

First-Order Logic and Inference



Berlin Chen
Department of Computer Science & Information Engineering
National Taiwan Normal University



References:

1. S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*, Chapters 7,8 and 9
2. S. Russell's teaching materials

Pros and Cons of Propositional Logic (PL)

- PL is **declarative**
 - Pieces of syntax correspond to facts
 - Knowledge and inference are separate and inference is entirely domain-independent
- PL is **compositional**
 - Meaning of a sentence is a function of the meaning of its parts
 - E.g., meaning of $B_{1,1} \wedge P_{1,2}$ is derived from meaning of $B_{1,1}$ and of $P_{1,2}$
- PL can **deal with partial information**
- The meaning of PL is **context-independent**
- PL has **very limited expressive power**
 - E.g., cannot say “pits cause breezes in adjacent squares” except by writing one sentence for each square

$$B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1}),$$
$$B_{2,1} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1}), \dots$$

The lack of concise representations

Natural Languages

- Natural Languages are
 - Very expressive
 - Mediums for communication rather than pure representation
 - Context-dependent
 - Not purely compositional (e.g., “和尚” in Chinese)
 - Ambiguous (e.g., 三人參加 in Chinese=> 三-人參-加 or 三-人-參加)
- Major elements of natural Languages
 - Nouns and noun phrases: refer to **objects**
 - E.g., people, houses, colors, ...
 - Verbs and verb phrases: refer to **relations** among objects
 - E.g., is red, is round, (properties)...; is brother of, has color, ...
 - Some of the relations are **functions** which return one value for a given input
 - E.g., father of, best friend, beginning of, ...

First-Order Logic (FOL)

- Whereas PL assumes world containing facts, FOL assumes the world contains objects and relations
 - FOL can express facts about some or all the objects in the universe
 - Such as “Squares neighboring the wumpus are smelly”
- Objects: things with individual identities
- Relations
 - Relations
 - Functions
 - Properties: distinguish objects from others

} Interrelations among objects

Logics in General

Language	Ontological Commitment	Epistemological Commitment
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 $\in [0, 1]$
Fuzzy logic	degree of truth $\in [0, 1]$	known interval value

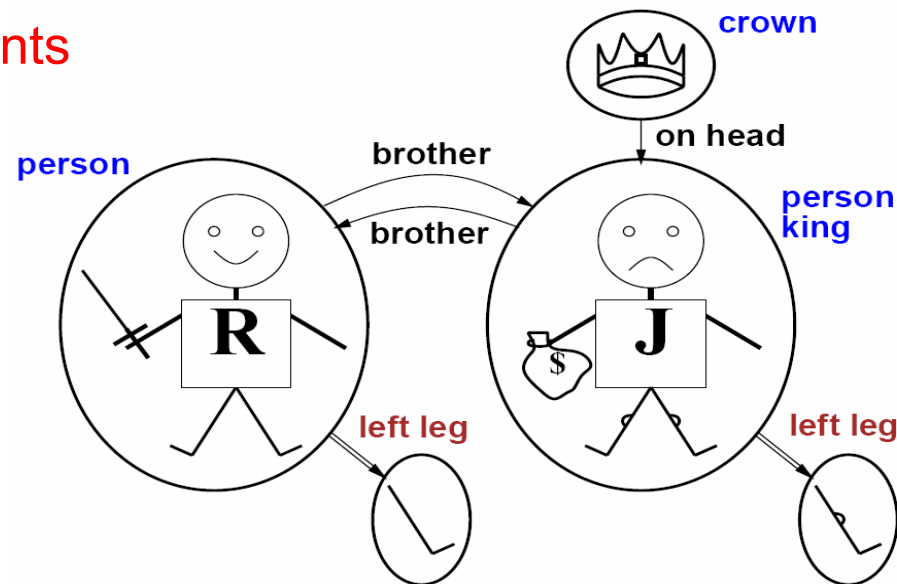
Examples

- One plus two equals three
 - Objects: one, two, three, one plus two
 - Relation: equals
 - Function: plus
- Squares neighboring the wumpus are smelly
 - Objects: wumpus, square
 - Property: smelly
 - Relation: neighboring
- Evil King John ruled England in 1200
 - Objects: John, England, 1200
 - Properties: evil, king
 - Relation: ruled

Models for FOL

- A model contains objects and relations among them
 - PL: models are sets of truth values for proposition symbols
- The domain of a model is the set of objects
 - Objects are **domain elements**

- Example



5 **objects**: Richard, John, Richard's left legs, John's left legs, crown

2 **binary relations**: brother, on head

3 **unary relations**: person, king, crown

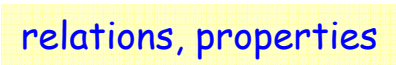
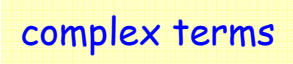
1 **unary function**: left leg

Syntax of FOL

- **BNF** (Backus-Naur Form) grammar for FOL

Sentence \rightarrow *AtomicSentence*
| (*Sentence* *Connective* *Sentence*)
| *Quantifier* *Variables*, *Sentence*
| \neg *Sentence*

AtomicSentence \rightarrow *Predicate*(*Term*, ...) | *Term* = *Term*

Term \rightarrow *Function*(*Term*, ...) 
| *Constant*
| *Variable* 

Connective \rightarrow \Rightarrow | \wedge | \vee | \Leftrightarrow

Quantifier \rightarrow \forall | \exists

Constant \rightarrow *A* | *X*₁ | *John* | ...

Variable \rightarrow *a* | *x* | *s* | ...

Predicate \rightarrow *Before* | *HasColor* | *Raining* | ...

Function \rightarrow *Mother* | *LeftLeg* | ...

Semantics of FOL

- The truth of any sentence is determined by a model and an interpretation for the sentence's symbols
- Interpretation specifies exactly which objects, relations and functions are referred by the constant, predicate, and function symbols
 - Constant symbols \rightarrow objects
 - Predicate symbols \rightarrow relations, properties
 - Function symbols \rightarrow functional relations
- An atomic sentence $Predicate(Term_1, \dots, Term_n)$ is true iff the objects referred to by $Term_1, \dots, Term_n$ are in the relation referred to by $Predicate$

Terms

- A term is a logic expression that refers to an object
- Simple term: e.g., constant/variable symbols
- Complex term: formed by a function symbol followed a parenthesized list of terms as arguments to the function symbol
 - The complex term refers to an object that is the value of the function (symbol) applied to the arguments

– E.g., *LeftLeg(John)*

function symbol argument/term

Atomic Sentences

- An atomic sentence is formed by
 - A predicate symbol followed by a parenthesized list of terms
 - $Predicate(Term_1, \dots, term_n)$
 - E.g., *Brother(Richard, John)*
 - Or just $term_1=term_2$
- Atomic sentences can have complex terms as arguments
 - E.g., *Married(Father(Richard), Mother(John))*

Complex Sentences

- An complex sentence is constructed using logical connectives

- Negation

$\neg \text{Brother}(\text{LeftLeg}(\text{Richard}), \text{John})$

- Conjunction

$\text{Brother}(\text{Richard}, \text{John}) \wedge \text{Brother}(\text{John}, \text{Richard})$

- Disjunction

$\text{King}(\text{Richard}) \vee \text{King}(\text{John})$

- Implication

$\neg \text{King}(\text{Richard}) \Rightarrow \text{King}(\text{John})$

The truth or falsehood of a complex sentence can be determined from the truth or falsehood of its component sentences

Universal Quantification

- The following sentence remains truth for **all** values of the variable

$\forall \langle \textit{variable} \rangle \langle \textit{sentence} \rangle$

- Variables are lowercase
- E.g., “Everyone in Taiwan is industrious”

$\forall x \textit{In}(x, \textit{Taiwan}) \Rightarrow \textit{Industrious}(x)$

$\forall x$ P is true in a model m iff P with x being each possible object in the model

- Equivalent to the **conjunction** of instantiations of P

$\textit{In}(\textit{Thomas}, \textit{Taiwan}) \Rightarrow \textit{Industrious}(\textit{Thomas})$

$\wedge \textit{In}(\textit{Rich}, \textit{Taiwan}) \Rightarrow \textit{Industrious}(\textit{Rich})$

$\wedge \textit{In}(\textit{Vicent}, \textit{Taiwan}) \Rightarrow \textit{Industrious}(\textit{Vicent})$

$\wedge \textit{In}(\textit{Eileen}, \textit{Taiwan}) \Rightarrow \textit{Industrious}(\textit{Eileen})$

