin{array}{c} [P] \\ \vdots \\ False \end{array}}{\neg P} \; notI$$

$\neg P = P \rightarrow False$

15.5 Negation

$\neg P = P \rightarrow False$

- Rule *notI*

$$\frac{\begin{array}{c} [P] \\ \vdots \\ False \end{array}}{P \rightarrow False} \; notI$$

15.5 Negation

$\neg P = P \rightarrow False$

- Rule *notI* and ND derivation

$$\frac{\begin{array}{c}[P]\\ \vdots \\ False\end{array}}{P \rightarrow False}\ impI$$

Isabelle rule (P ==> False) ==> ~P.

Negation (2)

- Rule *notE*

$$\frac{\neg P \qquad\qquad P}{R}\ \textit{notE}$$

Negation (2)

- Rule *notE*

$$\frac{P \rightarrow False \quad P}{R} \; notE$$

Negation (2)

- Rule *notE* and ND derivation

$$\cfrac{\cfrac{P \rightarrow False \quad P}{False} \; mp}{R} \; FalseE$$

Isabelle rule [| ∼P; P |] ==> R.

Negation (3)

- Rule *True_Not_False*

$$\overline{\neg(True = False)} \quad \textit{True\_Not\_False}$$

Negation (3)

- Rule *True_Not_False*

$$\frac{}{(True = False) \rightarrow False} \; \textit{True\_Not\_False}$$

Negation (3)

- Rule *True_Not_False* and ND derivation

$$\cfrac{\cfrac{[True = False]^1}{False} \;\; \textit{True\_neq\_False}}{(True = False) \rightarrow False} \;\; \textit{notI}^1$$

Isabelle rule True $\sim$= False.

15.6 Existential Quantification

$\exists P = P(\epsilon x.P(x))$

- Rule *existsI*

$$\frac{P(x)}{\exists P} \ \textit{existsI}$$

15.6 Existential Quantification

$\exists P = P(\epsilon x.P(x))$

- Rule *existsI*

$$\frac{P(x)}{P(\epsilon x.P(x))} \; \textit{existsI}$$

15.6 Existential Quantification

$\exists P = P(\epsilon x.P(x))$

- Rule *existsI* and ND derivation

$$\frac{P(x)}{P(\epsilon x.P(x))} \; \textit{selectI}$$

Isabelle rule `P(x) ==> ?x::'a.P(x)`.

Existential Quantification (Cont.)

- Rule *existsE*

$$\begin{array}{c} P(x) \\ \vdots \\ Q \end{array}$$

$$\frac{\exists P}{Q} \quad \textit{existsE}$$

Inherits side condition from *allI* (just like in FOL).

Isabelle rule `[| ?x.P(x); !!x.P(x) ==> Q |] ==> Q.`

Existential Quantification (Cont.)

- Rule *existsE*

$$P(x)$$
$$\vdots$$
$$Q$$

$$\frac{P(\epsilon x.P(x))}{Q} \quad \textit{existsE}$$

Inherits side condition from *allI* (just like in FOL).

Isabelle rule `[| ?x.P(x); !!x.P(x) ==> Q |] ==> Q`.

Existential Quantification (Cont.)

- Rule *existsE* and ND derivation

$$\cfrac{P(\epsilon x.P(x)) \quad \cfrac{\cfrac{\cfrac{\cfrac{[P(x)]^1}{\vdots}}{Q}}{P(x) \to Q} \; impl^1}{\cfrac{\forall x.(P(x) \to Q)}{P(\epsilon x.P(x)) \to Q} \; spec} \; alll}{Q} \; mp$$

Inherits side condition from *alll* (just like in FOL).

Isabelle rule `[| ?x.P(x); !!x.P(x) ==> Q |] ==> Q`.

15.7 Conjunction

$$P \wedge Q = \forall R.(P \to Q \to R) \to R$$

- Rule *conjI*

$$
\cfrac{\begin{array}{c} P \\ Q \end{array}}{P \wedge Q} \; \textit{conjI}
$$

15.7 Conjunction

$$P \wedge Q = \forall R.(P \rightarrow Q \rightarrow R) \rightarrow R$$

- Rule *conjI*

$$P$$

$$Q$$

$$\frac{}{\forall R.(P \rightarrow Q \rightarrow R) \rightarrow R} \; \textit{conjI}$$

15.7 Conjunction

$$P \wedge Q = \forall R.(P \to Q \to R) \to R$$

- Rule *conjI* and ND derivation

$$
\cfrac{
  \cfrac{
    \cfrac{[P \to Q \to R]^1 \quad P}{Q \to R} \; mp \qquad Q
  }{R} \; mp
}{
  \cfrac{(P \to Q \to R) \to R}{\forall R.(P \to Q \to R) \to R} \; allI
} \; impl^1
$$

Isabelle rule [| P; Q |] ==> P & Q.

Conjunction (Cont.)

- Rule *conjEL*

$$\frac{P \wedge Q}{P} \text{ conjEL}$$

Conjunction (Cont.)

- Rule *conjEL*

$$\forall R.(P \rightarrow Q \rightarrow R) \rightarrow R$$

$$\frac{}{P} \quad conjEL$$

Conjunction (Cont.)

- Rule *conjEL* and ND derivation

$$\cfrac{\cfrac{\forall R.(P \rightarrow Q \rightarrow R) \rightarrow R}{(P \rightarrow Q \rightarrow P) \rightarrow P} \; spec \qquad \cfrac{\cfrac{\cfrac{[P]^1}{Q \rightarrow P} \; impl}{P \rightarrow Q \rightarrow P} \; impl^1}{}}{P} \; mp$$

Isabelle rule `P & Q ==> P`.

Conjunction (Cont.)

- $P \wedge Q \implies Q$ (*conjER*)

- $[P \wedge Q;\ [P;Q] \implies R] \implies R$ (*conjE*) (rule analogous to *disjE*)

15.8 Disjunction

$$P \vee Q = \forall R.(P \to R) \to (Q \to R) \to R$$

- Rule *disjIL*

$$\frac{P}{P \vee Q} \; \textit{disjIL}$$

15.8 Disjunction

$$P \lor Q = \forall R.(P \to R) \to (Q \to R) \to R$$

- Rule *disjIL*

$$P$$

$$\frac{}{\forall R.(P \to R) \to (Q \to R) \to R} \; \textit{disjIL}$$

15.8 Disjunction

$$P \vee Q = \forall R.(P \to R) \to (Q \to R) \to R$$

- Rule *disjIL* and ND derivation

$$
\cfrac{
  \cfrac{
    \cfrac{
      \cfrac{[P \to R]^1 \quad P}{R} \; \textit{mp}
    }{(Q \to R) \to R} \; \textit{impl}
  }{(P \to R) \to (Q \to R) \to R} \; \textit{impl}^1
}{\forall R.(P \to R) \to (Q \to R) \to R} \; \textit{allI}
$$

Isabelle rule `P ==> P|Q`.

Disjunction (Cont.)

- $Q \implies P \vee Q$ (*disjIR*) similar
- Rule *disjE*

$$
\begin{array}{cccc}
& & [P] & [Q] \\
& & \vdots & \vdots \\
P \vee Q & & R & R \\
\hline
& R & & \text{\emph{disjE}}
\end{array}
$$

Disjunction (Cont.)

- $Q \implies P \vee Q$ (*disjIR*) similar
- Rule *disjE*

$$\forall R.(P \to R) \to (Q \to R) \to R$$

$$[P] \atop {\vdots \atop R}$$

$$[Q] \atop {\vdots \atop R}$$

$$\frac{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}{R} \; \textit{disjE}$$

Disjunction (Cont.)

- $Q \implies P \vee Q$ (*disjIR*) similar

- Rule *disjE* and ND derivation

$$
\cfrac{
  \cfrac{
    \cfrac{\forall R.(P \to R) \to (Q \to R) \to R}{(P \to R) \to (Q \to R) \to R} \; spec
    \qquad
    \cfrac{
      \cfrac{\genfrac{}{}{0pt}{}{[P]}{\;\vdots\;}{R}}{P \to R} \; impl
    }{} \; mp
  }{(Q \to R) \to R}
  \qquad
  \cfrac{\genfrac{}{}{0pt}{}{[Q]}{\;\vdots\;}{R}}{Q \to R} \; impl
}{R} \; mp
$$

Isabelle rule `[| P | Q; P ==> R; Q ==> R |] ==> R.`

411

Disjunction (Cont.)

- $Q \implies P \vee Q$ (*disjIR*) similar

- Rule *disjE* and ND derivation

$$
\cfrac{\cfrac{\forall R.(P \to R) \to (Q \to R) \to R}{(P \to R) \to (Q \to R) \to R}\ spec \quad \cfrac{\cfrac{\overset{[P]}{\vdots}}{R}}{P \to R}\ impl}{\cfrac{(Q \to R) \to R}{\qquad\qquad} }\ mp \quad \cfrac{\cfrac{\overset{[Q]}{\vdots}}{R}}{Q \to R}\ impl
$$

$$
\cfrac{\qquad\qquad\qquad\qquad\qquad\qquad}{R}\ mp
$$

  Isabelle rule `[| P | Q; P ==> R; Q ==> R |] ==>`
  `R`.

- $P \vee \neg P$ (*excluded middle*). Follows using *tof*.

15.9 Miscellaneous Definitions

See HOL.thy!

Typical example (if-then-else):

$$If = \lambda\phi^{bool}xy.\epsilon z. \ (\phi = True \rightarrow z = x)$$
$$\wedge \ (\phi = False \rightarrow z = y)$$

The way rules are derived should now be clear. E.g.,

$$\frac{P = True}{(If \ P \ x \ y) = x} \qquad \frac{P = False}{(If \ P \ x \ y) = y}$$

15.10 Summary on Deriving Rules

HOL is very powerful in terms of what we can represent/derive:

- All well-known inference rules can be derived.

- Other "logical" syntax (e.g. if-then-else) can be defined.

- Rich theories can be obtained by a method we see next lecture.

15.11 Mathematics and Software Engineering in HOL

In coming weeks, we will see how Isabelle/HOL can be used as foundation for mathematics and software engineering.
Outline:

- The central method for making HOL scale up: conservative extensions (< 1 week)

- How the different parts of mathematics are encoded in the Isabelle/HOL library (several weeks)

- How software systems are embedded in Isabelle/HOL (several weeks)

Outlook on Mathematics

After some historical background, we will look at how central parts of mathematics are encoded as Isabelle/HOL theories:

- Orders

- Sets

- Functions

- (Least) fixpoints and induction

- (Well-founded) recursion

- Arithmetic

- Datatypes

Outlook on Software Engineering

Some weeks from now, we will look at case studies of how HOL can be applied in software engineering, i.e. how software systems can be embedded in Isabelle/HOL:

- Foundations, functional languages and denotational semantics

- Imperative languages, Hoare logic

- Z^{395} and data-refinement, CSP and process-refinement

- Object-oriented languages (Java-Light ...)

Of the last three items, we want to treat only one in depth, depending on the audience's preferences.

---

[395]Z and CSP are specification languages. CSP stands for communicating sequential processes.

Conservative Extensions: Motivation

But first, conservative extensions.
   Stage of our course before studying HOL:

- fairly small theories,

- "intuitive" models, (e.g. naïve set theory),

- but inconsistent (due to foundational problems).

   How can we use HOL to

- reason about a reasonably large part of mathematics and software engineering;

- prevent inconsistencies?

What Is Needed for Scaling up?

Several well-known structuring mechanisms are known.

Modularization: Isabelle supports (class) polymorphism and theories.

What Is Needed for Scaling up?

Several well-known structuring mechanisms are known.

   Modularization: Isabelle supports (class) polymorphism and theories.

   Reuse: Isabelle supports libraries and retrieval utilities.

What Is Needed for Scaling up?

Several well-known structuring mechanisms are known.

Modularization: Isabelle supports (class) polymorphism and theories.

Reuse: Isabelle supports libraries and retrieval utilities.

Safe, well-understood integration mechanisms: Isabelle supports conservative theory extensions.

What Is Needed for Scaling up?

Several well-known structuring mechanisms are known.

Modularization: Isabelle supports (class) polymorphism and theories.

Reuse: Isabelle supports libraries and retrieval utilities.

Safe, well-understood integration mechanisms: Isabelle supports conservative theory extensions.

Topic of next lecture.

16 Conservative Theory Extensions

Outline

In the previous lecture, we have derived all well-known inference rules. There is now the need to scale up. Today we look at conservative theory extensions, an important method for this purpose.

16.1 Conservative Theory Extensions: Basics

Some definitions [GM93, Hué]

Definition (theory): *A (syntactic) theory T is a triple $(\mathcal{B}, \Sigma, A)$, where $\mathcal{B}$ is a type signature, Σ a signature and A a set of axioms[396].*

Definition (theory extension): *A theory $T' = (\mathcal{B}', \Sigma', A')$ is an extension of a theory $T = (\mathcal{B}, \Sigma, A)$ iff $\mathcal{B} \subseteq \mathcal{B}'$ and*

---

[396]Isabelle-technically, those axioms are `thm`'s, and they may include Isabelle's metalevel implication $\Longrightarrow$. For this reason, it is not required to mention rules explicitly.

But speaking more generally about HOL, not just its Isabelle implementation, one should better say "rules" here, i.e., objects with a horizontal line and zero or more formulas above the line and one formula below the line.

Outline

In the previous lecture, we have derived all well-known inference rules. There is now the need to scale up. Today we look at conservative theory extensions, an important method for this purpose.

In the weeks to come, we will look at how mathematics is encoded in the Isabelle/HOL library.

16.1 Conservative Theory Extensions: Basics

Some definitions [GM93, Hué]

Definition (theory): *A (syntactic) theory T is a triple $(\mathcal{B}, \Sigma, A)$, where $\mathcal{B}$ is a type signature, Σ a signature and A a set of axioms*[396].

Definition (theory extension): *A theory $T' = (\mathcal{B}', \Sigma', A')$ is an extension of a theory $T = (\mathcal{B}, \Sigma, A)$ iff $\mathcal{B} \subseteq \mathcal{B}'$ and*

---

[396]Isabelle-technically, those axioms are `thm`'s, and they may include Isabelle's metalevel implication $\Longrightarrow$. For this reason, it is not required to mention rules explicitly.

But speaking more generally about HOL, not just its Isabelle implementation, one should better say "rules" here, i.e., objects with a horizontal line and zero or more formulas above the line and one formula below the line.

$\Sigma \subseteq \Sigma'$ *and* $A \subseteq A'$.

Definitions (Cont.)

Definition (conservative extension): *A theory extension* $T' = (\mathcal{B}', \Sigma', A')$ *of a theory* $T = (\mathcal{B}, \Sigma, A)$ *is* conservative *iff for the set of derivable formulas*[397] Th *we have*

$$Th(T) = Th(T') \mid_{\Sigma},$$

where $\mid_{\Sigma}$ *filters away all formulas not belonging to* Σ.

Counterexample:

$$\frac{\rule{4cm}{0.4pt}}{\forall f^{\alpha \to \alpha}.\ Y\ f = f\ (Y\ f)}\ \text{fix}_{398}$$

---

[397]The derivable formulas are terms of type $bool$ derivable using the inference rules of HOL. We write $Th(T)$ for the derivable formulas of a theory T.

[398]Given a function $f : \alpha \to \alpha$, a fixpoint of f is a term t such that $f\ t = t$. Now Y is supposed to be a fixpoint combinator, i.e., for any function f, the term $Y\ f$ should be a fixpoint of f. This is what the rule

$$\frac{\rule{4cm}{0.4pt}}{\forall f^{\alpha \to \alpha}.Y\ f = f\ (Y\ f)}\ \text{fix}$$

says. Consider the example $f \equiv \neg$. Then the axiom allows us to infer $Y(\neg) = \neg(Y(\neg))$, and it is easy to derive $False$ from this. This axiom is a standard example of a non-conservative extension of a theory.

It is not surprising that this goes wrong: Not every function has a fixpoint, so there cannot be a combinator returning a fixpoint of any function.

Nevertheless, fixpoints are important and must be realized in some way, as we will see later.

Consistency Preserved

Definition (consistency): *If T' is a conservative extension of T, then*

$$False \notin Th(T) \Rightarrow False \notin Th(T').$$

Syntactic Schemata for Conservative Extensions

- Constant definition

- Type definition

- Constant specification

- Type specification

Will look at first two schemata now.
For the other two see [GM93].

16.2 Constant Definition

Definition (constant definition): *A theory extension* $T' = (\mathcal{B}', \Sigma', A')$ *of a theory* $T = (\mathcal{B}, \Sigma, A)$ *is a constant definition, iff*

- $\mathcal{B}' = \mathcal{B}$ *and* $\Sigma' = \Sigma \cup \{c : \tau\}$, *where* $c \notin dom^{399}(\Sigma)$;

- $A' = A \cup \{c = E\}$;

- E *does not contain* c *and is closed*[400];

- *no subterm of* E *has a type containing a type variable that is not contained in the type of* c.

---

[399]The **domain** of Σ, denoted $dom(\Sigma)$, is $\{c \mid c : A \in \Sigma$ for some $A\}$.

Likewise, the **domain** of Γ, denoted $dom(\Gamma)$, is $\{x \mid x : A \in \Gamma$ for some $A\}$.

Note the abuse of notation.

[400]A term is **closed** or **ground** if it does not contain any **free** variables.

Constant Definitions Are Conservative

Definition (constant definitions): *Constant definitions are conservative.*

Proof Sketch:

- $Th(T) \subseteq Th(T') \mid_\Sigma$: trivial.

- $Th(T) \supseteq Th(T') \mid_\Sigma$: let π' be a proof for $\phi \in Th(T') \mid_\Sigma$. We unfold any subterm in π' that contains c via $c = E$ into π. Then π must be a proof in T, implying $\phi \in Th(T)$.

Side Conditions

Where are those side conditions[401] needed? What goes wrong?

---

[401]By side conditions we mean

- E does not contain c and is closed;

- no subterm of E has a type containing a type variable that is not contained in the type of c;

in the definition.

The second condition also has a name: one says that the definition must be type-closed.

The notion of having a type is defined by the type assignment calculus. Since E is required to be closed, all variables occurring in E must be λ-bound, and so the type of those variables is given by the type superscripts.

[402]$True$ is defined as $\lambda x^{bool}.x = \lambda x.x$ and not $\lambda x^{\alpha}.x = \lambda x.x$. The definition must be type-closed.

Side Conditions

Where are those side conditions[401] needed? What goes wrong?

Very simple example: Let $E \equiv \exists x^{\alpha} y^{\alpha}. \; x \neq y$ and suppose σ is a type inhabited by only one term, and τ is a type inhabited by at least two terms. Then we would have:

$$
\begin{aligned}
c &= c & \text{holds by } \textit{refl} \\
\implies & (\exists x^{\sigma} y^{\sigma}. \, x \neq y) = (\exists x^{\tau} y^{\tau}. \, x \neq y) \\
\implies & \textit{False} = \textit{True} \\
\implies & \textit{False}
\end{aligned}
$$

This explains definition of $True^{402}$. Other (standard) example later.

---

[401] By side conditions we mean

- E does not contain c and is closed;

- no subterm of E has a type containing a type variable that is not contained in the type of c;

in the definition.

The second condition also has a name: one says that the definition must be type-closed.

The notion of having a type is defined by the type assignment calculus. Since E is required to be closed, all variables occurring in E must be λ-bound, and so the type of those variables is given by the type superscripts.

[402] $True$ is defined as $\lambda x^{bool}.x = \lambda x.x$ and not $\lambda x^{\alpha}.x = \lambda x.x$. The definition must be type-closed.

Constant Definition: Examples

Definitions of *True*, *False*, ∧, ∨, ∀ …
   Here the original Isabelle syntax (Ex_def changed)

```
True_def:  "True   == ((%x::bool. x) = (%x. x))"
All_def:   "All(P) == (P = (%x. True))"
Ex_def:    "Ex(P)  == P (SOME x. P x)"
False_def: "False  == (!P. P)"
not_def:   "~ P    == P-->False"
and_def:   "P & Q  == !R. (P-->Q-->R) --> R"
or_def:    "P | Q  == !R. (P-->R) --> (Q-->R)
                                     --> R"
```

More Constant Definitions in Isabelle

Function application (`Let`), if-then-else, unique existence[403]:

```
consts
 Let :: ['a, 'a => 'b] => 'b
 If  :: [bool, 'a, 'a] => 'a
defs
 Let_def "Let s f == f(s)"
 if_def  "If P x y == @z::'a.(P=True-->z=x) &
                            (P=False-->z=y)"
 Ex1_def "Ex1(P) == ?x. P(x) & (!y. P(y) --> y=x)"
```

Recall: `=>` is function type arrow; also recall `[]` syntax.

---

[403]We have never used unique existential quantification ($\exists!$) before. $\exists! x_1, \ldots, x_n . \phi(x_1, \ldots, x_n)$ is defined as $\exists x_1, \ldots, x_n . \phi(x_1, \ldots, x_n) \wedge (\forall y_1, \ldots, y_n . \phi(y_1, \ldots, y_n) \rightarrow x_1 = y_1 \wedge \ldots \wedge x_n = y_n)$.

Note that in general $\exists! x.(\exists! y.\phi)$ is not the same as $\exists! xy.\phi)$.

16.3 Type Definitions

Type definitions, explained intuitively: we have

- an existing type ρ;

---

[404]Although a set is formally a different object than a predicate, it is standard to interpret a predicate a set: the set of terms for which the predicate returns true. We have done this before.

16.3 Type Definitions

Type definitions, explained intuitively: we have

- an existing type ρ;

- a predicate $S : \rho \to bool$, defining a non-empty "subset"[404] of ρ;

---

[404]Although a set is formally a different object than a predicate, it is standard to interpret a predicate a set: the set of terms for which the predicate returns true. We have done this before.

16.3 Type Definitions

Type definitions, explained intuitively: we have

- an existing type ρ;

- a predicate $S : \rho \rightarrow bool$, defining a non-empty "subset" [404] of ρ;

- axioms stating an isomorphism between S and the new type τ.

---

[404] Although a set is formally a different object than a predicate, it is standard to interpret a predicate a set: the set of terms for which the predicate returns true. We have done this before.
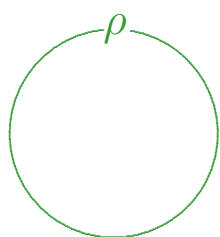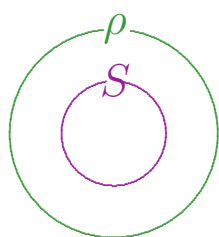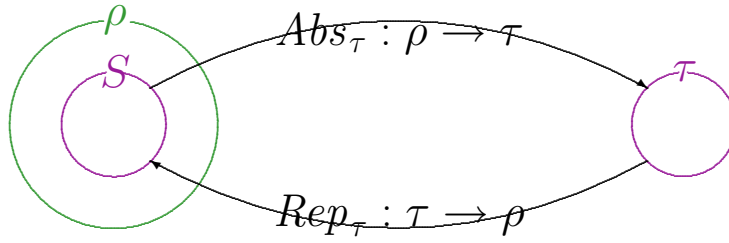
Type Definition: Definition

Definition (type definition): *Assume a theory $T = (\mathcal{B}, \Sigma, A)$ and a type ρ and a term S[405] such that $\Sigma \vdash S : \rho \to bool$.*

A theory extension $T' = (\mathcal{B}', \Sigma', A')$ of T is a type definition for type τ[406] (where τ fresh[407]), iff

---

[405]Here, S is any "predicate", i.e., term of type $\rho \to bool$, not necessarily a constant.

[406]A type definition is supposed to define a type constructor (where the arity and fixity are indicated in some way). We abuse notation here: we use τ to denote a type constructor, but also the type obtained by applying the type constructor to a vector of different type variables (as many as the type constructor requires).

So think of τ as either being a type constructor or a "generic" type (just a type constructor being applied to type variables).

We do the same in examples.

[407]The type constructor τ must not occur in $\mathcal{B}$.

$$
\begin{aligned}
\mathcal{B}' \;&=\; \mathcal{B} \;\uplus^{408}\; \{\tau\}, \\
\Sigma' \;&=\; \Sigma \;\cup\; \{Abs_\tau{}^{409} : \rho \to \tau,\, Rep_\tau : \tau \to \rho\} \\
A' \;&=\; A \;\cup\; \{\forall x. Abs_\tau(Rep_\tau\, x) = x^{410}, \\
&\qquad\qquad\quad \forall x. S\, x \to Rep_\tau(Abs_\tau\, x) = x\}
\end{aligned}
$$

Proof obligation[411] $\exists x.\, S\, x$ can be proven inside HOL!

---

[408]The symbol $\uplus$ denotes disjoint union, so the expression $A \uplus B$ is well-formed only when A and B have no elements in common. One thus uses this notation to indicate this fact.

[409]Of course we are giving a schematic definition here, so any letters we use are metanotation.

Notice that Abs_τ and Rep_τ stand for new constants. For any new type τ to be defined, two such constants must be added to the signature to provide a generic way of obtaining terms of the new type. Since the new type is isomorphic to the "subset" S, whose members are of type ρ, one can say that Abs_τ and Rep_τ provide a type conversion between (the subset S of) ρ and τ.

So we have a new type τ, and we can obtain members of the new type by applying Abs_τ to a term t of type ρ for which $S\, t$ holds.

[410]The formulas

$$
\forall x. Abs_\tau(Rep_\tau\, x) = x
$$
$$
\forall x. S\, x \to Rep_\tau(Abs_\tau\, x) = x
$$

Type Definitions Are Conservative

Definition (type definitions): *Type definitions are conservative.*

Proof see [GM93, pp.230].

state that the "set" S and the new type τ are isomorphic. Note that Abs_τ should not be applied to a term not in "set" S. Therefore we have the premise $S\,x$ in the above equation.

Note also that S could be the "trivial filter" $\lambda x.\,True$. In this case, Abs_τ and Rep_τ would provide an isomorphism between the entire type ρ and the new type τ.

[411]We have said previously that S should be a non-empty "subset" of τ. Therefore it must be proven that $\exists x.\,S\,x$. This is related to the semantics.

Whenever a type definition is introduced in Isabelle, the proof obligation must be shown inside Isabelle/HOL. Isabelle provides the `typedef` syntax for type definitions, as we will see later. Using this syntax, the "author" of a type definition can either explicitly provide a proof (see `Product_Type.thy`), or the proof is so easy that Isabelle can do it automatically (see `Sum_Type.thy`).

HOL Is Rich Enough!

This may seem fishy: if a new type is always isomorphic to a subset of an existing type, how is this construction going to lead to a "rich" collection of types for large-scale applications?

HOL Is Rich Enough!

This may seem fishy: if a new type is always isomorphic to a subset of an existing type, how is this construction going to lead to a "rich" collection of types for large-scale applications?

But in fact, due to ind and $\rightarrow$, the types in HOL are already very rich.

HOL Is Rich Enough!

This may seem fishy: if a new type is always isomorphic to a subset of an existing type, how is this construction going to lead to a "rich" collection of types for large-scale applications?

But in fact, due to ind and $\rightarrow$, the types in HOL are already very rich.

We now give three examples to convince you.

Example: Typed Sets

General scheme,

$$
\begin{aligned}
\mathcal{B}' &= \mathcal{B} \uplus \{\tau\}, \\
\Sigma' &= \Sigma \cup \{Abs_\tau : \rho \rightarrow \tau, \\
&\qquad\quad Rep_\tau : \tau \rightarrow \rho\} \\
A' &= A \cup \{\forall x.Abs_\tau \,(Rep_\tau \, x) = x, \\
&\qquad\quad \forall x.S\,x \rightarrow Rep_\tau \,(Abs_\tau \, x) = x\}
\end{aligned}
$$

---

[412]We have $S \equiv \lambda x^{\alpha \rightarrow bool}.True$, and so in $(\exists x.Sx)$, the variable x has type $\alpha \rightarrow bool$. The proposition $(\exists x.Sx)$ is true since the type $\alpha \rightarrow bool$ is inhabited, e.g. by the term $\lambda x^\alpha.True$ or $\lambda x^\alpha.False$.

Beware of a confusion: This does not mean that the new type $\alpha\,set$, defined by this construction, is the type of non-empty sets. There is a term for the empty set: The empty set is the term $Abs_{set}\,(\lambda x.False)$.

Recall a previous argument for the importance of inhabitation.

Example: Typed Sets

General scheme, substituting $\rho \equiv \alpha \to bool$ (α is any type variable),

$$
\begin{aligned}
\mathcal{B}' &= \mathcal{B} \uplus \{\tau\}, \\
\Sigma' &= \Sigma \cup \{Abs_\tau : (\alpha \to bool) \to \tau, \\
&\qquad\qquad Rep_\tau : \tau \to (\alpha \to bool)\} \\
A' &= A \cup \{\forall x.Abs_\tau (Rep_\tau x) = x, \\
&\qquad\qquad \forall x.S\,x \to Rep_\tau (Abs_\tau x) = x\}
\end{aligned}
$$

---

[412]We have $S \equiv \lambda x^{\alpha \to bool}.True$, and so in $(\exists x.Sx)$, the variable x has type $\alpha \to bool$. The proposition $(\exists x.Sx)$ is true since the type $\alpha \to bool$ is inhabited, e.g. by the term $\lambda x^\alpha.True$ or $\lambda x^\alpha.False$.

Beware of a confusion: This does not mean that the new type $\alpha\,set$, defined by this construction, is the type of non-empty sets. There is a term for the empty set: The empty set is the term $Abs_{set}(\lambda x.False)$.

Recall a previous argument for the importance of inhabitation.

Example: Typed Sets

General scheme, substituting $\rho \equiv \alpha \to bool$ (α is any type variable), $\tau \equiv \alpha\ set$ (or set),

$$\mathcal{B}' \;=\; \mathcal{B} \;\uplus\; \{set\},$$
$$\Sigma' \;=\; \Sigma \;\cup\; \{Abs_{set} : (\alpha \to bool) \to \alpha\ set,$$
$$Rep_{set} : \alpha\ set \to (\alpha \to bool)\}$$
$$A' \;=\; A \;\cup\; \{\forall x.Abs_{set}(Rep_{set}\,x) = x,$$
$$\forall x.S\,x \;\to\; Rep_{set}(Abs_{set}\,x) = x\}$$

---

[412]We have $S \equiv \lambda x^{\alpha \to bool}.True$, and so in $(\exists x.Sx)$, the variable x has type $\alpha \to bool$. The proposition $(\exists x.Sx)$ is true since the type $\alpha \to bool$ is inhabited, e.g. by the term $\lambda x^{\alpha}.True$ or $\lambda x^{\alpha}.False$.

Beware of a confusion: This does not mean that the new type $\alpha\ set$, defined by this construction, is the type of non-empty sets. There is a term for the empty set: The empty set is the term $Abs_{set}\,(\lambda x.False)$.

Recall a previous argument for the importance of inhabitation.

