

First Order Logic (FOL)

Testing, Quality Assurance, and Maintenance
Fall 2023

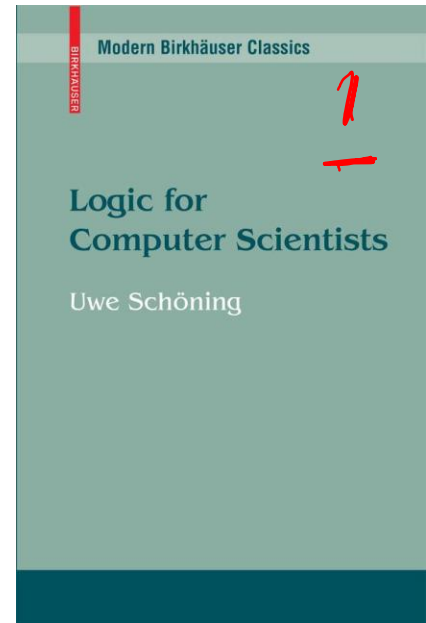
Prof. Arie Gurfinkel

based on slides by Prof. Ruzica Piskac, Nikolaj
Bjorner, and others

References

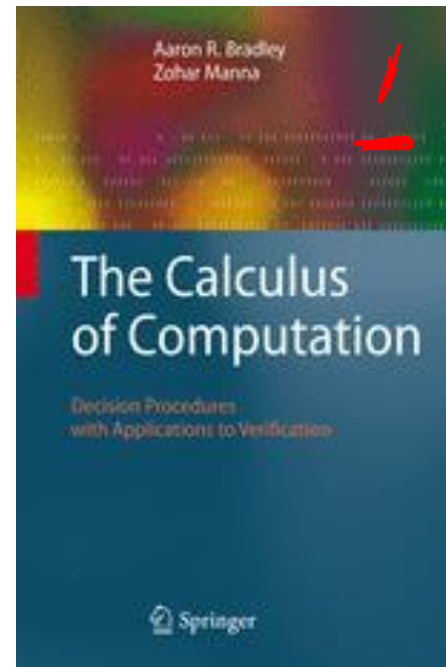
- Chapter 2 of Logic for Computer Scientists

<https://link.springer.com/book/10.1007/978-0-8176-4763-6>

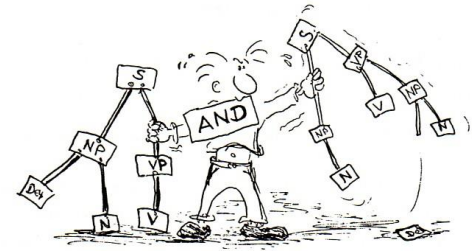


- Chapters 2 and 3 of Calculus of Computation

<https://link.springer.com/book/10.1007/978-3-540-74113-8>




Syntax and Semantics (Again)




Syntax

- MW: the way in which linguistic elements (such as words) are put together to form constituents (such as phrases or clauses)
- Determines and restricts how things are written

[[SEMANTICS]]
of a structure

[[

[[

Semantics

- MW: the study of meanings
- Determines how syntax is interpreted to give meaning

What's special about First Order Logic?

$\forall \exists$

(quantifiers)
and
uninterpreted functions,
predicates

Uninterpreted functions

Let's say f is an uninterpreted function

Then, f has an arity, e.g. $f_{/2}$ or $g_{/0}$

- that is, we can write $f(x, y)$ or $g()$ or g
- by convention, some functions are written in infix notation: $x + y$ for $+(x, y)$

That is all that f has! **IMPORTANT, DO NOT MISS**

- in programming language terms, we say that f has a declaration, but not a definition. We cannot assume anything about f (other than arity), even if f looks like a familiar function
- for example, we do not assume that $\text{plus}(x, y)$ will add two numbers. We only know that $\text{plus}()$ is a function, not what it is doing. $\text{plus}()$ is **uninterpreted**

(More later, when we talk about models.)

Predicates

Examples

- $P(x)$, $Q(y, z)$, R , $x = y$

Evaluate to **true** or **false**.

