

**CS 4700:**  
**Foundations of Artificial Intelligence**

**Bart Selman**  
**selman@cs.cornell.edu**

**Module: Knowledge, Reasoning, and Planning**

**First-Order Logic and Inference**  
**R&N: Chapters 8 and 9**

# First-Order Logic

Richer language. Closer match to ontology / conceptual structure  
**objects / properties / relations / functions**

Ex.:

**objects:** table, car, house, John...

**relations:** brother of, part of, has color...

**properties:** red, round, prime... [“unary relations”]

**functions:** father of, best friend, one more than...

Q. Contrast relation with function.

## Syntax and Semantics

Study section 8.2 of R&N carefully.

Note: Semantics can be defined more formally.

See e.g. “A course in mathematical logic” Bell & Machover.

R&N provide main ideas behind semantics.

Semantics give by **interpretations** (propositional  
analogue: truth assignment)

Sentence (“formula”) evaluates to True or False under a  
given interpretation. If True, interpretation is called  
a **model** of the sentence.

We hope that models of KB are close to actual state of affairs (world)

But often, we also have unexpected (non-standard) models.

Relation between mathematical notion of interpretation / model  
and actual physical world interesting philosophical issue.

We’ll ignore it.

Each interpretation is defined over a given **domain**  $U$  (set of  
individuals / objects).

**Constant symbols:**  $A, B, C, John, chair-1, house-10\dots$

In interpretation these symbols correspond to elements of  $U$ .  
(two constants can define the same element in  $U$ ).

(morning-star / evening-star)

**Predicate symbols:**  $Round, Brother, Part-of, \dots$

Each predicate symbol correspond to a relation on  $U$ .

E.g., a binary predicate, corresponds to a binary relation.

If  $U$  equals  $\{ car, tires, steering wheel, house \}$ .

$[tires, car]$ , and  $[steering wheel, car]$ .

could be the intended interpretation of “part-of”

Language	Ontological Commitment (What exists in the world)	Epistemological Commitment (What an agent believes about facts)
Propositional logic	facts	true/false/unknown
First-order logic	facts, objects, relations	true/false/unknown
Temporal logic	facts, objects, relations, times	true/false/unknown
Probability theory	facts	degree of belief 0...1
Fuzzy logic	degree of truth	degree of belief 0...1

**Function symbols:** *Cosine*, *FatherOf*, *LeftLegOf* ...

Correspond to functions defined on  $U$ .

Relation to predicate symbols?

Again, “what’s meant by embodying **knowledge** about the world?”

Example:

1)  $On(A, Fl) \Rightarrow Clear(B)$

2)  $(Clear(B) \wedge Clear(C)) \Rightarrow On(A, Fl)$

3)  $Clear(B) \vee Clear(A)$

4)  $Clear(B)$

5)  $Clear(C)$



B  
A C  
-----  
floor

One interpretation:

$U$  is the set  $\{ A, B, C, \text{Floor} \}$ .

1) mapping constant symbols to elements of  $U$ .

e.g.,  $A$  to  $A$ ,  $B$  to  $B$ ,  $C$  to  $C$

and  $Fl$  to  $\text{Floor}$

Could we have mapped  $Fl$  to  $A$ ??

2) mapping of relation symbol  $On$  to relation on  $U$ .

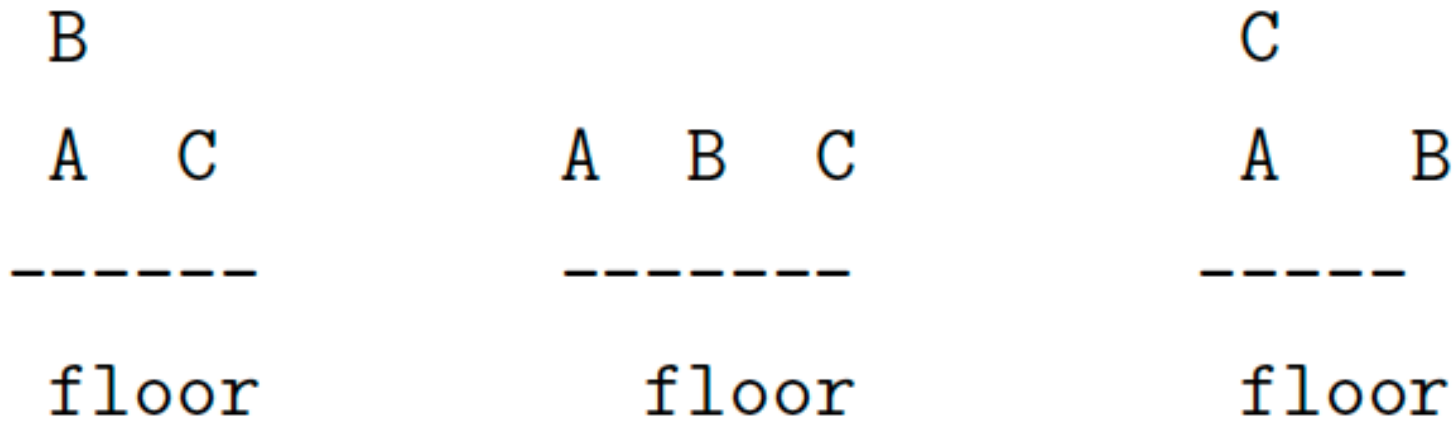
e.g.,  $On = \{ [ B, A ], [ A, \text{Floor} ], [ C, \text{Floor} ] \}$ .

