Computer Supported Modeling and Reasoning

David Basin, Achim D. Brucker, Jan-Georg Smaus, and Burkhart Wolff

April 2005

http://www.infsec.ethz.ch/education/permanent/csmr/

First-Order Logic: Natural Deduction

David Basin, Burkhart Wolff, and Jan-Georg Smaus

First-Order Logic: Deductive System

First-order logic is a generalization of propositional logic. All the rules of propositional logic carry over.

But we must introduce rules for the quantifiers.

Universal Quantification (∀): 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 $*$ means: x must be arbitrary.

Universal Quantification (∀): 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 $*$ means: x must be arbitrary.

Note that rules are schematic.

Universal Quantification: Side Condition

What does arbitrary mean? Consider the following "proof"

$$x = 0$$

Universal Quantification: Side Condition

What does arbitrary mean? Consider the following "proof"

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

Universal Quantification: Side Condition

What does arbitrary mean? Consider the following "proof"

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

Universal Quantification: Side Condition

What does arbitrary mean? Consider the following "proof"

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

Universal Quantification: Side Condition

What does arbitrary mean? Consider the following "proof"

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

Universal Quantification: Side Condition

What does arbitrary mean? Consider the following "proof"

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

Universal Quantification: Side Condition

What does arbitrary mean? Consider the following "proof"

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

Universal Quantification: Side Condition

What does arbitrary mean? Consider the following "proof"

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

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

Another Proof? (1)

Is the following a proof? Is the conclusion valid?

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

Another Proof? (1)

Is the following a proof? Is the conclusion valid?

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

Conclusion is not valid.

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

Another Proof? (1)

Is the following a proof? Is the conclusion valid?

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

Proof is incorrect.

Reason: Substitution must avoid capturing 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.

Another Proof? (2)

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

Another Proof? (2)

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

Another Proof? (2)

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

Another Proof? (2)

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

Another Proof? (2)

$$\dfrac{\dfrac{\dfrac{\dfrac{\forall x.\, A(x) \wedge B(x)}{A(x) \wedge B(x)}\,\forall\text{-}E}{A(x)}\,\wedge\text{-}EL}{\forall x.\, A(x)}\,\forall\text{-}I \qquad \dfrac{\dfrac{\dfrac{\dfrac{\forall x.\, A(x) \wedge B(x)}{A(x) \wedge B(x)}\,\forall\text{-}E}{B(x)}\,\wedge\text{-}ER}{\forall x.\, B(x)}\,\forall\text{-}I$$

Another Proof? (2)

$$\dfrac{\dfrac{\dfrac{\dfrac{\forall x.\, A(x) \wedge B(x)}{A(x) \wedge B(x)}\,\forall\text{-}E}{A(x)}\,\wedge\text{-}EL}{\forall x.\, A(x)}\,\forall\text{-}I \qquad \dfrac{\dfrac{\dfrac{\forall x.\, A(x) \wedge B(x)}{A(x) \wedge B(x)}\,\forall\text{-}E}{B(x)}\,\wedge\text{-}ER}{\forall x.\, B(x)}\,\forall\text{-}I}{(\forall x.\, A(x)) \wedge (\forall x.\, B(x))}\,\wedge\text{-}I$$

Another Proof? (2)

$$\dfrac{\dfrac{\dfrac{\dfrac{[\forall x.\, A(x) \wedge B(x)]^1}{A(x) \wedge B(x)}\,\forall\text{-}E}{A(x)}\,\wedge\text{-}EL}{\forall x.\, A(x)}\,\forall\text{-}I \qquad \dfrac{\dfrac{\dfrac{[\forall x.\, A(x) \wedge B(x)]^1}{A(x) \wedge B(x)}\,\forall\text{-}E}{B(x)}\,\wedge\text{-}ER}{\forall x.\, B(x)}\,\forall\text{-}I}{(\forall x.\, A(x)) \wedge (\forall x.\, B(x))}\,\wedge\text{-}I$$

$$\dfrac{}{(\forall x.\, A(x) \wedge B(x)) \rightarrow (\forall x.\, A(x)) \wedge (\forall x.\, B(x))}\,\rightarrow\text{-}I^1$$