Common arithmetic predicates like \leq and $=$ are written in infix notation

- $x = y$
- $x \leq y$
- $x \geq z$

Predicates are uninterpreted, just like functions

- it is ok to think of predicates as special Boolean functions

Interpretation Example

$$\Sigma = \{0, +, <\}, \text{ and } M \text{ such that } |M| = \{a, b, c\}$$

$$M(0) = a,$$

$$M(+) = \{\langle a, a \mapsto a \rangle, \langle a, b \mapsto b \rangle, \langle a, c \mapsto c \rangle, \langle b, a \mapsto b \rangle, \langle b, b \mapsto c \rangle, \\ \langle b, c \mapsto a \rangle, \langle c, a \mapsto c \rangle, \langle c, b \mapsto a \rangle, \langle c, c \mapsto b \rangle\}$$

$$M(<) = \{\langle a, b \rangle, \langle a, c \rangle, \langle b, c \rangle\}$$

If $M(x) = a, M(y) = b, M(z) = c$, then

$$M[\![+(+(x, y), z)]\!] =$$

$$M(+)(M(+)(M(x), M(y)), M(z)) = M(+)(M(+)(a, b), c) =$$

$$M(+)(b, c) = a$$

The language of First Order Logic

Functions , Variables, Predicates

- f, g, \dots x, y, z, \dots $P, Q, =, <, \dots$

Atomic formulas, Literals

- $P(x, f(y)), \neg Q(y, z)$

Quantifier free formulas

- $P(f(a), b) \wedge c = g(d)$

Formulas, sentences

- $\forall x . \forall y . [P(x, f(x)) \vee g(y, x) = h(y)]$

Language: Signatures

A *signature* Σ is a finite set of:

- Function symbols:

$$\Sigma_F = \{ f, g, +, \dots \}$$

- Predicate symbols:

$$\Sigma_P = \{ P, Q, =, \text{true}, \text{false}, \dots \}$$

- And an *arity* function:

$$\Sigma \rightarrow \mathbb{N}$$

Function symbols with arity 0 are *constants*

- notation: f_2 means a symbol with arity 2

A countable set V of *variables*

- disjoint from Σ

Language: Terms

The set of *terms* $T(\Sigma_F, V)$ is the smallest set formed by the syntax rules:

$$\begin{array}{lcl} \bullet t \in T & ::= & v \qquad v \in V \\ & / & f(t_1, \dots, t_n) \qquad f \in \Sigma_F, t_1, \dots, t_n \in T \end{array}$$

Ground terms are given by $T(\Sigma_F, \emptyset)$

- a term is *ground* if it contains no variables

Language: Atomic Formulas

$$a \in \text{Atoms} \quad ::= P(t_1, \dots, t_n)$$
$$P \in \Sigma_P \quad t_1, \dots, t_n \in T$$

An atom is *ground* if $t_1, \dots, t_n \in T(\Sigma_F, \emptyset)$

- ground atom contains no variables

Literals are atoms and negation of atoms:

$$l \in \text{Literals} \quad ::= a \mid \neg a \quad a \in \text{Atoms}$$

Language: Quantifier free formulas

The set $\text{QFF}(\Sigma, V)$ of *quantifier free formulas* is the smallest set such that:

$\varphi \in \text{QFF} ::=$	$a \in \text{Atoms}$	<i>atoms</i>
	$ \neg \varphi$	<i>negations</i>
	$ \varphi \leftrightarrow \varphi'$	<i>bi-implications</i>
	$ \varphi \wedge \varphi'$	<i>conjunction</i>
	$ \varphi \vee \varphi'$	<i>disjunction</i>
	$ \varphi \rightarrow \varphi'$	<i>implication</i>

Language: Formulas

The set of *first-order formulas* are obtained by adding the formation rules:

$$\begin{array}{lll} \varphi ::= & \dots & \\ & | \quad \forall x . \varphi & \text{universal quant.} \\ & | \quad \exists x . \varphi & \text{existential quant.} \end{array}$$

Free (occurrences) of *variables* in a formula are those not bound by a quantifier.

A *sentence* is a first-order formula with no free variables.