3) mapping of relation (property)  $Clear$  to a unary rel. on  $U$ .

e.g.,  $Clear = \{ [ B ], [ C ] \}$ .

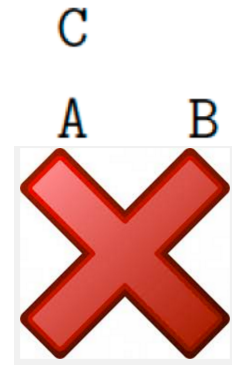
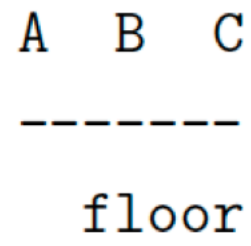
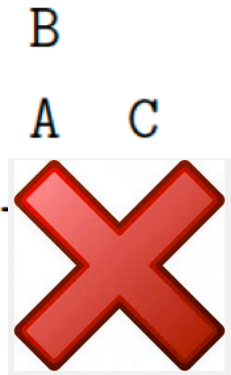
- 1)  $On(A, Fl) \Rightarrow Clear(B)$
- 2)  $(Clear(B) \wedge Clear(C)) \Rightarrow On(A, Fl)$
- 3)  $Clear(B) \vee Clear(A)$
- 4)  $Clear(B)$
- 5)  $Clear(C)$

Yet others ...



Including completely different interpretations!

E.g., use integers for domain. (Lowenheim 1915)



Try to add sufficient axioms (facts) to rule out unwanted models. E.g., add *clear(A)*.

**Terms** — a logical expressions that refers to an object. Constant symbols are terms. Functions applied to constant symbols. *FatherOf(John)*. Also, **variables** are terms (later) and functions applied to variables or other terms.

The interpretation is given by whatever the Constant or Function maps to in  $U$  (vars later).

If no vars, called **atomic terms**.

**Atomic sentences** — A predicate symbol applied to atomic terms

E.g.  $Married(FatherOf(Richard), MotherOf(John))$

Evaluated to **true** if predicate symbol holds between the objects referred to by the arguments.

**Complex sentences** — add **logical connectives**.

E.g.  $Older(John, 30) \Rightarrow Older(Jane, 29)$

# Quantifiers

Universal Quantification  $\forall$  —

E.g.,  $\forall x \text{ Cat}(x) \Rightarrow \text{Mammal}(x)$

Think of as:

$(\text{Cat}(\text{Spot}) \Rightarrow \text{Mammal}(\text{Spot})) \wedge$

$(\text{Cat}(\text{Felix}) \Rightarrow \text{Mammal}(\text{Felix})) \wedge$

$(\text{Cat}(\text{John}) \Rightarrow \text{Mammal}(\text{John})) \wedge$

...

Intuition: Expand over all object symbols.

## Existential Quantification $\exists$ —

E.g.,  $\exists x \text{ Sister}(x, \text{Spot}) \wedge \text{Cat}(x)$

Think of as:

$(\text{Sister}(\text{Spot}, \text{Spot}) \wedge \text{Cat}(\text{Spot})) \vee$   
 $(\text{Sister}(\text{Rebecca}, \text{Spot}) \wedge \text{Cat}(\text{Rebecca})) \vee$   
 $(\text{Sister}(\text{Felix}, \text{Spot}) \wedge \text{Cat}(\text{Felix})) \vee$

...

Intuition: Expand over all object symbols.

**Equality** = —

E.g.  $father(John) = Henry$

True iff refer to same object of  $U$  in interpretation.  
(identity relation)



See Chapter 8 of R&N for more discussion and fine details.

E.g. can't just switch quantifiers around.

Compare  $\forall x \exists y \text{Loves}(x, y)$  vs.

Compare  $\exists x \forall y \text{Loves}(x, y)$

# Reflex Agent

Directly connects percepts to actions:

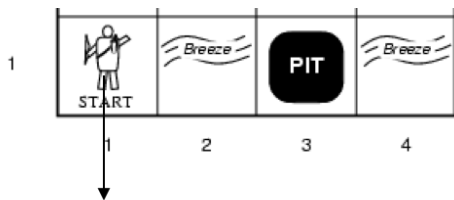
$$\forall s, b, u, c, t \quad \text{Percept}([s, b, \text{Glitter}, u, c], t) \Rightarrow \text{Action}(\text{Grab}, t)$$

Or, more indirectly:

$$\forall s, b, u, c, t \quad \text{Percept}([s, b, \text{Glitter}, u, c], t) \Rightarrow \text{AtGold}(t)$$

$$\forall t \quad \text{AtGold}(t) \Rightarrow \text{Action}(\text{Grab}, t)$$

Why more flexible? Limitations of reflex approach?