Example: Typed Sets

General scheme, substituting $\rho \equiv \alpha \to bool$ (α is any type variable), $\tau \equiv \alpha \, set$ (or set), $S \equiv \lambda x^{\alpha \to bool}.True$

$$
\begin{aligned}
\mathcal{B}' &= \mathcal{B} \uplus \{set\}, \\
\Sigma' &= \Sigma \cup \{Abs_{set} : (\alpha \to bool) \to \alpha \, set, \\
&\qquad\qquad Rep_{set} : \alpha \, set \to (\alpha \to bool)\} \\
A' &= A \cup \{\forall x.Abs_{set}(Rep_{set} \, x) = x, \\
&\qquad\qquad \forall x.True \to Rep_{set}(Abs_{set} \, x) = x\}
\end{aligned}
$$

---

[412]We have $S \equiv \lambda x^{\alpha \to bool}.True$, and so in $(\exists x.Sx)$, the variable x has type $\alpha \to bool$. The proposition $(\exists x.Sx)$ is true since the type $\alpha \to bool$ is inhabited, e.g. by the term $\lambda x^{\alpha}.True$ or $\lambda x^{\alpha}.False$.

Beware of a confusion: This does not mean that the new type $\alpha \, set$, defined by this construction, is the type of non-empty sets. There is a term for the empty set: The empty set is the term $Abs_{set} \, (\lambda x.False)$.

Recall a previous argument for the importance of inhabitation.

Example: Typed Sets

General scheme, substituting $\rho \equiv \alpha \to bool$ (α is any type variable), $\tau \equiv \alpha\ set$ (or set), $S \equiv \lambda x^{\alpha \to bool}.True$

$$
\begin{aligned}
\mathcal{B}' &= \mathcal{B} \uplus \{set\}, \\
\Sigma' &= \Sigma \cup \{Abs_{set} : (\alpha \to bool) \to \alpha\ set, \\
&\qquad\qquad Rep_{set} : \alpha\ set \to (\alpha \to bool)\} \\
A' &= A \cup \{\forall x. Abs_{set}(Rep_{set}\ x) = x, \\
&\qquad\quad \forall x. \qquad\qquad Rep_{set}(Abs_{set}\ x) = x\}
\end{aligned}
$$

Simplification since $S \equiv \lambda x.True$. Proof obligation: $(\exists x.Sx)$ trivial since $(\exists x.True) = True$. Inhabitation[412] crucial!

---

[412]We have $S \equiv \lambda x^{\alpha \to bool}.True$, and so in $(\exists x.Sx)$, the variable x has type $\alpha \to bool$. The proposition $(\exists x.Sx)$ is true since the type $\alpha \to bool$ is inhabited, e.g. by the term $\lambda x^{\alpha}.True$ or $\lambda x^{\alpha}.False$.

Beware of a confusion: This does not mean that the new type $\alpha\ set$, defined by this construction, is the type of non-empty sets. There is a term for the empty set: The empty set is the term $Abs_{set}\ (\lambda x.False)$.

Recall a previous argument for the importance of inhabitation.

Sets: Remarks

Any function $r : \alpha \to bool$ can be interpreted as a set of α; r is called characteristic function. That's what $Abs_{set}\ r$ does; Abs_{set} is a wrapper saying "interpret r as set".

---

[413]We said that in the general formalism for defining a new type, there is a term S of type $\rho \to bool$ that defines a "subset" of a type ρ. In other words, it filters some terms from type ρ. Thus the idea that a predicate can be interpreted as a set is present in the general formalism for defining a new type.

Now we are talking about a particular example, the type $\alpha\ set$. Having the idea "predicates are sets" in mind, one is tempted to think that in the particular example, S will take the role of defining particular sets, i.e., terms of type $\alpha\ set$. This is not the case!

Rather, S is $\lambda x.\,True$ and hence trivial in this example. Moreover, in the example, ρ is $\alpha \to bool$, and any term r of type ρ defines a set whose elements are of type α; $Abs_{set}\ r$ is that set.

Sets: Remarks

Any function $r : \alpha \to bool$ can be interpreted as a set of α; r is called characteristic function. That's what $Abs_{set}\ r$ does; Abs_{set} is a wrapper saying "interpret r as set".

$\quad S \equiv \lambda x.\,True$ and so S is trivial[413] in this case.

---

[413]We said that in the general formalism for defining a new type, there is a term S of type $\rho \to bool$ that defines a "subset" of a type ρ. In other words, it filters some terms from type ρ. Thus the idea that a predicate can be interpreted as a set is present in the general formalism for defining a new type.

Now we are talking about a particular example, the type $\alpha\,set$. Having the idea "predicates are sets" in mind, one is tempted to think that in the particular example, S will take the role of defining particular sets, i.e., terms of type $\alpha\,set$. This is not the case!

Rather, S is $\lambda x.\,True$ and hence trivial in this example. Moreover, in the example, ρ is $\alpha \to bool$, and any term r of type ρ defines a set whose elements are of type α; $Abs_{set}\ r$ is that set.

More Constants for Sets

For convenient use of sets, we define more constants:

$$\{x \mid f\,x\} = Collect^{414}\,f = Abs_{set}\,f$$
$$x \in A = (Rep_{set}\,A)^{415}\,x$$
$$A \cup B = \{x \mid x \in A \lor x \in B\}$$
$$\vdots$$

Consistent set theory[416] adequate for most of mathematics

---

[414]We have seen $Collect$ before in the theory file NSet.thy (naïve set theory).

$Collect\,f$ is the set whose characteristic function is f. There is also a concrete (i.e., according to mathematical practice) syntax $\{x \mid f\,x\}$. It is called set comprehension. The correspondence between the HOAS $Collect\,f$ and the concrete syntax $\{x \mid f\,x\}$ also makes it clear that set comprehension is a binding operator, as we learned some time ago.

Note also that $Collect$ is the same as Abs_{set} here.

The file Set.thy should be contained in your Isabelle distribution. Or, if you only have an Isabelle executable, you can find the sources here:

http://isabelle.in.tum.de/library/

[415]We define

$$x \in A = (Rep_{set}\,A)\,x$$

Since Rep_{set} has type $\alpha\,set \to (\alpha \to bool)$, this means that

and computer science.

In Isabelle/HOL however, sets are a special case[417].

Here, sets are just an example to demonstrate type definitions. Later we study them for their own sake.

---

x is of type α and A is of type $(\alpha \to bool)$. Therefore $\in$ is of type $\alpha \to (\alpha\ set) \to bool$ (but written infix).

In the Isabelle theory file `Set.thy`, you will indeed find that the constant : (Isabelle syntax for $\in$) has type $\alpha \to (\alpha\ set) \to bool$.

However, you will not find anything directly corresponding to Rep_{set}.

[416]Typed set theory is a conservative extension of HOL and hence consistent.

Recall the problems with untyped set theory.

[417]We presented a definition of $\alpha\ set$ according to the scheme of type definitions. However, in Isabelle/HOL (`Set.thy`), it is not done exactly like that. The reason will become clear later.

The type $\alpha\ set$ is defined in Isabelle/HOL in a way which essentially corresponds to the type definition scheme, but is different in the technical details. In particular, there are no constants Abs_{set} and Rep_{set}. Instead, we have $Collect$ and

Example: Pairs

Consider type $\alpha \rightarrow \beta \rightarrow bool$. We can regard a term $f : \alpha \rightarrow \beta \rightarrow bool$ as a representation of the pair (a, b), where $a : \alpha$ and $b : \beta$, iff $f\, x\, y$ is true exactly for $x = a$ and $y = b$. Observe:

---

the $\in$-sign. We will now explain how.

Concerning Abs_{set}, there is no worry, since it corresponds exactly to $Collect$.

Rep_{set} is related to the $\in$-sign via

$$x \in A = (Rep_{set}\ A)\ x$$

Let us see that this setup is equivalent to the scheme of type definitions. There are two axioms in `Set.thy`:

```
axioms
  mem_Collect_eq [iff]: "(a : {x. P(x)}) = P(a)"
  Collect_mem_eq [simp]: "{x. x:A} = A"
```

We translate these axioms using the definitions:

$$a \in \{x \mid P\, x\} = P\, a \rightsquigarrow$$
$$a \in (Collect\, P) = P\, a \rightsquigarrow$$
$$a \in (Abs_{set}\, P) = P\, a \rightsquigarrow$$
$$Rep_{set}(Abs_{set}\, P)\, a = P\, a \rightsquigarrow$$
$$Rep_{set}(Abs_{set}\, P) = P$$

- For given a and b, there is exactly one[418] such f (namely, $\lambda x^\alpha y^\beta.\, x = a \wedge y = b$).

- Some functions of type $\alpha \to \beta \to bool$ represent pairs and others don't (e.g., the function $\lambda xy.\, True$ does not represent a pair). The ones that do are exactly the ones that have the form $\lambda x^\alpha y^\beta.\, x = a \wedge y = b$, for some a and b.

The last step uses extensionality.

Now the second one:

$$\{x \mid x \in A\} = A \rightsquigarrow$$
$$\{x \mid (Rep_{set}\, A)\, x\} = A \rightsquigarrow$$
$$Collect(Rep_{set}\, A) = A$$

Ignoring some universal quantifications (these are implicit in Isabelle), these are the isomorphy axioms for set.

[418]When we say that there is "exactly one" f, this is meant modulo equality in HOL. This means that e.g. $\lambda x^\alpha y^\beta.\, y = b \wedge x = a$ is also such a term since $(\lambda x^\alpha y^\beta.\, x = a \wedge y = b) = (\lambda x^\alpha y^\beta.\, y = b \wedge x = a)$ is derivable in HOL.

Type Definition for Pairs

This gives rise to a type definition where S is non-trivial:

$$\rho \;\equiv\; \alpha \to \beta \to bool$$
$$S \;\equiv\; \lambda f^{\alpha \to \beta \to bool}.\exists ab.f = \lambda x^\alpha y^\beta.x = a \wedge y = b$$
$$\tau \;\equiv\; \alpha \times \beta \qquad\qquad\qquad\qquad (\times \text{ infix})$$

It is convenient to define a constant `Pair_Rep` (not to be confused with $Rep_\times$[419]) as $\lambda a^\alpha b^\beta.\lambda x^\alpha y^\beta.\; x = a \wedge y = b$[420]. Then `Pair_Rep` $a\,b = \lambda x^\alpha y^\beta.\; x = a \wedge y = b$.

---

[419] $Rep_\times$ would be the generic name for one of the two isomorphism-defining functions.

[420] Since $Rep_\times$ looks funny, the definition scheme for type definitions in Isabelle is such that it provides two names for a type, one if the type is used as such, and one for the purpose of generating the names of the isomorphism-defining functions.

[420] We write $\lambda a^\alpha b^\beta.\lambda x^\alpha y^\beta.x = a \wedge y = b$ rather than $\lambda a^\alpha b^\beta x^\alpha y^\beta.x = a \wedge y = b$ to emphasize the idea that one first applies $Pair\_Rep$ to a and b, and the result is a function representing a pair, wich can then be applied to x and y.

Now in Isabelle

Isabelle has a special set-based[421] syntax for type definitions:

```
typedef (T)
  ⟨typevars⟩ "T'" ⟨fixity⟩
  = "{x.ϕ}"
```

---

[421] The syntax $"\{x.\phi\}"$ does not just look like a set comprehension, it is one!

So, since the `typedef` syntax is based on sets, sets themselves could not have been defined using that syntax. This is the reason why in Isabelle/HOL, sets are a special case of a type definition.

See `Typedef.thy`, which should be contained in your Isabelle distribution. Or, if you only have an Isabelle executable, you can find the sources here:

http://isabelle.in.tum.de/library/

Now in Isabelle

Isabelle has a special set-based[421] syntax for type definitions:

$$\texttt{typedef}\ (T)$$
$$\langle typevars \rangle\ ''T'''\ \langle fixity \rangle$$
$$= ''\{x.\phi\}''$$

How is this linked to our scheme:

- the new type is called T';

- ρ is the type of x (inferred);

- S is $\lambda x.\phi$;

- constants $\texttt{Abs}\_T$ and $\texttt{Rep}\_T$ are automatically generated.

---

[421]The syntax $''\{x.\phi\}''$ does not just look like a set comprehension, it is one!

So, since the `typedef` syntax is based on sets, sets themselves could not have been defined using that syntax. This is the reason why in Isabelle/HOL, sets are a special case of a type definition.

See `Typedef.thy`, which should be contained in your Isabelle distribution. Or, if you only have an Isabelle executable, you can find the sources here:

http://isabelle.in.tum.de/library/

Isabelle Syntax for Pair Example

```
constdefs
 Pair_Rep :: ['a, 'b] => ['a, 'b] => bool
 "Pair_Rep == (%a b. %x y. x=a & y=b)"
typedef (Prod)
 ('a, 'b) "*" (infixr 20) =
   "{f.?a b. f=Pair_Rep(a::'a)(b::'b)}"
```

The keyword `constdefs`[422] introduces a constant definition. The definition and use of `Pair_Rep` is for convenience. There are "two names" $*$ and Prod.

See `Product_Type.thy`[423].

---

[422]In Isabelle theory files, `consts` is the keyword preceding a sequence of constant declarations (i.e., this is where the Σ is defined), and `defs` is the keyword preceding the axioms that define these constants (i.e., this is where the A is defined.

`constdefs` combines the two, i.e. it allows for a sequence of both constant declarations and definitions. When the `constdefs` syntax is used to define a constant c, then the identifier c_def is generated automatically. E.g.

```
        constdefs
          id :: "'a => 'a"
        "id == %x. x"
```

will bind `id_def` to $id \equiv \lambda x.x$.

[423]This file should be contained in your Isabelle distribution. Or, if you only have an Isabelle executable, you can find the sources here:

http://isabelle.in.tum.de/library/

Example: Sums

An element of (α, β) sum[424] is either $Inl\ a$ where $a : \alpha$ or $Inr\ b$ where $b : \beta$.

So think of $Inl\ a$ and $Inr\ b$ as syntactic objects that we want to represent.

Consider type $\alpha \rightarrow \beta \rightarrow bool \rightarrow bool$. We can regard $f : \alpha \rightarrow \beta \rightarrow bool \rightarrow bool$ as a

representation of . . .	iff $f\ x\ y\ i$ is true for . . .
$Inl\ a$	$x = a$, y arbitrary, and $i = True$
$Inr\ b$	x arbitrary, $y = b$, and $i = False$.

Similar to pairs.

[424]Idea of sum or union type: t is in the sum of τ and σ if t is either in τ or in σ. To do this formally in our type system, and also in the type system of functional programming languages like ML, t must be wrapped to signal if it is of type τ or of type σ.

For example, in ML one could define

$$\texttt{datatype}\ (\alpha, \beta)\ \texttt{sum} = Inl\ \alpha \mid Inr\ \beta$$

So an element of (α, β) sum is either $Inl\ a$ where $a : \alpha$ or $Inr\ b$ where $b : \beta$.

Isabelle Syntax for Sum Example

```
constdefs
  Inl_Rep :: ['a, 'a, 'b, bool] => bool
 "Inl_Rep == (%a. %x y p. x=a & p)"
  Inr_Rep :: ['b, 'a, 'b, bool] => bool
 "Inr_Rep == (%b. %x y p. y=b & ~p)"
typedef (Sum)
 ('a,'b)"+" =
   "{f. (?a. f = Inl_Rep(a::'a)) |
        (?b. f = Inr_Rep(b::'b))}"
```

See `Sum_Type.thy`[425].

How would you define[426] a type even based on nat?

---

[425]This file should be contained in your Isabelle distribution. Or, if you only have an Isabelle executable, you can find the sources here:

　　http://isabelle.in.tum.de/library/

[426]Suppose we have a type nat and a constant + with the expected meaning. We want to define a type even of even numbers. What is an even number?

Isabelle Syntax for Sum Example

```
constdefs
  Inl_Rep :: ['a, 'a, 'b, bool] => bool
 "Inl_Rep == (%a. %x y p. x=a & p)"
  Inr_Rep :: ['b, 'a, 'b, bool] => bool
 "Inr_Rep == (%b. %x y p. y=b & ~p)"
typedef (Sum)
 ('a,'b)"+" =
   "{f. (?a. f = Inl_Rep(a::'a)) |
        (?b. f = Inr_Rep(b::'b))}"
```

See Sum_Type.thy[425].

How would you define[426] a type even based on nat?

---

[425]This file should be contained in your Isabelle distribution. Or, if you only have an Isabelle executable, you can find the sources here:

> http://isabelle.in.tum.de/library/

[426]Suppose we have a type nat and a constant $+$ with the expected meaning. We want to define a type even of even numbers. What is an even number?

The following choice of S is adequate:

$$S \equiv \lambda x. \exists n. x = n + n$$

Using the Isabelle scheme, this would be

```
        typedef (Even)
           even = "{x. ?y. x=y+y}"
```

We could then go on by defining an operation PLUS on even,

16.4 Summary on Conservative Extensions

We have seen two schemata:

- Constant definition: new constant must be defined using old constants. No recursion! Subtle side conditions concerning types.

- Type definition: new type must be isomorphic to a "subset" S of an existing type ρ. Not possible to define any type that is "structurally" richer than the types one already has. But HOL is rich enough.

---

say as follows:

```
constdefs
  PLUS::[even,even] => even (infixl 56)
  PLUS_def "opPLUS ==
              %xy. Abs_Even (Rep_Even(x)+Rep_Even(x))"
```

Note that we chose to use names even and Even, but we could have used the same name twice as well.

17 Mathematics in the Isabelle/HOL Library: Introduction

Isabelle/HOL at Work

We have seen how the mechanism of conservative extensions works in principle.

For several lectures, we will now look at theories of the Isabelle/HOL library, all built by conservative extensions and modelling significant portions of mathematics.

Sets: The Basis of Principia Mathematica

Sets are ubiquitous in mathematics:

- 17th century: geometry can be reduced to numbers [Des16, vL16].

- 19th century: numbers can be reduced to sets [Can18, Pea18, Fre93, Fre03].

- 20th century: sets can be represented in logics [Zer19, Frä19, WR25, Göd31, Ber19, Chu40].

We call this the Principia Mathematica Structure [WR25].

The libraries of theorem provers follow this Principia Mathematica Structure — in reverse order![427]

---

[427]It is not surprising that the logical built-up of theorem prover is reversed w.r.t. to the historical development of mathematics and logics. Research usually starts from applications and the intuition and works its way back to the foundations.

The Roadmap

- Orders

- Sets

- Functions

- (Least) fixpoints and induction

- (Well-founded) recursion

- Arithmetic

- Datatypes

18 Orders

The Roadmap

We are looking at how the different parts of mathematics are encoded in the Isabelle/HOL library.

- Orders

- Sets

- Functions

- (Least) fixpoints and induction

- (Well-founded) recursion

- Arithmetic

- Datatypes

The Roadmap

We are looking at how the different parts of mathematics are encoded in the Isabelle/HOL library.

- Orders

- Sets

- Functions

- (Least) fixpoints and induction

- (Well-founded) recursion

- Arithmetic

- Datatypes

Three Order Classes

We first define a syntactic class ord. It is the class of types for which symbols $<$ and $<=$ exist.

Three Order Classes

We first define a syntactic class ord. It is the class of types for which symbols $<$ and $<=$ exist.

We then define two axiomatic classes order and linorder for which $<$ and $<=$ are required to have certain properties, that of being a partial order, or a linear order, resp.

Orders (in `HOL.thy`[428])

```
axclass
 ord < type
consts
 "op <"  :: ['a::ord, 'a] => bool
 "op <=" :: ['a::ord, 'a] => bool
constdefs
  min :: "['a::ord, 'a] => 'a"
  "min a b == (if a <= b then a else b)"
  max :: "['a::ord, 'a] => 'a"
  "max a b == (if a <= b then b else a)"
```

Recall `constdefs` syntax and note two uses of $<$[429].

---

[428] In previous versions of Isabelle, there used to be a theory file `Ord.thy`. Nowadays orders are defined in `HOL.thy`.

[429] The line

```
            axclass order < ord
```

in the theory file states that `order` is a subclass of `ord`.

The line

```
 "op <"  :: ['a::ord, 'a] => bool ("(_ < _)" [50, 51] 50)
```

in the theory file declares a constant $<$ with a certain type.

`type` is the class containing all types. In previous versions of Isabelle, it used to be called `term`.

Orders in `HOL.thy` (Cont.)

```
axclass order < ord
 order_refl     "x <= x"
 order_trans    "[|x <= y; y <= z|] ==> x <= z"
 order_antisym "[|x <= y; y <= x|] ==> x = y"
 order_less_le "x < y = (x <= y & x ~= y)"
%
axclass linorder < order
 linorder_linear "x <= y | y <= x"
```

Least Elements

In `Ord.thy`, least elements used to be defined as:

```
Least :: "('a::ord => bool) => 'a"
Least_def "Least P == @x. P(x) &
          (ALL y. P(y) => x <= y)"
```

Now it is done without using the Hilbert operator.

Monotonicity

In Ord.thy, monotonicity used to be defined as:

```
mono      :: ['a::ord => 'b::ord] => bool
mono_def    "mono(f)  ==
            (!A B. A <= B --> f(A) <= f(B))
```

Now it is done using a completely different syntax, but one can still use monotonicity as before.

Some Theorems[430] about Orders

`monoI`	$(\bigwedge AB. A \leq B \Longrightarrow f\,A \leq f\,B)$ $\Longrightarrow mono\,f$
`monoD`	$[\![mono\,f; A \leq B]\!] \Longrightarrow f\,A \leq f\,B$
`order_eq_refl`	$x = y \Longrightarrow x \leq y$
`order_less_irrefl`	$\neg x < x$
`order_le_less`	$(x \leq y) = (x < y \vee x = y)$
`linorder_less_linear`	$x < y \vee x = y \vee y < x$
`linorder_neq_iff`	$(x \neq y) = (x < y \vee y < x)$
`min_same`	$min\,x\,x = x$
`le_min_iff_conj`	$(z \leq min\,x\,y) = (z \leq x \wedge z \leq y)$

18.1 Summary on Orders

Type classes are a structuring mechanism in Isabelle:

---

[430]In the rest of the course, we will mostly be dealing with Isabelle HOL, and so when we speak of a theorem, we ususally mean an Isabelle theorem, i.e., a theorem in Isabelle's meta-logic, what we also call a `thm`. Such theorems may contain the meta-level implication $\Longrightarrow$ and universal quantifier $\bigwedge$.

So they are not theorems within HOL. Logically, this is not a big deal as one switch between object and meta-level by the introduction and elimination rules for $\rightarrow$ and $\forall$. But technically (for the proof procedures), it makes a difference.

To see a theorem displayed in Isabelle, simply type the name of the theorem followed by ";".

- Syntactic classes (e.g. $t :: \alpha :: ord$ as in Haskell [HHPW96]): merely a mechanism to structure visibility of operations.

- Syntactic classes (e.g. $t :: \alpha :: ord$ as in Haskell [HHPW96]): merely a mechanism to structure visibility of operations.

- Axiomatic classes (e.g. $t :: \alpha :: order$): a mechanism for structuring semantic knowledge[431] in types (foundation to be discussed later).

---

[431]The Isabelle type system records for any type variable what class constraints there are for this type variable. These class constraints may arise from the types of the constants used in an expression, or they may be given explicitly by the user in a goal. E.g. one might type

```
Goal "(x::'a::order)<y ==> x<=y";
```

to specify that x must be of a type in the type class order.

The axioms of an axiomatic class can only be applied if any constant declared in the axiomatic class (or a syntactic superclass) is applied to arguments of a type in the axiomatic class. E.g. order_refl can only be used to prove $y <= y$ if the type of y is in the type class order.

In this sense the type information (y is of type in class order) is semantic knowledge ($y <= y$ holds).

19 Sets

The Roadmap

We are still looking at how the different parts of mathematics are encoded in the Isabelle/HOL library.

- Orders

- Sets

- Functions

- (Least) fixpoints and induction

- (Well-founded) recursion

- Arithmetic

- Datatypes

The Roadmap

We are still looking at how the different parts of mathematics are encoded in the Isabelle/HOL library.

- Orders
- Sets
- Functions
- (Least) fixpoints and induction
- (Well-founded) recursion
- Arithmetic
- Datatypes

```
theory Set = HOL:
typedecl 'a set
instance set :: (type) ord ..
consts
 "{}"    :: 'a set ("{}")
 UNIV    :: 'a set
 insert  :: ['a, 'a set] => 'a set
 Collect :: ('a => bool) => 'a set
  "op :" :: "'a => 'a set => bool"
```

Note that Collect and ":" correspond to Abs_{set} and Rep_{set}.

Sets Are a Special Case

Recall that the `typedef` syntax is based on set comprehension. Therefore, sets are a special case of type definitions.

In deviation from our conservative approach, sets are axiomatized as follows:

```
axioms
  mem_Collect_eq [iff]: "(a : {x. P(x)}) = P(a)"
  Collect_mem_eq [simp]: "{x. x:A} = A"
```

One can see though that this is equivalent to the type definition scheme.

`Set.thy`: **More Constant Declarations**

```
Un, Int      :: ['a set, 'a set] => 'a set
Ball, Bex    :: ['a set, 'a => bool] => bool
UNION, INTER:: ['a set, 'a => 'b set] => 'b set
Union, Inter:: (('a set) set) => 'a set
Pow          :: 'a set => 'a set set
"image"      :: ['a => 'b, 'a set] => ('b set)
```

We use old syntax here but only since it is more concise.
  In what follows, recall that

$$\{x \mid f\,x\} = \textit{Collect } f = \textit{Abs}_{set}\, f$$

Set.thy: **Constant Definitions**

```
empty_def:                "{} == {x. False}"
UNIV_def:                "UNIV == {x. True}"
Un_def:              "A Un B == {x. x:A | x:B}"
Int_def:             "A Int B == {x. x:A & x:B}"
insert_def: "insert a B == {x. x=a} Un B"
Ball_def:          "Ball A P == ALL x. x:A --> P(x)"
Bex_def:            "Bex A P == EX x. x:A & P(x)"
```

Nice syntax:

$\{x, y, z\}$ for insert x (insert y (insert z {}))

ALL $x : A.\, Sx$ for Ball $A\ S$

EX $x : A.\, Sx$ for Bex $A\ S$

`Set.thy`: **Constant Definitions (2)**

```
subset_def:    "A <= B == ALL x:A. x:B"
Compl_def:      "- A == {x. ~x:A}"
set_diff_def:  "A - B == {x. x:A & ~x:B}"
UNION_def: "UNION A B == {y. EX x:A. y: B(x)}"
INTER_def: "INTER A B == {y. ALL x:A. y: B(x)}"
```

Note use of $<=$[432] instead of $\subseteq$!

Nice syntax:

UN $x : A.\, S\, x$ or $\bigcup_{x \in A} .\, S\, x$ for   UNION $A\, S$

INT $x : A.\, S\, x$ or $\bigcap_{x \in A} .\, S\, x$ for   INTER $A\, S$

---

[432]Sets are an instance of the type class ord, where the generic constant $<=$ is the subset relation in this particular case.

In fact, the subset relation is reflexive, transitive and anti-symmetric, and so sets are an instance of the axiomatic class order. This is non-obvious and must be proven, which is done not in `Set.thy` itself but in `Fun.thy`, later. This is a technicality of Isabelle.

Set.thy: **Constant Definitions (3)**

```
Union_def: "Union S == (UN x:S. x)"
Inter_def: "Inter S == (INT x:S. x)"
Pow_def:      "Pow A == {B. B <= A}"
image_def:     "f'A == {y. EX x:A. y = f(x)}"
```

Nice syntax:
$\bigcup S$ for   Union S
$\bigcap S$ for   Inter S

Some **Theorems** in `Set.thy`

`CollectI`	$P\,a \Longrightarrow a \in \{x.P\,x\}$
`CollectD`	$a \in \{x.P\,x\} \Longrightarrow P\,a$
`set_ext`	$(\bigwedge x.(x \in A) = (x \in B)) \Longrightarrow A = B$
`subsetI`	$(\bigwedge x.x \in A \Longrightarrow x \in B) \Longrightarrow A \subseteq B$
`eqset_imp_iff`	$A = B \Longrightarrow (x \in A) = (x \in B)$
`UNIV_I`	$x \in \mathtt{UNIV}$
`subset_UNIV`	$A \subseteq \mathtt{UNIV}$
`empty_subsetI`	$\{\} \subseteq A$
`Pow_iff`	$(A \in Pow\,B) = (A \subseteq B)$
`IntI`	$[\![c \in A; c \in B]\!] \Longrightarrow c \in A \cap B$

More Theorems in `Set.thy`

`insert_iff`	$(a \in insert\ b\ A) = (a = b \lor a \in A)$
`image_Un`	$f`(A \cup B) = f`A \cup f`B$
`Inter_lower`	$B \in A \Longrightarrow \bigcap A \subseteq B$
`Inter_greatest`	$(\bigwedge X.X \in A \Longrightarrow C \subseteq X) \Longrightarrow C \subseteq \bigcap A$

19.1 Summary on Sets

Rich and powerful set theory available in HOL:

- No problems with consistency

- Weaker than ZFC (since typed set-theory:) there is no "union of sets[433]"; but: complement-closed[434]

---

[433]In typed set theory (what we have here in HOL), it is not possible to form the union of two sets of different type. This is in contrast to ZFC.

[434]The complement of a typed set A, i.e.

$$\{x \mid x \notin A\}$$

is again a set, whose type is the same as the type of A. In ZFC, the complement construction is not generally allowed since it opens the door to Russell's Paradox.

- Good mechanical support for many set tautologies (`Fast_tac`, `fast_tac set_cs`, `fast_tac eq_cs`, . . . `simp_tac set_ss` . . .)

- Powerful basis for many problems in modeling

20 Functions

472

The Roadmap

We are still looking at how the different parts of mathematics are encoded in the Isabelle/HOL library.

- Orders

- Sets

- Functions

- (Least) fixpoints and induction

- (Well-founded) recursion

- Arithmetic

- Datatypes

The Roadmap

We are still looking at how the different parts of mathematics are encoded in the Isabelle/HOL library.

- Orders

- Sets

- Functions

- (Least) fixpoints and induction

- (Well-founded) recursion

- Arithmetic

- Datatypes

Fun.thy

The theory `Fun.thy`[435] defines some important notions on functions, such as concatenation, the identity function, the image of a function, etc.

We look at it briefly.

---

[435]This file should be contained in your Isabelle distribution. Or, if you only have an Isabelle executable, you can find the sources here:

http://isabelle.in.tum.de/library/

`Fun.thy` builds on `Set.thy`, and it is here that it is proven and used that sets are an instance of the type class `order`.

Two Extracts from `Fun.thy`

Composition and the identity function:

```
constdefs
 id :: "'a => 'a"
"id == %x. x"


 comp :: "['b => 'c, 'a => 'b, 'a] => 'c"
"f o g == %x. f(g(x))"
```

Recall `constdefs` syntax.

Instantiating an Axiomatic Class

Sets are partial orders: set is an instance of the axiomatic class order.

For some reason, this is proven in Fun.thy.

```
instance set :: (type) order
  by (intro_classes,
       (assumption | rule subset_refl
         subset_trans subset_antisym psubset_eq)+)
```

- Axiomatic classes result in proof obligations[436].

- These are discharged[437] whenever instance is stated.

- Type-checking has access to the established properties.

---

[436]To claim that a type is an instance of an axiomatic class, it has to be proven that the axioms (in the case of order: order_refl, order_trans, order_antisym, and order_less_le are indeed fulfilled by that type.
[437]The Isabelle mechanism is such that the line
```
 instance set :: (type) order
   by (intro_classes,
       (assumption | rule
        subset_refl subset_trans subset_antisym psubset_eq)+)
```
instructs Isabelle to prove the axioms using the previously proven theorems subset_refl, subset_trans, subset_antisym, and psubset_eq.

20.1 Conclusion of Orders, Sets, Functions

- Theory says: conservative extensions can be used to build consistent libraries.