Dreadbury Mansion Mystery

Someone who lived in Dreadbury Mansion killed Aunt Agatha. Agatha, the Butler and Charles were the only people who lived in Dreadbury Mansion. A killer always hates his victim, and is never richer than his victim. Charles hates no one that Aunt Agatha hates. Agatha hates everyone except the butler. The butler hates everyone not richer than Aunt Agatha. The butler also hates everyone Agatha hates. No one hates everyone. Agatha is not the butler.

Who killed Aunt Agatha?



Dreadbury Mansion Mystery

Someone who lived in Dreadbury Mansion **killed** Aunt **Agatha**. **Agatha**, the **Butler** and **Charles** were the only people who lived in Dreadbury Mansion. A killer always **hates** his victim, and is never **richer** than his victim. **Charles** **hates** no one that aunt **Agatha** **hates**. **Agatha** **hates** everyone except the **Butler**. The **Butler** **hates** everyone not **richer** than Aunt **Agatha**. The **Butler** also **hates** everyone **Agatha** **hates**. No one **hates** everyone. **Agatha** is not the **Butler**.

Who killed Aunt Agatha?

Constants are **blue**

Predicates are **purple**



Dreadbury Mansion Mystery

killed/₂, *hates*/₂, *richer*/₂, *a*/₀, *b*/₀, *c*/₀

$$\exists x \cdot \textit{killed}(x, a) \tag{1}$$

$$\forall x \cdot \forall y \cdot \textit{killed}(x, y) \implies (\textit{hates}(x, y) \wedge \neg \textit{richer}(x, y)) \tag{2}$$

$$\forall x \cdot \textit{hates}(a, x) \implies \neg \textit{hates}(c, x) \tag{3}$$

$$\textit{hates}(a, a) \wedge \textit{hates}(a, c) \tag{4}$$

$$\forall x \cdot \neg \textit{richer}(x, a) \implies \textit{hates}(b, x) \tag{5}$$

$$\forall x \cdot \textit{hates}(a, x) \implies \textit{hates}(b, x) \tag{6}$$

$$\forall x \cdot \exists y \cdot \neg \textit{hates}(x, y) \tag{7}$$

$$a \neq b \tag{8}$$



Solving Dreadbury Mansion in SMT

```
(declare-datatypes () ((Mansion (Agatha) (Butler) (Charles))))  
(declare-fun killed (Mansion Mansion) Bool)  
(declare-fun hates (Mansion Mansion) Bool)  
(declare-fun richer (Mansion Mansion) Bool)  
(assert (exists ((x Mansion)) (killed x Agatha)))  
(assert (forall ((x Mansion) (y Mansion))  
  (=> (killed x y) (hates x y))))  
(assert (forall ((x Mansion) (y Mansion))  
  (=> (killed x y) (not (richer x y)))))  
(assert (forall ((x Mansion))  
  (=> (hates Agatha x) (not (hates Charles x)))))  
(assert (hates Agatha Agatha))  
(assert (hates Agatha Charles))  
(assert (forall ((x Mansion))  
  (=> (not (richer x Agatha)) (hates Butler x)))))  
(assert (forall ((x Mansion))  
  (=> (hates Agatha x) (hates Butler x))))  
(assert (forall ((x Mansion)) (exists ((y Mansion)) (not (hates x y)))))  
  
(check-sat)  
(get-model)
```

Models (Semantics)

A model M is defined as:

- Domain S ; non-empty set of elements; often called the *universe*
- Interpretation, $f^M : S^n \rightarrow S$ for each $f \in \Sigma_F$ with $\text{arity}(f) = n$
- Interpretation $P^M \subseteq S^n$ for each $P \in \Sigma_P$ with $\text{arity}(P) = n$
- Assignment $x^M \in S$ for every variable $x \in V$

A formula φ is true in a model M if it evaluates to true under the given interpretations over the domain S .

M is a *model* for a set of sentences T if all sentences of T are true in M .