None, none, none, none, none

Stench, Breeze, Glitter, Bump, Scream

**Finite domains === “essentially propositional.” Also called: Propositional schema.**

## **Example: N-Queens, first-order**

- **1) no row with two queens**

$$\forall i, j, k \ (1 \leq i, j, k \leq N) \ (Queen(i, j) \wedge Queen(i, k)) \Rightarrow (j = k)$$

- **2) no column with two queens**

$$\forall i, j, k \ (1 \leq i, j, k \leq N) \ (Queen(j, i) \wedge Queen(k, i)) \Rightarrow (j = k)$$

- **3) no diagonal with two queens**

$$\forall i, j, k, l \quad (1 \leq i, j, k, l \leq N) \quad [(Queen(i, j) \wedge Queen(k, l) \wedge leftrightarrow ((i = j) \wedge (k = l))$$

similarly for right diagonal.

$$\forall i, j, k, l \quad (1 \leq i, j, k, l \leq N)$$

$$\text{leftdiagonal}(i, j, k, l) \Leftrightarrow [(i - j) = (k - l)]$$

Similarly, for *rightdiagonal*

**Complete?**

- 4) at least one queen per row  
 $\forall i \exists j (1 \leq i, j \leq N) \text{ Queen}(i, j)$

**Also discussed earlier. Here some additional axiom details. R&N Section 10.4.2.**

## Situation Calculus

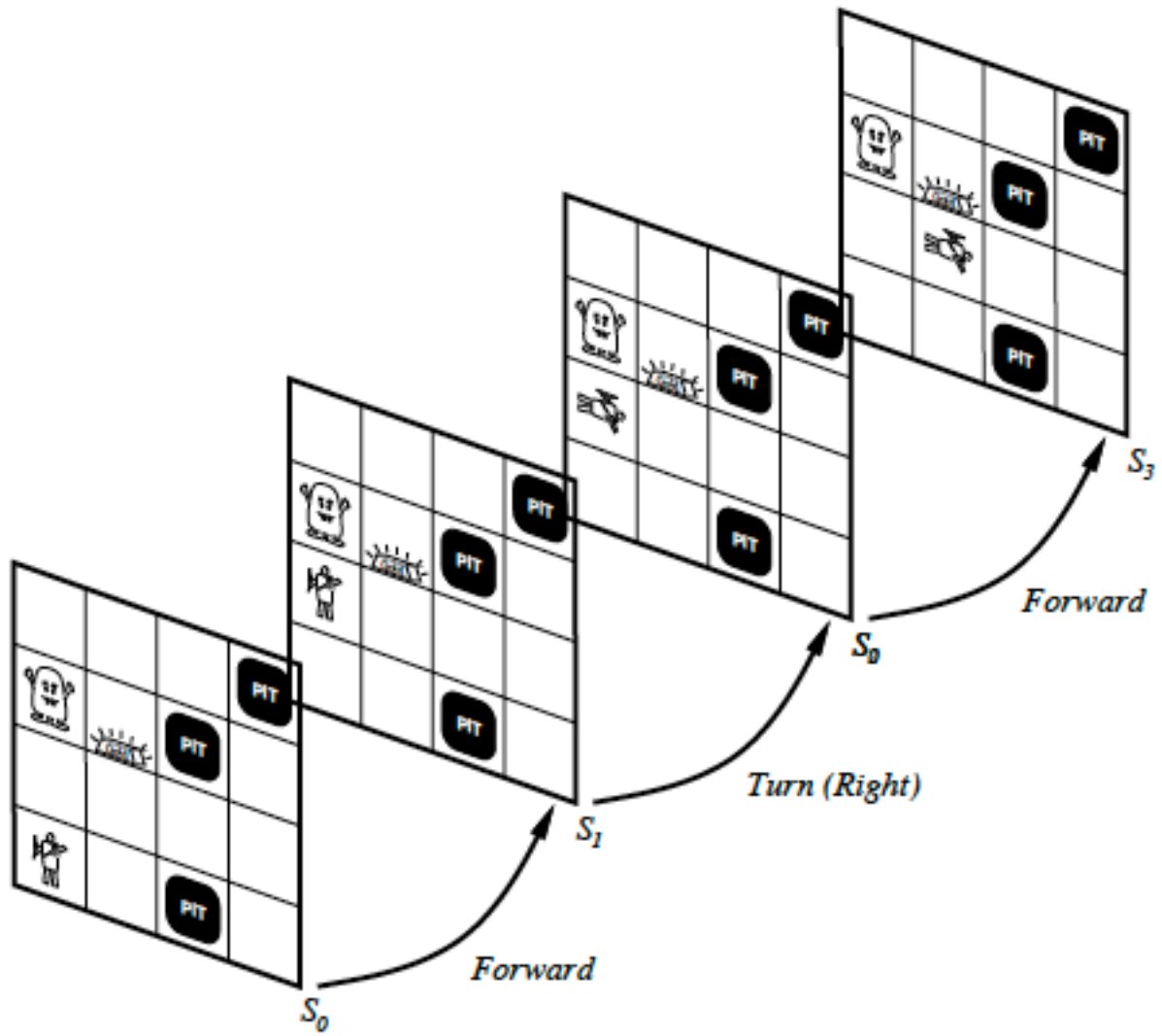
One approach: have time argument

We can be more concise, since we're only interested in when and how things **change**.

Focus on “**situations**”. (“**snapshots**”)

(John McCarthy 1963).

(Changing) World is represented by a series of *situations*.





E.g.:

$$At(\textit{Agent}, [1, 1], S_0) \wedge At(\textit{Agent}, [1, 2], S_1)$$

“talking” about change / actions:

$$Result(\textit{Forward}, S_0) = S_1$$

$$Result(\textit{Turn}(\textit{Right}), S_1) = S_2$$

$$Result(\textit{Forward}, S_2) = S_3$$

Is *Result* a relation or a function?