$\wedge \dots$

Universal Quantification: A Common Mistake

- Typically, \Rightarrow (implication) is the main connective with \forall
- Common mistake: using \wedge as the main connective with \forall
 $\forall x \text{ In}(x, \text{Taiwan}) \wedge \text{Industrious}(x)$

Means “ Everyone is in Taiwan and everyone is industrious”

an overly strong statement

Existential Quantification

- The following sentence remains true for **some** values of the variable

$\exists \langle \textit{variable} \rangle \langle \textit{sentence} \rangle$

- E.g., “Someone in Taiwan is industrious”

$\exists x \textit{In}(x, \textit{Taiwan}) \wedge \textit{Industrious}(x)$

$\exists x P$ is true in a model m iff P with x being each possible object in the model

- Equivalent to the **disjunction** of instantiations of P

$(\textit{In}(\textit{Thomas}, \textit{Taiwan}) \wedge \textit{Industrious}(\textit{Thomas}))$

$\vee (\textit{In}(\textit{Rich}, \textit{Taiwan}) \wedge \textit{Industrious}(\textit{Rich}))$

$\vee (\textit{In}(\textit{Vicent}, \textit{Taiwan}) \wedge \textit{Industrious}(\textit{Vicent}))$

$\vee (\textit{In}(\textit{Eileen}, \textit{Taiwan}) \wedge \textit{Industrious}(\textit{Eileen}))$

$\vee \dots\dots$

Existential Quantification : A Common Mistake

- Typically, \wedge is the main connective with \exists
- Common mistake: using \Rightarrow as the main connective with \exists

$$\exists x \text{ In}(x, \text{Taiwan}) \Rightarrow \text{Industrious}(x)$$

Is true if there is anyone who is not in Taiwan

a overly weak statement

Properties of Quantifiers

- Nested Quantifiers (commutative!)

$\forall x \forall y$ is the same as $\forall y \forall x$

$\exists x \exists y$ is the same as $\exists y \exists x$

$\exists x \forall y$ is the same as $\forall y \exists x$

- Examples:

- “There is a person who loves everyone in the world”

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

- “Everyone in the world is loved by at least one person”

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

- Quantifier Duality

- Each of the following sentences can be expressed using the other

$\forall x \text{ Likes}(x, \text{IceCream}) \iff \neg \exists x \neg \text{Likes}(x, \text{IceCream})$

$\exists x \text{ Likes}(x, \text{IceCream}) \iff \neg \forall x \neg \text{Likes}(x, \text{IceCream})$

Equality

- Make statements to the effect that two terms refer to the same object
 - Determine the truth of an equality sentence by seeing that the referents of the two terms are the same objects
 - E.g., state the facts about a given function

Father(John)=Henry

- E.g., insisting that two terms are not the same objects

$\exists x \exists y \text{ Brother}(x, \text{Richard}) \wedge \text{ Brother}(y, \text{Richard}) \wedge \neg(x=y)$

- Richard has at least two brothers

Review: De Morgan's Rules

$$\forall x \neg P \equiv \neg \exists x P$$

$$\neg \forall x P \equiv \exists x \neg P$$

$$\forall x P \equiv \neg \exists x \neg P$$

$$\exists x P \equiv \neg \forall x \neg P$$

$$\neg P \wedge \neg Q \equiv \neg (P \vee Q)$$

$$\neg (P \wedge Q) \equiv \neg P \vee \neg Q$$

$$P \wedge Q \equiv \neg (\neg P \vee \neg Q)$$

$$P \vee Q \equiv \neg (\neg P \wedge \neg Q)$$

Using First-Order Logic

- Assertions and Queries

- Assertions:

- Sentences are added to KB using **TELL**, such sentences are called assertions

TELL(KB, King(John))

TELL(KB, $\forall x \text{ King}(x) \Rightarrow \text{Person}(x)$)

- Queries

- Questions are asked using **ASK**, which are also called queries or goals

ASK(KB, King(John))



return **true**

ASK(KB, Person(John))



return **true**

ASK(KB, $\exists x \text{ Person}(x)$)



return **{x/John}**

A substitution
or binding list

Using First-Order Logic

- Example: The Kinship Domain

- One's mother is one's female parent

$$\forall m,c \text{ Mother}(m, c) \Leftrightarrow (\text{Female}(m) \wedge \text{Parent}(m, c))$$

- One's husband is one's male spouse

$$\forall w,h \text{ Husband}(h, w) \Leftrightarrow (\text{Male}(h) \wedge \text{Spouse}(h, w))$$

- A grandparent is a parent of one's parent

$$\forall g,c \text{ Grandparent}(g, c) \Leftrightarrow (\exists p \text{ Parent}(g, p) \wedge \text{Parent}(p, c))$$

- A sibling is another child of one's parents

$$\forall x,y \text{ Sibling}(x, y) \Leftrightarrow x \neq y \wedge (\exists p \text{ Parent}(p, x) \wedge \text{Parent}(p, y))$$

- A first cousin is a child of a parent's sibling

$$\forall x,y \text{ FirstCousin}(x, y) \Leftrightarrow \exists p,ps \text{ Parent}(p, x) \wedge \text{Sibling}(ps, p) \\ \wedge \text{Parent}(ps, y)$$

Interacting with FOL *KBs*

- Suppose a wumpus-world agent is using an FOL *KB* and perceives a stench and a breeze (but no glitter) at $t = 5$

stench breeze glitter bump scream
↓ ↓ ↓ ↓ ↓
Tell(*KB*, percept([Stench, Breeze, None, None, None], 5))
Ask(*KB*, $\exists a$ BestAction(a , 5))

- Does the *KB* entail any particular actions at $t = 5$?

Answer: Yes, $\{a/shoot\}$ A substitution or binding list

- Given a sentence *S* and a substitution θ , SUBST(θ , *S*) denotes the result of plugging θ into *S* ; e.g.,

$S = \text{Smarter}(x, y)$

$\theta = \{x/Vicent, y/Thomas\}$

SUBST(θ , *S*) = Smarter(Vicent, Thomas)

- ASK(*KB*, *S*) returns some/all θ (substitutions or binding lists) such that $KB \models \text{SUBST}(\theta, S)$

KB for the Wumpus World

- Perception

stench breeze glitter bump scream

↓ ↓ ↓ ↓ ↓

$$\forall t, s, g, m, c \text{ Percept}([s, \text{Breeze}, g, m, c], t) \Rightarrow \text{Breeze}(t)$$
$$\forall t, s, b, m, c \text{ Percept}([s, b, \text{Glitter}, m, c], t) \Rightarrow \text{Glitter}(t)$$

- Reflex

$$\forall t \text{ Glitter}(t) \Rightarrow \text{BestAction}(\text{Grab}, t)$$

- Environment

$$\forall x, y, a, b \text{ Adjacent}([x, y], [a, b]) \Leftrightarrow$$
$$[a, b] \in \{[x+1, y], [x-1, y], [x, y+1], [x, y-1]\}$$

- Properties of agent's locations

$$\forall s, t \text{ At}(\text{Agent}, s, t) \wedge \text{Breeze}(t) \Rightarrow \text{Breezy}(s)$$

KB for the Wumpus World

- Square are breezy near a pit
 - **Diagnostic rule** – infer hidden causes from observable effects
 - If a square is breezy , some adjacent square must contain a pit
$$\forall s \text{ Breezy}(s) \Rightarrow \exists r \text{ Adjacent}(r, s) \wedge \text{Pit}(r)$$
 - If a square is not breezy , no adjacent square contains a pit
$$\forall s \neg \text{Breezy}(s) \Rightarrow \neg \exists r \text{ Adjacent}(r, s) \wedge \text{Pit}(r)$$
 - Combined:
$$\forall s \text{ Breezy}(s) \Leftrightarrow \exists r \text{ Adjacent}(r, s) \wedge \text{Pit}(r)$$
 - **Causal rule** – infer observable effects from hidden causes
 - A pit causes all adjacent squares to be breezy
$$\forall r \text{ Pit}(r) \Rightarrow [\forall s \text{ Adjacent}(r, s) \Rightarrow \text{Breezy}(s)]$$
 - If all squares adjacent to a given square are pitless, the square will not be breezy
$$\forall s [\forall r \text{ Adjacent}(r, s) \Rightarrow \neg \text{Pit}(r)] \Rightarrow \neg \text{Breezy}(s)$$
 - Combined:
$$\forall s \text{ Breezy}(s) \Leftrightarrow \exists r \text{ Adjacent}(r, s) \wedge \text{Pit}(r)$$

model-based
reasoning

Inference Rules for Quantifiers

- Substitution $\text{SUBST}(\theta, \alpha)$
 - Refer to applying the *substitution* θ to the sentence α
 - θ is a set of variable/(ground)term pairs

$$\begin{array}{ccc}
 & \theta = \{x/\text{John}\} & \\
 \text{Person}(x) & \longrightarrow & \text{Person}(\text{John}) \\
 & & \text{SUBST}(\theta, \alpha)
 \end{array}$$