Models (Semantics)

A term t in a model M is interpreted as:

- Variable $x \in V$ is interpreted as x^M
- $f(t_1, \dots, t_n)$ is interpreted as $f^M(a_1, \dots, a_n)$,
 - where a_i is the current interpretation of t_i

$P(t_1, \dots, t_n)$ atom is *true* in a model M if and only if

- $(a_1, \dots, a_n) \in P^M$, where
- a_i is the current interpretation of t_i

Models (Semantics)

A formula φ is true in a model M if:

- $M \models \neg \varphi$ iff $M \not\models \varphi$ (i.e., M is not a model for φ)
- $M \models \varphi \leftrightarrow \varphi'$ iff $M \models \varphi$ is equivalent to $M \models \varphi'$
- $M \models \varphi \wedge \varphi'$ iff $M \models \varphi$ and $M \models \varphi'$
- $M \models \varphi \vee \varphi'$ iff $M \models \varphi$ or $M \models \varphi'$
- $M \models \varphi \rightarrow \varphi'$ iff if $M \models \varphi$ then $M \models \varphi'$
- $M \models \forall x. \varphi$ iff for all $s \in S$, $M[x:=s] \models \varphi$
- $M \models \exists x. \varphi$ iff exists $s \in S$, $M[x:=s] \models \varphi$

Interpretation Example

$$\Sigma = \{0, +, <\}, \text{ and } M \text{ such that } |M| = \{a, b, c\}$$

$$M(0) = a,$$

$$M(+) = \{\langle a, a \mapsto a \rangle, \langle a, b \mapsto b \rangle, \langle a, c \mapsto c \rangle, \langle b, a \mapsto b \rangle, \langle b, b \mapsto c \rangle, \\ \langle b, c \mapsto a \rangle, \langle c, a \mapsto c \rangle, \langle c, b \mapsto a \rangle, \langle c, c \mapsto b \rangle\}$$

$$M(<) = \{\langle a, b \rangle, \langle a, c \rangle, \langle b, c \rangle\}$$

If $M(x) = a, M(y) = b, M(z) = c$, then

$$M[\![+(+(x, y), z)]\!] =$$

$$M(+)(M(+)(M(x), M(y)), M(z)) = M(+)(M(+)(a, b), c) =$$

$$M(+)(b, c) = a$$

Interpretation Example

$$\Sigma = \{0, +, <\}, \text{ and } M \text{ such that } |M| = \{a, b, c\}$$

$$M(0) = a,$$

$$M(+) = \{\langle a, a \mapsto a \rangle, \langle a, b \mapsto b \rangle, \langle a, c \mapsto c \rangle, \langle b, a \mapsto b \rangle, \langle b, b \mapsto c \rangle, \\ \langle b, c \mapsto a \rangle, \langle c, a \mapsto c \rangle, \langle c, b \mapsto a \rangle, \langle c, c \mapsto b \rangle\}$$

$$M(<) = \{\langle a, b \rangle, \langle a, c \rangle, \langle b, c \rangle\}$$

$$M \models (\forall x : (\exists y : +(x, y) = 0))$$

$$M \not\models (\forall x : (\exists y : x < y))$$

$$M \models (\forall x : (\exists y : +(x, y) = x))$$

Dreadbury Mansion Mystery

killed/₂, *hates*/₂, *richer*/₂, *a*/₀, *b*/₀, *c*/₀

$$\exists x \cdot \textit{killed}(x, a) \tag{1}$$

$$\forall x \cdot \forall y \cdot \textit{killed}(x, y) \implies (\textit{hates}(x, y) \wedge \neg \textit{richer}(x, y)) \tag{2}$$

$$\forall x \cdot \textit{hates}(a, x) \implies \neg \textit{hates}(c, x) \tag{3}$$

$$\textit{hates}(a, a) \wedge \textit{hates}(a, c) \tag{4}$$

$$\forall x \cdot \neg \textit{richer}(x, a) \implies \textit{hates}(b, x) \tag{5}$$

$$\forall x \cdot \textit{hates}(a, x) \implies \textit{hates}(b, x) \tag{6}$$

$$\forall x \cdot \exists y \cdot \neg \textit{hates}(x, y) \tag{7}$$

$$a \neq b \tag{8}$$



Dreadbury Mansion Mystery: Model

$killed/2, hates/2, richer/2, a/0, b/0, c/0$

$$S = \{a, b, c\}$$

$$M(a) = a$$

$$M(b) = b$$

$$M(c) = c$$

$$M(killed) = \{(a, a)\}$$

$$M(richer) = \{(b, a)\}$$

$$M(hates) = \{(a, a), (a, c), (b, a), (b, c)\}$$



Semantics: Exercise

Drinker's paradox:

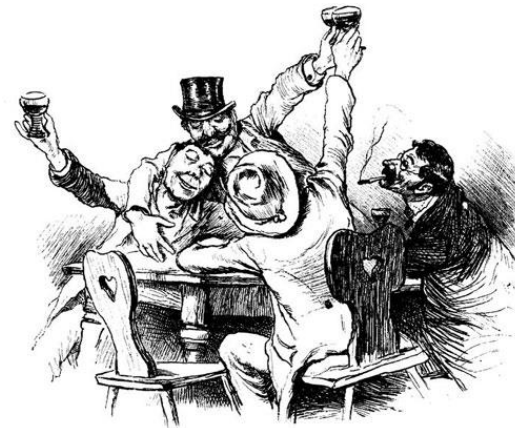
There is someone in the pub such that, if he is drinking, everyone in the pub is drinking.

- $\exists x. (D(x) \rightarrow \forall y. D(y))$

Is this logical formula valid?

Or unsatisfiable?

Or satisfiable but not valid?



Inference Rules for First Order Logic

We write $\vdash A$ when A can be inferred from basic axioms

We write $B \vdash A$ when A can be inferred from B

Natural deduction style rules

Notation: $A[a/x]$ means A with variable x replaced by term a

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

$$\frac{A}{A \vee B}$$

$$\frac{B}{A \vee B}$$

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

$$\frac{A[e/x]}{\exists x. A}$$

$$\frac{\forall x. A}{A[e/x]}$$

$$\frac{A[a/x]}{\forall x. A} \quad a \text{ is fresh}$$

$$\frac{A \vdash B}{A \rightarrow B}$$

$$\frac{\vdash \exists x. A \quad A[a/x] \vdash B}{\vdash B} \quad a \text{ is fresh}$$

Theories

A (first-order) *theory* T (over signature Σ) is a set of (deductively closed) sentences (over Σ and V)

Let $DC(\Gamma)$ be the deductive closure of a set of sentences Γ .

- For every theory T , $DC(T) = T$

A theory T is *consistent* if $false \notin T$

A theory captures the intended interpretation of the functions and predicates in the signature

- e.g., '+' is a plus, '0' is number 0, etc.

We can view a (first-order) theory T as the class of all *models* of T (due to completeness of first-order logic).

Theory of Equality T_E

Signature: $\Sigma_E = \{ =, a, b, c, \dots, f, g, h, \dots, P, Q, R, \dots \}$

$=$, a binary predicate, interpreted by axioms
all constant, function, and predicate symbols.

Axioms:

1. $\forall x . x = x$ (reflexivity)
2. $\forall x, y . x = y \rightarrow y = x$ (symmetry)
3. $\forall x, y, z . x = y \wedge y = z \rightarrow x = z$ (transitivity)

Theory of Equality T_E

Signature: $\Sigma_E = \{ =, a, b, c, \dots, f, g, h, \dots, P, Q, R, \dots \}$

$=$, a binary predicate, interpreted by axioms
all constant, function, and predicate symbols.