What about *Forward*?

**Result(action, situation) → situation  
(unique outcome)**

## Effect Axioms

*Portable(Gold)*

$\forall s \text{ AtGold}(s) \Rightarrow \text{Present}(\text{Gold}, s)$

$\forall x, s \ (\text{Present}(x, s) \wedge \text{Portable}(x)) \Rightarrow \text{Holding}(x, \text{Result}(\text{Grab}, s))$

$\forall x, s \ \neg \text{Holding}(x, \text{Result}(\text{Release}, s))$

Does this work?

**Again, as discussed in propositional case.**

## Frame Axioms

Need also to state what **doesn't** change!

$$\forall a, x, s \quad (\textit{Holding}(x, s) \wedge (a \neq \textit{Release})) \\ \Rightarrow \textit{Holding}(x, \textit{Result}(a, x))$$

$$\forall a, x, s \quad (\neg \textit{Holding}(x, s) \wedge (a \neq \textit{grab})) \\ \Rightarrow \neg \textit{Holding}(x, \textit{Result}(a, x))$$

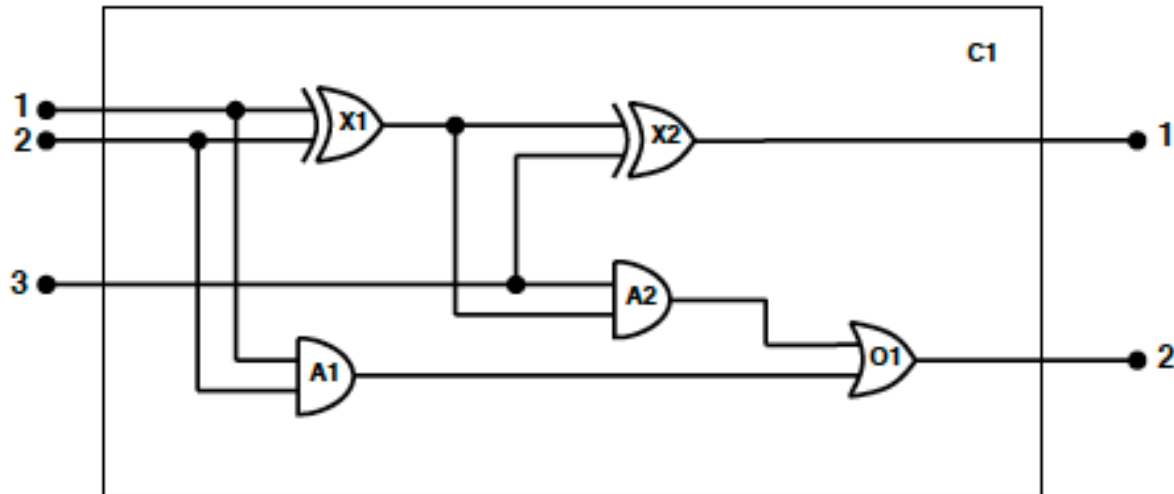
More compactly:

$$\forall a, x, s \text{ Holding}(x, \text{Results}(a, s)) \Leftrightarrow \\ [(a = \text{Grab} \wedge \text{Present}(x, s) \wedge \text{Portable}(x)) \\ \vee (\text{Holding}(x, s) \wedge a \neq \text{Release})]$$

**successor-state axioms:** need to list all the ways  
in which any predicate can become true / false.

## R&N 8.4.2.

# Another example: Electronic Circuits



**Further illustration of FOL formulation.**

A bit “tedious.” Check details at home.

## Formalization

one-bit full adder: two inputs and a carry / one output and carry

four gates: AND, OR, XOR and NOT.

Goal: analyze design to see if it matches specification.

Consider: circuits (gates and gate types), terminals, and signals.

formalization — keep task in mind.

e.g., for fault diagnosis: might want to specify “wires”

could be broken... (e.g.,  $Wire(x, y)$ )

**Always define first and carefully.**

## Vocabulary

pick: functions / predicates / constants.

constant symbols:  $X_1, X_2$ , etc.

type gate:  $Type(X_1) = XOR$ , note  $XOR$  new constant.

Alt.:  $Type(X_1, XOR)$ . Q. Advantage function?

terminals:  $Out(1, X_1), In(1, X_1), In(2, X_1)$ .

connectivity:  $Connected(Out(1, X_1), In(1, X_2))$ .

Note: we don't have to **name** the terminals explicitly.

the semantics of the function will assign some unique  
“object” to it.

(Skolemization can bring back name.)

signal values: function  $Signal(x)$ , e.g.,  $Signal(In(1, X_1))$ .

signal values:  $On$  and  $Off$ .



## General Rules

*how signals behave:*

$$1) \forall t_1, t_2 \text{ Connected}(t_1, t_2) \Rightarrow \text{Signal}(t_1) = \text{Signal}(t_2)$$

$$2a) \forall t \text{ Signal}(t) = \text{On} \vee \text{Signal}(t) = \text{Off}$$

$$2b) \text{On} \neq \text{Off}$$

$$3) \forall t_1, t_2 \text{ Connected}(t_1, t_2) \Leftrightarrow \text{Connected}(t_2, t_1)$$

*how gates behave:*

4)  $\forall g \text{ Type}(g) = OR \Rightarrow$

$$\text{Signal}(\text{Out}(1, g)) = On \Leftrightarrow \exists n \text{ Signal}(\text{In}(n, g)) = On.$$

5) how AND?