- Universal Instantiation (UI)
 - Infer any sentence obtained by substituting a ground term for the universally quantified variable
 - A ground term is a term without variable
 - could be a complex term

$$\frac{\forall v \alpha}{\text{SUBST}(\{v/g\}, \alpha)} \xrightarrow{\theta = \{x/\text{John}\}} \frac{\forall x \text{King}(x) \wedge \text{Greedy}(x) \Rightarrow \text{Evil}(x)}{\text{King}(\text{John}) \wedge \text{Greedy}(\text{John}) \Rightarrow \text{Evil}(\text{John})}$$

Inference Rules for Quantifiers

- Existential Instantiation (EI)
 - Infer any sentence obtained by substituting a new constant symbol that **does not appear elsewhere in the KB** for the existentially quantified variable

$$\frac{\exists v \alpha}{\text{SUBST}(\{v/k\}, \alpha)} \xrightarrow{\theta = \{x/c_1\}} \frac{\exists x \text{Crown}(x) \wedge \text{OnHead}(x, \text{John})}{\text{Crown}(C_1) \wedge \text{OnHead}(C_1, \text{John})}$$

- A new constant symbol called **Skolem constant**

Universal/Existential Instantiation

- Universal instantiation can be applied several times to add new sentences
 - The new *KB* is logically equivalent to the old one
- Existential instantiation can be applied just once to replace the existential sentence
 - The new *KB* is not equivalent to the old one (?)
 - But is satisfiable iff the old *KB* was satisfiable

Reduction to Propositional Inference

- Suppose the KB contains:

$\forall x \text{ King}(x) \wedge \text{Greedy}(x) \Rightarrow \text{Evil}(x)$
King(John)
Greedy(John)
Brother(Richard, John)

- Instantiate the universal sentence in all possible ways:

King(John) \wedge Greedy(John) \Rightarrow Evil(John)
King(Richard) \wedge Greedy(Richard) \Rightarrow Evil(Richard)
King(John)
Greedy(John)
Brother(Richard, John)

- The new KB is propositionalized
 - View the ground atomic sentences as propositional symbols

King(John), Greedy(John), Evil(John), King(Richard), etc.

Reduction to Propositional Inference

- Claims
 - A ground sentence is entailed by new *KB* iff entailed by original *KB*
 - Every FOL *KB* can be propositionalized so as to preserve entailment
- Idea
 - Propositionalize *KB* and query, apply **resolution**, return result
- Problem
 - When the *KB* includes a function symbol, there are infinitely many ground terms can be generated from substitutions
 - E.g., *Father(Father(Father(John)))*

Reduction to Propositional Inference

- Theorem: Herbrand (1930)
 - If a sentence is entailed by the original FOL *KB*, there is a proof involving just a **finite** subset of the propositionalized *KB*
- Idea:
 - for $n = 0$ to ∞ do
 - create a propositional *KB* by instantiating with depth- n terms
 - to see if the sentence α is entailed by this *KB*
- Problem
 - It works if α is entailed, and it loops if α is not entailed

$Father(John) \Rightarrow Father(Father(John)) \Rightarrow Father(Father(Father(John))) \Rightarrow \dots$

depth 1 depth 2 depth 3 depth n

Reduction to Propositional Inference

- Theorem: Turing (1936), Church (1936)
 - Entailment in FOL is **semidecidable**
 - Algorithms exists that say yes to every entailed sentence
 - The programs will halt
 - But no algorithm exists that also say no to every nonentailed sentence
 - The programs will stuck in a infinite loop
 - More deeply nested terms were generated

Problems with Propositionalization

- Propositionalization approach is rather inefficient
 - It seems to generate lots of irrelevant sentences
 - E.g., from

$\forall x \text{ King}(x) \wedge \text{Greedy}(x) \Rightarrow \text{Evil}(x)$
 $\text{King}(\text{John})$
 $\forall y \text{ Greedy}(y)$
 $\text{Brother}(\text{Richard}, \text{John})$

it seems obvious that $\text{Evil}(\text{John})$, but propositionalization produces lots of facts such as $\text{Greedy}(\text{Richard})$ that are irrelevant

- With p k -ary predicates and n constants, there are $p \cdot n^k$ instantiations

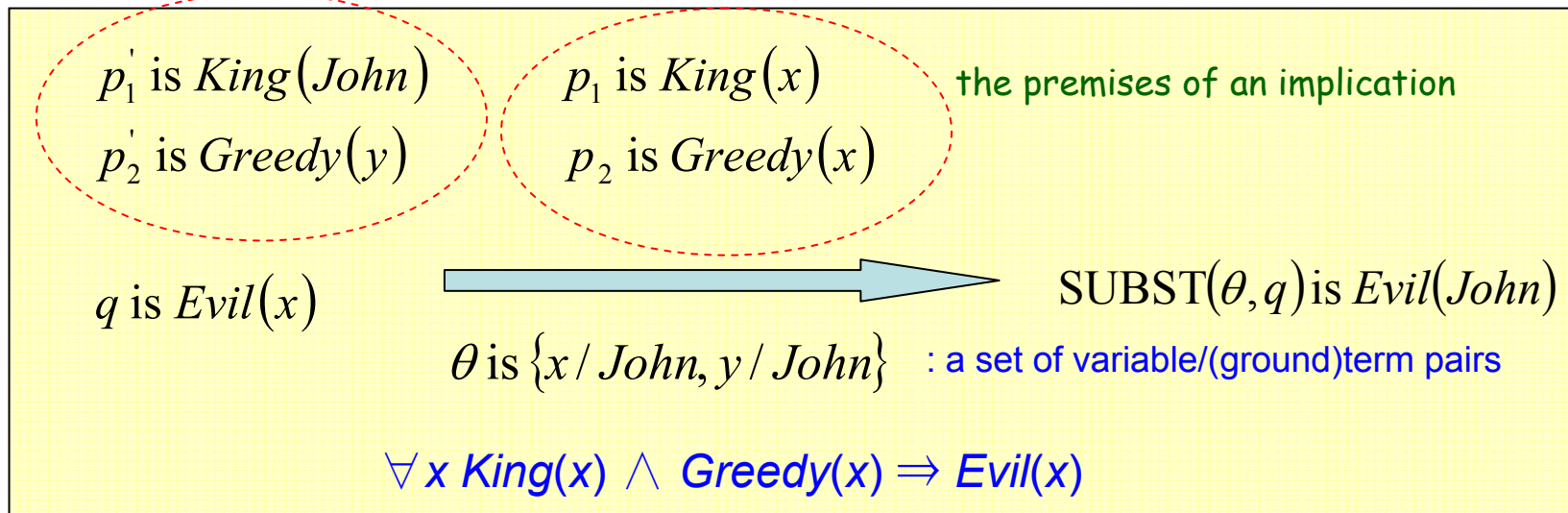
Generalized Modus Ponens (GMP)

- For atomic sentences p_i, p_i' , and q , where there is a substitution θ such that $\text{SUBST}(\theta, p_i') = \text{SUBST}(\theta, p_i)$ for all i

$$\frac{p_1', p_2', \dots, p_n', (p_1 \wedge p_2 \wedge \dots \wedge p_n \Rightarrow q)}{\text{SUBST}(\theta, q)}$$

n atomic sentences p_i'
1 implication

atomic sentences



- GMP used with *KB* of definite clauses (exactly one positive literal)
- All variables assumed universally quantified

Unification

- A process to find a substitution θ which can be applied to two sentences p and q to make them look the same

$$\text{UNIFY}(p, q) = \theta \text{ where } \text{SUBST}(\theta, p) = \text{SUBST}(\theta, q)$$

- The UNIFY algorithm returns a unifier (θ) for the two sentences

- Example

- Query: $\text{Knows}(\text{John}, x)$ KB: $\text{Knows}(\text{John}, \text{Jane})$
 $\text{Knows}(y, \text{Bill})$
 $\text{Knows}(y, \text{Mother}(y))$
 $\text{Knows}(x, \text{Elizabeth})$

p

q

θ

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

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

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

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

Standardizing Apart

- Eliminate overlap of variables to avoid clashes by renaming variables

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

 $\text{UNIFY}(\text{Knows}(\text{John}, x), \text{Knows}(z_{17}, \text{Elizabeth})) = \{x/\text{Elizabeth}, z_{17}/\text{John}\}$

Most Generalized Unifier (MGU)