Axioms:

4. for each positive integer n and n -ary function symbol f ,

$$\forall x_1, \dots, x_n, y_1, \dots, y_n. \bigwedge_i x_i = y_i \rightarrow f(x_1, \dots, x_n) = f(y_1, \dots, y_n) \quad (\text{congruence})$$

5. for each positive integer n and n -ary predicate symbol P

$$\forall x_1, \dots, x_n, y_1, \dots, y_n. \bigwedge_i x_i = y_i \rightarrow (P(x_1, \dots, x_n) \leftrightarrow P(y_1, \dots, y_n)) \quad (\text{equivalence})$$

Theory of Peano Arithmetic (Natural Number)

Signature: $\Sigma_{PA} = \{ 0, 1, +, *, = \}$

Axioms of T_{PA} : axioms for theory of equality, T_E , plus:

1. $\forall x. \neg (x + 1 = 0)$ (zero)
2. $\forall x, y. x + 1 = y + 1 \rightarrow x = y$ (successor)
3. $F[0] \wedge (\forall x. F[x] \rightarrow F[x + 1]) \rightarrow \forall x. F[x]$ (induction)
4. $\forall x. x + 0 = x$ (plus zero)
5. $\forall x, y. x + (y + 1) = (x + y) + 1$ (plus successor)
6. $\forall x. x * 0 = 0$ (times zero)
7. $\forall x, y. x * (y + 1) = x * y + x$ (times successor)

Note that induction (#3) is an axiom schema

- one such axiom is added for each predicate F in the signature

Peano arithmetic is undecidable!

Theory of Presburger Arithmetic

Signature: $\Sigma_{PA} = \{ 0, 1, +, = \}$

Axioms of T_{PA} : axioms for theory of equality, T_E , plus:

1. $\forall x. \neg (x + 1 = 0)$ (zero)
2. $\forall x, y. x + 1 = y + 1 \rightarrow x = y$ (successor)
3. $F[0] \wedge (\forall x. F[x] \rightarrow F[x + 1]) \rightarrow \forall x. F[x]$ (induction)
4. $\forall x. x + 0 = x$ (plus zero)
5. $\forall x, y. x + (y + 1) = (x + y) + 1$ (plus successor)

Note that induction (#3) is an axiom schema

- one such axiom is added for each predicate F in the signature

Can extend the signature to allow multiplication by a numeric constant

Presburger arithmetic is decidable

- linear integer programming (ILP)

McCarthy theory of Arrays T_A

Signature: $\Sigma_A = \{ \text{read}, \text{write}, = \}$

read(*a*, *i*) is a binary function:

- reads an array *a* at the index *i*
- alternative notations:
 - (select *a* *i*), and *a*[*i*]

write(*a*, *i*, *v*) is a ternary function:

- writes a value *v* to the index *i* of array *a*
- alternative notations:
 - (store *a* *i* *v*) , *a*[*i*:=*v*]
- side-effect free – results in new array, does not modify *a*

Axioms of T_A

Array congruence

- $\forall a, i, j. i = j \rightarrow \text{read}(a, i) = \text{read}(a, j)$

Read-Over-Write 1

- $\forall a, v, i, j. i = j \rightarrow \text{read}(\text{write}(a, i, v), j) = v$

Read-Over-Write 2

- $\forall a, v, i, j. i \neq j \rightarrow \text{read}(\text{write}(a, i, v), j) = \text{read}(a, j)$

Extensionality

- $a = b \leftrightarrow \forall i. \text{read}(a, i) = \text{read}(b, i)$

T-Satisfiability

A formula $\varphi(x)$ is T-satisfiable in a theory T if there is a model of $DC(T \cup \exists x.\varphi(x))$.

That is, there is a model M for T in which $\varphi(x)$ evaluates to true.

Notation:

$$M \models_T \varphi(x)$$

where, $DC(V)$ stands for deductive closure of V

T-Validity

A formula $\varphi(x)$ is *T-valid* in a theory T if
 $\forall x. \varphi(x) \in T$

That is, $\forall x. \varphi(x)$ evaluates to *true* in every
model M of T

T-validity:

$$\models_T \varphi(x)$$

Fragment of a Theory

Fragment of a theory T is a syntactically restricted subset of formulae of the theory

Example:

- Quantifier-free fragment of theory T is the set of formulae without quantifiers that are valid in T

Often *decidable* fragments for undecidable theories

Theory T is *decidable* if T -validity is decidable for every formula F of T

- There is an algorithm that always terminates with “yes” if F is T -valid, and “no” if F is T -unsatisfiable

Exercises (1/2)

Find a model for $P(f(x,y)) \Rightarrow P(g(x,y,x))$

Write an axiom that will restrict that every model has to have exactly three different elements.