6)  $\forall g \text{ Type}(g) = XOR \Rightarrow$

$$(\text{Signal}(\text{Out}(1, g)) = On \Leftrightarrow (\text{Signal}(\text{In}(1, g)) \neq \text{Signal}(\text{In}(2, g)))$$

7) NOT, similarly.

few rules (7): good ontology  
clear rules: good vocabulary  
what remains?

*atomic facts* — describing actual circuit under consideration.

types of gates:

$Type(X_1) = XOR$ ,  $Type(X_2) = XOR$ ,  $Type(A_1) = AND$ , ..

connectivity:

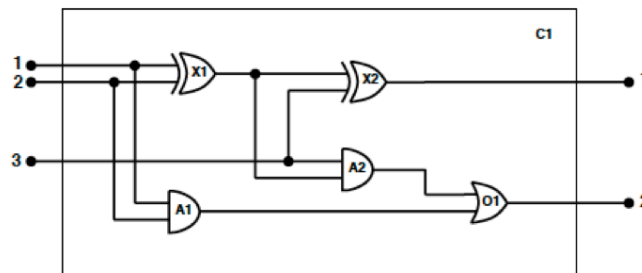
$Connected(Out(1, X_1), In(1, X_2))$ ,

$Connected(In(1, C_1), In(1, X_1))$ ,

$Connected(Out(1, X_1), In(1, A_2))$

$Connected(In(1, C_1), In(1, A_1))$ ,

etc.



# Queries

Our theory captures full behavior.

Can now ask many different queries about behavior etc.

E.g.

$\exists i_1, i_2, i_3 \text{ Signal}(In(1, C_1)) = i_1 \wedge \text{Signal}(In(2, C_1)) = i_2 \wedge \text{Signal}(In(3, C_1)) = i_3 \wedge \text{Signal}(Out(1, C_1)) = Off \wedge \text{Signal}(Out(2, C_1)) = On$

A.:  $(I_1 = On \wedge i_2 = On \wedge i_3 = Off) \vee$

$(I_1 = On \wedge i_2 = Off \wedge i_3 = On) \vee$

$(I_1 = Off \wedge i_2 = On \wedge i_3 = On)$

What is the advantage over direct simulation?

etc. **Aside: previously  $\text{KB} \models \alpha$  The same here but we want a bit more “detailed answer”. Inference will also give us variable bindings if existential query is entailed.**

Of course, formalization is somewhat facilitated by the “closeness” between logical formalisms and digital circuitry. Starting with Shannon, allows for very powerful design methods (but did not prevent Pentium bug...).

**Done with prop. logic. Just check for syntax and FOL form.**

## One more example: Graph Coloring

Graph:  $N$  nodes,  $K$  colors.

- 1)  $\forall i (1 \leq i \leq N) \exists j (1 \leq j \leq K) Color(i, j)$   
 $\forall i, j, l (1 \leq i \leq N) (1 \leq j, l \leq K)$   
 $[(Color(i, j) \wedge Color(i, l)) \Rightarrow (j = l)]$
- 2)  $\forall i, j (1 \leq i, j \leq N) [(i \neq j) \Rightarrow$   
 $(Edge(i, j) \Rightarrow [\neg \exists k (1 \leq k \leq K)$   
 $((Color(i, k) \wedge Color(j, k)))])]$

alternative:

- 3)  $\forall i, j \ (1 \leq i, j \leq N) \ [(i \neq j) \Rightarrow$   
 $(Edge(i, j) \Rightarrow [\forall k(1 \leq k \leq K)$   
 $(\neg Color(i, k) \vee \neg Color(j, k))])]$

Now actual graph given by, e.g.,:

- 4)  $Edge(1, 3), Edge(2, 4), Edge(5, 6) \dots$  etc.



reasoning: 3 & 4 gives e.g.:

$$\forall k(1 \leq k \leq K) (\neg Color(1, k) \vee \neg Color(3, k))$$

uses “unification”  $\{i/1, j/3\}$  with Modus Ponens.

For  $K = 5$ , we get:

$$\begin{aligned} &(\neg Color(1, 1) \vee \neg Color(3, 1)), (\neg Color(1, 2) \vee \neg Color(3, 2)) \\ &\dots (\neg Color(1, 5) \vee \neg Color(3, 5)) \end{aligned}$$

in propositional form.

uses Universal Elimination, e.g., substitute  $\{k/1\}$ , etc.

**See R&N p. 443**

**FOL formalizations can be challenging  
for “everyday” concepts.**

Defining **natural kinds** is much more difficult.

e.g. a *game*, or a *chair*.

difficulty with necessary and sufficient conditions.

problem with “strict definition” (Quine 1953)

“the Pope is a bachelor.”

Approaches?

**Probabilistic representations (extending prop. logic /  
FOL) can help!**

# Inference

# Resolution / Unification

## Chapter 9 R&N.

**NOTE:** for finite domains that are not too large, better to “ground to” propositional and use SAT solver.

**In hardware/software verification:**

**Use small numbers (eg int up to 10) OR**

**“bit blasting”, represent 64 bit int by 64 Boolean vars.**

# Inference

We've considered various first-order formalizations.