- Consider the following two unifications

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

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

- We say the first unifier is more general than the second
 - It places fewer restrictions on the values of variables
- For every unifiable pairs of expressions, there is a single **most generalized unifier** (MGU)
 - E.g., the former unifier, $\{y/\text{John}, x/z\}$, shown above

Unification Algorithm

```
function UNIFY( $x, y, \theta$ ) returns a substitution to make  $x$  and  $y$  identical
  inputs:  $x$ , a variable, constant, list, or compound
            $y$ , a variable, constant, list, or compound
            $\theta$ , the substitution built up so far (optional, defaults to empty)

  if  $\theta = \text{failure}$  then return failure
  else if  $x = y$  then return  $\theta$ 
  else if VARIABLE?( $x$ ) then return UNIFY-VAR( $x, y, \theta$ )
  else if VARIABLE?( $y$ ) then return UNIFY-VAR( $y, x, \theta$ )
  else if COMPOUND?( $x$ ) and COMPOUND?( $y$ ) then
    return UNIFY(ARGS[ $x$ ], ARGS[ $y$ ], UNIFY(OP[ $x$ ], OP[ $y$ ],  $\theta$ )) //Op: function symbol
  else if LIST?( $x$ ) and LIST?( $y$ ) then // Args: argument list
    return UNIFY(REST[ $x$ ], REST[ $y$ ], UNIFY(FIRST[ $x$ ], FIRST[ $y$ ],  $\theta$ ))
  else return failure
```

```
function UNIFY-VAR( $var, x, \theta$ ) returns a substitution
  inputs:  $var$ , a variable
            $x$ , any expression
            $\theta$ , the substitution built up so far

  if  $\{var/val\} \in \theta$  then return UNIFY( $val, x, \theta$ )
  else if  $\{x/val\} \in \theta$  then return UNIFY( $var, val, \theta$ )
  else if OCCUR-CHECK?( $var, x$ ) then return failure
  else return add  $\{var/x\}$  to  $\theta$ 
```

As matching a variable against a complex term, check whether the variable itself occurs inside the term.
If it does, the match fails.

Efficient Indexing and Retrieval

- Predicate Indexing

Query: *Knows(John, x)* //an instance of "Fetching"

Employs(x, Richard)

Employs(AIMA.org, y)

Employs(x, y)

KB: *Knows(John, Helen)*

Brother(John, Richard)

Employs(John, Richard)

.....

- Using a hash table
- Maintain indices on keys composed of a predicate plus (one to several) arguments

Forward Chaining

- Operations
 - Start with the atomic sentences (known facts) in the *KB* and apply **Generalized Modus Ponens** in the forward direction (trigger rules whose premises are satisfied)
 - Unification of literals (Predicates in FOL)
 - Adding new atomic sentences (conclusions of implications)
 - Not just a renaming of a known fact
 - Repeat until the query is answered or no further inferences can be made
- To apply *FC*, the *KB* should be converted into a set of definite clauses

Definite Clauses

- Are **disjunctions of literals**, and of which exactly one is positive
 - A definite clause is a Horn clause with exact one positive literal

E.g., $\neg P_1 \vee \neg P_2 \vee \dots \vee \neg P_n \vee Q$ can be converted to
 $(P_1 \wedge P_2 \wedge \dots \wedge P_n) \Rightarrow Q$

- More specifically, a definite clause is
 - Either an **atomic clause (an positive literal)**
 - Or an **implication** whose antecedent (premise/body) is a conjunction of positive literals and whose conclusion (head) is a single positive literal

King(John)

Greedy(y)

King(x) ∧ Greedy(x) ⇒ Evil(x)

- Variables are assumed to be universally quantified
- Not all *KB* can be converted into a set of definite clauses
 - Because of the single-positive-literal restriction

Example *KB*

- The law is that it is a crime for an American to sell weapon to hostile nations. The country Nono, an enemy of America, has some missiles, and all of its missiles were sold by Colonel West, who is American.
- Prove that West is a criminal

Criminal(West) true or false ?

Example *KB*

- It is a crime for an American to sell weapon to hostile nations

$$\text{American}(x) \wedge \text{Weapon}(y) \wedge \text{Sells}(x, y, z) \wedge \text{Hostile}(z) \Rightarrow \text{Criminal}(x) \quad \textcircled{1}$$

- The country Nono has some missiles

$$\exists x \text{ Owns}(\text{Nono}, x) \wedge \text{Missile}(x)$$

$$\longrightarrow \text{Owns}(\text{Nono}, M_1) \quad \textcircled{2}, \text{Missile}(M_1) \quad \textcircled{3}$$

existential elimination/instantiation
AND elimination

- All its (Nono's) missiles are sold to it by West

$$\text{Missile}(x) \wedge \text{Owns}(\text{Nono}, x) \Rightarrow \text{Sells}(\text{West}, x, \text{Nono}) \quad \textcircled{4}$$

- Missiles are weapons

$$\text{Missile}(x) \Rightarrow \text{Weapon}(x) \quad \textcircled{5}$$

- An enemy of America counts as "hostile"

$$\text{Enemy}(x, \text{America}) \Rightarrow \text{Hostile}(x) \quad \textcircled{6}$$

A datalog KB:
composed of a set of
FOL definite clauses
with no function symbols

background knowledge!

Example *KB*

- West, who is American

American(West) ⑦

- The country Nono, an enemy of America

Enemy(Nono, America) ⑧

Example *KB*: FC Proof

- Start with the atomic sentences (known facts) in the *KB*

Proof Tree

American(West)

Missile(M1)

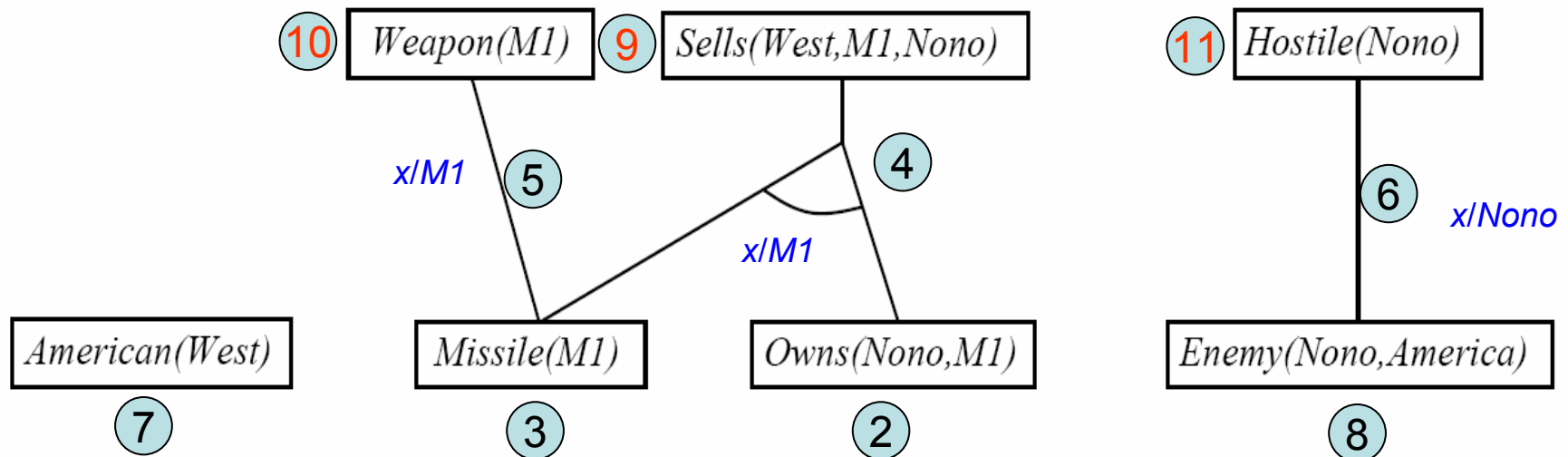
Owns(Nono,M1)

Enemy(Nono,America)

Example *KB*: FC Proof

- Apply Generalized Modus Ponens in the forward direction to trigger rules whose premises are satisfied
- Adding new atomic sentences (conclusions)