Write a FOL formula stating that i is the position of the minimal element of an integer array A

Write a FOL formula stating that v is the minimal element of an integer array A

Exercises (1/2)

Find a model for $P(f(x,y)) \Rightarrow P(g(x,y,x))$

Exercises (1/2)

Write an axiom that will restrict that every model has to have exactly three different elements.

Exercises (1/2)

Write a FOL formula stating that i is the position of the minimal element of an integer array A

Write a FOL formula stating that v is the minimal element of an integer array A

Exercises (1/2)

Find a model for $P(f(x,y)) \Rightarrow P(g(x,y,x))$

Write an axiom that will restrict that every model has to have exactly three different elements.

$$(\exists x, y, z \cdot x \neq y \wedge x \neq z \wedge y \neq z) \wedge (\forall a_0, a_1, a_2, a_3 \cdot \bigvee_{0 \leq i < j \leq 3} a_i = a_j)$$

Write a FOL formula stating that i is the position of the minimal element of an integer array A

$$\begin{aligned} & isIntArray(A) \wedge isInt(i) \wedge 0 \leq i < len(A) \\ & \forall j \cdot 0 \leq j < len(A) \wedge i \neq j \implies A[i] \leq A[j] \end{aligned}$$

Write a FOL formula stating that v is the minimal element of an integer array A

$$\begin{aligned} & isIntArray(A) \wedge isInt(v) \\ & \exists i \cdot 0 \leq i < len(A) \wedge A[i] = v \\ & \forall i \cdot 0 \leq i < len(A) \implies A[i] \leq v \end{aligned}$$

Exercises (2/2)

Show whether the following sentence is valid or not

$$(\exists x \cdot P(x) \vee Q(x)) \iff (\exists x \cdot P(x)) \vee (\exists x \cdot Q(x))$$

Show whether the following FOL sentence is valid or not

$$(\exists x \cdot P(x) \wedge Q(x)) \iff (\exists x \cdot P(x)) \wedge (\exists x \cdot Q(x))$$

Exercises (2/2)

Show whether the following sentence is valid or not

$$(\exists x \cdot P(x) \vee Q(x)) \iff (\exists x \cdot P(x)) \vee (\exists x \cdot Q(x))$$

Exercises (2/2)

Show whether the following FOL sentence is valid or not

$$(\exists x \cdot P(x) \wedge Q(x)) \iff (\exists x \cdot P(x)) \wedge (\exists x \cdot Q(x))$$

Exercises (2/2)

Show whether the following sentence is valid or not

$$(\exists x \cdot P(x) \vee Q(x)) \iff (\exists x \cdot P(x)) \vee (\exists x \cdot Q(x))$$

- Valid. Prove by contradiction that every model M of the LHS is a model of the RHS and vice versa.

Show whether the following FOL sentence is valid or not

$$(\exists x \cdot P(x) \wedge Q(x)) \iff (\exists x \cdot P(x)) \wedge (\exists x \cdot Q(x))$$

- Not valid. Prove by constructing a model M of the RHS that is not a model of the LHS. For example, $S = \{0,1\}$, $M(P) = \{0\}$, and $M(Q) = \{1\}$

Completeness, Compactness, Incompleteness

Gödel Completeness Theorem of FOL

- any (first-order) formula that is true in *all* models of a theory, must be logically deducible from that theory, and vice versa (every formula that is deducible from a theory is true in *all* models of that theory)

Corollary: Compactness Theorem

- A FOL theory G is SAT iff every finite subset G' of G is SAT
- A set G of FOL sentences is UNSAT iff exists a finite subset G' of G that is UNSAT

Incompleteness of FOL Theories

- A theory is *consistent* if it is impossible to prove both p and $\neg p$ for any sentence p in the signature of the theory
- A theory is *complete* if for every sentence p it includes either p or $\neg p$
- There are FOL theories that are consistent but incomplete

<https://terrytao.wordpress.com/2009/04/10/the-completeness-and-compactness-theorems-of-first-order-logic/>