But, how do we reason with them? Derive new info?

**A.** Use **resolution** as in propositional case

From  $(\alpha \vee p) \wedge (\neg p \vee \beta)$

conclude  $\alpha \vee \beta$  until you reach contradiction.

Need some extra “tricks” to deal with

**quantifiers and variables.**

# Resolution

## I put in clausal form

all variables universally quantified

main trick: “Skolemization” to remove existentials.

idea: invent names for unknown objects known to exist

## II use unification to match atomic sentences

## III apply resolution rule to the clausal set combined

with negated goal. Attempt to generate empty clause.

## Tricks

- **unification:** needed to match variables and terms between clauses that look similar  
See also R&N.
- **normalization:** put in **clausal form**  
move quantifiers /  $\wedge$  /  $\vee$  etc.  
and **Skolemization** — remove  $\exists$  by giving an arbitrary, but unique name to the object in question.  
E.g.  $D$  for the dog owned by Jack.

## Unification

UNIFY (P,Q) takes two atomic sentences P and Q and returns a substitution that makes P and Q **look the same**.

Rules for substitutions:

- Can replace a variable by a constant.
- Can replace a variable by a variable.
- Can replace a variable by a function expression, as long as the function expression does not contain the variable.

**Unifier:** a substitution that makes two clauses resolvable.

$$v_1 \rightarrow C; v_2 \rightarrow v_3; v_4 \rightarrow f(\dots)$$

## Unification — Purpose

Given:  $\forall x (\neg \mathbf{Knows}(\mathbf{John}, x) \vee \mathbf{Hates}(\mathbf{John}, x))$

$\mathit{Knows}(\mathit{John}, x) \rightarrow \mathit{Hates}(\mathit{John}, x)$

$\mathit{Knows}(\mathit{John}, \mathit{Jim})$

Derive

$\mathit{Hates}(\mathit{John}, \mathit{Jim})$

Need **unifier**  $\{x/\mathit{Jim}\}$  before resolution.

(simplest case)



$\neg \text{Knows}(\text{John}, x) \vee \text{Hates}(\text{John}, x)$  and  $\text{Knows}(\text{John}, \text{Jim})$

How do we resolve? First, match them.

Solution:

$\text{UNIFY}(\text{Knows}(\text{John}, x), \text{Knows}(\text{John}, \text{Jim})) = \{x/\text{Jim}\}$

Gives

$\neg \text{Knows}(\text{John}, \text{Jim}) \vee \text{Hates}(\text{John}, \text{Jim})$  and  
 $\text{Knows}(\text{John}, \text{Jim})$

Conclude by resolution

$\text{Hates}(\text{John}, \text{Jim})$

# Unification (example)

general rule:

$$\textit{Knows}(\textit{John}, x) \rightarrow \textit{Hates}(\textit{John}, x)$$

facts:

$$\textit{Knows}(\textit{John}, \textit{Jim})$$

$$\textit{Knows}(y, \textit{Leo})$$

$$\textit{Knows}(y, \textit{Mother}(y))$$

$$\textit{Knows}(x, \textit{Jane})$$

“matching facts to general rules”

**Can substitute in because original clause universally quantified.**

$\text{UNIFY}(\text{Knows}(\text{John}, x), \text{Knows}(\text{John}, \text{Jim})) = \{x/\text{Jim}\}$

$\text{UNIFY}(\text{Knows}(\text{John}, x), \text{Knows}(y, \text{Leo})) = \{x/\text{Leo}, y/\text{John}\}$

$\text{UNIFY}(\text{Knows}(\text{John}, x), \text{Knows}(y, \text{Mother}(y))) =$   
 $\{y/\text{John}, x/\text{Mother}(\text{John})\};$

$\text{UNIFY}(\text{Knows}(\text{John}, x), \text{Knows}(x, \text{Jane})) = \text{fail}$

- Last one fails because  $x$  can't take on both the value John and the value Jane. But intuitively we know that everyone John knows he hates and everyone knows Jane so we should be able to infer that John hates Jane.
- This is why we require, if possible, that every variable has a separate name.  $\text{Knows}(\text{John},x)$  and  $\text{Knows}(y,\text{Jane})$  works.

## Most General Unifier

In cases where there is more than one substitution choose the one that makes the least commitment (most general) about the bindings.

$\text{UNIFY}(\textit{Knows}(\textit{John}, x), \textit{Knows}(y, z))$

$= \{y/\textit{John}, x/z\}$

or  $\{y/\textit{John}, x/z, z/\textit{Freda}\}$

or  $\{y/\textit{John}, x/\textit{John}, z/\textit{John}\}$

or ....

# Normal form: Clausal

See also, R&N.

- **Eliminate implication**

$p \Rightarrow q$  becomes  $\neg p \vee q$

- **Move  $\neg$  inwards**

e.g.,  $\neg(p \vee q)$  becomes  $(\neg p \wedge \neg q)$

$\neg\exists x . p$  becomes  $\forall x \neg p$

$\neg\forall x . p$  becomes ...

- **Standardize variables**

rename variables to avoid conflicts.