Another Proof? (2)

$$\dfrac{\dfrac{\dfrac{\dfrac{[\forall x.\, A(x) \wedge B(x)]^1}{A(x) \wedge B(x)}\,\forall\text{-}E}{A(x)}\,\wedge\text{-}EL}{\forall x.\, A(x)}\,\forall\text{-}I \qquad \dfrac{\dfrac{\dfrac{[\forall x.\, A(x) \wedge B(x)]^1}{A(x) \wedge B(x)}\,\forall\text{-}E}{B(x)}\,\wedge\text{-}ER}{\forall x.\, B(x)}\,\forall\text{-}I}{(\forall x.\, A(x)) \wedge (\forall x.\, B(x))}\,\wedge\text{-}I$$

$$\dfrac{}{(\forall x.\, A(x) \wedge B(x)) \rightarrow (\forall x.\, A(x)) \wedge (\forall x.\, B(x))}\,\rightarrow\text{-}I^1$$

Yes (check side conditions of $\forall$-*I*).

Aside: $A \leftrightarrow B$

Define $A \leftrightarrow B$ as $A \rightarrow B \wedge B \rightarrow A$.

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!

Proof?

$$\dfrac{\dfrac{[A]^1}{\forall x.\, A}\,\forall\text{-}I \qquad \dfrac{[\forall x.\, A]^1}{A}\,\forall\text{-}E}{A \leftrightarrow \forall x.\, A}\,\leftrightarrow\text{-}I^1$$

Proof?

$$\dfrac{\dfrac{[A]^1}{\forall x.\, A}\,\forall\text{-}I \qquad \dfrac{[\forall x.\, A]^1}{A}\,\forall\text{-}E}{A \leftrightarrow \forall x.\, A}\,\leftrightarrow\text{-}I^1$$

Yes, but only if x not free in A.

Proof?

$$\dfrac{\dfrac{[A]^1}{\forall x.\, A}\,\forall\text{-}I \qquad \dfrac{[\forall x.\, A]^1}{A}\,\forall\text{-}E}{A \leftrightarrow \forall x.\, A}\,\leftrightarrow\text{-}I^1$$

Yes, but only if x not free in A.

Similar requirement arises in proving
$(\forall x.\, A \rightarrow B(x)) \leftrightarrow (A \rightarrow \forall x.\, B(x))$.

Existential Quantification

- We could define $\exists x.\, A$ 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: $\mathcal{A}(\exists x.\, A) = \mathcal{A}(\neg\forall x.\, \neg A)$

Where do the Rules for $\exists$ Come from?

- We can use definition $\exists x.\, A \equiv \neg\forall x.\, \neg A$ and the given rules for $\forall$ to derive ND proof rules.

Where do the Rules for $\exists$ Come from?

- We can 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.

∃-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.

$\exists$-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 $\neg\forall x.\, \neg A(x)$.

$\exists$-*I* as a Derived Rule

The rule:

$$\frac{A(t)}{\exists x.\, A(x)}\ \exists\text{-}I$$

$$\forall x.\, \neg A(x)$$

$$\vdots$$

$$\bot$$

$$\neg\forall x.\, \neg A(x)$$

We aim for applying $\rightarrow$-*I* in the last step (recall $\neg$-definition).

∃-*I* as a Derived Rule

The rule:

$$\frac{A(t)}{\exists x.\, A(x)}\ \exists\text{-}I$$

$$\frac{\dfrac{\forall x.\, \neg A(x)}{\neg A(t)}\ \forall\text{-}E \qquad A(t)}{\bot}$$

$$\neg\forall x.\, \neg A(x)$$

We apply ∀-*E*.

$\exists$-*I* as a Derived Rule

The rule:

$$\frac{A(t)}{\exists x.\, A(x)}\ \exists\text{-}I$$

$$\dfrac{\dfrac{\dfrac{\forall x.\, \neg A(x)}{\neg A(t)}\ \forall\text{-}E \qquad A(t)}{\bot}\ \rightarrow\text{-}E}{\neg \forall x.\, \neg A(x)}$$

Making assumption $A(t)$ allows us to use $\rightarrow$-*E* (recall $\neg$-definition).