20.1 Conclusion of Orders, Sets, Functions

- Theory says: conservative extensions can be used to build consistent libraries.

- Sets as one important package of Isabelle/HOL library:

  - Set theory is typed, but very rich and powerfully supported.

20.1 Conclusion of Orders, Sets, Functions

- Theory says: conservative extensions can be used to build consistent libraries.

- Sets as one important package of Isabelle/HOL library:

  - Set theory is typed, but very rich and powerfully supported.

  - Sets are instance of ord and order type class, demonstrates type classes as structuring mechanism in Isabelle.

20.1 Conclusion of Orders, Sets, Functions

- Theory says: conservative extensions can be used to build consistent libraries.

- Sets as one important package of Isabelle/HOL library:
  - Set theory is typed, but very rich and powerfully supported.
  - Sets are instance of `ord` and `order` type class, demonstrates type classes as structuring mechanism in Isabelle.

- Will see more examples: Isabelle/HOL contains some 10000 `thm`'s.

21 Background: Recursion, Induction, and Fixpoints

The Roadmap

We are still looking at how the different parts of mathematics are encoded in the Isabelle/HOL library.

- Orders

- Sets

- Functions

- (Least) fixpoints and induction

- (Well-founded) recursion

- Arithmetic

- Datatypes

The Roadmap

We are still looking at how the different parts of mathematics are encoded in the Isabelle/HOL library.

- Orders

- Sets

- Functions

- (Least) fixpoints and induction

- (Well-founded) recursion

- Arithmetic

- Datatypes

Recursion Based on Set Theory

Current stage of our course:

- On the basis of conservative extensions, set theory can be built safely.

- But: our mathematical world is still quite small and quite remote from computer science: we have no means of introducing recursive definitions (recursive programs, recursive set equations, ...).

How can we benefit from set theory to introduce recursion?

Recursion and General Fixpoints

Naïve Approach: One could axiomatize fixpoint combinator Y as

$$\frac{}{Y = \lambda F.F(YF)} \text{ fix}$$

This axiom is not a constant definition[438].

Then we could easily derive

$$\forall F^{\alpha \to \alpha}.Y\,F = F\,(Y\,F)^{439}.$$

---

[438]The axiom

$$Y = \lambda F.F(YF)$$

is not a constant definition, since Y occurs again on the right-hand side.

[439]In words, this says that $Y\,F$ is a fixpoint of F.

Recursion and General Fixpoints

Naïve Approach: One could axiomatize fixpoint combinator Y as

$$\frac{}{Y = \lambda F.F(YF)} \text{ fix}$$

This axiom is not a constant definition[438].

Then we could easily derive

$$\forall F^{\alpha \to \alpha}.Y \, F = F \, (Y \, F)^{[439]}.$$

- Why are we interested in Y?

- What is the problem with such a definition?

---

[438]The axiom

$$Y = \lambda F.F(YF)$$

is not a constant definition, since Y occurs again on the right-hand side.

[439]In words, this says that $Y \, F$ is a fixpoint of F.

Why Are We Interested in Y?

First, why are we interested in recursion (solutions to recursive equations[440])?

Why Are We Interested in Y?

First, why are we interested in recursion (solutions to recursive equations[440])?

- Recursively defined functions are solutions of such equations (example: fac[441]).

- Inductively defined sets are solutions of such equations

---

[440]By a recursive equation, we mean an equation of the form

$$f = e$$

where f occurs in e. A fortiori, such an equation does not qualify as constant definition.

[441]In the following explanations, any constants like 1 or $+$ or **if-then-else** are intended to have their usual meaning.

A fixpoint combinator is a function Y that returns a fixpoint of a function F, i.e., Y must fulfill the equation $YF = F(YF)$. Doing λ-abstraction over F on both sides and η-conversion (backwards) on the left-hand side, we have

$$Y = \lambda F.F(YF)$$

This is a recursive equation. We will now demonstrate how a definition of a function fac (factorial) using a recursive equation can be transformed to a definition that uses Y instead of using recursion directly.

In a functional programming language we might define

$$fac\ n = (\textbf{if}\ n = 0\ \textbf{then}\ 1\ \textbf{else}\ n * fac\ (n-1)).$$

We now massage this equation a bit. Doing λ-abstraction on both sides we get

$$\lambda n.\ fac\ n = (\lambda n.\ \textbf{if}\ n = 0\ \textbf{then}\ 1\ \textbf{else}\ n * fac(n-1))$$

which is the η-conversion of

$$fac = (\lambda n.\ \textbf{if}\ n = 0\ \textbf{then}\ 1\ \textbf{else}\ n * fac(n-1))$$

which in turn is a β-reduction of

$$fac = (\underline{(\lambda f.\ \lambda n.\ \textbf{if}\ n = 0\ \textbf{then}\ 1\ \textbf{else}\ n * f(n-1))\ fac}) \tag{3}$$

We are looking for a solution to (3). We abbreviate the underlined expression by Fac. We claim $fac = Y\ Fac$, i.e., it is a solution to (3). Simply replacing fac with $Y\ Fac$ in (3) we get

$$Y\ Fac = Fac\ (Y\ Fac)$$

(example: $Fin\ A^{442}$, all finite subsets of A).

(example: $Fin\ A$[442], all finite subsets of A).

We are interested in Y because it is the mother of all re-

which holds by the definition of Y.

Thus we see that a recursive definition of a function can be transformed so that the function is the fixpoint of an appropriate functional (a function taking a function as argument).

[442]We want to define a function Fin such that $Fin\ A$ is the set of all finite subsets of A.

How do you construct the set of all finite subsets of A? The following pseudo-code suggests what you have to do:

$$S := \{\{\}\};$$
forever do
   foreach $a \in A$ **do**
      foreach $B \in S$ **do**
         add $(\{a\} \cup B)$ **to** S
od od od

This means that you have to add new sets forever (however, when you actually do this construction for a finite set A, it will indeed reach a fixpoint, i.e., adding new sets won't change anything).

Generally (even if A is infinite), $Fin\ A$ is a set such that adding new sets as suggested by the pseudo-code won't change anything. Written as recursive equation:

$$Fin\ A = \{\{\}\} \cup \bigcup x \in A.((\texttt{insert}\ x)\ `\ (Fin\ A))$$

Recall that $`$ is nice syntax for $image$, defined in $\texttt{Set.thy}$.

The above is a β-reduction of

$$Fin\ A = \underline{(\lambda X.\ \{\{\}\} \cup \bigcup x \in A.((\texttt{insert}\ x)\ `\ X))}\ (Fin\ A) \tag{4}$$

We are looking for a solution to (4). We abbreviate the underlined expression by FA. We claim

$$Fin\ A = Y\ FA,$$

i.e., it is a solution to (4). Simply replacing $Fin\ A$ with $Y\ FA$ in (4) we get

$$Y\ FA = FA(Y\ FA),$$

which holds by the definition of Y.

cursions. With Y, recursive axioms can be converted[443] into constant definitions.

You should compare this to what we said about fac. Note that in this example, there is no such thing as a recursive call to a "smaller" argument as in fac example.

[443]Any recursive function can be defined by an expression (functional) which is not itself recursive, but instead relies on the recursive equation defining Y.

Consider fac or $Fin\ A$ as an example.

What's the Problem with such an Axiom?

Such a definition would lead to inconsistency.

This is not surprising because not all functions have a fixpoint.

Therefore we only consider special forms of fixpoint combinators.

We consider two approaches: Least fixpoints (Tarski) and well-founded orderings.

22 Least Fixpoints

The Roadmap

We are still looking at how the different parts of mathematics are encoded in the Isabelle/HOL library.

- Orders

- Sets

- Functions

- (Least) fixpoints and induction

- (Well-founded) recursion

- Arithmetic

- Datatypes

The Roadmap

We are still looking at how the different parts of mathematics are encoded in the Isabelle/HOL library.

- Orders

- Sets

- Functions

- (Least) fixpoints and induction

- (Well-founded) recursion

- Arithmetic

- Datatypes

22.1 First Approach: Least Fixpoints (Tarski)

- Restriction: F is not of arbitrary type $(\alpha \to \alpha)$ but of set type $(\alpha\ set \to \alpha\ set)$.

- Instead of Y define lfp by an equation which is not recursive.

- lfp is fixpoint combinator, but only under additional condition that F is monotone[444], and: this is not obvious (requires non-trivial proof)!

  This leads us towards recursion and induction.

---

[444]A function f is monotone w.r.t. a partial order $\leq$ if the following holds: $A \leq B$ implies $f(A) \leq f(B)$.

In particular, we consider the order given by the subset relation.

```
Lfp = Product_Type +
constdefs
 lfp :: ['a set => 'a set] => 'a set
 "lfp(f) == Inter({u. f(u) <= u})"
```

Note:

- => is function type arrow.

- <= ("⊆") is a partial order.

- Inter ("⋂") gives a "minimum": $\forall A \in S.(\bigcap S) \subseteq A$.

```
Lfp = Product_Type +
constdefs
 lfp :: ['a set => 'a set] => 'a set
 "lfp(f) == Inter({u. f(u) <= u})"
```

Note:

- => is function type arrow.

- <= ("⊆") is a partial order.

- Inter ("⋂") gives a "minimum": $\forall A \in S.(\bigcap S) \subseteq A$.

- Definition of lfp is conservative. That's fine. But is it a fixpoint combinator?

---

[445] These files should be contained in your Isabelle distribution. Or, if you only have an Isabelle executable, you can find the sources here:

http://isabelle.in.tum.de/library/

22.2 Tarski's Fixpoint Theorem

Definition (Tarski): *If f is monotone, then $lfp\ f = f\ (lfp\ f)$.*

In Isabelle, the theorem is shown in `Lfp.ML` and called `lfp_unfold`.

We show the theorem using mathematical notation and a graphical illustration to help intuition.

The proof has four steps.

Tarski's Fixpoint Theorem (1)

Claim 1 ("*lfp* lower bound"): If $f\ A \subseteq A$ then $lfp\ f \subseteq A$.

---

[446]α is not a set but a type (variable). But we can consider the set of all terms of that type (UNIV of type α).

The polymorphic constant UNIV was defined in `Set.thy`. UNIV of type τ set is the set containing all terms of type τ.
[447]In general, needless to say, there could be any number of such sets, but the picture is to be understood in the sense that the three circles are all the sets A with the property $f\ A \subseteq A$.

Tarski's Fixpoint Theorem (1)

Claim 1 ("*lfp* lower bound"): If $f\ A \subseteq A$ then $lfp\ f \subseteq A$.

The box denotes "the set" α[446]. The three circles[447] denote the sets A for which $f\ A \subseteq A$.

---

[446]α is not a set but a type (variable). But we can consider the set of all terms of that type (UNIV of type α).

The polymorphic constant UNIV was defined in `Set.thy`. UNIV of type τ set is the set containing all terms of type τ.

[447]In general, needless to say, there could be any number of such sets, but the picture is to be understood in the sense that the three circles are all the sets A with the property $f\ A \subseteq A$.

Tarski's Fixpoint Theorem (1)

Claim 1 ("*lfp* lower bound"): If $f\ A \subseteq A$ then $lfp\ f \subseteq A$.

The box denotes "the set" α^{446}. The three circles[447] denote the sets A for which $f\ A \subseteq A$.
By definition, $lfp\ f$ is the intersection.

---

[446]α is not a set but a type (variable). But we can consider the set of all terms of that type (`UNIV` of type α).

The polymorphic constant `UNIV` was defined in `Set.thy`. `UNIV` of type τ `set` is the set containing all terms of type τ.

[447]In general, needless to say, there could be any number of such sets, but the picture is to be understood in the sense that the three circles are all the sets A with the property $f\ A \subseteq A$.

Tarski's Fixpoint Theorem (1)

Claim 1 ("*lfp* lower bound"): If $f\ A \subseteq A$ then $lfp\ f \subseteq A$.

The box denotes "the set" α[446]. The three circles[447] denote the sets A for which $f\ A \subseteq A$.

By definition, $lfp\ f$ is the intersection.

Pick an A for which $f\ A \subseteq A$. Clearly, $lfp\ f \subseteq A$.

Or as proof tree.
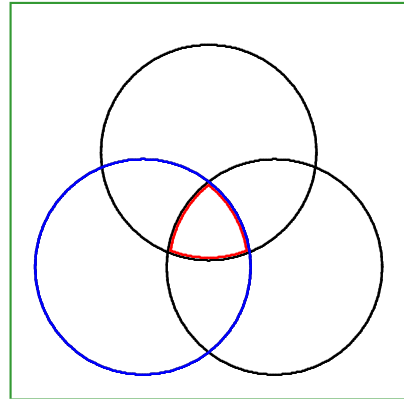
---

[446]α is not a set but a type (variable). But we can consider the set of all terms of that type (UNIV of type α).

The polymorphic constant UNIV was defined in Set.thy. UNIV of type τ set is the set containing all terms of type τ.

[447]In general, needless to say, there could be any number of such sets, but the picture is to be understood in the sense that the three circles are all the sets A with the property $f\ A \subseteq A$.

Tarski's Fixpoint Theorem (2)

Claim 2 ("*lfp* greatest"): For all A, if for all U, $f\ U \subseteq U$ implies $A \subseteq U$, then $A \subseteq lfp\ f$.

Tarski's Fixpoint Theorem (2)

Claim 2 ("*lfp* greatest"): For all A, if for all U, $f\ U \subseteq U$ implies $A \subseteq U$, then $A \subseteq lfp\ f$.

The three circles denote the sets U for which $f\ U \subseteq U$.
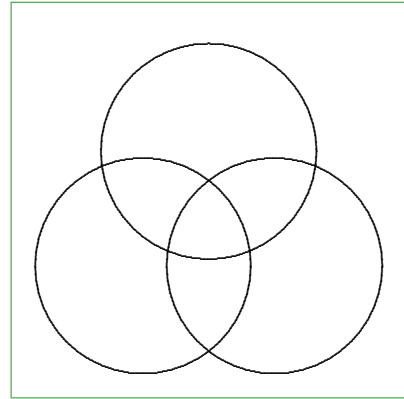
Tarski's Fixpoint Theorem (2)

Claim 2 ("*lfp* greatest"): For all A, if for all U, $f\ U \subseteq U$ implies $A \subseteq U$, then $A \subseteq lfp\ f$.

The three circles denote the sets U for which $f\ U \subseteq U$.
By <u>hypothesis</u>, $A \subseteq U$ for each U
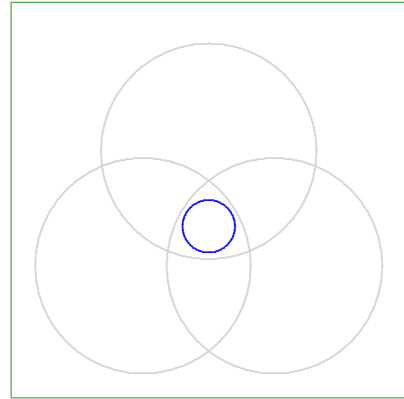
Tarski's Fixpoint Theorem (2)

Claim 2 ("*lfp* greatest"): For all A, if for all U, $f\ U \subseteq U$ implies $A \subseteq U$, then $A \subseteq lfp\ f$.

The three circles denote the sets U for which $f\ U \subseteq U$.

By hypothesis, $A \subseteq U$ for each U (1st,
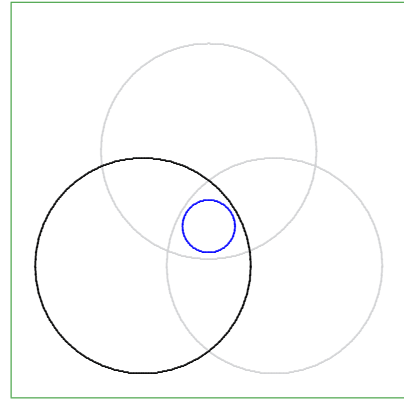
Tarski's Fixpoint Theorem (2)

Claim 2 ("*lfp* greatest"): For all A, if for all U, $f\ U \subseteq U$ implies $A \subseteq U$, then $A \subseteq lfp\ f$.

The three circles denote the sets U for which $f\ U \subseteq U$.

By hypothesis, $A \subseteq U$ for each U (1st, 2nd,
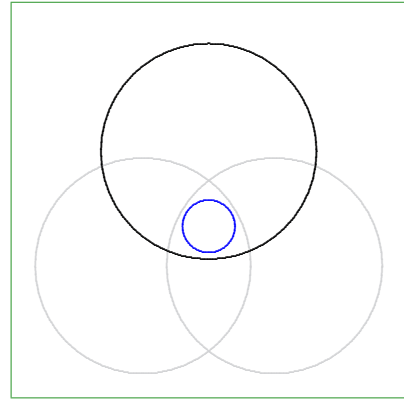
Tarski's Fixpoint Theorem (2)

Claim 2 ("*lfp* greatest"): For all A, if for all U, $f\ U \subseteq U$ implies $A \subseteq U$, then $A \subseteq lfp\ f$.

The three circles denote the sets U for which $f\ U \subseteq U$.
By hypothesis, $A \subseteq U$ for each U (1st, 2nd, 3rd ...).
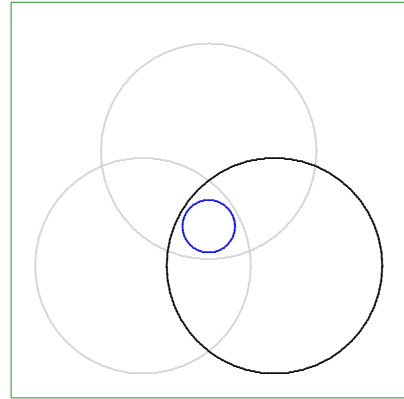
Tarski's Fixpoint Theorem (2)

Claim 2 ("*lfp* greatest"): For all A, if for all U, $f\ U \subseteq U$ implies $A \subseteq U$, then $A \subseteq lfp\ f$.

The three circles denote the sets U for which $f\ U \subseteq U$.

By hypothesis, $A \subseteq U$ for each U (1st, 2nd, 3rd ...).

By definition, $lfp\ f$ is the intersection.
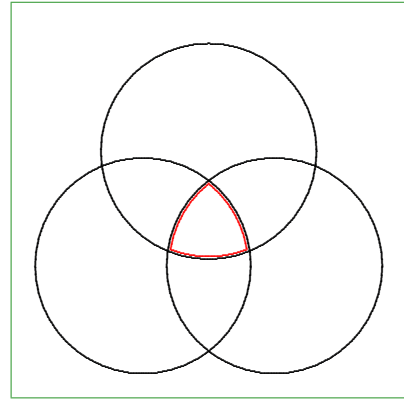
Tarski's Fixpoint Theorem (2)

Claim 2 ("*lfp* greatest"): For all A, if for all U, $f\ U \subseteq U$ implies $A \subseteq U$, then $A \subseteq lfp\ f$.

The three circles denote the sets U for which $f\ U \subseteq U$.

By hypothesis, $A \subseteq U$ for each U (1st, 2nd, 3rd . . .).

By definition, $lfp\ f$ is the intersection.
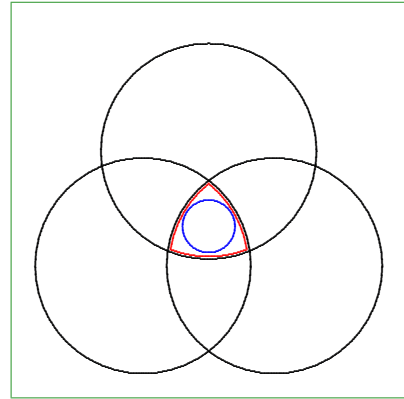
Clearly, $A \subseteq lfp\ f$.

Tarski's Fixpoint Theorem (2)

Claim 2 ("*lfp* greatest"): For all A, if for all U, $f\ U \subseteq U$ implies $A \subseteq U$, then $A \subseteq lfp\ f$.

The three circles denote the sets U for which $f\ U \subseteq U$.

By hypothesis, $A \subseteq U$ for each U (1st, 2nd, 3rd ...).

By definition, $lfp\ f$ is the intersection.

Clearly, $A \subseteq lfp\ f$.

Or as proof tree.
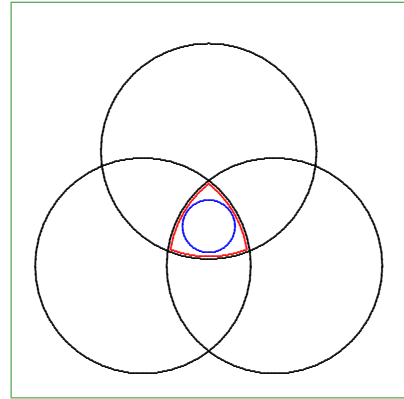
Tarski's Fixpoint Theorem (3)

Claim 3: If f is monotone then $f(\mathit{lfp}\ f) \subseteq \mathit{lfp}\ f$.

Tarski's Fixpoint Theorem (3)

Claim 3: If f is monotone then $f(\mathit{lfp}\ f) \subseteq \mathit{lfp}\ f$.

First show Claim 3*: $f\ U \subseteq U$ implies $f(\mathit{lfp}\ f) \subseteq U$.

Tarski's Fixpoint Theorem (3)

Claim 3: If f is monotone then $f(lfp\ f) \subseteq lfp\ f$.

First show Claim 3*: $f\ U \subseteq U$ implies $f(lfp\ f) \subseteq U$.

Let the circle be such a U. By Claim 1, $lfp\ f \subseteq U$.
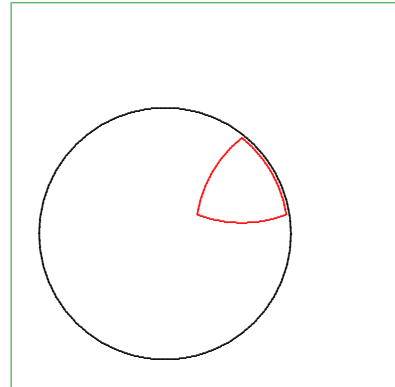
Tarski's Fixpoint Theorem (3)

Claim 3: If f is monotone then $f(\mathit{lfp}\ f) \subseteq \mathit{lfp}\ f$.

First show Claim 3*: $\underline{f\ U \subseteq U}$ implies $f(\mathit{lfp}\ f) \subseteq U$.

Let the circle be such a U. By Claim 1, $\mathit{lfp}\ f \subseteq U$.
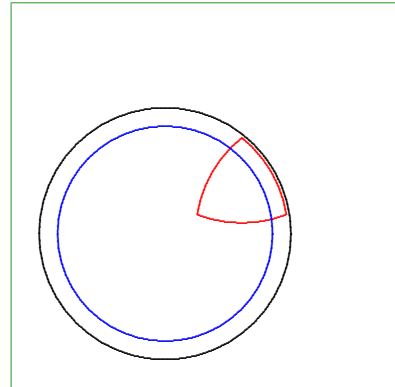
$f\ U \subseteq U$ (\underline{hypothesis}).

Tarski's Fixpoint Theorem (3)

Claim 3: If f is monotone then $f(lfp\ f) \subseteq lfp\ f$.

First show Claim 3*: $f\ U \subseteq U$ implies $f(lfp\ f) \subseteq U$.

Let the circle be such a U. By Claim 1, $lfp\ f \subseteq U$.

$f\ U \subseteq U$ (hypothesis).

$f(lfp\ f) \subseteq f\ U$ (monotonicity).
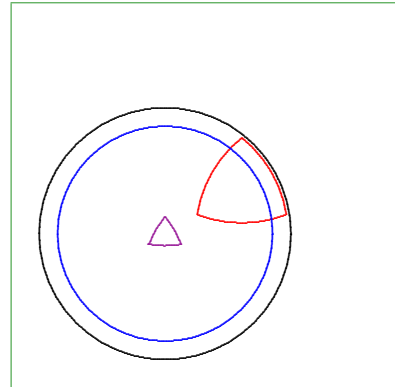
Tarski's Fixpoint Theorem (3)

Claim 3: If f is monotone then $f(\mathit{lfp}\ f) \subseteq \mathit{lfp}\ f$.

First show Claim 3*: $\underline{f\ U \subseteq U}$ implies $f(\mathit{lfp}\ f) \subseteq U$.

Let the circle be such a U. By Claim 1, $\mathit{lfp}\ f \subseteq U$.

$f\ U \subseteq U$ (hypothesis).

$f(\mathit{lfp}\ f) \subseteq f\ U$ (monotonicity).

$f(\mathit{lfp}\ f) \subseteq U$ (transitivity of $\subseteq$).
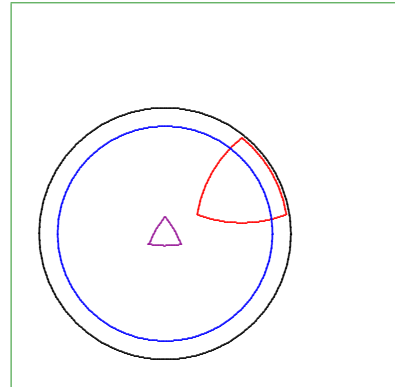
Claim 3* shown.

Tarski's Fixpoint Theorem (3)

Claim 3: If f is monotone then $f(lfp\ f) \subseteq lfp\ f$.

First show Claim 3*: $\underline{f\ U \subseteq U}$ implies $f(lfp\ f) \subseteq U$.

Let the circle be such a U. By Claim
1, $lfp\ f \subseteq U$.
$f\ U \subseteq U$ (hypothesis).
$f(lfp\ f) \subseteq f\ U$ (monotonicity).
$f(lfp\ f) \subseteq U$ (transitivity of $\subseteq$).
Claim 3* shown.
By Claim 2 (letting $A := f(lfp\ f)$),
$f(lfp\ f) \subseteq lfp\ f$.
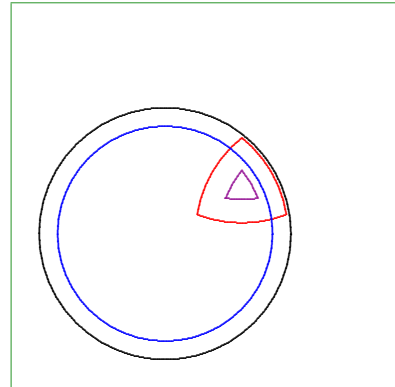
Tarski's Fixpoint Theorem (4)

Claim 4: If f is monotone then $lfp\ f \subseteq f(lfp\ f)$.

Tarski's Fixpoint Theorem (4)

Claim 4: If f is monotone then $lfp\ f \subseteq f(lfp\ f)$.

By Claim 3, $f(lfp\ f) \subseteq lfp\ f$.
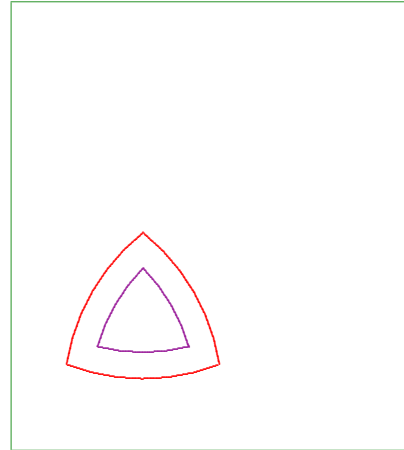
Tarski's Fixpoint Theorem (4)

Claim 4: If f is monotone then $lfp\ f \subseteq f(lfp\ f)$.

By Claim 3, $f(lfp\ f) \subseteq lfp\ f$.
By monotonicity, $f(f(lfp\ f)) \subseteq f(lfp\ f)$.
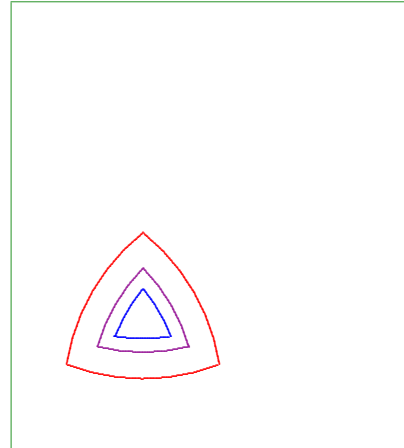
Tarski's Fixpoint Theorem (4)

Claim 4: If f is monotone then $lfp\ f \subseteq f(lfp\ f)$.

By Claim 3, $f(lfp\ f) \subseteq lfp\ f$.

By monotonicity, $f(f(lfp\ f)) \subseteq f(lfp\ f)$.

By Claim 1 (letting $A := f(lfp\ f)$), $lfp\ f \subseteq f(lfp\ f)$.
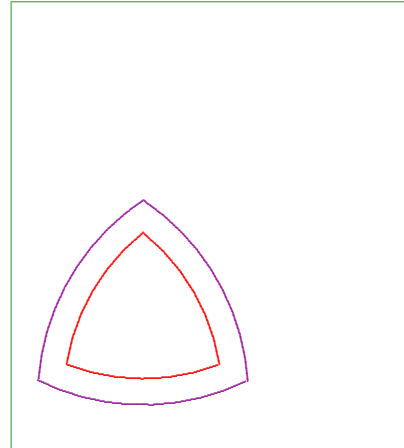
Tarski's Fixpoint Theorem (4)

Claim 4: If f is monotone then $lfp\ f \subseteq f(lfp\ f)$.

By Claim 3, $f(lfp\ f) \subseteq lfp\ f$.

By monotonicity, $f(f(lfp\ f)) \subseteq f(lfp\ f)$.

By Claim 1 (letting $A := f(lfp\ f)$), $lfp\ f \subseteq f(lfp\ f)$.

Or as proof tree.

Tarski's Fixpoint Theorem: QED

Claim 3 ($lfp\ f \subseteq f(lfp\ f)$) and Claim 4 ($f(lfp\ f) \subseteq lfp\ f$) together give the result:

If f is monotone, then $lfp\ f = f\ (lfp\ f)$.

So under appropriate conditions, lfp is a fixpoint combinator.

We will later reuse Claim 1.

Alternative: A Natural-Deduction Style Proof

The proof can also be presented in natural deduction style.