- **Move quantifiers left**

e.g.,  $p \vee \forall x q$  becomes  $\forall x (p \vee q)$

- **Skolemize** (remove existentials)

e.g.  $\forall x \text{ Person}(x) \Rightarrow \exists y \text{ Heart}(y) \wedge \text{Has}(x, y)$

consider:

$\forall x \text{ Person}(x) \Rightarrow \text{Heart}(H) \wedge \text{Has}(x, H)$

problem??

$\forall x \text{ Person}(x) \Rightarrow \text{Heart}(F(x)) \wedge \text{Has}(x, F(x))$

- **Distribute  $\wedge$  over  $\vee$**

$(a \wedge b) \vee c$  becomes  $(a \vee c) \wedge (b \vee c)$

- **Flatten nested conjunctions and disjunctions**

e.g.  $(a \vee b) \vee c$  becomes  $(a \vee b \vee c)$



## Example

Jack owns a dog.

**Natural language  
input.**

Every dog owner is an animal lover. **(From dog to more general.)**

No animal lover kills an animal. **(General statement.)**

Either Jack or Curiosity killed the cat, who is named Tuna.

Did Curiosity kill the cat? ***The query***

**YES!**

***What “hidden” background knowledge is being used?***

## Original Sentences (Plus Background Knowledge)

Jack owns a dog.

Every dog owner is an animal lover.

No animal lover kills an animal.

Either Jack or Curiosity killed the cat, who is named Tuna.

Did Curiosity kill the cat?

5.  $Cat(Tuna)$

Cats are animals.

6.  $\forall x \text{ } Cat(x) \rightarrow Animal(x)$

Not stated explicitly! **This is an example of *background***

**Query:  $KB \models Kills(Curiosity, Tuna) ??$  *knowledge* key to Natural Language Understanding.**

# Clausal Form

**D is a “fake name” for the dog.**

1.  $Dog(D)$

$\exists x : Dog(x) \wedge Owns(Jack, x)$

2.  $Owns(Jack, D)$

$\forall x (\exists y Dog(y) \wedge Owns(x, y)) \rightarrow AnimalLover(x)$

3.  $\neg Dog(S(x)) \vee \neg Owns(x, S(x)) \vee AnimalLover(x)$

$\forall x AnimalLover(x) \rightarrow (\forall y Animal(y) \rightarrow \neg Kills(x, y))$