$\exists$-*I* as a Derived Rule

The rule:

$$\frac{A(t)}{\exists x.\, A(x)}\ \exists\text{-}I$$

$$\dfrac{\dfrac{\dfrac{[\forall x.\, \neg A(x)]^1}{\neg A(t)}\ \forall\text{-}E \qquad A(t)}{\bot}\ \rightarrow\text{-}E}{\neg\forall x.\, \neg A(x)}\ \rightarrow\text{-}I^1$$

Finally we can apply $\rightarrow$-*I*. Note that the assumption $A(t)$ is still open.

$\exists$-E as a Derived Rule

The rule:

$$\dfrac{\exists x.\, A(x) \qquad \begin{array}{c}[A(x)]\\ \vdots \\ B\end{array}}{B}\ \exists\text{-}E$$

$\exists x.\, A(x)$

We will use $\exists x.\, A(x)$ as one assumption.

$\exists$-E as a Derived Rule

The rule:

$$\dfrac{\exists x.\, A(x) \qquad \begin{array}{c}[A(x)]\\ \vdots \\ B\end{array}}{B}\ \exists\text{-}E$$

$$\neg\forall x.\, \neg A(x)$$

But by definition that's $\neg\forall x.\, \neg A(x)$.

$\exists$-E as a Derived Rule

The rule:

$$\dfrac{\exists x.\, A(x) \qquad \begin{array}{c}[A(x)]\\ \vdots\\ B\end{array}}{B}\ \exists\text{-}E$$

$$\begin{array}{c}A(x)\\ \vdots\\ B\end{array}$$

$$\neg\forall x.\, \neg A(x)$$

We assume a hypothetical derivation.

$\exists$-E as a Derived Rule

The rule:

$$\dfrac{\exists x.\, A(x) \qquad \begin{array}{c}[A(x)]\\ \vdots\\ B\end{array}}{B}\ \exists\text{-}E$$

$$\neg\forall x.\, \neg A(x)$$

$$\dfrac{\neg B \qquad \begin{array}{c}A(x)\\ \vdots\\ B\end{array}}{\bot}\ \rightarrow\text{-}E$$

We make an additional assumption and apply $\rightarrow$-E (recall $\neg$-definition)

$\exists$-E as a Derived Rule

The rule:

$$\dfrac{\exists x.\, A(x) \qquad \begin{array}{c}[A(x)]\\ \vdots\\ B\end{array}}{B}\ \exists\text{-}E$$

$$\neg\forall x.\, \neg A(x)$$

$$\dfrac{\dfrac{\neg B \qquad \begin{array}{c}[A(x)]^2\\ \vdots\\ B\end{array}}{\bot}\ \rightarrow\text{-}E}{\neg A(x)}\ \rightarrow\text{-}I^2$$

Now we can discharge the assumption $A(x)$ made in the hypothetical derivation.

$\exists$-E as a Derived Rule

The rule:

$$\dfrac{\exists x.\, A(x) \qquad \begin{array}{c}[A(x)]\\ \vdots\\ B\end{array}}{B}\ \exists\text{-}E$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\neg B \qquad \begin{array}{c}[A(x)]^2\\ \vdots\\ B\end{array}}{\bot}\ \rightarrow\text{-}E}{\neg A(x)}\ \rightarrow\text{-}I^2}{\forall x.\, \neg A(x)}\ \forall\text{-}I}{\neg\forall x.\, \neg A(x)}$$

At this step, the side condition from $\forall$-I applies. $\exists$-E will inherit it!

$\exists$-E as a Derived Rule

The rule:

$$\dfrac{\exists x.\, A(x) \qquad \begin{array}{c}[A(x)]\\ \vdots\\ B\end{array}}{B}\ \exists\text{-}E$$

$$\dfrac{\dfrac{\neg\forall x.\, \neg A(x) \qquad \dfrac{\dfrac{\dfrac{\neg B \qquad \begin{array}{c}[A(x)]^2\\ \vdots\\ B\end{array}}{\bot}\ \to\text{-}E}{\neg A(x)}\ \to\text{-}I^2}{\forall x.\, \neg A(x)}\ \forall\text{-}I}{\bot}\ \to\text{-}E$$

We apply $\to$-E.