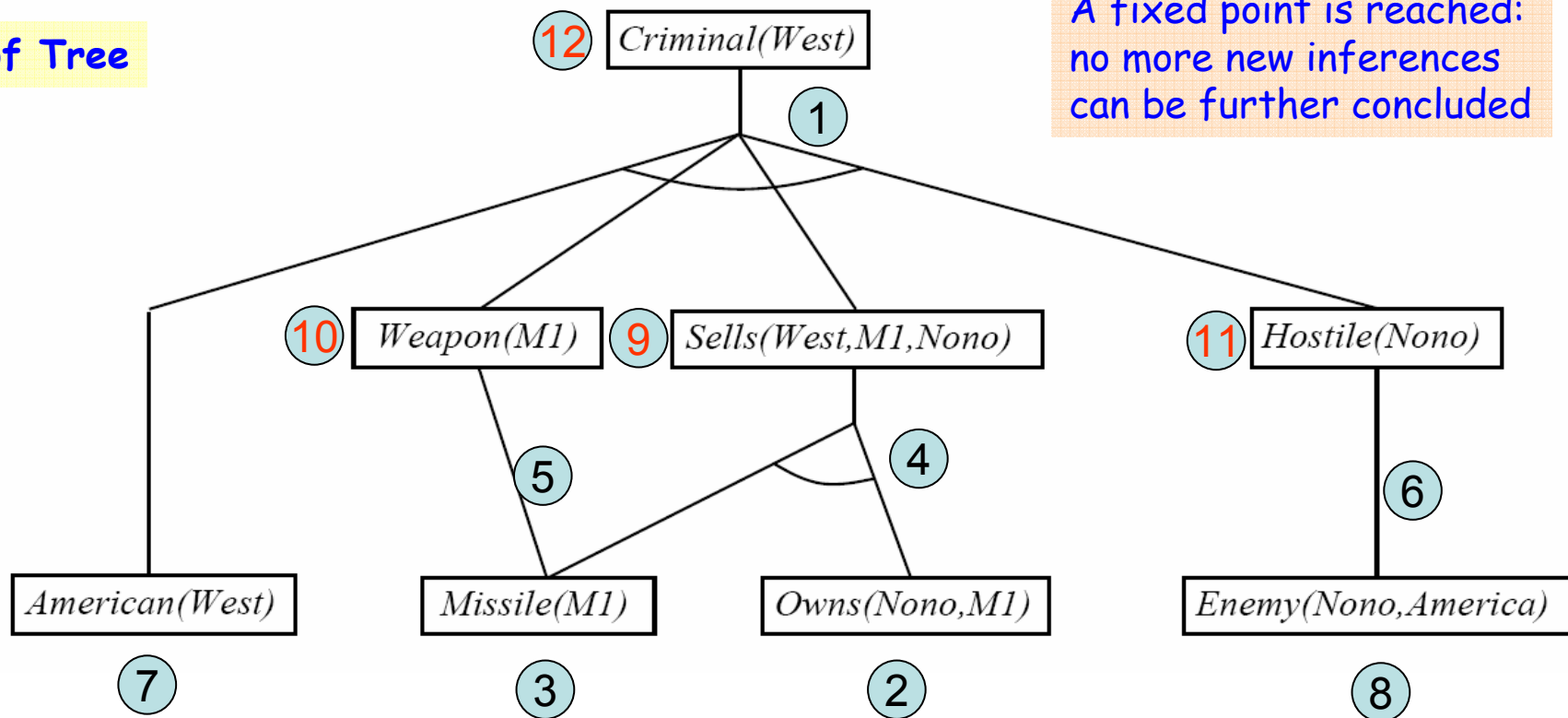
Proof Tree



Example *KB*: FC Proof

- Apply Generalized Modus Ponens in the forward direction to trigger rules whose premises are satisfied
- Adding new atomic sentences (conclusions)

Proof Tree



Forward Chaining Algorithm

function FOL-FC-ASK(KB, α) **returns** a substitution or *false*

inputs: KB , the knowledge base, a set of first-order definite clauses

α , the query, an atomic sentence

local variables: new , the new sentences inferred on each iteration

repeat until new is empty

$new \leftarrow \{ \}$

for each sentence r **in** KB **do**

renaming the variables

$(p_1 \wedge \dots \wedge p_n \Rightarrow q) \leftarrow \text{STANDARDIZE-APART}(r)$

pattern matching

for each θ such that $\text{SUBST}(\theta, p_1 \wedge \dots \wedge p_n) = \text{SUBST}(\theta, p'_1 \wedge \dots \wedge p'_n)$

for some p'_1, \dots, p'_n in KB atomic sentences

$q' \leftarrow \text{SUBST}(\theta, q)$

if q' is not a renaming of some sentence already in KB or new **then do**

add q' to new

$\phi \leftarrow \text{UNIFY}(q', \alpha)$

the new fact unified with the query

if ϕ is not *fail* **then return** ϕ

add new to KB

return *false*

Forward Chaining Algorithm

- Problems
 - The inner loop (pattern matching) is very expensive
 - Rules will be rechecked on every iteration to see if its premises are satisfied
 - Many facts generated are irrelevant to the goal

Incremental Forward Chaining

- Every new fact inferred on iteration t must be derived from at least one new fact from iteration $t-1$
 - Check a rule only if its premise include a conjunct p_i can be unified with a fact p_i' newly inferred at iteration $t-1$
 - If so, fix p_i to match with p_i' and allow the other conjuncts of the rule to match with facts from any previous iteration

Properties of Forward Chaining

- FC is **sound** and **complete** for first-order definite clauses
- FC terminates for Datalog in poly iterations: (at most $p \cdot n^k$)
 - Datalog = first-order definite clauses + **no functions**
- May not terminate in general if α is not entailed
 - Entailment with datalog is **decidable**
 - Entailment with definite clauses is **semi-decidable**
 - When KB with functional symbols

With p k -ary predicates and n constants

$NatNum(0)$

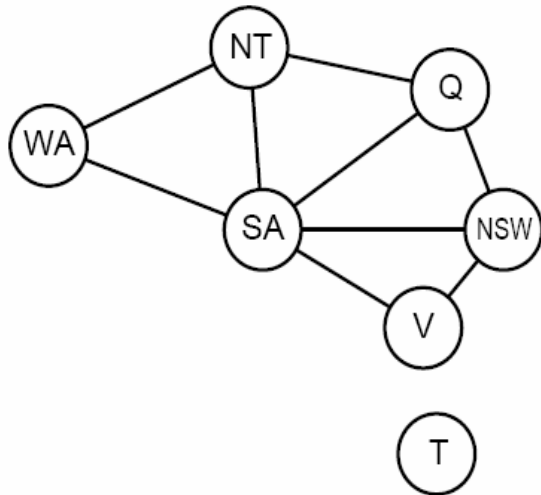
$\forall n \text{ NatNum}(n) \Rightarrow \text{NatNum}(S(n))$

Will add:

$\text{NatNum}(S(0)), \text{NatNum}(S(S(0))), \text{NatNum}(S(S(S(0))))$, ...

Hard Matching Example

- Express a finite-domain CSP as a single definite clause together with some associated ground facts
 - E.g., the map coloring problem



rule

$$\begin{aligned}
 &Diff(wa, nt) \wedge Diff(wa, sa) \wedge \\
 &Diff(nt, q) \wedge Diff(nt, sa) \wedge \\
 &Diff(q, nsw) \wedge Diff(q, sa) \wedge \\
 &Diff(nsw, v) \wedge Diff(nsw, sa) \wedge \\
 &Diff(v, sa) \Rightarrow Colorable()
 \end{aligned}$$

$$\begin{aligned}
 &Diff(Red, Blue) \quad Diff(Red, Green) \\
 &Diff(Green, Red) \quad Diff(Green, Blue) \\
 &Diff(Blue, Red) \quad Diff(Blue, Green)
 \end{aligned}$$

- Matching a definite clause against a set of facts is *NP-hard* Known facts

Backward Chaining

- Work backward from the goal (query), chaining through rules to find known facts that support the proof
 - Put the query on a stack
 - Pop it and find the set of all substitutions that satisfies the query
 - Find all implications in KB whose heads (conclusions) can be unified with the goals and put their bodies (premises) on the stack as new goals
 - Goals unified with known facts generate no new goals
 - If all goals on the stack are satisfied, (the current branch of) the proof succeeds