Tarski's Fixpoint Theorem (1)

Claim 1 ("lfp lower bound"): If $f\ A \subseteq A$ then $lfp\ f \subseteq A$.

$$\cfrac{\cfrac{\cfrac{[f\ A \subseteq A]^1}{A \in \{u.fu \subseteq u\}}\ \text{CollectI}}{\bigcap\{u.fu \subseteq u\} \subseteq A}\ \text{Inter\_lower}}{\cfrac{lfp\ f \subseteq A}{f\ A \subseteq A \to lfp\ f \subseteq A}\ \text{Def.}\ lfp}\ \to\text{-}I^1$$

Tarski's Fixpoint Theorem (2)

Claim 2 ("*lfp* greatest"): For all A, if for all U, $f\ U \subseteq U$ implies $A \subseteq U$, then $A \subseteq lfp\ f$.

$$\cfrac{\cfrac{\cfrac{[\forall x.fx \subseteq x \to A \subseteq x]^1}{\forall x.x \in \{u.fu \subseteq u\} \to A \subseteq x}\ \textit{subst},\ \mathsf{CollectI}}{A \subseteq \cap\{u.fu \subseteq u\}}\ \mathsf{Inter\_greatest}}{\cfrac{A \subseteq lfp\,f}{(\forall x.fx \subseteq x \to A \subseteq x) \to A \subseteq lfp\,f}}\ \text{Def. } lfp}{\quad} \to\text{-}I^1$$

Tarski's Fixpoint Theorem (3)

Claim 3: If f is monotone then $f(lfp\ f) \subseteq lfp\ f$.

$$\cfrac{\cfrac{[mono\ f]^1 \quad \cfrac{[fx \subseteq x]^2}{lfp\ f \subseteq x}}{\cfrac{f(lfp\ f) \subseteq f\ x \qquad [fx \subseteq x]^2}{\cfrac{\cfrac{f(lfp\ f) \subseteq x}{\forall x. fx \subseteq x \to f(lfp\ f) \subseteq x}\ \forall\text{-}I,\ \to\text{-}I^2}{f(lfp\ f) \subseteq lfp\ f}\ \mathsf{lfp\_greatest},\ \to\text{-}E}\ \mathsf{order\_trans}}}{mono\ f \to f(lfp\ f) \subseteq lfp\ f}\ \to\text{-}I^1$$

Tarski's Fixpoint Theorem (4)

Claim 4: If f is monotone then $lfp\ f \subseteq f(lfp\ f)$.

$$\cfrac{\cfrac{[mono\ f]^1 \quad \cfrac{[mono\ f]^1}{f(lfp\ f) \subseteq lfp\ f}\ \text{Claim 3}, \rightarrow\text{-}E}{\cfrac{f(f(lfp\ f)) \subseteq f(lfp\ f)}{\text{monoD}}}{\cfrac{lfp\ f \subseteq f(lfp\ f)}{\text{lfp\_lowerbound}, \rightarrow\text{-}E}}}{mono\ f \rightarrow lfp\ f \subseteq f(lfp\ f)}\ \rightarrow\text{-}I^1$$

Completing Proof Tree

$$\dfrac{\dfrac{[mono\ f]^1}{lfp\ f \subseteq f(lfp\ f)}\ \text{Claim 4} \qquad \dfrac{[mono\ f]^1}{f(lfp\ f) \subseteq lfp\ f}\ \text{Claim 3}}{\dfrac{lfp\ f = f(lfp\ f)}{mono\ f \to lfp\ f = f(lfp\ f)}\ \to\text{-}I^1}\ \text{equalityI}$$

22.3 Induction Based on `Lfp.thy`

Definition (lfp induction): *If*

- f *is* *monotone, and*

- $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq \{x \mid P\,x\}$,

then $lfp\ f \subseteq \{x \mid P\,x\}$.

22.3 Induction Based on `Lfp.thy`

Definition (lfp induction): *If*

- f *is monotone, and*

- $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq \{x \mid P\,x\}$,

then $lfp\ f \subseteq \{x \mid P\,x\}$.

In Isabelle[448], it is called `lfp_induct`:

$$[\![a \in lfp\ f;\, mono\ f;\, \bigwedge x.x \in f(lfp\ f \cap \{x.P\,x\}) \Longrightarrow P\,x]\!]$$
$$\Longrightarrow P\,a$$

We now show the theorem similarly as Tarski's Theorem.

---

[448]The theorem is phrased a bit differently in the "mathematical" version we give here and in the Isabelle version (see `Lfp.ML`). This is convenient for the graphical illustration of the proof.

The "mathematical phrasing" corresponding closely to the Isabelle version would be the following:

Definition (Induct (alternative)): *If*

- $a \in lfp\ f$, *and*

- f *is monotone, and*

- *for all* x, $x \in f(lfp\ f \cap \{x \mid P\,x\})$ *implies* $P\,x$

then $P\,a$ *holds.*

Other phrasings, which may help to get some intuition about the theorem:

Definition (Induct (alternative)): *If*

- $a \in lfp\ f$, *and*

- f *is monotone, and*

Showing `lfp_induct`

Circles denote $lfp\ f$ and $\{x \mid P\ x\}$.

- $f(lfp\ f \cap \{x \mid P\ x\}) \subseteq \{x \mid P\ x\}$

then $P\ a$ holds.

Definition (Induct (alternative)): *If*

- *f is monotone, and*

- $f(lfp\ f \cap \{x \mid P\ x\}) \subseteq \{x \mid P\ x\}$

then for all x in lfp f, we have $P\ x$.

[449]$lfp\ f \cap \{x \mid P\ x\} \subseteq lfp\ f$, so by monotonicity, $f(lfp\ f \cap \{x \mid P\ x\}) \subseteq f(lfp\ f)$.

[450]We have just seen $f(lfp\ f \cap \{x \mid P\ x\}) \subseteq lfp\ f \cap \{x \mid P\ x\}$.

By Claim 1 (setting $A := lfp\ f \cap \{x \mid P\ x\}$), this implies $lfp(f) \subseteq lfp\ f \cap \{x \mid P\ x\}$.

[451]We have $lfp\ f \cap \{x \mid P\ x\} \subseteq lfp(f)$ and $lfp(f) \subseteq lfp\ f \cap \{x \mid P\ x\}$, and so $lfp(f) = lfp\ f \cap \{x \mid P\ x\}$ by the antisymmetry of $\subseteq$.

Showing `lfp_induct`

Circles denote $lfp\ f$ and $\{x \mid P\ x\}$.

By monotonicity[449],

$f(lfp\ f \cap \{x \mid P\ x\}) \quad \subseteq \quad f(lfp\ f)$.

By Tarski, $lfp\ f \quad = \quad f(lfp\ f)$.

- $f(lfp\ f \cap \{x \mid P\ x\}) \subseteq \{x \mid P\ x\}$

then $P\ a$ holds.

Definition (Induct (alternative)): If

- f is monotone, and

- $f(lfp\ f \cap \{x \mid P\ x\}) \subseteq \{x \mid P\ x\}$

then for all x in $lfp\ f$, we have $P\ x$.

[449]$lfp\ f \cap \{x \mid P\ x\} \subseteq lfp\ f$, so by monotonicity, $f(lfp\ f \cap \{x \mid P\ x\}) \subseteq f(lfp\ f)$.

[450]We have just seen $f(lfp\ f \cap \{x \mid P\ x\}) \subseteq lfp\ f \cap \{x \mid P\ x\}$.

By Claim 1 (setting $A := lfp\ f \cap \{x \mid P\ x\}$), this implies $lfp(f) \subseteq lfp\ f \cap \{x \mid P\ x\}$.

[451]We have $lfp\ f \cap \{x \mid P\ x\} \subseteq lfp(f)$ and $lfp(f) \subseteq lfp\ f \cap \{x \mid P\ x\}$, and so $lfp(f) = lfp\ f \cap \{x \mid P\ x\}$ by the antisymmetry of $\subseteq$.

Showing `lfp_induct`

Circles denote $lfp\ f$ and $\{x \mid P\,x\}$.

By monotonicity[449],
$f(lfp\ f \cap \{x \mid P\,x\}) \quad \subseteq \quad f(lfp\ f)$.
By Tarski, $lfp\ f \ = \ f(lfp\ f)$. Hence
$f(lfp\ f \cap \{x \mid P\,x\}) \subseteq lfp\ f$.

- $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq \{x \mid P\,x\}$

then $P\ a$ holds.

Definition (Induct (alternative)): *If*

- *f is monotone, and*

- $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq \{x \mid P\,x\}$

then for all x in lfp f, we have $P\ x$.

[449]$lfp\ f \cap \{x \mid P\,x\} \subseteq lfp\ f$, so by monotonicity, $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq f(lfp\ f)$.

[450]We have just seen $f(lfp\ \ f \cap \{x \ \mid \ P\,x\}) \ \subseteq \ lfp\ f \cap \{x \mid P\,x\}$.

By Claim 1 (setting $A := lfp\ f \cap \{x \mid P\,x\}$), this implies $lfp(f) \subseteq lfp\ f \cap \{x \mid P\,x\}$.

[451]We have $lfp\ f \cap \{x \mid P\,x\} \ \subseteq \ lfp(f)$ and $lfp(f) \ \subseteq \ lfp\ f \cap \{x \mid P\,x\}$, and so $lfp(f) = lfp\ f \cap \{x \mid P\,x\}$ by the antisymmetry of $\subseteq$.

Showing `lfp_induct`

Circles denote $lfp\ f$ and $\{x \mid P\,x\}$.
By monotonicity[449],
$f(lfp\ f \cap \{x \mid P\,x\}) \subseteq f(lfp\ f)$.
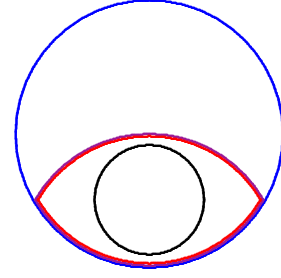By Tarski, $lfp\ f = f(lfp\ f)$. Hence
$f(lfp\ f \cap \{x \mid P\,x\}) \subseteq lfp\ f$.
By hypothesis, $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq \{x \mid P\,x\}$, and so we must adjust picture:
$f(lfp\ f \cap \{x \mid P\,x\}) \subseteq lfp\ f \cap \{x \mid P\,x\}$.

- $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq \{x \mid P\,x\}$

then $P\ a$ holds.

Definition (Induct (alternative)): *If*

- f *is monotone, and*

- $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq \{x \mid P\,x\}$

then for all x in lfp f, we have $P\ x$.
[449]$lfp\ f \cap \{x \mid P\,x\} \subseteq lfp\ f$, so by monotonicity, $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq f(lfp\ f)$.
[450]We have just seen $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq lfp\ f \cap \{x \mid P\,x\}$.

By Claim 1 (setting $A := lfp\ f \cap \{x \mid P\,x\}$), this implies $lfp(f) \subseteq lfp\ f \cap \{x \mid P\,x\}$.
[451]We have $lfp\ f \cap \{x \mid P\,x\} \subseteq lfp(f)$ and $lfp(f) \subseteq lfp\ f \cap \{x \mid P\,x\}$, and so $lfp(f) = lfp\ f \cap \{x \mid P\,x\}$ by the antisymmetry of $\subseteq$.

Showing `lfp_induct`

Circles denote $lfp\ f$ and $\{x \mid P\,x\}$.

By monotonicity[449],
$f(lfp\ f \cap \{x \mid P\,x\}) \subseteq f(lfp\ f)$.
By Tarski, $lfp\ f = f(lfp\ f)$. Hence
$f(lfp\ f \cap \{x \mid P\,x\}) \subseteq lfp\ f$.
By hypothesis, $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq \{x \mid P\,x\}$, and so we must adjust picture:
$f(lfp\ f \cap \{x \mid P\,x\}) \subseteq lfp\ f \cap \{x \mid P\,x\}$.
By Claim 1[450], $lfp\ f \subseteq lfp\ f \cap \{x \mid P\,x\}$
and so[451] $lfp\ f = lfp\ f \cap \{x \mid P\,x\}$.

- $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq \{x \mid P\,x\}$

then $P\,a$ holds.

Definition (Induct (alternative)): *If*

- *f is monotone, and*

- $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq \{x \mid P\,x\}$

then for all x in $lfp\ f$, we have $P\,x$.

[449]$lfp\ f \cap \{x \mid P\,x\} \subseteq lfp\ f$, so by monotonicity, $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq f(lfp\ f)$.

[450]We have just seen $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq lfp\ f \cap \{x \mid P\,x\}$.

By Claim 1 (setting $A := lfp\ f \cap \{x \mid P\,x\}$), this implies $lfp(f) \subseteq lfp\ f \cap \{x \mid P\,x\}$.

[451]We have $lfp\ f \cap \{x \mid P\,x\} \subseteq lfp(f)$ and $lfp(f) \subseteq lfp\ f \cap \{x \mid P\,x\}$, and so $lfp(f) = lfp\ f \cap \{x \mid P\,x\}$ by the antisymmetry of $\subseteq$.

Showing `lfp_induct`

Circles denote $lfp\ f$ and $\{x \mid P\,x\}$.

By monotonicity[449], $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq f(lfp\ f)$. By Tarski, $lfp\ f = f(lfp\ f)$. Hence $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq lfp\ f$.

By hypothesis, $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq \{x \mid P\,x\}$, and so we must adjust picture: $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq lfp\ f \cap \{x \mid P\,x\}$.

By Claim 1[450], $lfp\ f \subseteq lfp\ f \cap \{x \mid P\,x\}$ and so[451] $lfp\ f = lfp\ f \cap \{x \mid P\,x\}$.

Conclusion: $lfp\ f \subseteq \{x \mid P\,x\}$.

- $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq \{x \mid P\,x\}$

then $P\ a$ holds.

Definition (Induct (alternative)): *If*

- *f is monotone, and*

- $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq \{x \mid P\,x\}$

then for all x in $lfp\ f$, we have $P\ x$.

[449]$lfp\ f \cap \{x \mid P\,x\} \subseteq lfp\ f$, so by monotonicity, $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq f(lfp\ f)$.

[450]We have just seen $f(lfp\ f \cap \{x \mid P\,x\}) \subseteq lfp\ f \cap \{x \mid P\,x\}$.

By Claim 1 (setting $A := lfp\ f \cap \{x \mid P\,x\}$), this implies $lfp(f) \subseteq lfp\ f \cap \{x \mid P\,x\}$.

[451]We have $lfp\ f \cap \{x \mid P\,x\} \subseteq lfp(f)$ and $lfp(f) \subseteq lfp\ f \cap \{x \mid P\,x\}$, and so $lfp(f) = lfp\ f \cap \{x \mid P\,x\}$ by the antisymmetry of $\subseteq$.

Approximating Fixpoints

Looking ahead: Suppose we have the set $\mathbb{N}$ of natural numbers (the type is formally introduced later). The theorem `approx`

$$(\forall S.\ f(\bigcup S) = \bigcup (f \text{ ' } S)) \implies \bigcup_{n \in \mathbb{N}} (f^n\{\})) = \textit{lfp } f$$

shows a way of approximating $\textit{lfp}$, which is important for algorithmic solutions[452] (e.g. in program analysis).

There will be an exercise on this.

---

[452]The theorem

$$(\forall S.\ f(\bigcup S) = \bigcup (f \text{ ' } S)) \implies \bigcup_{n \in \mathbb{N}} (f^n\{\})) = \textit{lfp } f$$

says that under a certain condition, $\textit{lfp } f$ can be computed by applying f to the empty set over and over again:

- although the expression uses the union over all natural numbers, which is an infinite set, this can sometimes effectively be computed. Under certain conditions, there exists a k such that $f^k\ \{\} = f^{k+1}\{\}$.

- Even if $\bigcup n \in \mathbb{N}.f^n\ \{\}$ cannot be effectively computed, it can still be approximated: for any k, we know that $\bigcup i \leq k.f^i\ \{\} \subseteq \bigcup n \in \mathbb{N}.f^n\ \{\}$.

506

Where Are We Going? Induction and Recursion

Let's step back: What is an inductive definition of a set S?

Where Are We Going? Induction and Recursion

Let's step back: What is an inductive definition of a set S?
It has the form: S is the smallest set such that:

- $\emptyset \subseteq S$ (just mentioned for emphasis);

- if $S' \subseteq S$ then $F(S') \subseteq S$ (for some appropriate F).

Where Are We Going? Induction and Recursion

Let's step back: What is an inductive definition of a set S?
It has the form: S is the smallest set such that:

- $\emptyset \subseteq S$ (just mentioned for emphasis);

- if $S' \subseteq S$ then $F(S') \subseteq S$ (for some appropriate F).

At the same time, S is the smallest solution of the recursive equation $S = F(S)$.

Where Are We Going? Induction and Recursion

Let's step back: What is an inductive definition of a set S?
   It has the form: S is the smallest set such that:

- $\emptyset \subseteq S$ (just mentioned for emphasis);

- if $S' \subseteq S$ then $F(S') \subseteq S$ (for some appropriate F).

   At the same time, S is the smallest solution of the recursive equation $S = F(S)$.
   Induction and recursion are two faces of the same coin.

`Lfp.thy` **for Inductive Definitions**

Least fixpoints are for building inductive definitions of sets in a definitional way[453]: $S := lfp\ F$.

This is obviously well-defined, so why this fuss about monotonicity and Tarski?

---

[453]Recall why we were interested in fixpoints.

The problem with Y is that it leads to inconsistency (and of course, the definition of Y is not a constant definition/conservative extension.).

The definition of lfp is conservative.

And in appropriate situations, it can be used to define recursive functions.

Compared to Y, the type of lfp is restricted.

This restriction means that there is no obvious way to use lfp for defining recursive numeric functions such as fac.

`Lfp.thy` **for Inductive Definitions**

Least fixpoints are for building inductive definitions of sets in a definitional way[453]: $S := lfp\ F$.

This is obviously well-defined, so why this fuss about monotonicity and Tarski?

Tarski allows us to exploit the equation $lfp\ f = f(lfp\ f)$ in proofs about S! That's what lfp is all about.

---

[453]Recall why we were interested in fixpoints.

The problem with Y is that it leads to inconsistency (and of course, the definition of Y is not a constant definition/conservative extension.).

The definition of lfp is conservative.

And in appropriate situations, it can be used to define recursive functions.

Compared to Y, the type of lfp is restricted.

This restriction means that there is no obvious way to use lfp for defining recursive numeric functions such as fac.

Example **(Revisited)**

The set of all finite subsets of a set A:

$$Fin\ A = lfp\ F$$

where $F = \lambda X.\{\{\}\} \cup \bigcup x \in A.((\texttt{insert}\ x)\text{'}X).$

---

[454] This proof is of course done in Isabelle.

[455] This file should be contained in your Isabelle distribution. Or, if you only have an Isabelle executable, you can find the sources here:

http://isabelle.in.tum.de/library/

[456] Above, we defined the set of finite subsets of a set A. Alternatively, one could define "the set of all finite sets whose elements have type τ". In this case, no fixed set A is involved, and it is closer to what actually happens in Isabelle. In `Finite_Set.thy` a constant $Finites$ is defined. It has polymorphic type $\alpha\ set\ set$. We have $A \in Finites$ if and only if A is a finite set. However, it would be wrong to think of $Finites$ as one single set that contains all finite sets. Instead, for each τ, there is a polymorphic instance of $Finites$ of type $\tau\ set\ set$ containing all finite sets of element type τ.

In `Finite_Set.thy` we find the lines

Example **(Revisited)**

The set of all finite subsets of a set A:

$$Fin\ A = lfp\ F$$

where $F = \lambda X.\{\{\}\} \cup \bigcup x \in A.((\texttt{insert}\ x)\ {}^\backprime X)$.

Thus we can do using lfp what we would have wanted to do using Y.

To show: F is monotone[454]!

In the Isabelle library[455], this is done a bit differently[456].

There will be an exercise on this.

---

[454]This proof is of course done in Isabelle.

[455]This file should be contained in your Isabelle distribution. Or, if you only have an Isabelle executable, you can find the sources here:

> http://isabelle.in.tum.de/library/

[456]Above, we defined the set of finite subsets of a set A. Alternatively, one could define "the set of all finite sets whose elements have type τ". In this case, no fixed set A is involved, and it is closer to what actually happens in Isabelle. In `Finite_Set.thy` a constant $Finites$ is defined. It has polymorphic type $\alpha\ set\ set$. We have $A \in Finites$ if and only if A is a finite set. However, it would be wrong to think of $Finites$ as one single set that contains all finite sets. Instead, for each τ, there is a polymorphic instance of $Finites$ of type $\tau\ set\ set$ containing all finite sets of element type τ.

In `Finite_Set.thy` we find the lines

22.4 The Package for Inductive Sets

Since monotonicity proofs can be automated, Isabelle has special proof support for inductive definitions. Example:

```
consts Fin :: 'a set => 'a set set
inductive "Fin(A)"
  intrs
    emptyI  "{} : Fin(A)"
    insertI "[| a: A;  b: Fin(A) |] ==>
                     insert a b : Fin(A)"
```

Translated into expression using *lfp*.

```
inductive "Finites"
  intros
    emptyI [simp, intro!]: "{} : Finites"
    insertI [simp, intro!]: "A : Finites ==>
         insert a A : Finites"
```

The Isabelle mechanism of interpreting the keyword inductive translates this into the following definition: $Finites = lfp\ G$ where

$$G \equiv \lambda S.\ \{x \mid x = \{\} \vee (\exists A\ a.\ x = insert\ a\ A \wedge A \in S)\}$$

You can see this by typing in your proof script:

```
open Finites;
defs;
```

Talking (ML-)technically, Finites is a structure (module), and defs is a value (component) of this structure.

510

As a sanity-check, consider the type of this expression. The expression $insert\,a\,A$ forces A to be of type $\tau\,set$ for some τ and a to be of type τ. Next, $insert\,a\,A$ is of type $\tau\,set$, and hence x is also of type $\tau\,set$. Moreover, the expression $A \in S$ forces S to be of type $\tau\,set\,set$. The expression $\{x \mid x = \{\} \vee (\exists A\,a.\ x = insert\,a\,A \wedge A \in S)\}$ is of type $\tau\,set\,set$. Next, G is of type $\tau\,set\,set \to \tau\,set\,set$, and so finally, $Finites$ is of type $\tau\,set\,set$. But actually, since τ is arbitrary, we can replace it by a type variable α.

Note that there is a convenient syntactic translation

```
translations "finite A" == "A : Finites"
```

When does Isabelle generate ML-structures, and what are the names of those structures?

This question is highly Isabelle-technical, related to different formats used for writing theory files, which is in turn partly due to mere historic reasons.

It used to be the case that for a theory file called F.thy,

511

a structure F would be generated. Certain keywords in F.thy such as `inductive`, `recursive`, and `datatype`, would trigger the creation of substructures, so for example `inductive` I would call for the creation of a substructure I.

For a newer format of theory files, this is no longer the case.

The treatment of the keyword `constdefs`, followed by the declaration and definition of a constant C, also depends on the format used for writing theory files.

- Sometimes (when an older format is used), it will automatically generate a `thm` called $C\_\texttt{def}$ which is the definition of C.

- Sometimes (when a newer format is used), it will insert the definition of C into a database which can be accessed by a function called `thm` taking a string as argument. In this case, not $C\_\texttt{def}$ would be the definition of C, but rather

$$\texttt{thm } "C\_\texttt{def}"$$

Package relies on proven lemma[457] `lfp_unfold`.

You should be aware of such problems, but we do not treat them in this course.

[457]If you look around in the `ML`-files of the Isabelle/HOL library, you might not find any uses of `lfp_unfold`, so you may wonder: why is it important then? But you must bear in mind that the package for inductive sets relies on these lemmas.

This is a general insight about proven results in the library: Even though you might not find them being used in other `ML`-files, special packages of Isabelle/HOL might use those results.

Technical Support for Inductive Definitions

Support important in practice since many constructions are based on inductively defined sets (datatypes, ...). Support provided for:

- Automatic proof of monotonicity

- Automatic proof of induction rule, for example[458]:
  $$[\![xa \in Fin\ A;\ P\ \{\};\ \bigwedge a\,b.[\![a \in A;\ b \in Fin\ A;\ P\ b]\!] \implies P\,(insert\ a\,b)]\!] \implies P\,xa$$

---

[458] The theorem
$$[\![xa \in Fin\ A;\ P\ \{\};\ \bigwedge ab.[\![a \in A;\ b \in Fin\ A;\ P\ b]\!] \implies P\,(insert\ a\,b)]\!]$$
$$\implies P\,xa$$
is an instance of the general induction scheme. That is to say, if we take the general induction scheme `lfp_induct`
$$[\![a \in lfp\ f;\ mono\ f;\ \bigwedge x.x \in f(lfp\ f \cap \{x.P\,x\}) \implies P\,x]\!] \implies P\,a$$
and instantiate f to $\lambda X.\{\{\}\} \cup \bigcup x \in A.((\text{insert}\ x)\,{}^{\backprime}X)$ then some massaging using the definitions will give us the first theorem.

Note here that monotonicity has disappeared from the assumptions. This is because the monotonicity of F is shown by Isabelle once and for all. This is one aspect of what we mean by special proof support for inductive definitions.

The least fixpoint of the functional is $Fin\ A$ (the set of finite subsets of A) in this case.

This works also for mutually recursive[459] definitions, co-inductive[460] definitions, . . .

---

[459]Two functions f and g are mutually recursive if f is defined in terms of g and vice versa.

[460]Co-induction is a construction analogous to induction but using greatest fixpoints.

Example: Program Semantics

This is an Isabelle/HOL add-on: Can reason about programming languages in Isabelle/HOL.

In imperative programming, program statements such as assignments or if-statements transform a program state (which consists of the current values of the program variables) into a new state.

This transformation is defined inductively.

We will later study a small imperative programming language.

22.5 Summary on Least Fixpoints

We are interested in recursion because inductively defined sets and recursively defined functions are solutions to recursive equations.

We cannot have general fixpoint operator Y, but we have, by conservative extension, least fixpoints for defining sets.

There is an induction scheme (lfp induction) for proving theorems about an inductively defined set.

Restriction of F to set type ($\alpha\ set \rightarrow \alpha\ set$) means that least fixpoints are not generally suitable for defining functions . . .

23 Well-Founded Recursion

The Roadmap

We are still looking at how the different parts of mathematics are encoded in the Isabelle/HOL library.

- Orders

- Sets

- Functions

- (Least) fixpoints and induction

- (Well-founded) recursion

- Arithmetic

- Datatypes

The Roadmap

We are still looking at how the different parts of mathematics are encoded in the Isabelle/HOL library.

- Orders

- Sets

- Functions

- (Least) fixpoints and induction

- (Well-founded) recursion

- Arithmetic

- Datatypes

Well-Founded Recursion

After least fixpoints, well-founded recursion is our second concept of recursion (and fixpoint combinator).

Idea: Modeling "terminating" recursive functions, i.e. recursive definitions that use "smaller" arguments for the recursive call.

23.1 Prerequisite: Relations

We need some standard operations on binary relations (sets of pairs), such as converse, composition, image of a set and a relation, the identity relation, . . .

These are provided by `Relation.thy`[461].

---

[461] This file should be contained in your Isabelle distribution. Or, if you only have an Isabelle executable, you can find the sources here:

http://isabelle.in.tum.de/library/

Relation.thy (Fragment)

```
constdefs
  converse :: "('a * 'b) set => ('b * 'a) set"
  "r^-1 == {(y, x). (x, y):r}"
  rel_comp  :: "[('b * 'c) set, ('a * 'b) set] =>
                                    ('a * 'c) set"
  "r O s == {(x,z). EX y. (x, y):s & (y, z):r}"
  Image :: "[('a * 'b) set, 'a set] => 'b set"
  "r '' s == {y. EX x:s. (x,y):r}"
  Id      :: "('a * 'a) set"
  "Id == {p. EX x. p = (x,x)}"
```

Somewhat similar to Fun.thy.

23.2 Prerequisite: Closures

We need the transitive, as well as the reflexive transitive closure of a relation.

These are provided by `Transitive_Closure.thy`[462].

How would you define those inductively, ad-hoc?[463]

---

[462] This file should be contained in your Isabelle distribution. Or, if you only have an Isabelle executable, you can find the sources here:

> `http://isabelle.in.tum.de/library/`

[463]r^* is the smallest set such that:

- $Id \subseteq r^*$;

- if $r' \subseteq r^*$ then $r' \cup r \circ r' \subseteq r'$.

Or, in line with the schema for inductive definitions:

- $\emptyset \subseteq r^*$;

- if $r' \subseteq r^*$ then $(\lambda s.\, Id \cup (r \circ s))r' \subseteq r^*$.

The latter form corresponds to the definition in `Transitive_Closure.thy`.

The definition of r^+ is similar.

Transitive_Closure.thy (Fragment)

```
consts
 rtrancl :: "('a * 'a) set => ('a * 'a) set"
                              ("(_^*)" [1000] 999)
inductive "r^*"
 intros
  rtrancl_refl [...]: "(a, a) : r^*"
  rtrancl_into_rtrancl [...]: "(a, b) : r^* ==>
                (b, c) : r ==> (a, c) : r^*"
```

```
consts
 trancl :: "('a * 'a) set => ('a * 'a) set"
                            ("(_^+)" [1000] 999)
inductive "r^+"
 intros
  r_into_trancl [...]: "(a, b) : r ==>
                       (a, b) : r^+"
   trancl_into_trancl [...]: "(a, b) : r^+ ==>
               (b, c) : r ==> (a,c) : r^+"
```

23.3 Well-Founded Orderings

Defined in `Wellfounded_Recursion.thy`[464].

```
Wellfounded_Recursion = Transitive_Closure +
constdefs
   wf          :: "('a * 'a) set => bool"
   "wf(r) == (!P. (!x. (!y. (y,x):r --> P(y))
                   --> P(x)) --> (!x. P(x)))"
```

What does this mean?

23.3 Well-Founded Orderings

Defined in `Wellfounded_Recursion.thy`[464].

```
Wellfounded_Recursion = Transitive_Closure +
constdefs
  wf          :: "('a * 'a) set => bool"
  "wf(r) == (!P. (!x. (!y. (y,x):r --> P(y))
                   --> P(x)) --> (!x. P(x)))"
```

What does this mean? A relation r is well-founded if well-founded (Noetherian) induction based on r is a valid proof scheme[465].

---

[464] This file should be contained in your Isabelle distribution. Or, if you only have an Isabelle executable, you can find the sources here:

> http://isabelle.in.tum.de/library/

In older versions the file used to be called `WF.thy`.

[465] For a moment, forget everything you have ever heard about proofs using induction! The definition of wf has the form

$$wf(r) \equiv \forall P.\phi(r, P) \rightarrow \forall x.P(x)$$

That is, it says: a relation r is well-founded if a certain scheme ϕ can be used to show a property P that holds for all x.

By the fact that this is a constant definition (conservative extension), it is immediately clear that this gives us a correct method of proving $\forall x.P(x)$. To prove $\forall x.P(x)$ for some given P, find some r such that $\forall P.\phi(r, P) \rightarrow \forall x.P(x)$ holds, and show $\phi(r, P)$.

Once again, this method is correct regardless of what ϕ is. Forget about induction!

Example: Is $\emptyset$ well-founded[466]?

But how is that possible? How is it ensured that only true statements can be proven, if the method is correct for any old ϕ?

The point is this: The method is correct in principle, but it will typically not work unless ϕ is something sensible, e.g. an induction scheme as in the actual definition of wf. It will not work simply because we will fail to show either $\forall P.\phi(r, P) \rightarrow \forall x.P(x)$ or $\phi(r, P)$.

---

[466] The definition of wf is:

$$wf(r) \equiv (\forall P.(\forall x.(\forall y.(y, x) \in r \rightarrow P(y)) \rightarrow P(x)) \rightarrow (\forall x.P(x)))$$

Example: Is $\emptyset$ well-founded[466]?

But how is that possible? How is it ensured that only true statements can be proven, if the method is correct for any old ϕ?

The point is this: The method is correct in principle, but it will typically not work unless ϕ is something sensible, e.g. an induction scheme as in the actual definition of wf. It will not work simply because we will fail to show either $\forall P.\phi(r, P) \rightarrow \forall x.P(x)$ or $\phi(r, P)$.

---

[466] The definition of wf is:

Let's instantiate r to $\emptyset$.

$$wf(\emptyset) \equiv (\forall P.(\forall x.(\forall y.(y, x) \in \emptyset \rightarrow P(y)) \rightarrow P(x)) \rightarrow (\forall x.P(x)))$$

Example: Is $\emptyset$ well-founded[466]?

But how is that possible? How is it ensured that only true statements can be proven, if the method is correct for any old ϕ?

The point is this: The method is correct in principle, but it will typically not work unless ϕ is something sensible, e.g. an induction scheme as in the actual definition of wf. It will not work simply because we will fail to show either $\forall P. \phi(r, P) \to \forall x. P(x)$ or $\phi(r, P)$.