$\exists$-E as a Derived Rule

The rule:

$$\dfrac{\exists x.\, A(x) \qquad \begin{array}{c}[A(x)]\\ \vdots\\ B\end{array}}{B}\ \exists\text{-}E$$

$$\dfrac{\dfrac{\neg\forall x.\, \neg A(x) \qquad \dfrac{\dfrac{\dfrac{[\neg B]^1 \qquad \begin{array}{c}[A(x)]^2\\ \vdots\\ B\end{array}}{\bot}\ \to\text{-}E}{\neg A(x)}\ \to\text{-}I^2}{\forall x.\, \neg A(x)}\ \forall\text{-}I}{\bot}\ \to\text{-}E}{B}\ RAA^1$$

We are done. Note that this proof uses classical reasoning.

Sample Derivation

Assumption: x does not occur free in B

$$\dfrac{\dfrac{\dfrac{[\exists x.\ A(x)]^2 \qquad \dfrac{\dfrac{[\forall x.\ A(x) \to B]^1}{A(x) \to B}\ \forall\text{-}E \qquad [A(x)]^3}{B}\ \to\text{-}E}{B}\ \exists\text{-}E^3}{(\exists x.\ A(x)) \to B}\ \to\text{-}I^2}{(\forall x.\ A(x) \to B) \to ((\exists x.\ A(x)) \to B)}\ \to\text{-}I^1$$

Conclusion on FOL

- Propositional logic is good for modeling simple patterns of reasoning like "if . . . then . . . else".

Conclusion on FOL

- Propositional logic is good for modeling simple patterns of reasoning like "if . . . then . . . else".
- In first-order logic, one has "elements in a universe of discourse" and relations on / properties of them. One can quantify over the elements of the universe. Powerful!
- Some people advocate intuitionistic, relevant, and other "deviant" logics.
- Limitation: cannot quantify over predicates.
- "A" world or "the" world is modeled in first-order logic using so-called first-order theories. This will be studied

next lecture.

More Detailed Explanations

Boolean Functions

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

Relations/Functions, Infinity

In propositional logic, there is no notation for writing "element x has property p" or "element x and y are related as follows" or for denoting the "element obtained from element 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^n) different propositional formulae.

What is a Domain?

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

Any set of elements of the universe of discourse that we want to reason about.

Alternating State

Say, of a pedestrian traffic light which alternates between red and non-red (=green).

The State Alternates over all Time

"The state alternates over all time" could be expressed by the first-order logic formula $\forall t.\, x(t) \leftrightarrow \neg x(t+1)$. We explain it intuitively.

x is some property which may or may not hold at a given point of time t, expressed as a number. So we write $x(t)$ if x holds at t. The formula says that x is true at point t if and only if x is not true at point $t+1$.

To be more precise, one would have to axiomatize the fact that t is a number etc.

"Constant Elements"?

As opposed to a variable which ranges over all values (elements of the universe of discourse).

Function Notation

So a function symbol f denotes an operation that takes n elements of the universe of discourse and returns a another one: $f(t_1, \ldots, t_n)$ is an element that depends on $t_1, \ldots, t_n$.

The generic notation for function application is like this: $f(t_1, \ldots, t_n)$, but the brackets are omitted for nullary functions (= constant symbols), 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.

Equational Axiomatization

How do we formalize/axiomatize the fundamental truths of our domain of interest?

It turns out that much of this is done by stating which terms are equal to which other terms.

Here it is assumed that $=$ is interpreted as the identity on the domain.

"Some values"?

Just like a constant, a variable stands for a value (element of discourse).

The most important difference between a constant and a (free) variable is that the latter ranges over all values.

Variables can be bound by quantifiers, so one can make statements such as "for all x . . . " or "there exists x such that . . . ".

What is Satisfiability? Validity?

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

More formally, this will be defined later.

Syntactic Categories

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

A term is an expression that stands for an element in the universe of discourse called a value.

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

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

In first-order logic the answer is: no!

Signatures

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 formalised 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.

One often calls the pair $(\mathcal{F}, \mathcal{P})$ 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.

A Language

$Term$ 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.

Free, Bound, and Binding Occurrences

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 ($\circ \in \{\wedge, \vee, \rightarrow\}$);

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

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

Structures

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

A universe (of discourse) is sometimes also called domain.

As you saw, a structure gives a meaning to functions, predicates, 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].

Omitted Cases

The following cases have been omitted:

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

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

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

These cases are the same as in propositional logic. This can also be found in any textbook on logic [vD80].