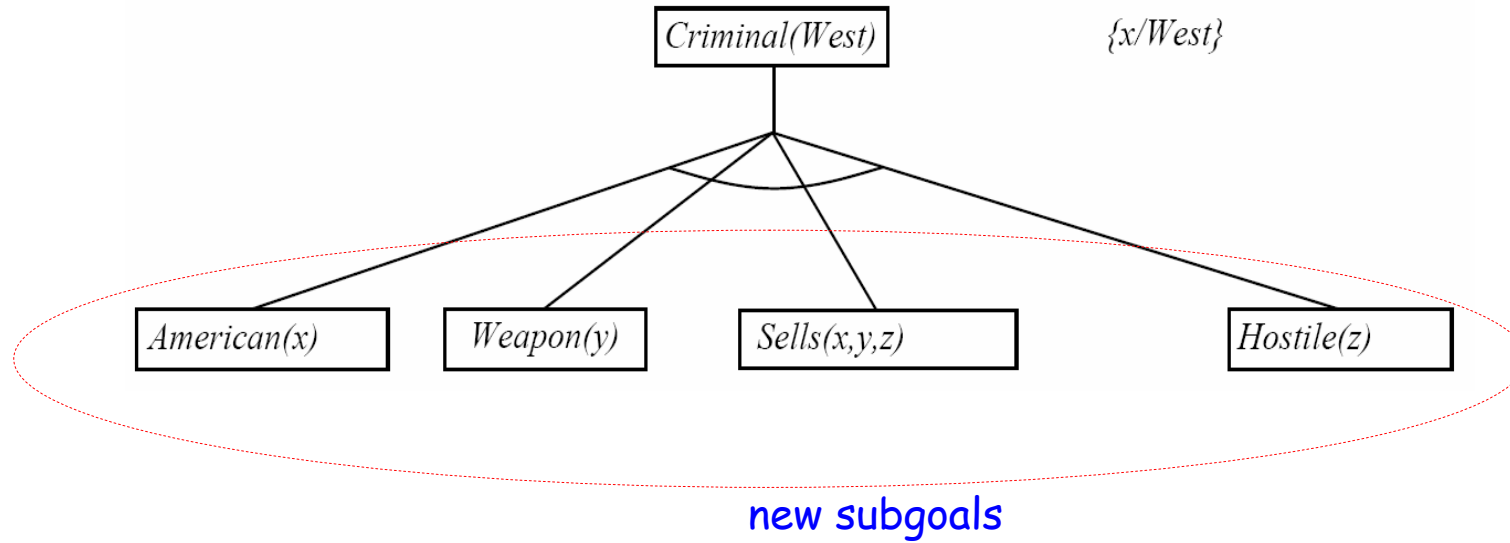
Example *KB*: BC Proof

Criminal(West)