---

[466] The definition of wf is:

Let's instantiate r to $\emptyset$.

$$wf(\emptyset) \equiv (\forall P.(\forall x.(\forall y. False \qquad \to P(y)) \to P(x)) \to (\forall x. P(x)))$$

Example: Is $\emptyset$ well-founded[466]?

But how is that possible? How is it ensured that only true statements can be proven, if the method is correct for any old ϕ?

The point is this: The method is correct in principle, but it will typically not work unless ϕ is something sensible, e.g. an induction scheme as in the actual definition of wf. It will not work simply because we will fail to show either $\forall P.\phi(r, P) \rightarrow \forall x.P(x)$ or $\phi(r, P)$.

---

[466] The definition of wf is:

Let's instantiate r to $\emptyset$.

$$wf(\emptyset) \equiv (\forall P.(\forall x.(\forall y.\mathit{True} \qquad) \rightarrow P(x)) \rightarrow (\forall x.P(x)))$$

Example: Is $\emptyset$ well-founded[466]?

But how is that possible? How is it ensured that only true statements can be proven, if the method is correct for any old ϕ?

The point is this: The method is correct in principle, but it will typically not work unless ϕ is something sensible, e.g. an induction scheme as in the actual definition of wf. It will not work simply because we will fail to show either $\forall P.\phi(r, P) \rightarrow \forall x.P(x)$ or $\phi(r, P)$.

---

[466] The definition of wf is:

Let's instantiate r to $\emptyset$.

$$wf(\emptyset) \equiv (\forall P.(\forall x. \quad True \quad \rightarrow P(x)) \rightarrow (\forall x.P(x)))$$

Example: Is $\emptyset$ well-founded[466]?

But how is that possible? How is it ensured that only true statements can be proven, if the method is correct for any old ϕ?

The point is this: The method is correct in principle, but it will typically not work unless ϕ is something sensible, e.g. an induction scheme as in the actual definition of wf. It will not work simply because we will fail to show either $\forall P.\phi(r, P) \rightarrow \forall x.P(x)$ or $\phi(r, P)$.

---

[466]The definition of wf is:

Let's instantiate r to $\emptyset$.

$$wf(\emptyset) \equiv (\forall P.(\forall x. \qquad\qquad\qquad\qquad\qquad P(x)) \rightarrow (\forall x.P(x)))$$

Example: Is $\emptyset$ well-founded[466]?

But how is that possible? How is it ensured that only true statements can be proven, if the method is correct for any old ϕ?

The point is this: The method is correct in principle, but it will typically not work unless ϕ is something sensible, e.g. an induction scheme as in the actual definition of wf. It will not work simply because we will fail to show either $\forall P.\phi(r, P) \rightarrow \forall x.P(x)$ or $\phi(r, P)$.

[466] The definition of wf is:

Let's instantiate r to $\emptyset$.

$$wf(\emptyset) \equiv (\forall P.\,True \qquad\qquad\qquad\qquad)$$

Example: Is $\emptyset$ well-founded[466]?

But how is that possible? How is it ensured that only true statements can be proven, if the method is correct for any old ϕ?

The point is this: The method is correct in principle, but it will typically not work unless ϕ is something sensible, e.g. an induction scheme as in the actual definition of wf. It will not work simply because we will fail to show either $\forall P.\phi(r, P) \rightarrow \forall x.P(x)$ or $\phi(r, P)$.

[466]The definition of wf is:

Let's instantiate r to $\emptyset$.

$$wf(\emptyset) \equiv \quad True$$

So the empty set is well-founded.

Intuition of Well-Foundedness

Intuition of wf: All descending chains are finite.

But: Cannot express infinity; must look for alternatives[467].

- Not symmetric:

---

[467]We will now try some ideas, work out their formalization as a formula, and then illustrate why the condition is either too weak or too strong, using an example. Finally, we will give the correct condition.

[468]In this attempt, we formalized the "minimal element in p" as an x such that there is no y with $(x, y) \in p$. But this is a bad formalization since an isolated element, i.e., one that is completely unrelated to p, or even to r, would meet the definition.

In fact, this problem was already present for the previous attempt where we just required $\exists x. \forall y. (y, x) \notin r$ (i.e., r has a minimal element).

Intuition of Well-Foundedness

Intuition of wf: All descending chains are finite.

But: Cannot express infinity; must look for alternatives[467].

- Not symmetric: $(x, y) \in r \rightarrow (y, x) \notin r$?

---

[467]We will now try some ideas, work out their formalization as a formula, and then illustrate why the condition is either too weak or too strong, using an example. Finally, we will give the correct condition.

[468]In this attempt, we formalized the "minimal element in p" as an x such that there is no y with $(x, y) \in p$. But this is a bad formalization since an isolated element, i.e., one that is completely unrelated to p, or even to r, would meet the definition.

In fact, this problem was already present for the previous attempt where we just required $\exists x. \forall y. (y, x) \notin r$ (i.e., r has a minimal element).

Intuition of Well-Foundedness

Intuition of wf: All descending chains are finite.

  But: Cannot express infinity; must look for alternatives[467].

  • Not symmetric: $(x, y) \in r \rightarrow (y, x) \notin r$?

---

[467]We will now try some ideas, work out their formalization as a formula, and then illustrate why the condition is either too weak or too strong, using an example. Finally, we will give the correct condition.

[468]In this attempt, we formalized the "minimal element in p" as an x such that there is no y with $(x, y) \in p$. But this is a bad formalization since an isolated element, i.e., one that is completely unrelated to p, or even to r, would meet the definition.

  In fact, this problem was already present for the previous attempt where we just required $\exists x. \forall y. (y, x) \notin r$ (i.e., r has a minimal element).

Intuition of Well-Foundedness

Intuition of wf: All descending chains are finite.

    But: Cannot express infinity; must look for alternatives[467].

- Not symmetric: $(x, y) \in r \rightarrow (y, x) \notin r$?

- No cycles:

---

[467]We will now try some ideas, work out their formalization as a formula, and then illustrate why the condition is either too weak or too strong, using an example. Finally, we will give the correct condition.

[468]In this attempt, we formalized the "minimal element in p" as an x such that there is no y with $(x, y) \in p$. But this is a bad formalization since an isolated element, i.e., one that is completely unrelated to p, or even to r, would meet the definition.

In fact, this problem was already present for the previous attempt where we just required $\exists x. \forall y. (y, x) \notin r$ (i.e., r has a minimal element).

Intuition of Well-Foundedness

Intuition of wf: All descending chains are finite.

But: Cannot express infinity; must look for alternatives[467].

- Not symmetric: $(x, y) \in r \rightarrow (y, x) \notin r$?

- No cycles: $(x, x) \notin r^+$?

---

[467]We will now try some ideas, work out their formalization as a formula, and then illustrate why the condition is either too weak or too strong, using an example. Finally, we will give the correct condition.

[468]In this attempt, we formalized the "minimal element in p" as an x such that there is no y with $(x, y) \in p$. But this is a bad formalization since an isolated element, i.e., one that is completely unrelated to p, or even to r, would meet the definition.
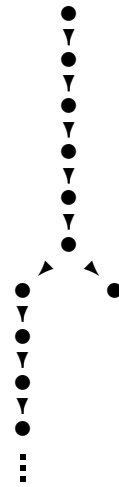
In fact, this problem was already present for the previous attempt where we just required $\exists x. \forall y. (y, x) \notin r$ (i.e., r has a minimal element).

Intuition of Well-Foundedness

Intuition of wf: All descending chains are finite.

But: Cannot express infinity; must look for alternatives[467].

- Not symmetric: $(x, y) \in r \rightarrow (y, x) \notin r$?

- No cycles: $(x, x) \notin r^+$?

---

[467]We will now try some ideas, work out their formalization as a formula, and then illustrate why the condition is either too weak or too strong, using an example. Finally, we will give the correct condition.

[468]In this attempt, we formalized the "minimal element in p" as an x such that there is no y with $(x, y) \in p$. But this is a bad formalization since an isolated element, i.e., one that is completely unrelated to p, or even to r, would meet the definition.

In fact, this problem was already present for the previous attempt where we just required $\exists x. \forall y. (y, x) \notin r$ (i.e., r has a minimal element).

Intuition of Well-Foundedness

Intuition of wf: All descending chains are finite.

But: Cannot express infinity; must look for alternatives[467].

- Not symmetric: $(x, y) \in r \rightarrow (y, x) \notin r$?

- No cycles: $(x, x) \notin r^+$?

- r has minimal element:

---

[467]We will now try some ideas, work out their formalization as a formula, and then illustrate why the condition is either too weak or too strong, using an example. Finally, we will give the correct condition.

[468]In this attempt, we formalized the "minimal element in p" as an x such that there is no y with $(x, y) \in p$. But this is a bad formalization since an isolated element, i.e., one that is completely unrelated to p, or even to r, would meet the definition.

In fact, this problem was already present for the previous attempt where we just required $\exists x. \forall y. (y, x) \notin r$ (i.e., r has a minimal element).

Intuition of Well-Foundedness

Intuition of wf: All descending chains are finite.

But: Cannot express infinity; must look for alternatives[467].

- Not symmetric: $(x, y) \in r \rightarrow (y, x) \notin r$?

- No cycles: $(x, x) \notin r^+$?

- r has minimal element: $\exists x. \forall y. (y, x) \notin r$?
  Note: Trivial for $r = \emptyset$.

---

[467]We will now try some ideas, work out their formalization as a formula, and then illustrate why the condition is either too weak or too strong, using an example. Finally, we will give the correct condition.

[468]In this attempt, we formalized the "minimal element in p" as an x such that there is no y with $(x, y) \in p$. But this is a bad formalization since an isolated element, i.e., one that is completely unrelated to p, or even to r, would meet the definition.

In fact, this problem was already present for the previous attempt where we just required $\exists x. \forall y. (y, x) \notin r$ (i.e., r has a minimal element).

Intuition of Well-Foundedness

Intuition of wf: All descending chains are finite.

   But: Cannot express infinity; must look for alternatives[467].

- Not symmetric: $(x, y) \in r \rightarrow (y, x) \notin r$?

- No cycles: $(x, x) \notin r^+$?

- r has minimal element: $\exists x. \forall y. (y, x) \notin r$?
   Note: Trivial for $r = \emptyset$.

---

[467]We will now try some ideas, work out their formalization as a formula, and then illustrate why the condition is either too weak or too strong, using an example. Finally, we will give the correct condition.

[468]In this attempt, we formalized the "minimal element in p" as an x such that there is no y with $(x, y) \in p$. But this is a bad formalization since an isolated element, i.e., one that is completely unrelated to p, or even to r, would meet the definition.
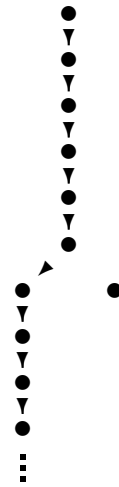
   In fact, this problem was already present for the previous attempt where we just required $\exists x. \forall y. (y, x) \notin r$ (i.e., r has a minimal element).

Intuition of Well-Foundedness

Intuition of wf: All descending chains are finite.

But: Cannot express infinity; must look for alternatives[467].

- Not symmetric: $(x, y) \in r \rightarrow (y, x) \notin r$?

- No cycles: $(x, x) \notin r^+$?

- r has minimal element: $\exists x. \forall y. (y, x) \notin r$?
  Note: Trivial for $r = \emptyset$.

- Any subrelation must have minimal element:

---

[467]We will now try some ideas, work out their formalization as a formula, and then illustrate why the condition is either too weak or too strong, using an example. Finally, we will give the correct condition.

[468]In this attempt, we formalized the "minimal element in p" as an x such that there is no y with $(x, y) \in p$. But this is a bad formalization since an isolated element, i.e., one that is completely unrelated to p, or even to r, would meet the definition.

In fact, this problem was already present for the previous attempt where we just required $\exists x. \forall y. (y, x) \notin r$ (i.e., r has a minimal element).

Intuition of Well-Foundedness

Intuition of wf: All descending chains are finite.

    But: Cannot express infinity; must look for alternatives[467].

- Not symmetric: $(x, y) \in r \rightarrow (y, x) \notin r$?

- No cycles: $(x, x) \notin r^+$?

- r has minimal element: $\exists x. \forall y. (y, x) \notin r$?
  Note: Trivial for $r = \emptyset$.

- Any subrelation must have minimal element:
  $\forall p. p \subseteq r \rightarrow \exists x. \forall y. (y, x) \notin p$?

---

[467]We will now try some ideas, work out their formalization as a formula, and then illustrate why the condition is either too weak or too strong, using an example. Finally, we will give the correct condition.

[468]In this attempt, we formalized the "minimal element in p" as an x such that there is no y with $(x, y) \in p$. But this is a bad formalization since an isolated element, i.e., one that is completely unrelated to p, or even to r, would meet the definition.

In fact, this problem was already present for the previous attempt where we just required $\exists x. \forall y. (y, x) \notin r$ (i.e., r has a minimal element).

Intuition of Well-Foundedness

Intuition of wf: All descending chains are finite.
  But: Cannot express infinity; must look for alternatives[467].

- Not symmetric: $(x, y) \in r \to (y, x) \notin r$?

- No cycles: $(x, x) \notin r^+$?

- r has minimal element: $\exists x. \forall y. (y, x) \notin r$?
  Note: Trivial for $r = \emptyset$.

- Any subrelation must have minimal element:
  $\forall p. p \subseteq r \to \exists x. \forall y. (y, x) \notin p$? "Minimal element" badly formalized[468] (already in previous point).

---

[467]We will now try some ideas, work out their formalization as a formula, and then illustrate why the condition is either too weak or too strong, using an example. Finally, we will give the correct condition.

[468]In this attempt, we formalized the "minimal element in p" as an x such that there is no y with $(x, y) \in p$. But this is a bad formalization since an isolated element, i.e., one that is completely unrelated to p, or even to r, would meet the definition.

In fact, this problem was already present for the previous attempt where we just required $\exists x. \forall y. (y, x) \notin r$ (i.e., r has a minimal element).

The Characterization

All these attempts are just necessary but not sufficient conditions for well-foundedness.

Here is a characterization[469]:

$$(\forall Q\, x.\ x \in Q \rightarrow (\exists z \in Q.\forall y.(y, z) \in r \rightarrow y \notin Q))$$

Here is an alternative characterization (exercise):

$$(\forall r.r \neq \{\} \wedge r \subseteq p \rightarrow (\exists x \in Domain\ r.\forall y.(y, x) \notin r))$$

Let's see some theorems to confirm our intuition, including the statements just shown.

---

[469] The final condition

$$(\forall Q\, .\ x \in Q \rightarrow (\exists z \in Q.\forall y.(y, z) \in r \rightarrow y \notin Q))$$

expresses the absence of infinite descending chains without explicitly using the concept of infinity.

It is a characterization of well-foundedness. One could say that the above formula expresses what well-foundedness is, while the "official" definition is somewhat indirect since it defines well-foundedness by what one can do with it.

A Theorem[470] on the Empty Set

   wf_empty $wf\,\{\}$

   Proof sketch: wf_empty: substitute r into definition, simplify.

---

[470]The   theorems   we   present   here   are   proven   in Wellfounded_Recursion.ML.

  This file should be contained in your Isabelle distribution. Or, if you only have an Isabelle executable, you can find the sources here:

   http://isabelle.in.tum.de/library/

but in older versions the file used to be called WF.ML

A Theorem for Induction

By massage[471] of the definition of well-foundedness

$$\forall P.(\forall x.(\forall y.(y,x) \in r \rightarrow P\,y) \rightarrow P\,x) \rightarrow (\forall x.P\,x)$$

one obtains the theorem `wf_induct`

$$[\![wf\ r;\ \bigwedge x.\forall y.(y,x) \in r \rightarrow P\,y \Longrightarrow P\,x]\!] \Longrightarrow P\,a.$$

This is a form suitable for doing induction proofs in Isabelle.

---

[471]As far as the induction principle is concerned, `induct_wf` states the same as the very definition of wf. All that happens is that some explicit universal object-level quantifiers are removed and the according variables are (implicitly) universally quantified on the meta-level, and some shifting from object-level implications to meta-level implications using `mp`. This is why we dare say "logical massage". See `Wellfounded_Recursion.ML`.

Induction Theorem as Proof Rule

The Isabelle theorem `wf_induct`

$$\llbracket \mathit{wf}\ r;\ \bigwedge x.\forall y.(y,x) \in r \to P\,y \Longrightarrow P\,x \rrbracket \Longrightarrow P\,a.$$

as proof rule:

$$\cfrac{\mathit{wf}\ r \qquad \begin{array}{c} [\forall y.(y,x) \in r \to P\,y] \\ \vdots \\ P\,x \end{array}}{P\,a}\ \texttt{wf\_induct}$$

A Theorem on Antisymmetry

`wf_not_sym` $[\![wf\ r; (a,x) \in r]\!] \Longrightarrow (x,a) \notin r$

Proof sketch:

$$[\forall y.(y,x) \in r \rightarrow (\forall z.(y,z) \in r \rightarrow (z,y) \notin r)]$$
$$\vdots$$
$$\frac{wf\ r \qquad\qquad \forall z.(x,z) \in r \rightarrow (z,x) \notin r}{\forall z.(a,z) \in r \rightarrow (z,a) \notin r}\ \texttt{wf\_induct}$$

The induction part needs classical reasoning. We will now give it at a level of detail that shows the essential reasoning but hides all the swapping involved in the Isabelle proof.

A variation will be done as exercise.

The Induction Part in More Detail

$$\dfrac{\dfrac{\forall y.(y,x) \in r \to (\forall z.(y,z) \in r \to (z,y) \notin r)}{(w,x) \in r \to (\forall z.(w,z) \in r \to (z,w) \notin r)}\ \forall\text{-}E}{(w,x) \notin r \vee (\forall z.(w,z) \in r \to (z,w) \notin r)}\ (\text{c})^{472} \equiv \phi$$

"(c)" stands for classical reasoning steps.

$$\dfrac{\phi \quad \dfrac{[(w,x) \notin r]^1}{(x,w) \in r \to (w,x) \notin r}\ impI^2 \quad \dfrac{\dfrac{\dfrac{[\forall z.(w,z) \in r \to (z,w) \notin r]^1}{\forall z.(z,w) \in r \to (w,z) \notin r}\ (\text{c})^{473}}{(x,w) \in r \to (w,x) \notin r}\ \forall\text{-}E}{}\ disjE^1}{\dfrac{(x,w) \in r \to (w,x) \notin r}{\forall z.(x,z) \in r \to (z,x) \notin r}\ \forall\text{-}I}$$

Theorems on Absence of Cycles

`wf_not_refl`	$wf\ r \implies (a, a) \notin r$
`wf_trancl`	$wf\ r \implies wf(r^+)$
`wf_acyclic`	$wf\ r \implies acyclic\ r$

$(acyclic\ r \equiv \forall x.(x, x) \notin r^+)$

	`wf_not_refl`:	Corollary of `wf_not_sym`.
Proof sketch:	`wf_trancl`:	Uses induction.
	`wf_acyclic`:	Apply `wf_not_refl` and `wf_trancl`

Ergo: Definition of wf really meets our intuition of "no cycles".

Another **Theorem** ("Exists Minimal Element")

`wf_minimal` $wf\ r \implies \exists x.\forall y.(y,x) \notin r^+$

  Proof sketch, writing $\phi \equiv (\exists x.\forall y.(y,x) \notin r^+)$:

$\texttt{wf}(r)$

$$\frac{\rule{0pt}{0pt}}{\phi} \hspace{2cm} \texttt{wf\_minimal}$$

  This is what we must construct.

---

[474]In the proof of $\exists x.\forall y.(y,x) \notin r^+$ we had the sub-proof

$$\frac{\neg\phi \quad \forall w.(w,v) \in r^+ \to \phi}{\neg\exists w.(w,v) \in r^+}$$

  This sub-proof does not actually depend on ϕ, it would hold
no matter what ϕ is (unlike the entire proof)
  In detail, the sub-proof looks as follows:

$$\frac{\forall w.(w,v) \in r^+ \to \phi}{(w,v) \in r^+ \to \phi}\ \textit{spec}$$

Another Theorem ("Exists Minimal Element")

`wf_minimal` $wf\ r \implies \exists x. \forall y.(y,x) \notin r^+$

Proof sketch, writing $\phi \equiv (\exists x.\forall y.(y,x) \notin r^+)$:

$$\begin{array}{c} \forall w.(w,v) \\ \in r^+ \to \phi \end{array}$$

$$\dfrac{\texttt{wf}(r)}{\texttt{wf}(r^+)} \qquad\qquad \phi$$

$$\dfrac{\texttt{wf}(r^+) \qquad\qquad\qquad\qquad \phi}{\phi} \ \texttt{wf\_induct}$$

Note "special case": w and v do not occur in ϕ!

---

[474]In the proof of $\exists x.\forall y.(y,x) \notin r^+$ we had the sub-proof

$$\dfrac{\neg\phi \quad \forall w.(w,v) \in r^+ \to \phi}{\neg\exists w.(w,v) \in r^+}$$

This sub-proof does not actually depend on ϕ, it would hold no matter what ϕ is (unlike the entire proof)

In detail, the sub-proof looks as follows:

$$(w,v) \in r^+ \qquad \dfrac{\forall w.(w,v) \in r^+ \to \phi}{(w,v) \in r^+ \to \phi} \ spec$$

Another Theorem ("Exists Minimal Element")

`wf_minimal` $wf\ r \implies \exists x.\forall y.(y,x) \notin r^+$

Proof sketch, writing $\phi \equiv (\exists x.\forall y.(y,x) \notin r^+)$:

$$\forall w.(w,v) \in r^+ \to \phi$$

$$\frac{\dfrac{\texttt{wf}(r)}{\texttt{wf}(r^+)} \quad \bullet \qquad \phi}{\phi} \quad \textit{wf\_induct}$$

This is `wf_trancl`.

---

[474]In the proof of $\exists x.\forall y.(y,x) \notin r^+$ we had the sub-proof

$$\frac{\neg\phi \quad \forall w.(w,v) \in r^+ \to \phi}{\neg\exists w.(w,v) \in r^+}$$

This sub-proof does not actually depend on ϕ, it would hold no matter what ϕ is (unlike the entire proof)

In detail, the sub-proof looks as follows:

$$\frac{(w,v) \in r^+ \qquad \dfrac{\forall w.(w,v) \in r^+ \to \phi}{(w,v) \in r^+ \to \phi}\ \textit{spec}}{\phi}\ \textit{mp}$$

Another Theorem ("Exists Minimal Element")

`wf_minimal` $wf\ r \Longrightarrow \exists x.\forall y.(y,x) \notin r^+$

Proof sketch, writing $\phi \equiv (\exists x.\forall y.(y,x) \notin r^+)$:

$$
\cfrac{
  \cfrac{\texttt{wf}(r)}{\texttt{wf}(r^+)} \quad\bullet\quad
  \cfrac{
    \phi \vee \neg\phi \qquad \phi \qquad
    \cfrac{
      \cfrac{}{\neg\phi} \qquad
      \cfrac{}{\substack{\forall w.(w,v) \\ \in r^+ \to \phi}}
    }{\phi}
  }{\phi}\ disjE
}{\phi}\ \texttt{wf\_induct}
$$

We now try a proof by case distinction on ϕ.

---

[474]In the proof of $\exists x.\forall y.(y,x) \notin r^+$ we had the sub-proof

$$\cfrac{\neg\phi \quad \forall w.(w,v) \in r^+ \to \phi}{\neg\exists w.(w,v) \in r^+}$$

This sub-proof does not actually depend on ϕ, it would hold no matter what ϕ is (unlike the entire proof)

In detail, the sub-proof looks as follows:

$$
\cfrac{
  \exists w.(w,v) \in r^+ \qquad
  \cfrac{
    (w,v) \in r^+ \qquad
    \cfrac{\forall w.(w,v) \in r^+ \to \phi}{(w,v) \in r^+ \to \phi}\ spec
  }{\phi}\ mp
}{}
$$

Another Theorem ("Exists Minimal Element")

`wf_minimal` $wf\ r \implies \exists x.\forall y.(y,x) \notin r^+$

Proof sketch, writing $\phi \equiv (\exists x.\forall y.(y,x) \notin r^+)$:

$$
\cfrac{
  \cfrac{\mathtt{wf}(r)}{\mathtt{wf}(r^+)}
  \quad\bullet\quad
  \cfrac{
    \cfrac{\rule{2em}{0.4pt}\;\bullet}{\phi \vee \neg\phi}
    \qquad
    \cfrac{\neg\phi}{\phi}
    \qquad
    \cfrac{\neg\phi \qquad \cfrac{}{\forall w.(w,v) \in r^+ \to \phi}}{\phi}
  }{\phi}\ \textit{disjE}
}{\phi}\ \mathtt{wf\_induct}
$$

Classical reasoning.

---

[474]In the proof of $\exists x.\forall y.(y,x) \notin r^+$ we had the sub-proof

$$
\cfrac{\neg\phi \quad \forall w.(w,v) \in r^+ \to \phi}{\neg\exists w.(w,v) \in r^+}
$$

This sub-proof does not actually depend on ϕ, it would hold no matter what ϕ is (unlike the entire proof)

In detail, the sub-proof looks as follows:

$$
\cfrac{
  \exists w.(w,v) \in r^+
  \qquad
  \cfrac{
    [(w,v) \in r^+]^4
    \qquad
    \cfrac{\forall w.(w,v) \in r^+ \to \phi}{(w,v) \in r^+ \to \phi}\ \textit{spec}
  }{\phi}\ \textit{mp}
}{\phi}\ \textit{existsE}^4
$$

Another Theorem ("Exists Minimal Element")

`wf_minimal` $wf\ r \implies \exists x.\forall y.(y,x) \notin r^+$

Proof sketch, writing $\phi \equiv (\exists x.\forall y.(y,x) \notin r^+)$:

$$\cfrac{\cfrac{\cfrac{\dfrac{\overline{\hspace{1.5cm}}\;\bullet}{\texttt{wf}(r)}}{\texttt{wf}(r^+)} \;\bullet \quad \cfrac{\phi \vee \neg\phi \quad \cfrac{\cfrac{\neg\phi}{\forall x.\exists y.(y,x) \in r^+}\;\cdots \quad \cfrac{\neg\phi \quad \begin{array}{c}\forall w.(w,v)\\ \in r^+ \to \phi\end{array}}{\phi}}{\phi} \quad \phi}{\phi}\; disjE}{\phi}}{\phi}\; \texttt{wf\_induct}$$

Using some elementary equivalences[475].

---

[474]In the proof of $\exists x.\forall y.(y,x) \notin r^+$ we had the sub-proof

$$\cfrac{\neg\phi \quad \forall w.(w,v) \in r^+ \to \phi}{\neg\exists w.(w,v) \in r^+}$$

This sub-proof does not actually depend on ϕ, it would hold no matter what ϕ is (unlike the entire proof)

In detail, the sub-proof looks as follows:

$$\cfrac{\neg\phi \quad \cfrac{\exists w.(w,v) \in r^+ \quad \cfrac{[(w,v) \in r^+]^4 \quad \cfrac{\forall w.(w,v) \in r^+ \to \phi}{(w,v) \in r^+ \to \phi}\; spec}{\phi}\; mp}{\phi}\; existsE^4}{\phi}$$

Another Theorem ("Exists Minimal Element")

`wf_minimal` $wf\ r \implies \exists x.\forall y.(y,x) \notin r^+$

Proof sketch, writing $\phi \equiv (\exists x.\forall y.(y,x) \notin r^+)$:

$$\dfrac{\neg\phi}{\forall x.\exists y.(y,x) \in r^+} \quad \cdots \quad \dfrac{\neg\phi \quad \dfrac{\forall w.(w,v)}{\in r^+ \to \phi}}{\neg\exists w.(w,v) \in r^+}\ \bullet^{474}$$

$$\dfrac{\mathtt{wf}(r)}{\mathtt{wf}(r^+)}\ \bullet \quad \dfrac{\overline{}\ \bullet \quad \dfrac{\phi \vee \neg\phi \qquad \phi \qquad \qquad \qquad \qquad \phi}{\phi}\ disjE}{\phi}\ \mathtt{wf\_induct}$$

This subproof works for any ϕ. Think semantically or check (5 rule applications)!

---

[474]In the proof of $\exists x.\forall y.(y,x) \notin r^+$ we had the sub-proof

$$\dfrac{\neg\phi \quad \forall w.(w,v) \in r^+ \to \phi}{\neg\exists w.(w,v) \in r^+}$$

This sub-proof does not actually depend on ϕ, it would hold no matter what ϕ is (unlike the entire proof)

In detail, the sub-proof looks as follows:

$$\dfrac{\neg\phi \quad \dfrac{\exists w.(w,v) \in r^+ \qquad \dfrac{[(w,v) \in r^+]^4 \quad \dfrac{\forall w.(w,v) \in r^+ \to \phi}{(w,v) \in r^+ \to \phi}\ spec}{\phi}\ mp}{\phi}\ existsE^4}{False}\ notE$$

535

Another Theorem ("Exists Minimal Element")

`wf_minimal` $wf\ r \implies \exists x.\forall y.(y,x) \notin r^+$

Proof sketch, writing $\phi \equiv (\exists x.\forall y.(y,x) \notin r^+)$:

$$
\cfrac{
  \cfrac{\mathtt{wf}(r)}{\mathtt{wf}(r^+)}\ \bullet \quad
  \cfrac{
    \cfrac{}{\phi \vee \neg\phi}\ \bullet \quad \phi \quad
    \cfrac{
      \cfrac{\neg\phi}{\forall x.\exists y.(y,x) \in r^+}\ \cdots \quad
      \cfrac{\neg\phi \quad \cfrac{\forall w.(w,v) \in r^+ \to \phi}{\neg\exists w.(w,v) \in r^+}}{}\ \bullet^{474} \cdots
    }{\cfrac{False}{\phi}\ FalseE}
  }{\phi}\ disjE
}{\phi}\ \mathtt{wf\_induct}
$$

It is routine to derive $False$.

---

[474] In the proof of $\exists x.\forall y.(y,x) \notin r^+$ we had the sub-proof

$$\cfrac{\neg\phi \quad \forall w.(w,v) \in r^+ \to \phi}{\neg\exists w.(w,v) \in r^+}$$

This sub-proof does not actually depend on ϕ, it would hold no matter what ϕ is (unlike the entire proof)

In detail, the sub-proof looks as follows:

$$
\cfrac{
  \neg\phi \quad
  \cfrac{
    [\exists w.(w,v) \in r^+]^3 \quad
    \cfrac{
      [(w,v) \in r^+]^4 \quad
      \cfrac{
        \cfrac{\forall w.(w,v) \in r^+ \to \phi}{(w,v) \in r^+ \to \phi}\ spec
      }{\phi}\ mp
    }{\phi}\ existsE^4
  }{\phi}\ notE
}{
  \cfrac{False}{\neg\exists w.(w,v) \in r^+}\ notI^3
}
$$

Another Theorem ("Exists Minimal Element")

`wf_minimal` $wf\ r \Longrightarrow \exists x.\forall y.(y,x) \notin r^+$

Proof sketch, writing $\phi \equiv (\exists x.\forall y.(y,x) \notin r^+)$:

$$
\cfrac{
  \cfrac{\mathtt{wf}(r)}{\mathtt{wf}(r^+)}\ \bullet
  \quad
  \cfrac{
    \phi \vee \neg\phi
    \quad
    [\phi]^2
    \quad
    \cfrac{
      \cfrac{[\neg\phi]^2}{\forall x.\exists y.(y,x) \in r^+}\ \cdots
      \quad
      \cfrac{[\neg\phi]^2 \quad \cfrac{\forall w.(w,v)}{\in r^+ \to \phi}}{\neg\exists w.(w,v) \in r^+}\ \bullet 474\ \cdots
    }{
      \cfrac{\cfrac{False}{\phi}\ FalseE}{ }
    }
  }{\phi}\ disjE^2
}{\phi}\ \mathtt{wf\_induct}
$$

This completes the proof by case distinction . . .