In mathematics and computer science, the word or is almost always meant to be inclusive. If it is meant to be exclusive (A or B hold but not both) this is usually mentioned explicitly.

The Notation $\mathcal{A}_{[x/u]}$

$\mathcal{A}_{[x/u]}$ is the structure $\mathcal{A}'$ identical to $\mathcal{A}$, except that $x^{\mathcal{A}'} = u$.

Models

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 constant 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, $\mathcal{I}$ 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 a $\mathcal{I}$ 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.

Suitable Structures

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.

$\mathcal{N}$

$\mathcal{N}$ denotes the natural numbers.

Confusion of Syntax and Semantics?

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 $m, n \in \mathcal{N}$, in this context.

Why is this a Model?

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

Why is this not a Model?

The identity function maps every object to itself.

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

Implicit Quantifiers

In the statement

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

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

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

Adapt the Proof for Reals

The proof for natural numbers exploits the fact that $y > 0$ implies $y \geq 1$. This is of course not true in the reals.

A proof for the reals:
Consider an arbitrary x where $x > 2$. Then $x = 2 + y$ for some $y > 0$ and hence

$$x^2 = (2 + y)^2 = 4 + 4y + y^2 > 4$$

The proof looks even simpler than for the naturals, and moreover, it also holds for the naturals, since those are a subset of the reals. However, such observations may be deceptive.

Tacitly, the proof uses the following assumptions (which we know are true in the usual interpretation of mathematics):

$$y > 0 \quad \text{implies} \quad 4y > 0$$

$$y > 0 \quad \text{implies} \quad y^2 > 0$$

$$z > 0 \text{ and } z' > 0 \quad \text{implies} \quad z + z' > 0$$

$$z > 0 \quad \text{implies} \quad w + z > w$$

Inheriting Rules

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

Schematic Rules

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.

Reflexivity

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

Side Condition Violated!

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

Why is $(\forall x.\, \neg\forall y.\, x = y) \rightarrow \neg\forall y.\, y = y$ False?

Here we assume that the predicate symbol $=$ is interpreted by $\mathcal{A}$ as equality on $U_{\mathcal{A}}$. Suppose $U_{\mathcal{A}}$ contains two elements α and β and $I_{\mathcal{A}}(x) = \alpha$ and $I_{\mathcal{A}}(y) = \beta$. Then $\mathcal{A}(x = y) = 0$, hence $\mathcal{A}(\forall y.\, x = y) = 0$, hence $\mathcal{A}(\neg\forall y.\, x = y) = 1$. Now one can see that $\mathcal{A}_{[x/u]}(\neg\forall y.\, x = y) = 1$ for all $u \in U_{\mathcal{A}}$, and hence $\mathcal{A}(\forall x.\, \neg\forall y.\, x = y) = 1$. On the other hand, $\mathcal{A}'(y = y) = 1$ for any $\mathcal{A}'$ and hence $\mathcal{A}(\forall y.\, y = y) = 1$ and hence $\mathcal{A}(\neg\forall y.\, y = y) = 0$. Therefore, $\mathcal{A}((\forall x.\, \neg\forall y.\, x = y) \rightarrow \neg\forall y.\, y = y) = 0$.

Substitutions in FOL

The notation $s[x \leftarrow t]$ denotes the term obtained by substituting t for x in s. However, a substitution $[x \leftarrow t]$ replaces only the free occurrences of x in the term that it is applied to. A substitution is defined as follows:

1. $x[x \leftarrow t] = t$;
2. $y[x \leftarrow t] = y$ if y is a variable other than x;
3. $f(t_1, \ldots, t_n)[x \leftarrow t] = f(t_1[x \leftarrow t], \ldots, t_n[x \leftarrow t])$ (where f is a function symbol, $n \geq 0$);
4. $p(t_1, \ldots, t_n)[x \leftarrow t] = p(t_1[x \leftarrow t], \ldots, t_n[x \leftarrow t])$ (where p is a predicate symbol, possibly $\bot$);
5. $(\neg\psi)[x \leftarrow t] = \neg(\psi[x \leftarrow t])$
6. $(\psi \circ \phi)[x \leftarrow t] = (\psi[x \leftarrow t] \circ \phi[x \leftarrow t])$ (where $\circ \in \{\wedge, \vee, \rightarrow\}$);
7. $(\mathsf{Q}x.\psi)[x \leftarrow t] = \mathsf{Q}x.\psi$ (where $\mathsf{Q} \in \{\forall, \exists\}$);