4.  $\neg AnimalLover(w) \vee \neg Animal(y) \vee \neg Kills(w, y)$

$Kills(Jack, Tuna) \vee Kills(Curiosity, Tuna)$

5.  $Kills(Jack, Tuna) \vee Kills(Curiosity, Tuna)$

$Cat(Tuna)$

6.  $Cat(Tuna)$

7.  $\neg Cat(z) \vee Animal(z)$

$\forall x Cat(x) \rightarrow Animal(x)$

**Negation of query!**

**Missing?**

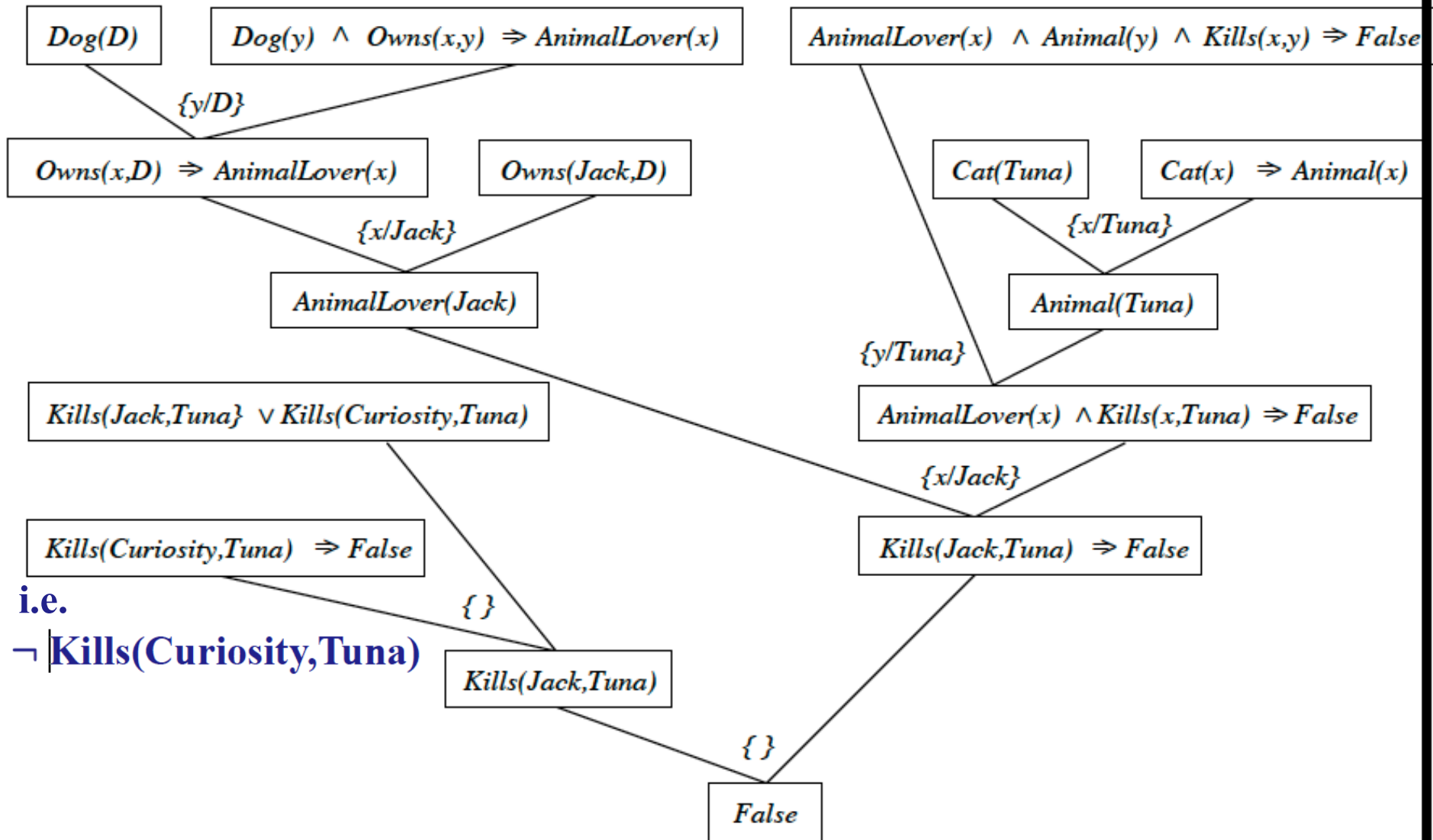
**8.  $\neg Kills(Curiosity, Tuna)$**

**Proof by contradiction  
for resolution.**

59

**Translation to clausal form automatic.**

# Proof by Resolution with Refutation



*Warning: Non-standard notation!!*

## First-order resolution proof (more carefully)

1.  $Dog(D)$

3.  $\neg Dog(S(x)) \vee \neg Owns(x, S(x)) \vee AnimalLover(x)$

9.  $\neg Owns(x, D) \vee AnimalLover(x)$  using  $S(x)/D$

2.  $Owns(Jack, D)$

10.  **$AnimalLover(Jack)$**  using  $x/Jack$

6.  $Cat(Tuna)$

7.  $\neg Cat(z) \vee Animal(z)$

11.  **$Animal(Tuna)$**  using  $z/Tuna$

11. **Animal (Tuna)**

$$4. \neg \text{AnimalLover}(w) \vee \neg \text{Animal}(y) \vee \neg \text{Kills}(w, y)$$

12.  $\neg \text{AnimalLover}(w) \vee \neg \text{Kills}(w, \text{Tuna})$  using  $y/\text{Tuna}$

10. **AnimalLover(Jack)**

13.  $\neg \text{Kills}(\text{Jack}, \text{Tuna})$  using  $w/\text{Jack}$

$$5. \text{Kills}(\text{Jack}, \text{Tuna}) \vee \text{Kills}(\text{Curiosity}, \text{Tuna})$$

14. **Kills(Curiosity, Tuna)**

*15 lines; trivial with  
modern solvers.*

8.  $\neg \text{Kills}(\text{Curiosity}, \text{Tuna})$

*Can do 1+ billion lines!*

15  $\square$  (contradiction reached)

So,  $\text{KB} \models \text{Kills}(\text{Curiosity}, \text{Tuna})$

Jack owns a dog.

Every dog owner is an animal lover.

No animal lover kills an animal.

Either Jack or Curiosity killed the cat, who is named Tuna.

Did Curiosity kill the cat?

**So, we answered a natural language query using**

- |                                     |                              |
|-------------------------------------|------------------------------|
| <b>(1) Natural language parsing</b> | <b>(almost there)</b>        |
| <b>(2) Background knowledge</b>     | <b>(much work remaining)</b> |
| <b>(3) Reasoning</b>                | <b>(works fine now)</b>      |

**We will see much progress in this kind of natural language question answering in next decade.**

*Eg executable semantic parsing.  
Persi Liang, Stanford.*

# Completeness

**F.O. Resolution** with **unification** applied to **clausal form**,  
is **refutation** complete.

Interesting proof! Based on building an “artificial” domain  
of interpretation, called the **Herbrand universe**.



# Schubert Steamroller

Wolves, foxes, birds, caterpillars, and snails are animals, and there are some of each of them.

Also there are some grains, and grains are plants.

Every animal either likes to eat all plants or all animals much smaller than itself that like to eat some plants.

Caterpillars and snails are much smaller than birds, which are much smaller than foxes, which are much smaller than wolves.

Wolves do not like to eat foxes or grains, while birds like to eat caterpillars but not snails.

Caterpillars and snails like to eat some plants.

Some logical forms:

$$\forall x (Wolf(x) \Rightarrow animal(x))$$

$$\forall x \forall y ((Caterpillar(x) \wedge Bird(y)) \Rightarrow Smaller(x, y)).$$

$$\exists x bird(x)$$

To prove:

There is an animal that likes to eat  
a grain-eating animal.

**Requires almost 150 resolution steps (minimal)**

Significant challenge for early systems.

Certain sets of first-order statements can overwhelm general resolution solvers, e.g. about infinite sets (natural numbers).

As a consequence, many **practical** Knowledge Representation formalisms in AI use a **restricted form** and **specialized inference**.

Can often understand them in terms of standard first-order logic! (clear **syntax & semantics**)

**Or, better yet, for finite domains, fall back to SAT solvers.**

**Concludes propositional and first-order logic for knowledge representation and reasoning.**

**Next “Big Picture Slide”**