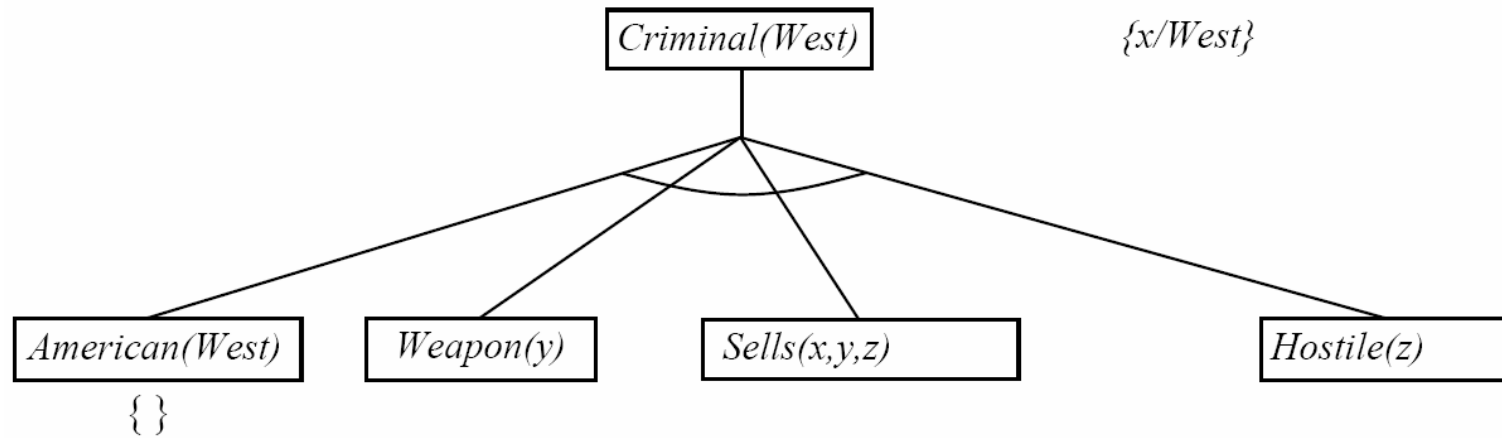
Put the query on a stack

Pop it and find the set of all substitutions that satisfies the query

Example *KB*: BC Proof

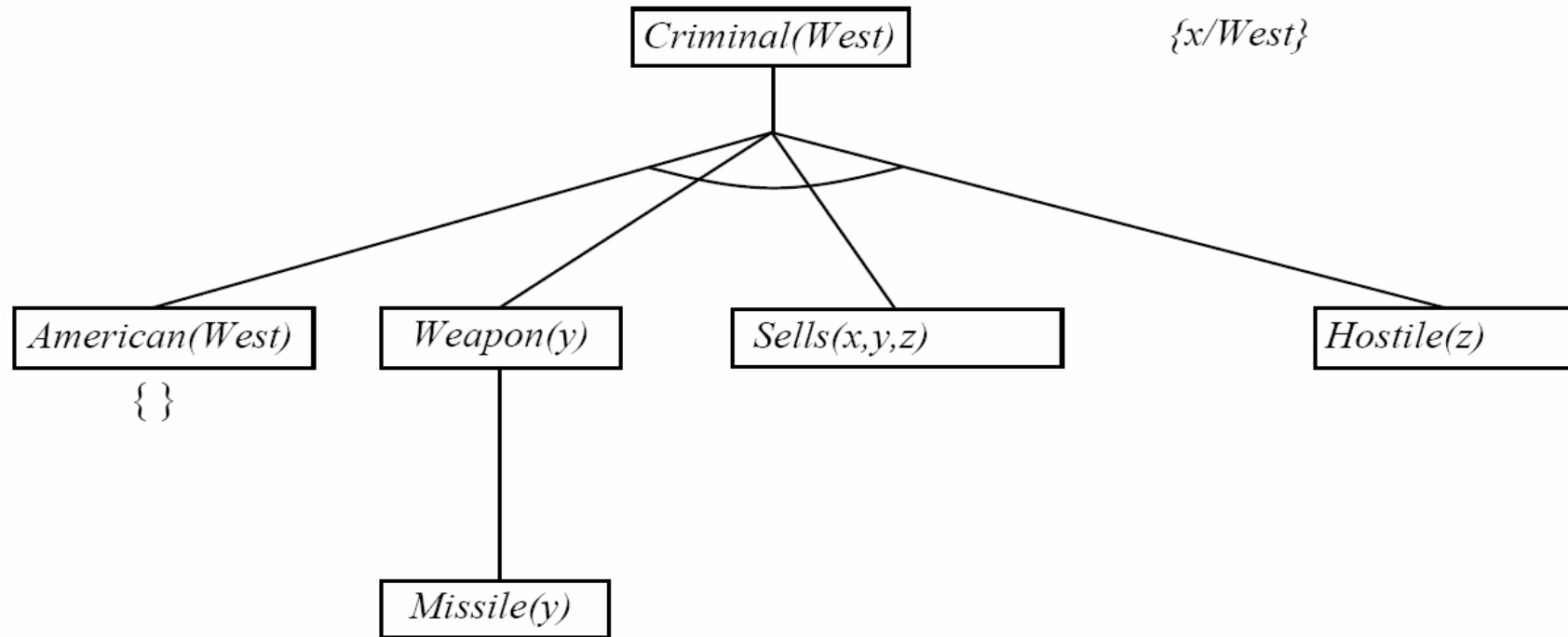


Example *KB*: BC Proof

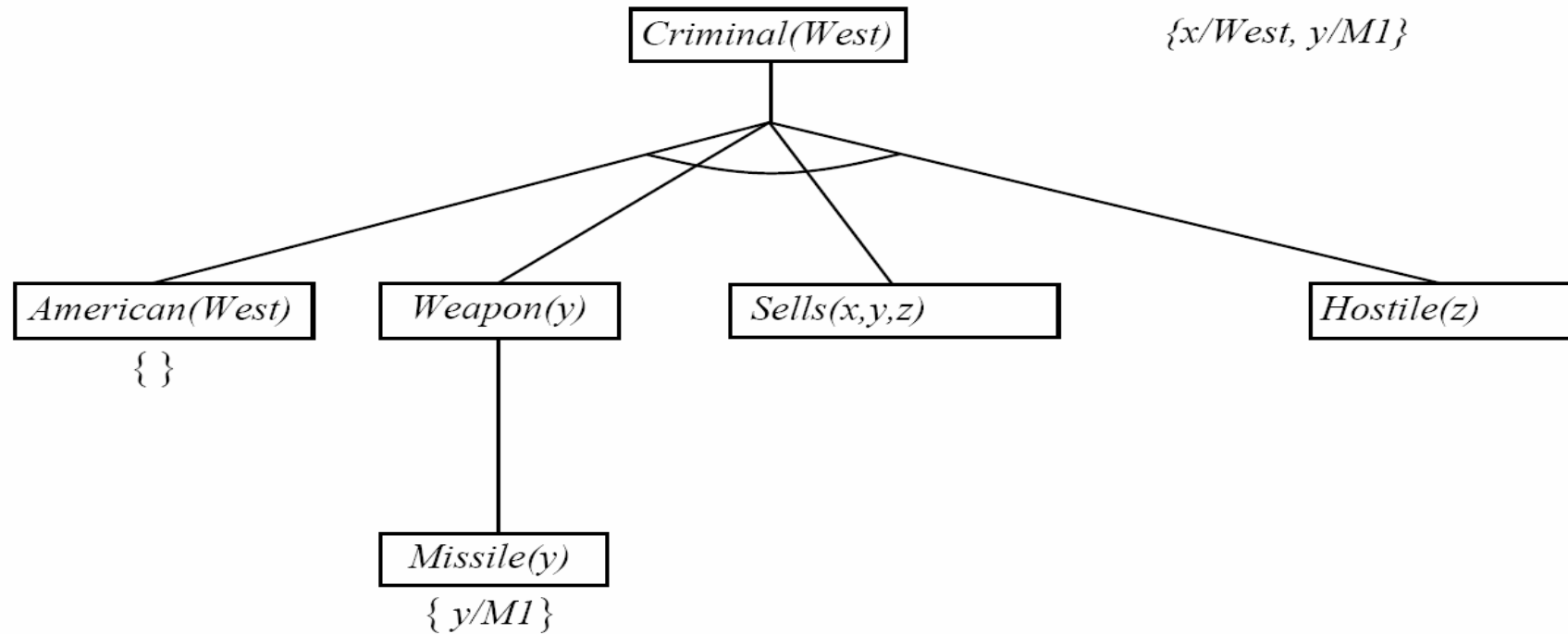


Unified with
known fact

Example *KB*: BC Proof



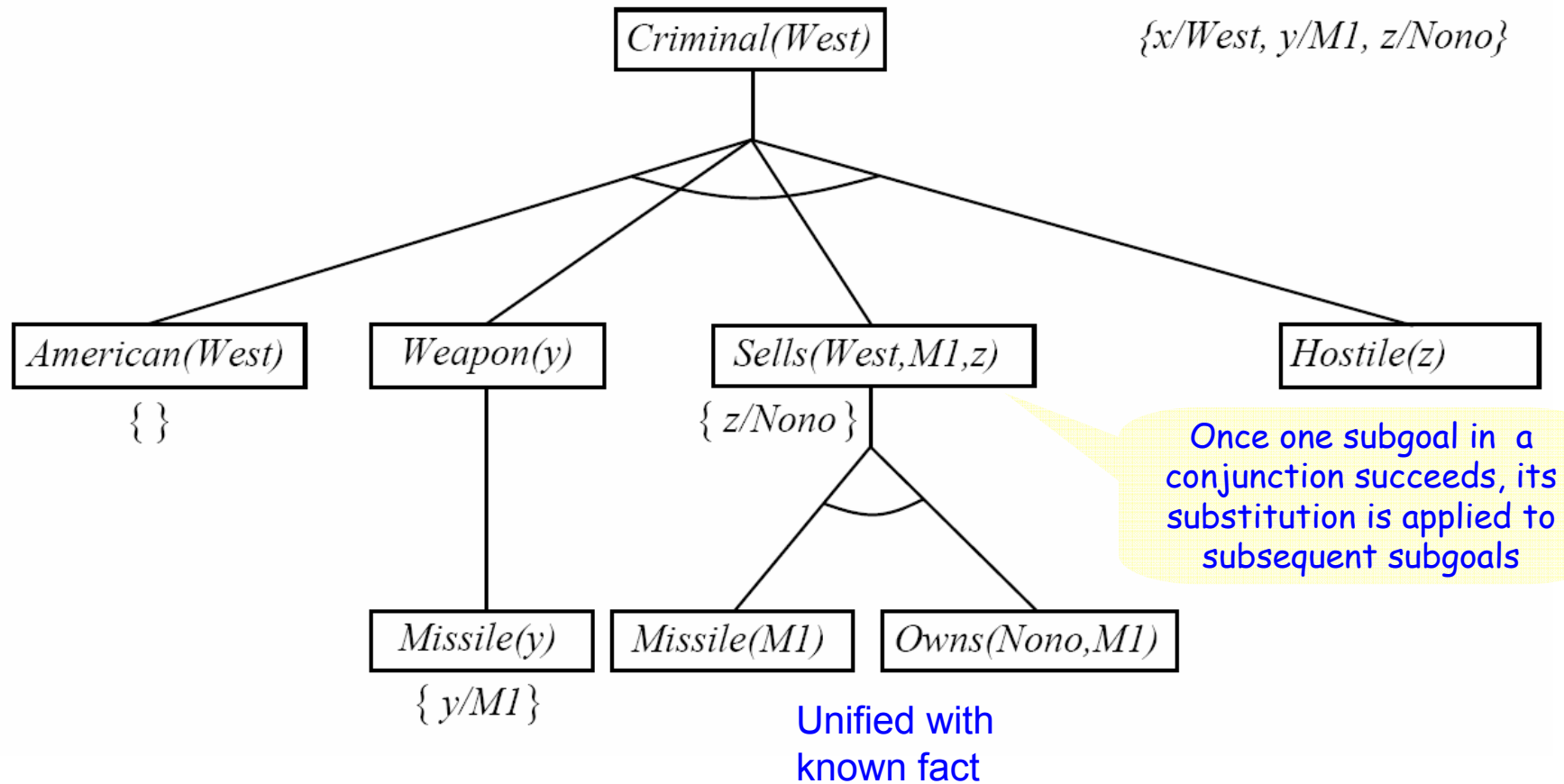
Example *KB*: BC Proof



Unified with
known fact

Proof Tree

Example *KB*: BC Proof



$$\text{SUBST}(\text{COMPOSE}(\theta_1, \theta_2), p) = \text{SUBST}(\theta_2, \text{SUBST}(\theta_1, p))$$

Backward Chaining Algorithm

```
function FOL-BC-ASK(KB, goals,  $\theta$ ) returns a set of substitutions
  inputs: KB, a knowledge base
           goals, a list of conjuncts forming a query ( $\theta$  already applied)
            $\theta$ , the current substitution, initially the empty substitution { }
  local variables: answers, a set of substitutions, initially empty

  if goals is empty then return { $\theta$ }
   $q' \leftarrow$  SUBST( $\theta$ , FIRST(goals))
  for each sentence r in KB where STANDARDIZE-APART(r) = ( $p_1 \wedge \dots \wedge p_n \Rightarrow q$ )
    and  $\theta' \leftarrow$  UNIFY(q,  $q'$ ) succeeds
    new_goals  $\leftarrow$  [ $p_1, \dots, p_n$  | REST(goals)]   stack: LIFO
    answers  $\leftarrow$  FOL-BC-ASK(KB, new_goals, COMPOSE( $\theta'$ ,  $\theta$ ))  $\cup$  answers
  return answers
```

Properties of Backward Chaining

- Depth-first recursive proof search
 - Space is linear in size of proof
- Incomplete due to infinite loops
 - Can be fixed by checking current goal against every goal on stack
- Inefficient due to repeated subgoals
 - Can be fixed by using caching of previous results (extra space !)