---

[474] In the proof of $\exists x.\forall y.(y,x) \notin r^+$ we had the sub-proof

$$\cfrac{\neg\phi \quad \forall w.(w,v) \in r^+ \to \phi}{\neg\exists w.(w,v) \in r^+}$$

This sub-proof does not actually depend on ϕ, it would hold no matter what ϕ is (unlike the entire proof)

In detail, the sub-proof looks as follows:

$$(w,v) \in r^+$$

$$\exists w.(w,v) \in r^+$$

$$\neg\phi$$

Another Theorem ("Exists Minimal Element")

`wf_minimal` $\quad wf\ r \implies \exists x.\forall y.(y,x) \notin r^+$

Proof sketch, writing $\phi \equiv (\exists x.\forall y.(y,x) \notin r^+)$:

$$
\cfrac{\cfrac{\dfrac{[\neg\phi]^2}{\forall x.\exists y.(y,x)\in r^+}\ \cdots\ \dfrac{[\neg\phi]^2 \quad [\ \genfrac{}{}{0pt}{}{\forall w.(w,v)}{\in r^+ \to \phi}\ ]^1}{\neg\exists w.(w,v)\in r^+}\ \bullet^{474}\ \cdots}{\dfrac{False}{\phi}\ \textit{FalseE}}}{}
$$

$$
\cfrac{\dfrac{\mathtt{wf}(r)}{\mathtt{wf}(r^+)}\ \bullet \quad \dfrac{\phi\vee\neg\phi \quad [\phi]^2 \quad \quad \quad \phi}{\phi}\ \textit{disjE}^2}{\phi}\ \mathtt{wf\_induct}^1
$$

...and the proof by induction.

---

[474]In the proof of $\exists x.\forall y.(y,x)\notin r^+$ we had the sub-proof

$$
\dfrac{\neg\phi \quad \forall w.(w,v)\in r^+ \to \phi}{\neg\exists w.(w,v)\in r^+}
$$

This sub-proof does not actually depend on ϕ, it would hold no matter what ϕ is (unlike the entire proof)

In detail, the sub-proof looks as follows:

$$
(w,v)\in r^+
$$

$$
\exists w.(w,v)\in r^+
$$

$$
\neg\phi
$$

Remarks on the Proof

We used an instance of `wf_induct`, where we instantiated x by v, y by w, and P by $\lambda w.(\exists x.\forall y.(y,x) \notin r^+)$. I.e., ϕ does not contain the "induction variables" w and v.

Still this is a "proper" induction proof: Although ϕ does not contain the "induction variables", the proof does depend on the actual form of ϕ! (Try doing it without induction ...)

Scoping of quantifiers (e.g., in general $(\forall w.(w,v) \in r^+ \rightarrow \phi) \not\equiv (\forall w.(w,v) \in r^+) \rightarrow \phi$) and side conditions are very subtle in this proof. Underlines the importance of machine-checked proofs.

Remarks on `wf_minimal`

Ergo: Definition of `wf` fulfills the condition corresponding to our first attempt of characterizing well-foundedness using minimal elements.

However, this formalization had a problem: there could be local minima, and
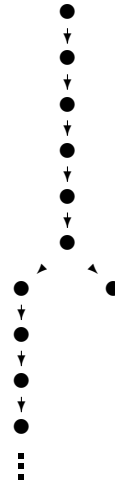
Remarks on `wf_minimal`

Ergo: Definition of `wf` fulfills the condition corresponding to our first attempt of characterizing well-foundedness using minimal elements.

However, this formalization had a problem: there could be local minima, and isolated points are also always minima. In particular, if r is empty, then any element is trivially a minimum.
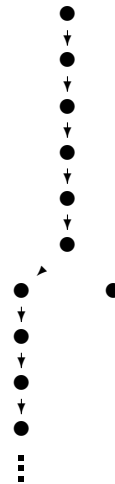
A Characterization

The theorem `wf_eq_minimal` is characterization of well-foundedness.:

$$wf\ r = (\forall Q\ .\ x \in Q \rightarrow (\exists z \in Q.\forall y.(y,z) \in r \rightarrow y \notin Q))$$

Proof uses split $=$[476], `wf_def`, rest routine.

Ergo: Definition of `wf` meets textbook definitions "every non-empty set Q has a minimal element in r".

---

[476]By this we simply mean to split a proof of $\phi = \psi$ into two proofs $\phi \implies \psi$ and $\psi \implies \phi$.

A **Theorem** on Subsets

`wf_subset` $[\![wf\ r; p \subseteq r]\!] \Longrightarrow wf\ p$

Proof sketch: `wf_subset`: simplification tactic using `wf_eq_minimal`.

A Theorem on Subrelations

```
wf_subrel
```

$$wf\ r \implies \forall p.p \subseteq r \to \exists x.\forall y.(y,x) \notin p^+$$

Proof sketch:

Combine `wf_minimal` and `wf_subset`.

This implies $wf\ r \implies \forall p.p \subseteq r \to \exists x.\forall y.(x,y) \notin p.$

Ergo: Definition of `wf` fulfills the condition corresponding to our second attempt of characterizing well-foundedness using minimal elements.

A Theorem on Subrelations

`wf_subrel`

$$wf\ r \implies \forall p.p \subseteq r \to \exists x.\forall y.(y,x) \notin p^+$$

Proof sketch:

Combine `wf_minimal` and `wf_subset`.

This implies $wf\ r \implies \forall p.p \subseteq r \to \exists x.\forall y.(x,y) \notin p$.

Ergo: Definition of `wf` fulfills the condition corresponding to our second attempt of characterizing well-foundedness using minimal elements.

However, this formalization still had a problem: The minimum could be an isolated element, unrelated to the subrelation.

23.4 Defining Recursive Functions

Idea of well-founded recursion: Wish to define f by recursive equation $f = e$, e.g.

$$fac = (\lambda n.\ \textbf{if}\ n = 0\ \textbf{then}\ 1\ \textbf{else}\ n * fac(n-1))$$

Define $F = \lambda f.e$, e.g.

$$Fac = (\lambda fac.\ \lambda n.\ \textbf{if}\ n = 0\ \textbf{then}\ 1\ \textbf{else}\ n * fac(n-1))$$

23.4 Defining Recursive Functions

Idea of well-founded recursion: Wish to define f by recursive equation $f = e$, e.g.

$$fac = (\lambda n.\ \textbf{if}\ n = 0\ \textbf{then}\ 1\ \textbf{else}\ n * fac(n-1))$$

Define $F = \lambda f.e$, e.g. (α-conversion of what you have seen)

$$Fac = (\lambda f\ \ .\ \lambda n.\ \textbf{if}\ n = 0\ \textbf{then}\ 1\ \textbf{else}\ n * f\ \ (n-1))$$

We say: F is the functional defining f.

Recall that $Y\ F$ would solve $f = e$, but we don't have Y, so what can we do?

Coherent Functionals

A functional F is coherent w.r.t. $<$ if all recursive calls are with arguments "smaller" than the original argument. This means that if F has the form

$$\lambda f.\lambda n.e'$$

then for any $(f\ m)$ occurring in e', we have $m < n$.

Here $<$ could be any relation (although the idea is that it should be a well-founded ordering).

(Simplification, assumes that recursion is on the first argument of f.)

Using Bad f's

Let $f|_{<a}$ be a function that is like f on all values $< a$, and arbitrary elsewhere. $f|_{<a}$ is an approximation, a "bad" f.

If F is coherent, then we would expect that for any a,

$$f\, a = (F\, f)\, a = (F\, f|_{<a})\, a. \tag{5}$$

It's not that we are ultimately interested in constructing such a "bad" f, but our formalization of well-founded recursion defines coherence by the fact that one could use such a "bad" f, i.e., via (5).

"Bad" f's: Example

Consider fac. On the right-hand side, we show one possibility[477] for $fac|_{<4}$):

---
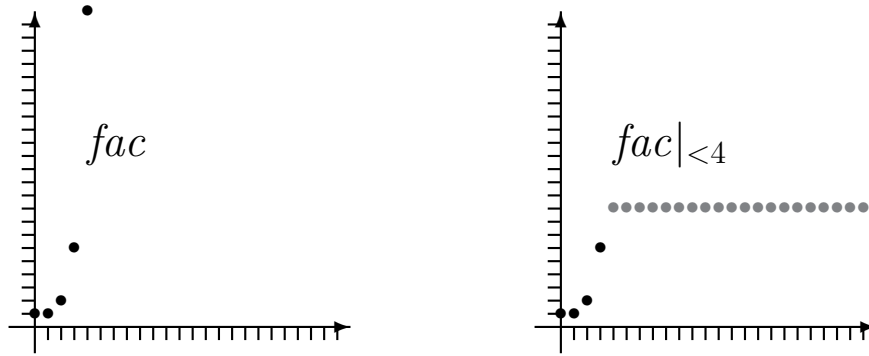
[477]For the construction we have in mind, it would be fine that $f|_{<a}$ be a function that is like f on all values $< a$, and arbitrary elsewhere. E.g., $fac|_{<4}$ could be
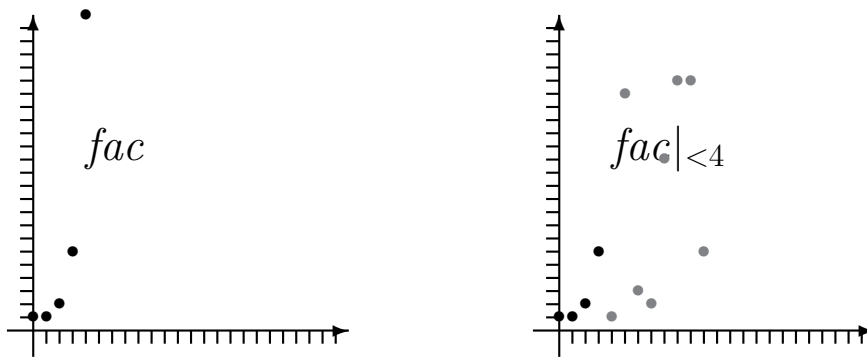
However, such a $fac|_{<4}$ could not be in a model for HOL (with the extensions we consider here). The way that arbitrary elements are formalized in HOL.thy, it turns out that in any model and for each type, there must be one specific domain element for the constant arbitrary (you don't have to understand why this is so). That is, in different models we could have different ones, but within each model the element must be a specific one. Since the value of $fac|_{<4}$ is "arbi-

cut **(in** `Wellfounded_Recursion.thy`**)**

```
constdefs
 cut    :: "('a => 'b) => ('a * 'a) set =>
                                'a => 'a => 'b"
  "cut f r x ==
    (%y. if (y,x):r then f y else arbitrary)"
```

cut $f\ r\ x$ is what we denoted by $f|_{<x}$ (taking $<$ for r).
`arbitrary` is defined in `HOL.thy`.

---

trary" for all arguments ≥ 4, this means that in each model, this value must be the same for all arguments ≥ 4, ruling out the function above.

Of course, these are considerations taking place only in our heads. In the actual deduction machinery, one never constructs these "arbitrary" terms.

cut (in `Wellfounded_Recursion.thy`)

```
constdefs
 cut    :: "('a => 'b) => ('a * 'a) set =>
                              'a => 'a => 'b"
 "cut f r x ==
   (%y. if (y,x):r then f y else arbitrary)"
```

cut $f\ r\ x$ is what we denoted by $f|_{<x}$ (taking $<$ for r).
`arbitrary` is defined in `HOL.thy`.

The function cut $f\ r\ x$ is unspecified for arguments y
where $(y, x) \notin r$, but for each such argument, $(\text{cut } f\ r\ x)\ y$
must be the same (in any particular model).

___

trary" for all arguments ≥ 4, this means that in each model,
this value must be the same for all arguments ≥ 4, ruling out
the function above.

Of course, these are considerations taking place only in our
heads. In the actual deduction machinery, one never con-
structs these "arbitrary" terms.

Theorems Involving `cut`

`cuts_eq`
$$(cut\ f\ r\ x = cut\ g\ r\ x) =$$
$$(\forall y.(y, x) \in r \to f\ y = g\ y)$$

`cut_apply` $\quad (x, a) \in r \implies cut\ f\ r\ a\ x = f\ x$

Or, using the more intuitive notation:

`cuts_eq` $\quad (f|_{<x} = g|_{<x}) = (\forall y.y < x \to f\ y = g\ y)$

`cut_apply` $\quad x < a \implies f|_{<a}\ x = f\ x$

wfrec_rel (in `Wellfounded_Recursion.thy`)

Auxiliary construction: "approximate" f by a relation *wfrec_rel* $R\,F$.

```
  wfrec_rel :: "('a * 'a) set =>
   (('a => 'b) => 'a => 'b) => ('a * 'b) set"
inductive "wfrec_rel R F"
intrs
  wfrecI
   "ALL z. (z, x) : R -->
           (z, g z) : wfrec_rel R F
     ==> (x, F g x) : wfrec_rel R F"
```

wfrec_rel **Explained**

$$\forall z.(z, x) \in R \rightarrow (z, g\, z) \in \mathit{wfrec\_rel}\ R\, F \Longrightarrow$$
$$(x, F\, g\, x) \in \mathit{wfrec\_rel}\ R\, F$$

- For R and F arbitrary, $\mathit{wfrec\_rel}\ R\, F$ is defined but we wouldn't want to know what it is.

- But if R is well-founded and F is coherent, $\mathit{wfrec\_rel}\ R\, F$ defines a recursive "function"[478].

Show that $(4, 24) \in (\mathit{wfrec\_rel}\ {'}{<}{'}\ \mathit{Fac})$!

Now let us really turn $\mathit{wfrec\_rel}\ R\, F$ into a function . . .

---

[478]When we say that a binary relation $r : \tau \times \sigma$ is in fact a function, we mean that for $t : \tau$, there is exactly one $s : \sigma$ such that $(t, s) \in r$.

wfrec (in `Wellfounded_Recursion.thy`)

```
wfrec :: "('a * 'a) set =>
    (('a => 'b) => 'a => 'b) => 'a => 'b"
"wfrec R F == %x. THE y.
  (x, y) : wfrec_rel R (%f x. F (cut f R x) x)"
```

THE $x. P\, x$[479] picks the unique a such that $P\, a$ holds, if it exists. We don't care what it does otherwise (see `HOL.thy`).

---

[479] The operator THE is similar to the Hilbert operator, but it returns the unique element having a certain property rather than an arbitrary one. The Isabelle formalization of HOL nowadays heavily relies on THE rather than the Hilbert operator.

wfrec **Explained**

$$wfrec\, R\, F \equiv$$
$$\lambda x.\text{THE}\, y.(x, y) \in wfrec\_rel\, R\, (\lambda f x.F\, (cut\, f\, R\, x)\, x)$$

We don't care what this means for arbitrary R and F.

But if R is well-founded and F is coherent, then $F\, (cut\, f\, R\, x)\, x = F\, f\, x$ (by (5)), and so $\lambda f x.F\, (cut\, f\, R\, x)\, x = F$, and so $\lambda x.\text{THE}\, y.(x, y) \in wfrec\_rel\, R\, (\lambda f x.F\, (cut\, f\, R\, x)\, x)$ is the function defined by $wfrec\_rel\, R\, F$ in the obvious way.

`wfrec` $R\, F$ is the recursive function defined by functional F.

The "Fixpoint" Theorem

`wfrec` $\quad wf\ r \implies wfrec\ r\ H\ a = H\,(cut\,(wfrec\ r\ H)\ r\ a)\ a$

Note that `wfrec` is used here both as a name of a constant (defined above) and a theorem.

So if r is well-founded and H is coherent, we have (by (5))

$$wfrec\ r\ H\ a = H\,(wfrec\ r\ H)\ a$$

Theorem states that $wfrec$ is like a fixpoint combinator (disregarding the additional argument r).

Thus we can do using $wfrec$ what we would have liked to do using Y.

23.5 Example for *wfrec*: Natural Numbers

The constant *wfrec* provides the mechanism/support for defining recursive functions. We illustrate this using `nat`, the type of natural numbers (pretending we have it).

wfrec is applied to a well-founded order and a functional to define a function.

23.5 Example for *wfrec*: Natural Numbers

The constant *wfrec* provides the mechanism/support for defin-
ing recursive functions. We illustrate this using nat, the type
of natural numbers (pretending we have it).

wfrec is applied to a well-founded order and a functional
to define a function.

First, define predecessor relation:

```
constdefs
  pred_nat :: "(nat * nat) set"
  pred_nat_def "pred_nat == {(m,n). n = Suc m}"
```

Defining Addition and Subtraction

```
add :: [nat, nat] => nat    (infixl 70)
"m add n == wfrec (pred_nat^+)
   (%f j. if j=0 then n else Suc (f (pred j))) m"
```

Recursive in first argument[480].

```
subtract :: [nat, nat] => nat    (infixl 70)
"m subtract n == wfrec (pred_nat^+)
   (%f j. if j=0 then m else pred (f (pred j))) n"
```

Recursive in second argument.

480

```
add :: [nat, nat] => nat    (infixl 70)
"m add n == wfrec (pred_nat^+)
      (%f j. if j=0 then n else Suc (f (pred j))) m"
```

Here we suppose that we have a predecessor function pred. The implementation in Isabelle is different, but conceptually, the above is a definition of the add function.

Note that add is a function of type $nat \rightarrow nat \rightarrow nat$ (written infix), but it is only recursive in one argument, namely the first one.

You may be confused about this and wonder: how do I know that it is the first? Is this some Isabelle mechanism saying that it is always the first? The answer is: no. You must look at the two sides in isolation. On the right-hand side, we have

```
wfrec (pred_nat^+)
   (%f j. if j=0 then n else Suc (f (pred j)))
```

By the definitions (of *wfrec* most importantly), this expression is a function of type $nat \rightarrow nat$, namely the function that

Defining Division and Modulus

```
div :: ['a::div, 'a] => 'a          (infixl 70)
"m div n == wfrec (pred_nat^+)
(%f j. if j<n | n=0 then 0 else Suc (f (j-n))) m"


mod :: ['a::div, 'a] => 'a          (infixl 70)
"m mod n == wfrec (pred_nat^+)
  (%f j. if j<n | n=0 then j else f (j-n)) m"
```

Here, `div` is a syntactic class for which division is defined (don't worry about it). We know how to define $-$.

The functions are recursive in one argument (just like add).

---

adds n (which is not known looking at this expression alone; it occurs on the left-hand side) to its argument. The function is recursive in its argument (and hence not in n). Now, this function is applied to m. Therefore we say that the final function add is recursive in m but not in n.

Now look at subtraction:

```
subtract :: [nat, nat] => nat    (infixl 70)
"m subtract n == wfrec (pred_nat^+)
    (%f j. if j=0 then m else pred (f (pred j))) n"
```

Note that `subtract` is recursive in its second argument, simply because the right-hand side of the defining equation was constructed in a different way than for add.

Similar considerations apply for other binary functions defined by recursion in one argument.

Theorems of the Example

wf_pred_nat   *wf pred_nat*

mod_if
$$m \bmod n = $$
$$(if \ m < n \ then \ m \ else \ (m - n) \bmod n)$$

div_if
$$0 < n \implies m \ div \ n = $$
$$(if \ m < n \ then \ 0 \ else \ Suc((m - n) \ div \ n))$$

wf_pred_nat   *wf pred_nat*

Theorems of the Example

`wf_pred_nat`	$wf\ pred\_nat$
`mod_if`	$m\ mod\ n =$ $(if\ m < n\ then\ m\ else\ (m - n)\ mod\ n)$
`div_if`	$0 < n \implies m\ div\ n =$ $(if\ m < n\ then\ 0\ else\ Suc((m - n)\ div\ n))$

This is very similar to functional programming code and hence lends itself to real computations (rewriting), as opposed to only doing proofs.

23.6 Conclusion on Well-founded Recursion

Well-founded recursion allows us to define recursive functions in HOL and thus reason about computations.

We can derive recursive theorems that can be used for rewriting just like in a functional programming language.

Isabelle Package for Primitive Recursion

For primitive recursion[481], finding a well-founded ordering is simple enough for automation[482]!

---

[481]A function is primitive recursive if the recursion is based on the immediate predecessor w.r.t. the well-founded order used (e.g., the predecessor on the natural numbers, as opposed to any arbitrary smaller numbers).

This is not the same concept as used in the context of computation theory, where primitive recursive is in contrast to μ-recursive [LP81].

[482]The `primrec` syntax provides a convenient front-end for defining primitive recursive functions.

Isabelle will guess a well-founded ordering to use. E.g. for functions on the natural numbers, it will use the usual $<$ ordering.

Isabelle Package for Primitive Recursion

For primitive recursion[481], finding a well-founded ordering is simple enough for automation[482]!

Examples (use `nat` and `case`-syntax): ...

---

[481] A function is primitive recursive if the recursion is based on the immediate predecessor w.r.t. the well-founded order used (e.g., the predecessor on the natural numbers, as opposed to any arbitrary smaller numbers).

This is not the same concept as used in the context of computation theory, where primitive recursive is in contrast to μ-recursive [LP81].

[482] The `primrec` syntax provides a convenient front-end for defining primitive recursive functions.

Isabelle will guess a well-founded ordering to use. E.g. for functions on the natural numbers, it will use the usual $<$ ordering.

Recursion and Arithmetic

```
primrec
  add_0:    "0 + n = n"
  add_Suc:  "Suc m + n = Suc (m + n)"
primrec
  diff_0:   "m - 0 = m"
  diff_Suc: "m - Suc n =
    (case m - n of 0 => 0 | Suc k => k)"
primrec
  mult_0:   "0 * n = 0"
  mult_Suc: "Suc m * n = n + (m * n)"
```

23.7 Conclusion on Recursion and Induction

We are interested in recursion because inductively defined sets and recursively defined functions are solutions to recursive equations.

We cannot have general fixpoint operator Y, but we have, by conservative extension:

- Least fixpoints for defining sets;

- well-founded orders for defining functions.

Both concepts come with induction schemes (lfp induction and definition of well-foundedness) for proving properties of the defined objects.

Summary: Proof Support

The methodological overhead can be faced by powerful mechanical support, since many proof-tasks are routine.

Isabelle offers extensibility for such routine task constructions.

24 Arithmetic

The Roadmap

We are still looking at how the different parts of mathematics are encoded in the Isabelle/HOL library.

- Orders

- Sets

- Functions

- (Least) fixpoints and induction

- (Well-founded) recursion

- Arithmetic

- Datatypes

The Roadmap

We are still looking at how the different parts of mathematics are encoded in the Isabelle/HOL library.

- Orders

- Sets

- Functions

- (Least) fixpoints and induction

- (Well-founded) recursion

- Arithmetic

- Datatypes

Current Stage of our Course

- On the basis of conservative embeddings, set theory can be built safely.

- Inductive sets can be defined using least fixpoints and suitably supported by Isabelle.

- Well-founded orderings can be defined without referring to infinity. Recursive functions can be based on these. Needs inductive sets though. Support by Isabelle provided.

Current Stage of our Course

- On the basis of conservative embeddings, set theory can be built safely.

- Inductive sets can be defined using least fixpoints and suitably supported by Isabelle.

- Well-founded orderings can be defined without referring to infinity. Recursive functions can be based on these. Needs inductive sets though. Support by Isabelle provided.

Next important topic: arithmetic.

Which Approach to Take?

- Purely definitional?

---

[483]Our intuition/knowledge about arithmetics clearly requires that there are infinite sets, e.g., the set of infinite numbers. Technically, the HOL model of the set of natural numbers must be an infinite set, otherwise we would not be willing to say that have "modeled" arithmetic.

[484]The Peano axioms include

- $0 \in nat$
- $\forall x.x \in nat \rightarrow Suc(x) \in nat$
- $\forall x.Suc(x) \neq 0$
- $\forall P.(P(0) \land \forall n.(P(n) \rightarrow P(Suc(n)))) \rightarrow \forall n.P(n)$

Which Approach to Take?

- Purely definitional?

  Not possible with eight basic rules (cannot enforce infinity[483] of HOL model)!

---

[483]Our intuition/knowledge about arithmetics clearly requires that there are infinite sets, e.g., the set of infinite numbers. Technically, the HOL model of the set of natural numbers must be an infinite set, otherwise we would not be willing to say that have "modeled" arithmetic.

[484]The Peano axioms include

- $0 \in nat$
- $\forall x.x \in nat \rightarrow Suc(x) \in nat$
- $\forall x.Suc(x) \neq 0$
- $\forall P.(P(0) \wedge \forall n.(P(n) \rightarrow P(Suc(n)))) \rightarrow \forall n.P(n)$

Which Approach to Take?

- Purely definitional?

  Not possible with eight basic rules (cannot enforce infinity[483] of HOL model)!

- Heavily axiomatic? I.e., we state natural numbers by Peano axioms[484] and claim analogous axioms for any other number type?

[483]Our intuition/knowledge about arithmetics clearly requires that there are infinite sets, e.g., the set of infinite numbers. Technically, the HOL model of the set of natural numbers must be an infinite set, otherwise we would not be willing to say that have "modeled" arithmetic.

[484]The Peano axioms include

- $0 \in nat$
- $\forall x. x \in nat \rightarrow Suc(x) \in nat$
- $\forall x. Suc(x) \neq 0$
- $\forall P.(P(0) \wedge \forall n.(P(n) \rightarrow P(Suc(n)))) \rightarrow \forall n.P(n)$

Which Approach to Take?

- Purely definitional?

  Not possible with eight basic rules (cannot enforce infinity[483] of HOL model)!

- Heavily axiomatic? I.e., we state natural numbers by Peano axioms[484] and claim analogous axioms for any other number type?

  Insecure!

---

[483]Our intuition/knowledge about arithmetics clearly requires that there are infinite sets, e.g., the set of infinite numbers. Technically, the HOL model of the set of natural numbers must be an infinite set, otherwise we would not be willing to say that have "modeled" arithmetic.

[484]The Peano axioms include

- $0 \in nat$
- $\forall x. x \in nat \rightarrow Suc(x) \in nat$
- $\forall x. Suc(x) \neq 0$
- $\forall P.(P(0) \wedge \forall n.(P(n) \rightarrow P(Suc(n)))) \rightarrow \forall n. P(n)$

Which Approach to Take?

- Purely definitional?

  Not possible with eight basic rules (cannot enforce infinity[483] of HOL model)!

- Heavily axiomatic? I.e., we state natural numbers by Peano axioms[484] and claim analogous axioms for any other number type?

  Insecure!

- Minimally axiomatic? We construct an infinite set, and define numbers etc. as inductive subset?

---

[483]Our intuition/knowledge about arithmetics clearly requires that there are infinite sets, e.g., the set of infinite numbers. Technically, the HOL model of the set of natural numbers must be an infinite set, otherwise we would not be willing to say that have "modeled" arithmetic.

[484]The Peano axioms include

- $0 \in nat$
- $\forall x.x \in nat \rightarrow Suc(x) \in nat$
- $\forall x.Suc(x) \neq 0$
- $\forall P.(P(0) \wedge \forall n.(P(n) \rightarrow P(Suc(n)))) \rightarrow \forall n.P(n)$

Which Approach to Take?

- Purely definitional?

  Not possible with eight basic rules (cannot enforce infinity[483] of HOL model)!

- Heavily axiomatic? I.e., we state natural numbers by Peano axioms[484] and claim analogous axioms for any other number type?

  Insecure!

- Minimally axiomatic? We construct an infinite set, and define numbers etc. as inductive subset?

  Yes. Finally use infinity axiom.

---

[483]Our intuition/knowledge about arithmetics clearly requires that there are infinite sets, e.g., the set of infinite numbers. Technically, the HOL model of the set of natural numbers must be an infinite set, otherwise we would not be willing to say that have "modeled" arithmetic.

[484]The Peano axioms include

- $0 \in nat$
- $\forall x. x \in nat \to Suc(x) \in nat$
- $\forall x. Suc(x) \neq 0$
- $\forall P. (P(0) \land \forall n. (P(n) \to P(Suc(n)))) \to \forall n. P(n)$

24.1 What is Infinity? Cantor's Hotel
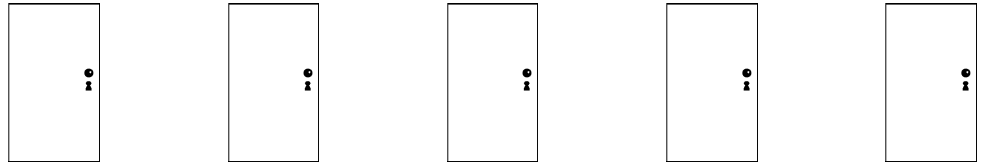
Cantor's hotel has infinitely many rooms.

24.1 What is Infinity? Cantor's Hotel
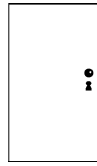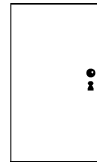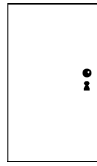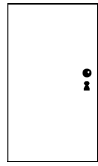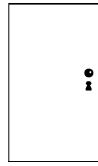
Cantor's hotel has infinitely many rooms. New guest arrives.

24.1 What is Infinity? Cantor's Hotel
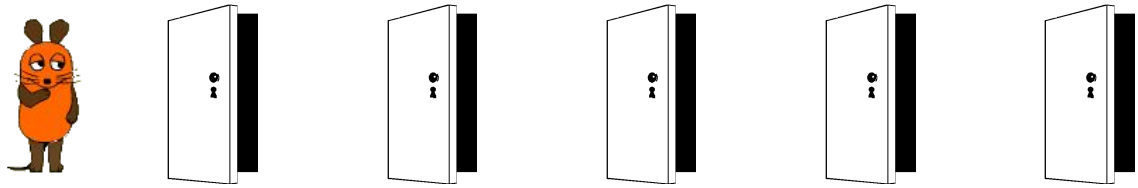
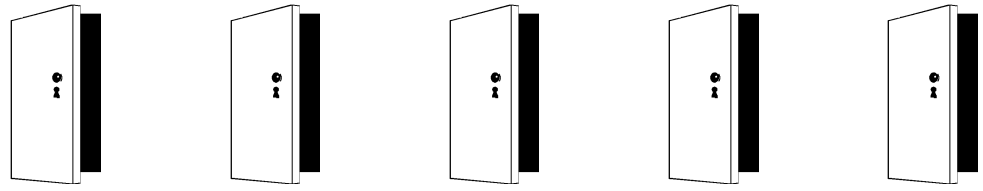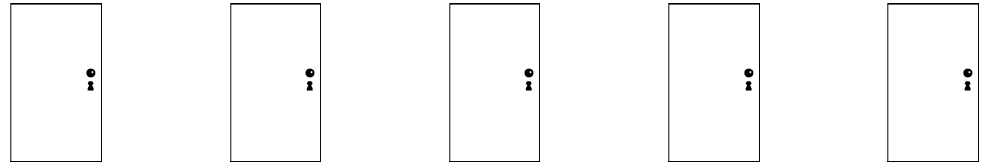Cantor's hotel has infinitely many rooms. New guest arrives.
The doors open,

24.1 What is Infinity? Cantor's Hotel
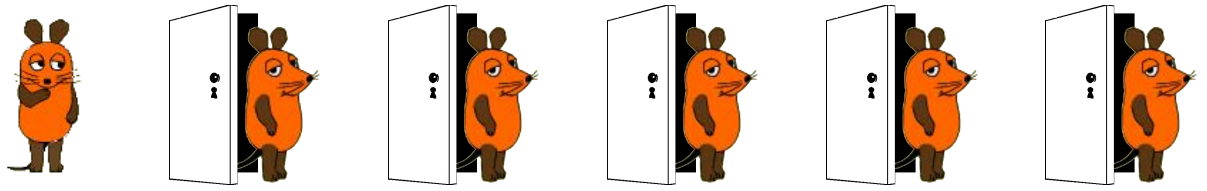
Cantor's hotel has infinitely many rooms. New guest arrives.
The doors open, and all guests come out of their rooms.