8. $(\mathsf{Q}y.\psi)[x \leftarrow t] = \mathsf{Q}y.(\psi[x \leftarrow t])$ (where $\mathsf{Q} \in \{\forall, \exists\}$) if $y \neq x$ and $y \notin FV(t)$;
9. $(\mathsf{Q}y.\psi)[x \leftarrow t] = \mathsf{Q}z.(\psi[y \leftarrow z][x \leftarrow t])$ (where $\mathsf{Q} \in \{\forall, \exists\}$) if $y \neq x$ and $y \in FV(t)$ where z is a variable such that $z \notin FV(t)$ and $z \notin FV(\psi)$.

Avoiding Capture of Variables

A substitution (replacement of a variable by a term) must not replace bound occurrences of variables, and if we replace x with t in an expression ϕ, then this replacement should not turn free occurrences of variables in t into bound occurrences in ϕ. It is possible to avoid this by renaming variables.

This is part of the standard definition of a substitution. The problem is not related to $\forall$-E in particular.

The definition can be found in any textbook on logic [vD80]. We will also give a formal definition later, in the context of the λ-calculus.

Check Side Conditions

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.

Defining $\leftrightarrow$

By defining we mean, use $A \leftrightarrow B$ as shorthand for $A \rightarrow B \wedge B \rightarrow A$, in the same way as we regard negation as a shorthand.

Defining $\exists$

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 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.

Hypothetical Derivation

We are constructing here a "schematic fragment" of a derivation tree. Within this construction, we simply assume a hypothetical derivation of B from assumption $A(x)$. When we are done with the construction of this fragment, we will collapse the fragment by throwing away all the nodes in the middle and only keep the root and leaves.

Note two points:

- We assume a hypothetical derivation of B from assumption $A(x)$. Somewhere in the middle of the constructed fragment, we will discharge the assumption $A(x)$. In the final rule $\exists$-*E*, this means an application of $\exists$-*E* involves discharging $A(x)$. Therefore $\exists$-*E* has brackets around the $A(x)$.
- The hypothetical derivation of B may contain other assumptions than $A(x)$. These are not discharged in the constructed fragment, and so in

the final rule $\exists$-E, we must also read the notation

$$\begin{array}{c} A(x) \\ \vdots \\ B \end{array}$$

as a derivation of B where one of the assumptions is $A(x)$. There may be other assumptions, but these are not discharged. This is no different from previous rules involving discharging.

Inheriting a Side Condition

$\exists$-*E* will inherit the side condition from $\forall$-*I*. Hence, the side condition for $\exists$-*E* is:

x must not be free in B or in hypotheses of the subderivation of B other than $A(x)$ (occurrences in $(A(x)$ are allowed because the assumption $A(x)$ was discharged before the application of $\forall$-*I*).

Classical Reasoning

Defining $\exists x.\, A$ as $\neg\forall x.\, \neg A$ is only sensible in classical reasoning, since the derivation of the rule $\exists$-E requires the RAA rule.

The Power of First-Order Logic

In first-order logic, one has values(elements of the universe of discourse) and relations/properties that may or may not hold for these values.
Quantifiers are used to speak about "all values" and "some value".

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. [Tho95]:

> **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 [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.

Other Logics

There are still controversies about what the best logic is for reasoning about values (=elements of a universe of discourse) 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 "ϕ always holds". But many other interpretations are possible, e.g., $\Box_A\phi$ could mean "A knows that ϕ holds" [HC68].

In relevance logics, it is not true that $A \longrightarrow B$ holds whenever A is false. Rather, A must somehow be "relevant" for B.

Limitations of First-Order Logic

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.

References

[Ari] Aristotle. *Analytica priora I*, chapter 4.

[HC68] George E. Hughes and Maxwell John Cresswell. *An Introduction to Modal Logic*. Muthuen and Co. Ltd, London, 1968.

[Tho95] Della Thompson, editor. *The Concise Oxford Dictionary*. Clarendon Press, 1995.

[vD80] Dirk van Dalen. *Logic and Structure*. Springer-Verlag, 1980. An introductory textbook on logic.