Conjunctive Normal Form (CNF) for FOL

- A *CNF* sentence in *FOL*
 - A conjunction (via \wedge 's operations) of clauses
 - Each clause is a disjunction (via \vee 's operations) of literals, where literals contain variables which are assumed to be universally quantified

$\forall x \text{ American}(x) \wedge \text{Weapon}(y) \wedge \text{Sells}(x, y, z) \wedge \text{Hostile}(z) \Rightarrow \text{Criminal}(x)$

CNF



$\neg \text{American}(x) \vee \neg \text{Weapon}(y) \vee \neg \text{Sells}(x, y, z) \vee \neg \text{Hostile}(z) \vee \text{Criminal}(x)$

- Every sentence of *FOL* can be converted into an inferentially equivalent *CNF*
 - The *CNF* sentence will be unsatisfiable if the original one is unsatisfiable

Conversion to CNF

- Example :

Everyone who loves all animals is loved by someone

$$\forall x [\forall y \text{ Animal}(y) \Rightarrow \text{loves}(x, y)] \Rightarrow \exists y \text{ loves}(y, x)$$

- Eliminate implications

$$\forall x [\neg \forall y \neg \text{Animal}(y) \vee \text{loves}(x, y)] \vee \exists y \text{ loves}(y, x)$$

- Move negation (\neg) inwards

$$\forall x [\exists y \neg(\neg \text{Animal}(y) \vee \text{loves}(x, y))] \vee \exists y \text{ loves}(y, x)$$

$$\forall x [\exists y \text{ Animal}(y) \wedge \neg \text{loves}(x, y)] \vee \exists y \text{ loves}(y, x)$$

- Standardize apart (renaming)

$$\forall x [\exists y \text{ Animal}(y) \wedge \neg \text{loves}(x, y)] \vee \exists z \text{ loves}(z, x)$$

Conversion to CNF

- Skolemize (remove existential quantifier)

$$\forall x [Animal(A) \wedge \neg loves(x, A)] \vee loves(B, x) \quad ?$$

$$\forall x [Animal(F(x)) \wedge \neg loves(x, F(x))] \vee loves(G(x), x)$$

- Existential variables replaced by **skolem functions**
- The skolemized sentence is satisfiable when the original one is satisfiable

- Drop universal quantifiers

$$[Animal(F(x)) \wedge \neg loves(x, F(x))] \vee loves(G(x), x)$$

- Distribute conjunction (\wedge) over disjunction (\vee)

$$[Animal(F(x)) \vee loves(G(x), x)] \wedge [\neg loves(x, F(x)) \vee loves(G(x), x)]$$

Resolution

- The binary resolution rule for FOL can be express as

$$\frac{l_1 \vee \dots \vee l_k, m_1 \vee \dots \vee m_n}{\text{SUBST}(\theta, l_1 \vee \dots \vee l_{i-1} \vee l_{i+1} \dots \vee l_k \vee m_1 \vee \dots \vee m_{j-1} \vee m_{j+1} \dots \vee m_n)}$$

where $\text{UNIFY}(l_i, \neg m_j) = \theta$

Or

$$\frac{\begin{array}{l} l_1 \wedge \dots \wedge l_k \Rightarrow p_1 \wedge \dots \wedge p_s \\ q_1 \wedge \dots \wedge q_r \Rightarrow m_1 \wedge \dots \wedge m_n \end{array}}{\text{SUBST}(\theta, l_1 \wedge \dots \wedge l_{i-1} \wedge l_{i+1} \dots \wedge l_k \wedge q_1 \wedge \dots \wedge q_r \Rightarrow p_1 \wedge \dots \wedge p_s \wedge m_1 \wedge \dots \wedge m_{j-1} \wedge m_{j+1} \dots \wedge m_n)}$$

where $\text{UNIFY}(l_i, m_j) = \theta$

Resolution

- The combination of binary resolution rule and factoring is complete
 - Factoring: remove multiple copies of literals if they are unifiable (the unifier must be applied to the entire clause)

$[Animal(F(x)) \vee loves(G(x), x)]$ and $[\neg loves(u, v) \vee \neg Kills(u, v)]$

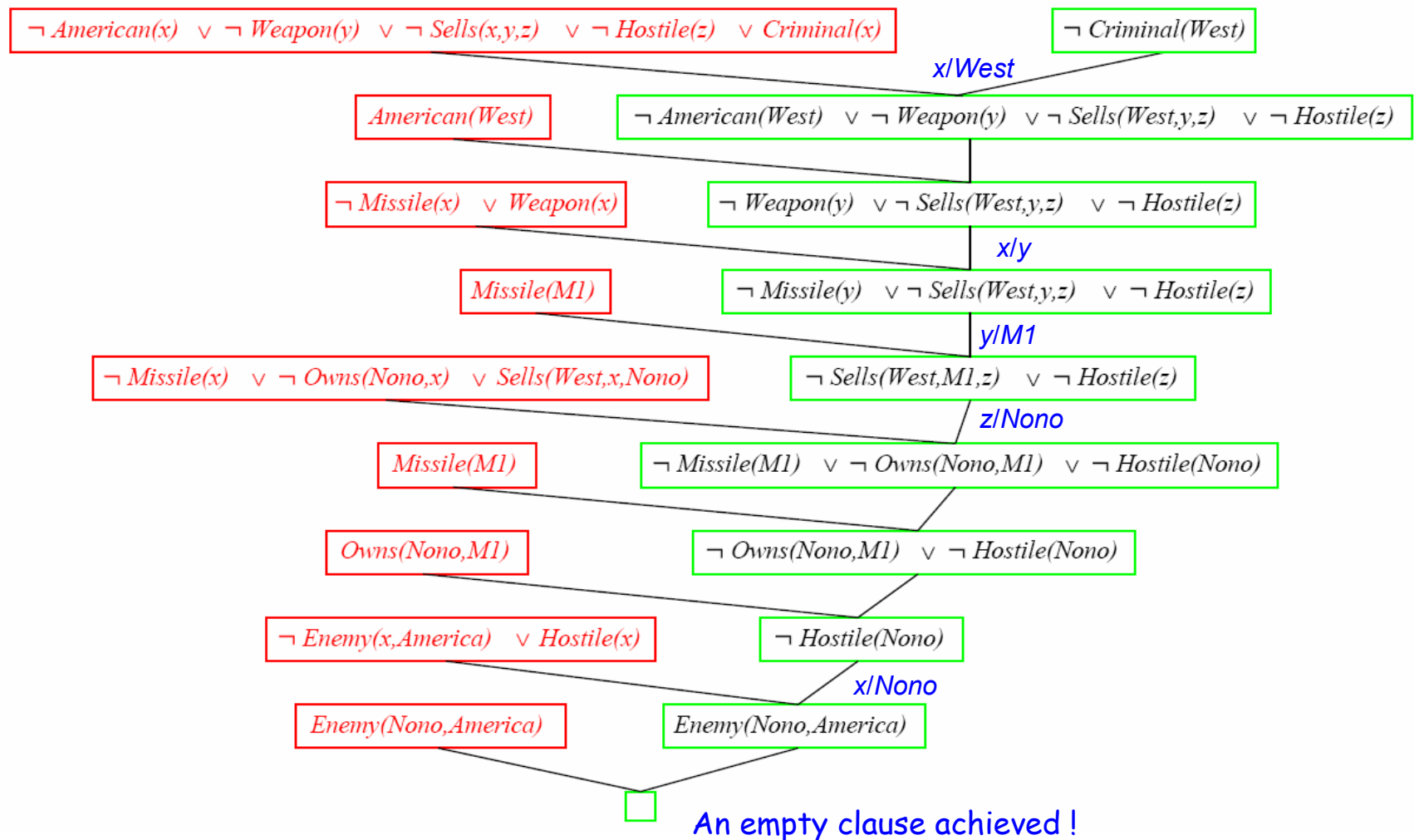
$\theta = \{ u/G(x), v/x \}$



$[Animal(F(x)) \vee \neg Kills(G(x), x)]$

Resolution: Example Proof 1

- Proved by refutation



Resolution: Example Proof 2

Problem:

Everyone who loves all animals is loved by someone.

Anyone who kills an animal is loved by no one.

Jack loves all animals.

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

Did Curiosity kill the cat?

- A. $\forall x [\forall y \text{ Animal}(y) \Rightarrow \text{loves}(x, y)] \Rightarrow \exists y \text{ loves}(y, x)$
- B. $\forall x [\forall y \text{ Animal}(y) \Rightarrow \text{Kills}(x, y)] \Rightarrow \forall z \neg \text{loves}(z, x)$
- C. $\forall x \text{ Animal}(x) \Rightarrow \text{loves}(\text{Jack}, x)$
- D. $\text{Kills}(\text{Jack}, \text{Tuna}) \vee \text{Kills}(\text{Curiosity}, \text{Tuna})$
- E. $\text{Cat}(\text{Tuna})$
- F. $\forall x \text{ Cat}(x) \Rightarrow \text{Animal}(x)$ (background knowledge!)
- \neg G. $\neg \text{Kills}(\text{Curiosity}, \text{Tuna})$

Resolution: Example Proof 2

- A1. $Animal(F(x)) \vee loves(G(x), x)$
- A2. $\neg loves(x, F(x)) \vee loves(G(x), x)$
- B. $\neg Animal(y) \vee \neg Kills(x, y) \vee \neg loves(z, x)$
- C. $\neg Animal(x) \vee loves(Jack, x)$
- D. $Kills(Jack, Tuna) \vee Kills(Curiosity, Tuna)$
- E. $Cat(Tuna)$
- F. $\neg Cat(x) \vee Animal(x)$ (background knowledge!)

- $\neg G.$ $\neg Kills(Curiosity, Tuna)$

Resolution: Example Proof 2