24.1 What is Infinity? Cantor's Hotel
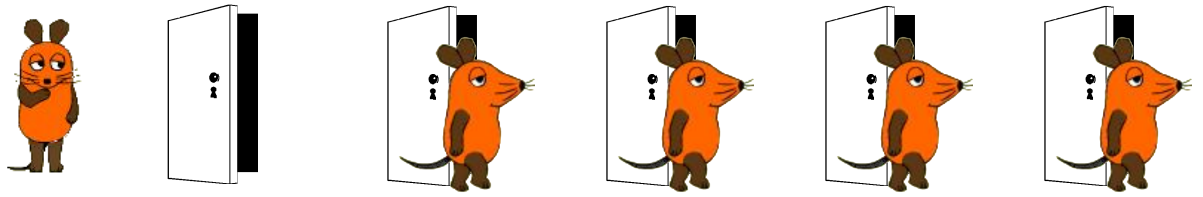
Cantor's hotel has infinitely many rooms. New guest arrives.

The doors open, and all guests come out of their rooms. They move one room forward[485],

24.1 What is Infinity? Cantor's Hotel

Cantor's hotel has infinitely many rooms. New guest arrives.

The doors open, and all guests come out of their rooms. They move one room forward[485], the new guest walks towards the first room,

24.1 What is Infinity? Cantor's Hotel

Cantor's hotel has infinitely many rooms. New guest arrives.

The doors open, and all guests come out of their rooms. They move one room forward[485], the new guest walks towards the first room, they turn around,

24.1 What is Infinity? Cantor's Hotel

Cantor's hotel has infinitely many rooms. New guest arrives.

The doors open, and all guests come out of their rooms. They move one room forward[485], the new guest walks towards the first room, they turn around, enter their new rooms.

24.1 What is Infinity? Cantor's Hotel

Cantor's hotel has infinitely many rooms. New guest arrives.

The doors open, and all guests come out of their rooms. They move one room forward[485], the new guest walks towards the first room, they turn around, enter their new rooms. The doors close, all guests are accomodated.

---

[485]This means, there must be a successor function on rooms. To each room, it assigns the "next" room.

Axiom of Infinity

The axiomatic core[486] of datatypes (and hence, numbers[487]):

$$\frac{}{\exists f :: (ind \to ind).\ injective\ f \wedge \neg surjective\ f}\ infty$$

where

$$
\begin{aligned}
injective^{488}\ f &= \forall xy.\ f\,x = f\,y \to x = y \\
surjective\ f &= \forall y. \exists x.\ y = f\,x
\end{aligned}
$$

Forces ind to be "infinite type" (called "I" in [Chu40]).
We will see soon how this is done in Isabelle.

---

[486]Note that theoretically, it is not needed to add the infinity axiom (or some equivalent formulation) to HOL. Instead one could add the infinity axiom as premise to each arithmetic theorem that one wants to prove.

However this would not be a viable approach since the resulting formulas would be very, very complicated.

[487]The natural numbers can be built as an algebraic datatype by having a constant 0 and a term constructor Suc (for successor).

[488]These constants (actually called inj and sur) are defined in Fun.thy.

24.2 Type-Closed Conservative Extensions

Why must conservative extensions be type-closed? (the standard example)

Consider $H \equiv \exists f :: \alpha \Rightarrow \alpha.injective\ f \wedge \neg surjective\ f$

24.2 Type-Closed Conservative Extensions

Why must conservative extensions be type-closed? (the standard example)

Consider $H \equiv \exists f :: \alpha \Rightarrow \alpha. injective\ f \wedge \neg surjective\ f$

Then the type of H is $bool$, but H contains a subterm of type $\alpha \Rightarrow \alpha$ (H is not type-closed).

Then we could reason as follows ...

Type-Closed Conservative Extensions (2)

$(H \equiv \exists f :: \alpha \Rightarrow \alpha.injective\ f \wedge \neg surjective\ f)$

---

[489]We use inj and sur as abbreviations for $injective$ and $surjective$.

Type-Closed Conservative Extensions (2)

$(H \equiv \exists f :: \alpha \Rightarrow \alpha.injective\ f \wedge \neg surjective\ f)$

$$H = H \qquad \text{holds by } \textit{refl}$$
$$\Rightarrow\ \exists f :: bool \Rightarrow bool.inj^{489}\ f \wedge \neg sur\ f =$$
$$\exists f :: ind \Rightarrow ind.inj\ f \wedge \neg sur\ f$$
$$\Rightarrow\ False = True$$
$$\Rightarrow\ False$$

(unfolding H using two different type instantiations, and then using axiom of infinity and the fact that there are only finitely many functions on $bool$).

---

[489]We use inj and sur as abbreviations for $injective$ and $surjective$.

Types Affect the Semantics

Type instantiations may change semantic values, and hence cause inconsistency!

This example was somewhat more concrete than our previous simpler example.

24.3 Natural Numbers: `Nat.thy`

```
consts
  Zero_Rep        :: ind
  Suc_Rep         :: "ind => ind"
axioms
  inj_Suc_Rep:             "inj Suc_Rep"
  Suc_Rep_not_Zero_Rep: "Suc_Rep x ~= Zero_Rep"
```

So the axiom of infinity is formulated by defining a constant
Suc_Rep having the two required properties.

    inj and *sur* are defined in `Fun.thy`.

    Think of Zero_Rep, Suc_Rep as provisional 0, successor.

Defining the Set Nat

Want to define **new type** nat. How?

Defining the Set Nat

Want to define new type nat. How?

Must define a set isomorphic to the natural numbers. How?

Defining the Set Nat

Want to define new type $nat.$ How?

Must define a set isomorphic to the natural numbers. How?

By induction using the `inductive` syntax:

```
inductive Nat
intros
  Zero_RepI: "Zero_Rep : Nat"
  Suc_RepI: "i : Nat ==> Suc_Rep i : Nat"
```

Translated by Isabelle to:
$$Nat = lfp\,(\lambda X.\{Zero\_Rep\} \cup (Suc\_Rep\,`\,X))$$

Defining the Type nat

Now we have the set Nat. What next?

490

Note the two ingredients for defining the type `nat`:

- An inductively defined set `Nat`, i.e., a set defined as fix-point of a monotone function. In Isabelle (`Nat.thy`), the `inductive` syntax is used for this purpose. This automatically generates an induction rule for the set.

- A type definition based on this set, defined using the `typedef` syntax.

  Recall that this process automatically generates the two constants `Abs_Nat` and `Rep_Nat`.

But note: the induction theorem is not inherited automatically. More precisely, the `typedef` syntax does not cause the type `nat` to inherit the inductive theorem of the set `Nat`. The theorem `nat_induct` is explicitly proven in `Nat.thy`.

572

Defining the Type nat

Now we have the set Nat. What next?

Define the type nat, isomorphic to Nat, using the `typedef` syntax:

```
typedef (open Nat)
  nat = "Nat" by (rule exI, rule Nat.Zero_RepI)
```

After these two steps[490] we have the type nat.

---

[490]

Note the two ingredients for defining the type `nat`:

- An inductively defined set `Nat`, i.e., a set defined as fixpoint of a monotone function. In Isabelle (`Nat.thy`), the `inductive` syntax is used for this purpose. This automatically generates an induction rule for the set.

- A type definition based on this set, defined using the `typedef` syntax.

  Recall that this process automatically generates the two constants `Abs_Nat` and `Rep_Nat`.

But note: the induction theorem is not inherited automatically. More precisely, the `typedef` syntax does not cause the type `nat` to inherit the inductive theorem of the set `Nat`. The theorem `nat_induct` is explicitly proven in `Nat.thy`.

Constants in nat

Moreover, define[491]:

```
consts
  Suc :: "nat => nat"
  pred_nat :: "(nat * nat) set"

defs
  Zero_nat_def: "0 == Abs_Nat Zero_Rep"
  Suc_def:      "Suc ==
           (%n. Abs_Nat (Suc_Rep (Rep_Nat n)))"
  pred_nat_def: "pred_nat == {(m, n). n = Suc m}"
```

---

[491]Based on the generic constants Abs_Nat and Rep_Nat, we define all the constants that we need to work conveniently with nat, most importantly, 0 and Suc.

Some **Theorems** in `Nat.thy`[492]

$$\texttt{nat\_induct} \quad [\![P\,0; \bigwedge n.P\,n \Longrightarrow P\,(Suc\,n)]\!] \Longrightarrow P\,n$$

$$\texttt{diff\_induct} \quad \begin{aligned}&[\![\bigwedge x.P\,x\,0; \bigwedge y.P\,0\,(Suc\,y);\\ &\bigwedge xy.P\,x\,y \Longrightarrow P\,(Suc\,x)\,(Suc\,y)]\!]\\ &\Longrightarrow P\,m\,n\end{aligned}$$

We can now exploit that nat is defined based on a set defined using least fixpoints. In particular, `nat_induct` follows (but not "automatically"!) from the `induct` theorem of `Lfp`.

---

[492]This file should be contained in your Isabelle distribution. Or, if you only have an Isabelle executable, you can find the sources here:

http://isabelle.in.tum.de/library/

Nat and Well-Founded Orders

Examples of theorems involving well-founded orders:

`wf_pred_nat`	$wf\ pred\_nat$
`less_linear`	$m < n \lor m = n \lor n < m$
`Suc_less_SucD`	$Suc\,m < Suc\,n \implies m < n$

Using Primitive Recursion

`Nat.thy` defines rich theory on nat. Uses `primrec` syntax for defining recursive functions, and case[493] construct.

```
primrec
   add_0    "0 + n = n"
   add_Suc  "Suc m + n = Suc(m + n)"
primrec
   diff_0   "m - 0 = m"
   diff_Suc "m - Suc n =
      (case m - n of 0 => 0 | Suc k => k)"
primrec
   mult_0   "0 * n = 0"
   mult_Suc "Suc m * n = n + (m * n)"
```

---

[493]The case statement for nat is a function of type $nat \Rightarrow (nat \Rightarrow nat) \Rightarrow nat \Rightarrow nat$. case z f n is defined as follows (using a common mathemtical notation):

$$\text{case } z \ f \ n = \begin{cases} z & \text{if } n = 0 \\ f \ k & \text{if } n = Suc \ k \end{cases}$$

The syntax

```
 diff_Suc "m - Suc n = (case m - n of 0 => 0 | Suc k => k)
```

used on the slide is a paraphrasing ("concrete syntax") of the original ("abstract") syntax. In the original syntax it would read case 0 $(\lambda x.x)$ $(n - m)$.

Some **Theorems** in `Nat`

$$\texttt{add\_0\_right} \quad m + 0 = m$$
$$\texttt{add\_ac} \qquad m + n + k = m + (n + k)$$
$$m + n = n + m$$
$$x + (y + z) = y + (x + z)$$
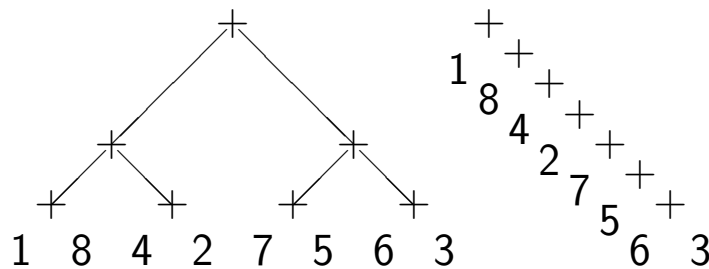$$\texttt{mult\_ac} \qquad m * n * k = m * (n * k)$$
$$m * n = n * m$$
$$x * (y * z) = y * (x * z)$$

Note third part[494] of `add_ac`, `mult_ac`, respectively.
Technically, `add_ac` and `mult_ac` are lists of `thm`'s.

---

[494] The theorems $x + (y + z) = y + (x + z)$ and $x * (y * z) = y * (x * z)$ are called left-commutation laws and are crucial for (ordered) rewriting.

Suppose we have the term shown below.

Some **Theorems** in `Nat`

`add_0_right`	$m + 0 = m$
`add_ac`	$m + n + k = m + (n + k)$
	$m + n = n + m$
	$x + (y + z) = y + (x + z)$
`mult_ac`	$m * n * k = m * (n * k)$
	$m * n = n * m$
	$x * (y * z) = y * (x * z)$

Note third part[494] of `add_ac`, `mult_ac`, respectively.
Technically, `add_ac` and `mult_ac` are lists of `thm`'s.

---

[494] The theorems $x + (y + z) = y + (x + z)$ and $x * (y * z) = y * (x * z)$ are called left-commutation laws and are crucial for (ordered) rewriting.

Suppose we have the term shown below. Using associativity $(m + n + k = m + (n + k))$ this will be rewritten to the second term.

Some **Theorems** in `Nat`

`add_0_right`	$m + 0 = m$
`add_ac`	$m + n + k = m + (n + k)$
	$m + n = n + m$
	$x + (y + z) = y + (x + z)$
`mult_ac`	$m * n * k = m * (n * k)$
	$m * n = n * m$
	$x * (y * z) = y * (x * z)$

Note third part[494] of `add_ac`, `mult_ac`, respectively.
Technically, `add_ac` and `mult_ac` are lists of `thm`'s.

---

[494] The theorems $x + (y + z) = y + (x + z)$ and $x * (y * z) = y * (x * z)$ are called left-commutation laws and are crucial for (ordered) rewriting.

Suppose we have the term shown below. Using associativity ($m + n + k = m + (n + k)$) this will be rewritten to the second term. Using left-commutation, this will be rewritten to the third term. This is a so-called AC-normal form, for an appropriately chosen term ordering.

Proof of `add_0_right`

578
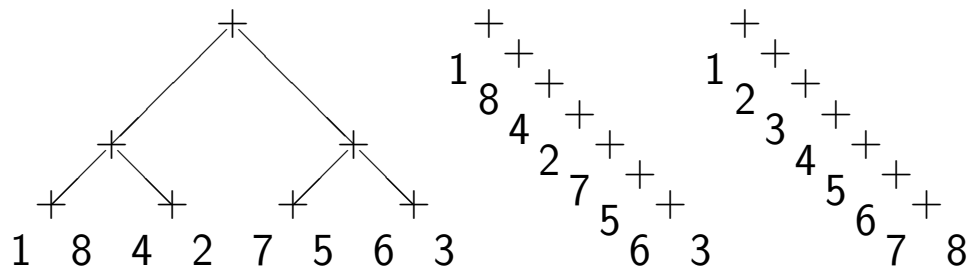
$$m + 0 = m$$

add_0_right

Proof of `add_0_right`

$$n + 0 = n$$

$$\cfrac{\dfrac{\text{add\_0}}{0 + 0 = 0} \qquad\qquad Suc\,n + 0 = Suc\,n}{m + 0 = m}\;\text{nat\_induct}$$

Proof of `add_0_right`

$$\dfrac{\dfrac{\text{add\_0}}{0+0=0} \qquad \dfrac{\dfrac{\dfrac{\text{add\_Suc}}{Suc\,n + 0 = Suc(n+0)}}{Suc(n+0) = Suc\,n + 0}\ \textit{sym} \qquad \dfrac{n + 0 = n}{Suc(n+0) = Suc\,n}\ \textit{arg\_cong}}{Suc\,n + 0 = Suc\,n}\ \textit{subst}}{m + 0 = m}\ \text{nat\_induct}$$

Note that $Suc\,n + 0 = Suc(n+0)$ is an instance of $Suc\,m + n = Suc(m+n)$.

Proof of `add_0_right`

$$
\cfrac{
  \cfrac{}{0 + 0 = 0} \text{ add\_0}
  \qquad
  \cfrac{
    \cfrac{
      \cfrac{}{Suc\, n + 0 = Suc(n + 0)} \text{ add\_Suc}
    }{Suc(n + 0) = Suc\, n + 0} \text{ sym}
    \qquad
    \cfrac{
      [n + 0 = n]^1
    }{Suc(n + 0) = Suc\, n} \text{ arg\_cong}
  }{Suc\, n + 0 = Suc\, n} \text{ subst}
}{m + 0 = m} \text{ nat\_induct}^1
$$

Note that $Suc\, n + 0 = Suc(n + 0)$ is an instance of $Suc\, m + n = Suc(m + n)$.

24.4 Integers

The integers are implemented[495] as equivalence classes[496] over $nat \times nat$.

```
IntDef = Equiv + NatArith +
constdefs
  intrel :: "((nat * nat) * (nat * nat)) set"
  "intrel == {p. EX x1 y1 x2 y2.
   p=((x1::nat,y1),(x2,y2)) & x1+y2 = x2+y1}"


typedef (Integ)
  int = "UNIV//intrel"  (quotient_def)
```

---

[495]The file should be contained in your Isabelle distribution. Or, if you only have an Isabelle executable, you can find the sources here:

http://isabelle.in.tum.de/library/

[496]Recall the general concept of an equivalence relation. Generally, for a set S and an equivalence relation R defined on the set, one can define $S//R$, the quotient of S w.r.t. R.

$$S//R = \{A \mid A \subseteq S \land \forall x, y \in A.(x, y) \in R\}$$

That is, one partitions the set S into subsets such that each subset collects equivalent elements. This is a standard mathematical concept.

We do not go into the Isabelle details here, but we explain how this works for the integers. One can view a pair (n, m) of natural numbers as representation of the integer $n - m$. But then (n, m) and (n', m') represent the same integer if and only if $n - m = n' - m'$, or equivalently, $n + m' = n' + m$.

Some **Theorems** in `IntArith`

`zminus_zadd_distrib`	$-(z + w) = -z + -w$
`zminus_zminus`	$-(-z) = z$
`zadd_ac`	$z1 + z2 + z3 = z1 + (z2 + z3)$
	$z + w = w + z$
	$x + (y + z) = y + (x + z)$
`zmult_ac`	$z1 * z2 * z3 = z1 * (z2 * z3)$
	$z * w = w * z$
	$z1 * (z2 * z3) = z2 * (z1 * z3)$

Compare to nat theorems.

In this case (n, m) and (n', m') are said to be equivalent. The construction of the integer type is based on this equivalence relation, called `intrel`. More precisely, the definition of the integers will be based on the set of all pairs of naturals (which corresponds to the `UNIV` constant on the type $nat \times nat$) modulo the equivalence `intrel`. In other words, it will be based on the quotient of the set of pairs of naturals w.r.t. `intrel`.

24.5 Further Number Theories

- Binary Integers (`Integ/Bin.thy`, for fast computation)

- Rational Numbers (`Real/PRat.thy`)

- Reals[497] (`Real/PReal.thy`: based on Dedekind-cuts of rationals [Ded])

- Hyperreals (`Real/RealDef.thy` for non-standard analysis)[498]

- Machine numbers (floats); see work for Intel's PentiumIV; built in HOL-light [Har98, Har00]

---

[497]The reals have been axiomatized by Dedekind by stating that a set R is partitioned into two sets A and B such that $R = A \cup B$ and for all $a \in A$ and $b \in B$, we have $a < b$. Now there is a number s such that $a \leq s \leq b$ for all $a \in A$ and $b \in B$. The irrational numbers are characterised by the fact that there exists exactly one such s. This axiomatization has been used as a basis for formalizing real numbers in Isabelle/HOL.
[498]In non-standard analysis, one works with sequences that are not necessarily converging. This is a relatively new field in mathematics and Isabelle/HOL has been successfully applied in it [FP98]. We just mention this here to say that Isabelle/HOL is used for "cutting-edge" mathematics and not just toy examples.

24.6 Conclusion on Arithmetic

Using conservative extensions in HOL, we can build

- the naturals (as type definition based on ind), and

- higher number theories (via equivalence construction).

24.6 Conclusion on Arithmetic

Using conservative extensions in HOL, we can build

- the naturals (as type definition based on ind), and

- higher number theories (via equivalence construction).

Potential for

- analysis of processor arithmetic units, and

- function analysis in HOL (combination with computer algebra systems such as Mathematica).

Future: analysis of hybrid systems[499].

The methodological overhead can be tackled by powerful mechanical support, since many proof-tasks are routine.

---

[499]Hybrid systems is a field in software engineering concerned with using finite automata for controlling physical systems such as ABS in cars etc.

25 Datatypes

The Roadmap

We are still looking at how the different parts of mathematics are encoded in the Isabelle/HOL library.

- Orders

- Sets

- Functions

- (Least) fixpoints and induction

- (Well-founded) recursion

- Arithmetic

- Datatypes

The Roadmap

We are still looking at how the different parts of mathematics are encoded in the Isabelle/HOL library.

- Orders

- Sets

- Functions

- (Least) fixpoints and induction

- (Well-founded) recursion

- Arithmetic

- Datatypes

What Are Datatypes?

We have seen types, but what are data[500]types?

---

[500]We have seen types, but what are datatypes?

First of all, a datatype must be of order 0, so it must be a non-functional type. Note that if we do not have polymorphism, this means that a datatype must be a in $\mathcal{B}$. But if we have polymorphism, it just means that the type must not contain $\rightarrow$. E.g., $\alpha\ list$ could be a datatype. However, when one describes a datatype, one would usually speak about generic instances such as $\alpha\ list$, and not about, say, $nat\ list$.

Secondly, the terms that inhabit a datatype τ must be defined using a finite set of term constructors that have τ as result type. At least one term constructor should just have type τ. E.g., $Nil : \alpha\ list$ and $Cons : \alpha \rightarrow (\alpha\ list) \rightarrow \alpha\ list$ are the term constructors that define the list datatype. One also finds a syntax where Nil is written $[]$ and $Cons$ is written $::$. Intuitively, we could say: the terms of a datatype are exactly the terms that can be constructed by some finite syntactic construction rule.

Whenever we have a term constructor that has τ as argu-

What Are Datatypes?

We have seen types, but what are data[500]types?

- Order 0 (no $\rightarrow$ in type).

- Terms defined by finite set of term constructors.

- Typically inductive definition.

- Term constructed by syntactic rule is unique.

Counterexample: $\alpha\ set$.

---

[500]We have seen types, but what are datatypes?

First of all, a datatype must be of order 0, so it must be a non-functional type. Note that if we do not have polymorphism, this means that a datatype must be a in $\mathcal{B}$. But if we have polymorphism, it just means that the type must not contain $\rightarrow$. E.g., $\alpha\ list$ could be a datatype. However, when one describes a datatype, one would usually speak about generic instances such as $\alpha\ list$, and not about, say, $nat\ list$.

Secondly, the terms that inhabit a datatype τ must be defined using a finite set of term constructors that have τ as result type. At least one term constructor should just have type τ. E.g., $Nil : \alpha\ list$ and $Cons : \alpha \rightarrow (\alpha\ list) \rightarrow \alpha\ list$ are the term constructors that define the list datatype. One also finds a syntax where Nil is written $[]$ and $Cons$ is written $::$. Intuitively, we could say: the terms of a datatype are exactly the terms that can be constructed by some finite syntactic construction rule.

Whenever we have a term constructor that has τ as argu-

Datatypes: Motivation

Last lecture: Construction of natural numbers.

How can we build "datatypes" (as in ML [Pau96])?

Can we base the construction of arbitrary datatypes on the natural numbers?

Datatypes: Motivation

: Construction of natural numbers.

How can we build "datatypes" (as in ML [Pau96])?

Can we base the construction of arbitrary datatypes on the natural numbers?

---

ment as well as result, the construction rule is inductive. E.g., we have

- Nil is a list;

- if t is a list h is of type α, then $Cons(h, t)$ is a list.

This is an inductive construction of lists. Usually, when one speaks about datatypes, one has inductively defined ones in mind. Examples are lists, natural numbers, trees. One could say that e.g. $bool$ is also a datatype defined by the constants $True$ and $False$, but it is not particularly interesting in this context.

At the same time, each term constructed by such a syntactic rule is unique. So if we say: lists are defined by the above inductive construction, then we imply that $Cons(1, Nil)$ must not be equal to $Cons(1, Cons(1, Nil))$.

To understand better the distinction of a datatype from another type, consider the following counterexample: $\alpha\ set$. Sets are not a datatype:

1. While the type $\alpha\ set$ does not contain an $\rightarrow$, it is isomorphic to $\alpha \rightarrow bool$ which does contain an $\rightarrow$.

2. The most basic way of defining "what a set is" is: if f is of type $\tau \rightarrow bool$, then $Abs_{set}\ f$ (alternatively: $Collect\ f$) is a set. This is not an inductive syntactic construction rule.

3. One could define sets similarly to lists by an inductive rule saying: $\{\}$ is a set; if S is a set and h is some term of type α, then $Insert(h, S)$ is a set. But then $Insert(1, \{\})$ would be different from $Insert(1, Insert(1, \{\}))$, which is not what we want! Moreover, we could not define infinite sets this way.

4. In point **??** we say: the definition of the terms called "sets" is not an inductive definition. This is not in contradiction to the inductive definition of particular sets. These inductive definitions have the form: If *foo* is in the set then *bar* is in the set, e.g., if n is in the set then $Suc\ n$

- Well, yes — by Gödelization[501]. But that wouldn't be automatable easily ...

---

is in the set. This is in contrast to what is suggested in point **??**, where we say: If *foo* is a set then *bar* is a set.

[501]In computation theory, Gödelization is the process of encoding words over arbitrary alphabets, or any arbitrary datastructures, for that matter, as natural numbers.

As an example, how could we encode a list of natural numbers as a single natural number. Prime factors are one technique for that. Suppose the number n has a factorization $2^i \cdot 3^j \cdot 5^k \cdot 7^l$. Then it would encode the list $[i, j, k, l]$.

But such an encoding for datatypes would be very cumbersome. In particular, there is no way of realizing polymorphism.
[502]This file should be contained in your Isabelle distribution. Or, if you only have an Isabelle executable, you can find the sources here:

http://isabelle.in.tum.de/library/

- Well, yes — by Gödelization[501]. But that wouldn't be automatable easily ...

- Instead: We use S-expressions [**?**] rather than nat.

Caveat: We will only sketch the construction and we will simplify, meaning that the technical details will not be strictly correct! See `Datatype_Universe.thy`[502] and [Wen99].

---

is in the set. This is in contrast to what is suggested in point **??**, where we say: If *foo* is a set then *bar* is a set.

[501]In computation theory, Gödelization is the process of encoding words over arbitrary alphabets, or any arbitrary datastructures, for that matter, as natural numbers.

As an example, how could we encode a list of natural numbers as a single natural number. Prime factors are one technique for that. Suppose the number n has a factorization $2^i \cdot 3^j \cdot 5^k \cdot 7^l$. Then it would encode the list $[i, j, k, l]$.

But such an encoding for datatypes would be very cumbersome. In particular, there is no way of realizing polymorphism.

[502]This file should be contained in your Isabelle distribution. Or, if you only have an Isabelle executable, you can find the sources here:

http://isabelle.in.tum.de/library/

S-Expressions as Basis

In the end we want to have datatypes such as lists and trees.

It turns out that LISP-like S-expressions are a datatype that is so rich that other datatypes can nicely be embedded in it.

Since we have not got the concept of datatype yet, we must first represent S-expressions using constructs we already have.

25.1 S-Expressions

LISP-like S-expressions[503] are a kind of of binary trees. We call the type $\alpha\ dtree$. This uses $\alpha + nat$.

---
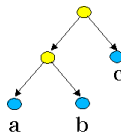
[503]The datastructure we have in mind here consists of binary trees where the inner nodes are not labeled, and the leaves are labeled

- either with a term of arbitrary type, in which case the leaf would be an actual "piece of content" in the datastructure,

- or with a natural number, in which case the leaf serves special purposes for organizing our datastructure, as we will see later.

I.e., such binary trees have a type parametrized by a type variable α, the type of the latter kind of leaves. Let us call the type of such trees $\alpha\ dtree$.

As always with parametric polymorphism, when we consider how the datastructure as such works, we are not interested in what the values in the former kind of leaves are. This is just like the type and values of list elements are irrelevant for concatenating two lists. Of course, α could, by coincidence,

This is encoded as a set of "nodes"[504] (defined by their path from the root and a value in the leaves), e.g.:

$$\{(\langle 0,0\rangle, a), (\langle 0,1\rangle, b), (\langle 1\rangle, c)\}$$

The type definition of α $dtree$ uses such an encoding.

be instantiated to type nat.

Think of a label of the first kind as content label and a label of the second kind as administration label.

Technically, if something is either of this type or of that type, we are talking about a sum type. So a leaf label has type $\alpha + nat$ (written (α, nat) sum before), and it has the form either $Inl(a)$ for some $a :: \alpha$, or $Inr(n)$ for some $n :: nat$.

[504] The set

$$\{(\langle 0,0\rangle, a), (\langle 0,1\rangle, b), (\langle 1\rangle, c)\}$$

represents the tree
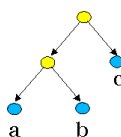
The path $\langle 0,0\rangle$ means: from the root take left subtree, then again left subtree. The path $\langle 1\rangle$ means: take right subtree.
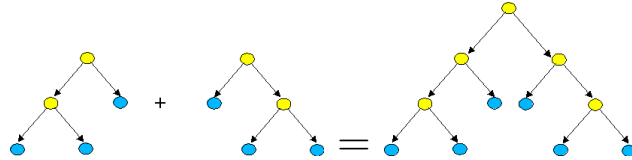
How can a path $\langle p_0, \ldots, p_n\rangle$ be represented? One idea is

Building Trees

- $Atom(n)^{505}$

  n

- $Scons\ X\ Y^{506}$

  + =
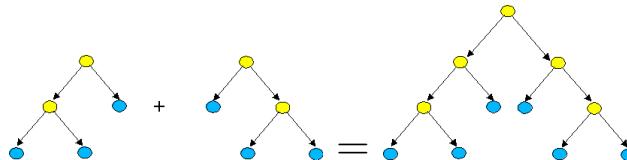
to use the function $f :: nat \Rightarrow nat$ defined by

$$f\ i = \begin{cases} p_i & \text{if } i \leq n \\ 2 & \text{otherwise} \end{cases}$$

as representation of $\langle p_0, \ldots, p_n \rangle$.

[505] $Atom$ takes a leaf label and turns it into a (simplest possible) S-expression (tree).

So it has type $\alpha + nat \Rightarrow \alpha\ dtree$.

[506] $Scons$ takes two S-expressions and creates a new S-expression as illustrated below:
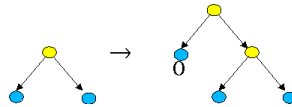
  + =

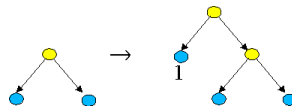So it has type $[\alpha\ dtree, \alpha\ dtree] \Rightarrow \alpha\ dtree$.

Tagging Trees

We want to tag an S-expression by either 0 or 1. This can be done by *"Scons"*-ing it with an S-expression consisting of an administration label. By convention, the tag is to the left.

- **In0_def** $In0(X) \equiv Scons\ Atom(Inr(0))(X)$

- **In1_def** $In1(X) \equiv Scons\ Atom(Inr(1))(X)$

Products and Sums on Sets of S-Expressions

Product of two sets A and B of S-expressions: All $Scons$-trees where left subtree from A, right subtree from B.

$$\texttt{uprod\_def} \quad uprod\ A\ B \equiv \bigcup_{x \in A} \bigcup_{y \in B} \{(Scons\ x\ y)\}$$

---

[507] Recall that ' denotes the image of a function applied to a set.

Products and Sums on Sets of S-Expressions

Product of two sets A and B of S-expressions: All $Scons$-trees where left subtree from A, right subtree from B.

$$\texttt{uprod\_def} \quad uprod\ A\ B \equiv \bigcup_{x \in A} \bigcup_{y \in B} \{(Scons\ x\ y)\}$$

Sum of two sets A and B of S-expressions: union of A and B after S-expressions in A have been tagged 0 and S-expressions in B have been tagged 1, so that one can tell where they come from.

$$\texttt{usum\_def} \quad usum\ A\ B \equiv In0\ {}^{\iota\,507}A \cup In1\ {}^{\iota}B$$

---

[507] Recall that ι denotes the image of a function applied to a set.

Some Properties of Trees and Tree Sets

- $Atom$, $In0$, $In1$, $Scons$ are[508] injective.

- $Atom$ and $Scons$ are pairwise distinct. $In0$ are $In1$ pairwise distinct.

---

[508]This means that any of $Atom$, $In0$, $In1$, $Scons$ applied to different S-expressions will return different S-expressions.

Moreover, a term with root $Scons$ is definitely different from a term with root $Atom$, and a term with root $In0$ is definitely different from a term with root $In1$.

Why is this important? It is an inherent characteristic of a datatype. A datatype consists of terms constructed using term constructors and is uiquely defined by what it is syntactically (one also says that terms are generated freely using the constructors). For example, injectivity of Suc and pairwise-distinctness of 0 and Suc mean for any two numbers m and n, the terms $\underbrace{Suc(\ldots Suc}_{m \text{ times}}(0)\ldots)$ and $\underbrace{Suc(\ldots Suc}_{n \text{ times}}(0)\ldots)$ are different.

[509]Given a set T of trees (S-expressions), the closure of T under $Atom$, $In0$, $In1$, $Scons$, $usum$, $uprod$ is the smallest set T' such that $T \subseteq T'$ and given any tree (or two trees, as applicable) from T', any tree constructable using $Atom$,

594

Some Properties of Trees and Tree Sets

- $Atom$, $In0$, $In1$, $Scons$ are[508] injective.

- $Atom$ and $Scons$ are pairwise distinct. $In0$ are $In1$ pairwise distinct.

- Tree sets represent a universe that is closed under products and sums: $usum$, $uprod$ have type $[(\alpha\ dtree)\ set, (\alpha\ dtree)\ set] \Rightarrow (\alpha\ dtree)\ set$.

- $uprod$ and $usum$ are monotone.

---

[508]This means that any of $Atom$, $In0$, $In1$, $Scons$ applied to different S-expressions will return different S-expressions.

Moreover, a term with root $Scons$ is definitely different from a term with root $Atom$, and a term with root $In0$ is definitely different from a term with root $In1$.

Why is this important? It is an inherent characteristic of a datatype. A datatype consists of terms constructed using term constructors and is uiquely defined by what it is syntactically (one also says that terms are generated freely using the constructors). For example, injectivity of Suc and pairwise-distinctness of 0 and Suc mean for any two numbers m and n, the terms $\underbrace{Suc(\ldots Suc}_{m \text{ times}}(0)\ldots)$ and $\underbrace{Suc(\ldots Suc}_{n \text{ times}}(0)\ldots)$ are different.

[509]Given a set T of trees (S-expressions), the closure of T under $Atom$, $In0$, $In1$, $Scons$, $usum$, $uprod$ is the smallest set T' such that $T \subseteq T'$ and given any tree (or two trees, as applicable) from T', any tree constructable using $Atom$,

Some Properties of Trees and Tree Sets

- $Atom$, $In0$, $In1$, $Scons$ are[508] injective.

- $Atom$ and $Scons$ are pairwise distinct. $In0$ are $In1$ pairwise distinct.

- Tree sets represent a universe that is closed under products and sums: $usum$, $uprod$ have type $[(\alpha\ dtree)\ set, (\alpha\ dtree)\ set] \Rightarrow (\alpha\ dtree)\ set$.

- $uprod$ and $usum$ are monotone.

- Tree sets represent a universe that is closed under products and sums[509] combined with arbitrary applications of lfp.

Reminder: we simplified!

---

[508]This means that any of $Atom$, $In0$, $In1$, $Scons$ applied to different S-expressions will return different S-expressions.

Moreover, a term with root $Scons$ is definitely different from a term with root $Atom$, and a term with root $In0$ is definitely different from a term with root $In1$.

Why is this important? It is an inherent characteristic of a datatype. A datatype consists of terms constructed using term constructors and is uiquely defined by what it is syntactically (one also says that terms are generated freely using the constructors). For example, injectivity of Suc and pairwise-distinctness of 0 and Suc mean for any two numbers m and n, the terms $\underbrace{Suc(\ldots Suc}_{m \text{ times}}(0)\ldots)$ and $\underbrace{Suc(\ldots Suc}_{n \text{ times}}(0)\ldots)$ are different.

[509]Given a set T of trees (S-expressions), the closure of T under $Atom$, $In0$, $In1$, $Scons$, $usum$, $uprod$ is the smallest set T' such that $T \subseteq T'$ and given any tree (or two trees, as applicable) from T', any tree constructable using $Atom$,

594

25.2 Lists in Isabelle

Similar to the construction of nat, we first construct a set of S-expressions having the "structure of lists". We start by defining "provisional" list constructors:

```
constdefs
  NIL :: 'a dtree
  "NIL == In0(Atom(Inr(0)))"
  CONS ::  ['a dtree, 'a dtree] => 'a dtree
  "CONS M N == In1(Scons M N)"
```

What type do you expect[510] $Cons$ to have, and how does $CONS$ compare?

25.2 Lists in Isabelle

Similar to the construction of nat, we first construct a set of S-expressions having the "structure of lists". We start by defining "provisional" list constructors:

```
constdefs
   NIL :: 'a dtree
   "NIL == In0(Atom(Inr(0)))"
   CONS ::  ['a dtree, 'a dtree] => 'a dtree
   "CONS M N == In1(Scons M N)"
```

What type do you expect[510] $Cons$ to have, and how does $CONS$ compare? Must wrap list elements by $Atom \circ Inl$.

$In0$, $In1$, $Scons$, $usum$, $uprod$ is also contained in T'.

Remembering the construction of inductively defined sets, the closure is the least fixpoint of a monotone function adding trees to a tree set. This function must be constructed using $Atom$, $In0$, $In1$, $Scons$, $usum$, $uprod$. We do not go into the details, but note that it is crucial that $uprod$ and $usum$ are monotone, and note as well that slight complications arise from the fact that $usum$ and $uprod$ have type $[(\alpha\ dtree)\ set, (\alpha\ dtree)\ set] \Rightarrow (\alpha\ dtree)\ set$ rather than $(\alpha\ dtree)\ set \Rightarrow (\alpha\ dtree)\ set$.

[510] $Cons$ should have the polymorphic type $[\alpha, \alpha\ list] \Rightarrow \alpha\ list$. The important point is: the first argument is of different type than the second argument. If the first is of type τ, then the second must be of type $\tau\ list$.

In contrast, $CONS$ is of type $[(\alpha\ dtree), (\alpha\ dtree)] \Rightarrow \alpha\ dtree$.

In order to apply $CONS$ to a "list" (in fact an S-expression) and a "list element", we must first wrap the list element by

Lists as S-Expressions: Intuition

Examples of how lists would be represented as S-expressions:

Nil^{511}	$[]$
$Cons(7, Nil)$	$[7]$
$Cons(5, Cons(7, Nil))$	$[5, 7]$

$Atom \circ Inl$, so that it becomes an S-expression.

[511] Nil, $Cons(7, Nil)$, $Cons(5, Cons(7, Nil))$ are lists written according to what some programming languages introduce as the first, "official" syntax for lists.

For convenience, programming languages typically allow for the same lists to be written as $[]$, $[7]$, $[5, 7]$.

Lists as S-Expressions: Intuition

Examples of how lists would be represented as S-expressions:

Nil[511]	$[]$
$In0(Atom(Inr\ 0))$	
$Cons(7, Nil)$	$[7]$
$Cons(5, Cons(7, Nil))$	$[5, 7]$

$Atom \circ Inl$, so that it becomes an S-expression.

[511] Nil, $Cons(7, Nil)$, $Cons(5, Cons(7, Nil))$ are lists written according to what some programming languages introduce as the first, "official" syntax for lists.

For convenience, programming languages typically allow for the same lists to be written as $[]$, $[7]$, $[5, 7]$.

Lists as S-Expressions: Intuition

Examples of how lists would be represented as S-expressions:

Nil^{511} 　　　　　　　　　　　　$[]$

$In0(Atom(Inr\ 0))$

$Cons(7, Nil)$ 　　　　　　　$[7]$

$CONS\ (Atom(Inl\ 7))\ In0(Atom(Inr\ 0))$

$Cons(5, Cons(7, Nil))$ 　$[5, 7]$

$Atom \circ Inl$, so that it becomes an S-expression.

[511] $Nil,\ Cons(7, Nil),\ Cons(5, Cons(7, Nil))$ are lists written according to what some programming languages introduce as the first, "official" syntax for lists.

For convenience, programming languages typically allow for the same lists to be written as $[], [7], [5, 7]$.

Lists as S-Expressions: Intuition

Examples of how lists would be represented as S-expressions:

Nil^{511} $\qquad\qquad\qquad$ $[]$

$\qquad In0(Atom(Inr\ 0))$

$Cons(7, Nil)$ $\qquad\qquad$ $[7]$

$\qquad CONS\ (Atom(Inl\ 7))\ In0(Atom(Inr\ 0))$

$Cons(5, Cons(7, Nil))$ $\ \ [5, 7]$

$\qquad CONS\ (Atom(Inl\ 5))$

$\qquad\quad (CONS\ (Atom(Inl\ 7))\ In0(Atom(Inr\ 0)))$

Now let's construct the S-expressions having this form.

$Atom \circ Inl$, so that it becomes an S-expression.

[511] $Nil,\ Cons(7, Nil),\ Cons(5, Cons(7, Nil))$ are lists written according to what some programming languages introduce as the first, "official" syntax for lists.

For convenience, programming languages typically allow for the same lists to be written as $[],\ [7],\ [5, 7]$.

Lists as S-Expressions: Inductive Construction

Idea: let $A :: (\alpha\ dtree)\ set$ be the set of all "wrapped" elements, e.g. for $\alpha = nat$, the set $\{(Atom\ Inl\ 0), (Atom\ Inl\ 1), \ldots\}$. Then define $list(A)$, the set of S-expressions that represent lists of element type α:

```
list       :: "'a dtree set => 'a dtree set"
inductive "list(A)"
 intrs
  NIL_I   "NIL : list(A)"
  CONS_I  "[|a : A; M : list(A) |] ==>
          CONS a M : list(A)"
```

See `SList.thy`[512] for how it's really done!

---

[512]This file should be contained in your Isabelle distribution. Or, if you only have an Isabelle executable, you can find the sources here:

http://isabelle.in.tum.de/library/

Defining the "Real" List Type

We now apply the type definition mechanism using the `typedef` syntax. How do we define A formally?

Defining the "Real" List Type

We now apply the type definition mechanism using the `typedef`
syntax. How do we define A formally?

```
typedef (List)
  'a list =
    "list(range (Atom o Inl)) :: 'a dtree set"
  by ...
```

Choosing A as $range\,(Atom \circ Inl)$ together with the ex-
plicit type declaration forces A to be the set containing all
$Atom\,(Inl\,t)$, for each $t :: \alpha$.

   Example of a definition of a polymorphic type.

List Constructors

We define the real constructor names for lists:

```
Nil_def  "Nil::'a list == Abs_list(NIL)"
Cons_def "x#(xs::'a list) ==
   Abs_list(CONS (Atom(Inl(x))) (Rep_list xs))"
```

We then forget about *NIL* and *CONS*.

Isabelle's Datatype Package

Similar to the `typedef syntax`, Isabelle provides the `datatype` syntax to support the construction of a datatype:

```
datatype 'a list = Nil | Cons 'a ('a list)
```

In particular, this automates the proofs of:

- the induction theorem;

- distinctness;

- injectivity of constructors.

---

[513] The `datatype` syntax is very convenient since the complex construction we have seen today is transparent to the normal user.

In particular, proofs of the induction theorem are automated. This is in contrast to the construction of nat where this theorem was not generated automatically.

So why didn't we use the `datatype` syntax to define nat, since it is so much more convenient?

The reason is that we needed nat to define S-expressions, so the type nat must exist before there can be a datatype package, and so the datatype package cannot be used to define nat.

Isabelle's Datatype Package

Similar to the `typedef` syntax, Isabelle provides the `datatype` syntax to support the construction of a datatype:

```
datatype ’a list = Nil | Cons ’a (’a list)
```

In particular, this automates the proofs of:

- the induction theorem;

- distinctness;

- injectivity of constructors.

The package also works for mutually and indirectly recursive datatype definitions.

Question: Why didn't we use this package to define nat[513]?

---

[513]The `datatype` syntax is very convenient since the complex construction we have seen today is transparent to the normal user.

In particular, proofs of the induction theorem are automated. This is in contrast to the construction of nat where this theorem was not generated automatically.

So why didn't we use the `datatype` syntax to define nat, since it is so much more convenient?

The reason is that we needed nat to define S-expressions, so the type nat must exist before there can be a datatype package, and so the datatype package cannot be used to define nat.

26 Summary of HOL Library / Outlook on Modeled Systems

Summary

In the previous weeks, we looked at how the different parts of mathematics are encoded in the Isabelle/HOL library:

- Orders

- Sets

- Functions

- (Least) fixpoints and induction

- (Well-founded) recursion

- Arithmetic

- Datatypes

Summary (Cont.)

We conclude: HOL is a logical framework for theoretical computer science. Its features are:

- a clean methodology, which can be supported automatically to a surprising extent;

- a powerful set theory and proof support;

- adequate theories for arithmetics (proof-support: not quite satisfactory so far);

- a package for induction;

- a package for recursion;

- a package for datatypes.

Outline

We will now look at how various formalisms (specification and programming languages) can be embedded in HOL:

- Z and data-refinement

- Imperative languages

- Denotational semantics and functional languages

- Object-oriented languages (Java-Light ...)

27 A Taste of some Isabelle and HOL Applications

Just a few Isabelle or HOL Applications

We briefly introduce two Isabelle/HOL applications, and one application of HOL Light:

- Java bytecode verification;

- floating-point arithmetic;

- red-black trees.

This is just to stimulate you to look for more applications on your own!

27.1 Java Bytecode Verification

Typically, Java programs are delivered as bytecode, as opposed to source code on the one hand and machine code on the other hand. Bytecode is machine-independent.

A Java runtime system provides the Java Virtual Machine, i.e., an interpreter for Java bytecode.

Java is a typed language: the type system forbids things like pointer arithmetic, thus preventing illegal[514] memory access.

However, bytecode is not type-safe by itself. For various reasons, bytecode could be corrupted. This is obviously critical for security and possibly safety.

---

[514]By "illegal memory access", we mean access to regions not assigned to the program.

Ensuring Type Safety

The loader of a typical JVM has a bytecode verifier: A program that checks whether bytecode is type-safe.

Klein and Nipkow have specified a JVM and a bytecode verifier in Isabelle and proved its correctness using Isabelle [KN03, Nip03].

Such applications may have big impact since they are concerned with the correctness of not just some particular program, but rather the programming language (implementation) itself.

JavaCard

JavaCard is a subset of Java employed on smart cards. Aspects in contrast to full Java:

- Memory on smart cards is limited[515].

- Security is vital for smart card applications (banking etc.).

Project Verificard concerned with ensuring reliability of smart card applications.

Verificard @ Munich have applied the work on bytecode verification (using Isabelle) to JavaCard.

End user panel includes Ericsson, France Télécom R&D, and Gemplus.

---

[515]The memory on smart cards is limited. A full-fledged bytecode verifier would be too large/slow. One approach to tackling this problem is to work with bytecode programs with type annotations. Checking if a bytecode program is consistent with its type annotations is a much simpler task than computing these type annotations, which is what a bytecode verifier is supposed to do. The task can therefore be performed on a smart card more easily than full bytecode verification.

27.2 Floating Point Arithmetic

John Harrison has done much work on verifying arithmetic functions operating on various number types adhering to certain standards [Har98, Har99, Har00].

He has used HOL Light, not Isabelle. This means: no metalogic, specialized theorem prover for HOL.

He formally proved that the floating point operations of an Intel processor behave according to the IEEE standard 754 [IEE85]. First machine-checked proof of this kind.

We briefly review his work [Har99] using an Isabelle-like syntax where helpful.

What Are Floats?

Conventionally: floats have the form $\pm 2^e \cdot k$.

e is called exponent, $E_{min} \leq e \leq E_{max}$.

k is called mantissa, can be represented with p bits.

Floats in HOL

For formalization in HOL, equivalent representation

$$(-1)^s \cdot 2^{e-N} \cdot k$$

with $k < 2^p$ and $0 \le e < E$.

Thus a particular float format is characterized by maximal exponent E, precision p, and exponent offset ("ulpscale") N. The set of real numbers representable by a triple is:

$$format\ (E, p, N) =$$
$$\{x \mid \exists s\, e\, k.\ s < 2 \wedge e < E \wedge k < 2^p \wedge x = (-1)^s \cdot 2^e \cdot k/2^N\}$$

Rounding

Rounding takes a real to a representable real nearby. E.g. rounding up:

$$round\ fmt\ x = \epsilon a.\ a \in format\ fmt \wedge a \leq x \wedge$$
$$\forall b \in format\ fmt.\ b \leq x \rightarrow b \leq a$$

Formalization of the Standard [IEE85].

Useful lemmas such as:

$$x \leq y \implies round\ fmt\ x \leq round\ fmt\ y$$
$$a \in format\ fmt \wedge b \in format\ fmt \wedge 0.5 \leq \frac{a}{b} \leq 2 \implies$$
$$(b - a) \in format\ fmt$$

Operations

For operations such as addition, multiplication etc., it is proven in HOL that they behave as if they computed the exact result and rounded afterwards.

However, there are some debatable questions related to the sign of zeros.

27.3 Red-Black Trees

Red-black trees are trees that can be used for implementing sets/dictionaries, just like AVL trees. To formulate "balanced-ness" invariants, nodes are colored:

1. Every red node has a black parent.

2. Each path from the root to a leaf has the same number of black nodes.

Together these invariants ensure that maximal paths can differ in length by at most factor 2.

These invariants must be maintained by insertion and deletion operations.

Red-Black Trees in SML

Red-black trees provided in New Jersey SML library [Pau96].

Angelika Kimmig[516] tried to verify the insertion operation of red-black trees using Isabelle. Findings?

---

[516]Angelika Kimmig is a student who took this course in Wintersemester 02/03 in Freiburg. She then continued working with Isabelle in a Studienarbeit (a project required by computer science students in Freiburg).

Red-Black Trees in SML

Red-black trees provided in New Jersey SML library [Pau96].
   Angelika Kimmig[516] tried to verify the insertion operation of red-black trees using Isabelle. Findings?

- There is a mistake in the implementation of red-black trees in New Jersey SML! Insertion may lead to a violation of the first invariant, since the root may become red.

- As long as one just inserts, this is just a slight constant deterioration.

- Angelika has suggested a fix and proven the correctness of red-black tree insertion using Isabelle.

---

[516]Angelika Kimmig is a student who took this course in Wintersemester 02/03 in Freiburg. She then continued working with Isabelle in a Studienarbeit (a project required by computer science students in Freiburg).

Node Deletion

- **Deletion** is also wrongly implemented!

- With deletion, not just the root can become red, but the tree coloring can become completely wrong.

- Angelika has an idea for fixing deletion as well, but no proof (yet?).

Read the Studienarbeit for more details [Kim03]!

References

[Acz77] Peter Aczel. *Handbook of Mathematical Logic*, chapter An Introduction to Inductive Definitions, pages 739–782. North-Holland, 1977.

[AHMP92] Arnon Avron, Furio Honsell, Ian A. Mason, and Robert Pollack. Using typed lambda calculus to implement formal systems on a machine. *Journal of Automated Reasoning*, 9(3):309–354, 1992.

[And02]   Peter B. Andrews. *An Introduction to Mathematical Logic and Type Theory: To Truth Through Proofs*. Kluwer Academic Publishers, 2002. Second Edition.

[Apt97]   Krzysztof R. Apt. *From Logic Programming to Prolog*. Prentice Hall, 1997.

[Ari]   Aristotle. *Analytica priora I*, chapter 4.

[Ber19]   Paul Bernays. *?? ??*, 19??

[BM00]   David A. Basin and Seàn Matthews. Structuring metatheory on inductive definitions. *Information and Computation*, 162(1-2):80–95, 2000. Download.

[BN98]   Franz Baader and Tobias Nipkow. *Term Rewriting and* All That. Cambridge University Press, 1998.

[Can18] Georg Cantor. *?? ??*, 18??

[Chu40] Alonzo Church. A formulation of the simple theory of types. *Journal of Symbolic Logic*, 5:56–68, 1940.

[dB80]   Nicolaas G. de Bruijn. A survey of the project AUTOMATH. In *Essays in Combinatory Logic, Lambda Calculus, and Formalism*. Academic Press, 1980.

[Ded]   Richard Dedekind. *Cuts of Rationals*. ??, ??

[Des16]   Rene Descartes. *??* ??, 16??

[Dev93]   Keith Devlin. *The Joy of Sets. Fundamentals of Contemporary Set Theory*. Undergraduate Texts in Mathematics. Springer-Verlag, 1993.

[Ebb94]   Heinz-Dieter Ebbinghaus. *Einführung in die Mengenlehre*. BI-Wissenschaftsverlag, 1994.

[Fit96]   M. Fitting. *First-order Logic and Automated Theorem Proving*. Springer-Verlag, 1996.

[FP98]   Jacques D. Fleuriot and Lawrence C. Paulson. A combination of nonstandard analysis and geometry theorem proving, with application to newton's principia. In Claude Kirchner and Hélène Kirch-

ner, editors, *Proceedings of the 15th CADE*, volume 1421 of *LNCS*, pages 3–16. Springer-Verlag, 1998.

[Frä19]   Adolf Fränkel. *?? ??*, 19??

[Fre93]   Gottlob Frege. *Grundgesetze der Arithmetik*, volume I. Verlag Hermann Pohle, 1893. Translated in part in [Fur64].

[Fre03]   Gottlob Frege. *Grundgesetze der Arithmetik*, volume II. Verlag Hermann Pohle, 1903. Translated in part in [Fur64].

[Fur64]   Montgomery Furth. *The Basic Laws of Arithmetic*. Berkeley: University of California Press, 1964. Translation of [Fre03].

[Gen35]   Gerhard Gentzen. Untersuchungen über das lo-
            gische Schliessen. *Mathematische Zeitschrift*,
            39:176–210, 405–431, 1935. English translation
            in [Sza69].

[GLT89]   Jean-Yves Girard, Yves Lafont, and Paul Taylor.
            Proofs and Types. Cambridge University Press,
            1989.

[GM93]   Michael J. C. Gordon and Tom F. Melham, edi-
            tors. *Introduction to HOL*. Cambridge University
            Press, 1993.

[Göd31]   Kurt Gödel. Über formal unentscheidbare Sätze
            der Principia Mathematica und verwandter Sys-
            teme. *Monatshefte für Mathematik und Physik*,
            38:173–198, 1931.

[Har98]   John Harrison. *Theorem Proving with the Real Numbers*. Springer-Verlag, 1998.

[Har99]   John Harrison. A machine-checked theory of floating point arithmetic. In Yves Bertot, Gilles Dowek, André Hirschowitz, C. Paulin, and Laurent Théry, editors, *Proceedings of the 12th TPHOLs*, volume 1690 of *LNCS*, pages 113–130. Springer-Verlag, 1999.

[Har00]   John Harrison. Formal verification of the IA/64 division algorithms. In Mark Aagaard and John Harrison, editors, *Proceedings of the 13th TPHOLs*, volume 1869 of *LNCS*, pages 233–251. Springer-Verlag, 2000.

[HC68]   George E. Hughes and Maxwell John Cresswell.

An Introduction to Modal Logic. Muthuen and Co. Ltd, London, 1968.

[Hen50]   Henkin. ?? *??*, 1950.

[HHP93]   Robert Harper, Furio Honsell, and Gordon D. Plotkin. A framework for defining logics. *JACM*, 40(1):143–184, 1993.

[HHPW96] Cordelia V. Hall, Kevin Hammond, Simon L. Peyton Jones, and Philipp Wadler. Type classes in Haskell. *ACM Transactions on Programming Languages and Systems*, 18(2):109–138, 1996.

[Höl90]   Steffen Hölldobler. Conditional equational theories and complete sets of transformations. *Theoretical Computer Science*, 75(1&2):85–110, 1990.

[HP93]   G. Huet and G. Plotkin, editors. *Logical Environments*. Cambridge University Press, 1993.

[HR04]   Michael Huth and Mark Ryan. *Logic in Computer Science. Modelling and Reasoning about Systems*. Cambridge University Press, 2nd edition edition, 2004.

[HS86]   J. Roger Hindley and Jonathan P. Seldin. *Introduction to Combinators and λ-Calculus*. Cambridge University Press, 1986. Republished in [HS90].

[HS90]   J. Roger Hindley and Jonathan P. Seldin. *Introduction to Combinators and λ-Calculus*. Cambridge University Press, 1990.

[Hué]   Gerard Huét. ?? *??*, ??

[IEE85]   The Institute of Electrical and Electronic Engineers, Inc. *IEEE. Standard for binary floating point arithmetic. ANSI/IEEE Standard 754-1985*, 1985.

[Kim03]   Angelika Kimmig. Red-black trees of slmnj. Studienarbeit at Universität Freiburg, Download, 2003.

[Klo93]   Jan Willem Klop. *Handbook of Logic in Computer Science*, chapter "Term Rewriting Systems". Oxford: Clarendon Press, 1993.

[KN03]   Gerwin Klein and Tobias Nipkow. Verified bytecode verifiers. *Theoretical Computer Science*, 3(298):583–626, 2003.

[LP81]   Harry R. Lewis and Christos H. Papadimitriou. *Elements of the Theory of Computation*. Prentice-Hall, 1981.

[Mil78]   Robin Milner. A theory of type polymorphism in programming. *Journal of Computer and System Sciences*, 17(3):348–375, 1978.

[Min00]   Grigori Mints. *A Short Introduction to Intuitionistic Logic*. Kluwer Academic/Plenum Publishers, 2000.

[Nip93]   Tobias Nipkow. *Order-Sorted Polymorphism in Isabelle*, pages 164–188. Cambridge University Press, 1993. In [HP93].

[Nip03]   Tobias Nipkow. Java bytecode verification. *Journal of Automated Reasoning*, 30(3-4):233–233, 2003.

[NN99]   Wolfgang Naraschewski and Tobias Nipkow. Type inference verified: Algorithm $\mathcal{W}$ in Isabelle/HOL. *Journal of Automated Reasoning*, 23(3-4):299–318, 1999.

[NP93]   Tobias Nipkow and Christian Prehofer. Type checking type classes. In *Proceedings of the 20th ACM Symposium Principles of Programming Languages*, pages 409–418. ACM Press, 1993.

[Pau89]   Lawrence C Paulson. The foundation of a generic theorem prover. *Journal of Automated Reasoning*, 5(3):363–397, 1989.

[Pau94]   Lawrence C. Paulson. *Isabelle: A Generic Theorem Prover*, volume 828 of *LNCS*. Springer, 1994.

[Pau96]   Lawrence C. Paulson. *ML for the Working Programmer*. Cambridge University Press, 1996.

[Pau97]   Lawrence C. Paulson. Generic automatic proof tools. In Robert Veroff, editor, *Automated Reasoning and its Applications: Essays in Honor of Larry Wos*, chapter 3. MIT Press, 1997.

[Pau05]   Lawrence C. Paulson. *The Isabelle Reference Manual*. Computer Laboratory, University of Cambridge, October 2005.

[Pea18]   Guiseppe Peano. *??* ??, 18??

[PM68]   Dag Prawitz and Per-Erik Malmnäs. A survey of some connections between classical, intuitionistic and minimal logic. In A. Schmidt and H. Schütte, editors, *Contributions to Mathematical Logic*, pages 215–229. North-Holland, 1968.

[Pra65]   Dag Prawitz. *Natural Deduction: A proof theoretical study*. Almqvist and Wiksell, 1965.

[Pra71]   Dag Prawitz. Ideas and results in proof theory. In Jens Erik Fenstad, editor, *Proceedings of the Second Scandinavian Logic Symposium*, pages 235–308. North-Holland, 1971.

[SH84]   Peter Schroeder-Heister. A natural extension of natural deduction. *Journal of Symbolic Logic*, 49(4):1284–1300, 1984.

[Sza69]   M. E. Szabo. *The Collected Papers of Gerhard Gentzen*. North-Holland, 1969.

[Tho91]   Simon Thompson. *Type Theory and Functional Programming*. Addison-Wesley, 1991.

[Tho95a]   Della Thompson, editor. *The Concise Oxford Dictionary*. Clarendon Press, 1995.

[Tho95b]   Simon Thompson. *Miranda: The Craft of Functional Programming*. Addison-Wesley, 1995.

[Tho99]   Simon Thompson. *Haskell: The Craft of Functional Programming*. Addison-Wesley, 1999. Second Edition.

[vD80]   Dirk van Dalen. *Logic and Structure*. Springer-Verlag, 1980. An introductory textbook on logic.

[Vel94] Daniel J. Velleman. *How to Prove It*. Cambridge University Press, 1994.

[vH67] Jean van Heijenoort, editor. *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-193*. Harvard University Press, 1967. Contains translations of original works by David Hilbert.

[vL16] Gottfried Wilhelm von Leibniz. *?? ??*, 16??

[WB89] Phillip Wadler and Stephen Blott. How to make ad-hoc polymorphism less ad-hoc. In *Conference Record of the 16th ACM Symposium on Principles of Programming Languages*, pages 60–76, 1989.

[Wen99] Markus Wenzel. Inductive datatypes in HOL - lessons learned in formal-logic engineering. In Yves Bertot, Gilles Dowek, André Hirschowitz,

and and Laurent Théry C. Paulin, editors, *Proceedings of TPHOLs*, volume 1690 of *LNCS*, pages 19–36. Springer-Verlag, 1999.

[WR25]   Alfred N. Whitehead and Bertrand Russell. *Principia Mathematica*, volume 1. Cambridge University Press, 1925. 2nd edition.

[Zer19]   Ernst Zermelo. *?? ??*, 19??