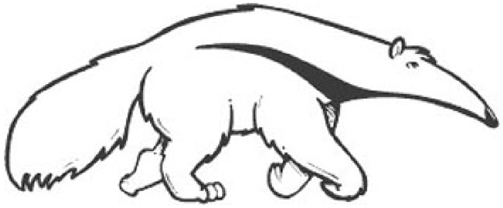
First Order Logic B: Semantics, Inference, Proof

Introduction to Artificial Intelligence
Prof. Richard Lathrop



Read Beforehand: R&N 8, 9.1-9.2, 9.5.1-9.5.5

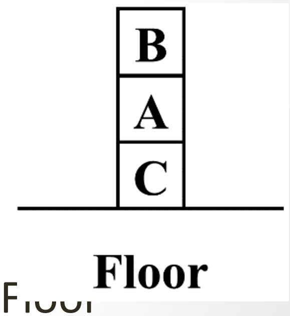
BREN:ICS
INFORMATION AND COMPUTER SCIENCES
UNIVERSITY of CALIFORNIA IRVINE

Semantics: Worlds

- The **world** consists of **objects** that have **properties**.
  - There are **relations** and **functions** between these objects
  - Objects in the world, individuals: people, houses, numbers, colors, baseball games, wars, centuries
    - Clock A, John, 7, the-house in the corner, Tel-Aviv
  - Functions on individuals:
    - father-of, best friend, third inning of, one more than
    - a function returns an object
  - Relations (terminology: same thing as a predicate):
    - brother-of, bigger than, inside, part-of, has color, occurred after
    - a relation/predicate returns a truth value
  - Properties (a relation of arity 1):
    - red, round, bogus, prime, multistoried, beautiful

Semantics: Interpretation

- An interpretation of a sentence is an assignment that maps
  - Object constants to objects in the worlds,
  - n-ary function symbols to n-ary functions in the world,
  - n-ary relation symbols to n-ary relations in the world
- Given an interpretation, an atomic sentence has the value "true" if it denotes a relation that holds for those individuals denoted in the terms. Otherwise it has the value "false."
  - Example: Block world:
    - A, B, C, Floor, On, Clear
  - World:
  - On(A,B) is false, Clear(B) is true, On(C,Floor) is true…
    - Under an interpretation that maps symbol A to block A, symbol B to block B, symbol C to block C, symbol Floor to the Floor
    - Some other interpretation might result in different truth values.



Truth in first-order logic

- Sentences are true with respect to a model and an interpretation
- Model contains objects (domain elements) and relations among them
- Interpretation specifies referents for
  constant symbols → objects

  predicate symbols → relations (a relation yields a truth value)

  function symbols → functions (a function yields an object)
- An atomic sentence *predicate*$(term_1,\ldots,term_n)$ is true
  iff the objects referred to by $term_1,\ldots,term_n$
  are in the relation referred to by *predicate*

Review: Models (and in FOL, Interpretations)

- Models are formal worlds within which truth can be evaluated
- Interpretations map symbols in the logic to the world
  - Constant symbols in the logic map to objects in the world
  - n-ary functions/predicates map to n-ary functions/predicates in the world

- We say *m* is a model given an interpretation i of a sentence α if and only if α is true in the world *m* under the mapping *i.*

- *M(α)* is the set of all models of α

- Then KB $\models$ α iff $M(KB) \subseteq M(\alpha)$
  - E.g. *KB,* = "Mary is Sue's sister and Amy is Sue's daughter."
  - α = "Mary is Amy's aunt." (Must Tell it about mothers/daughters)

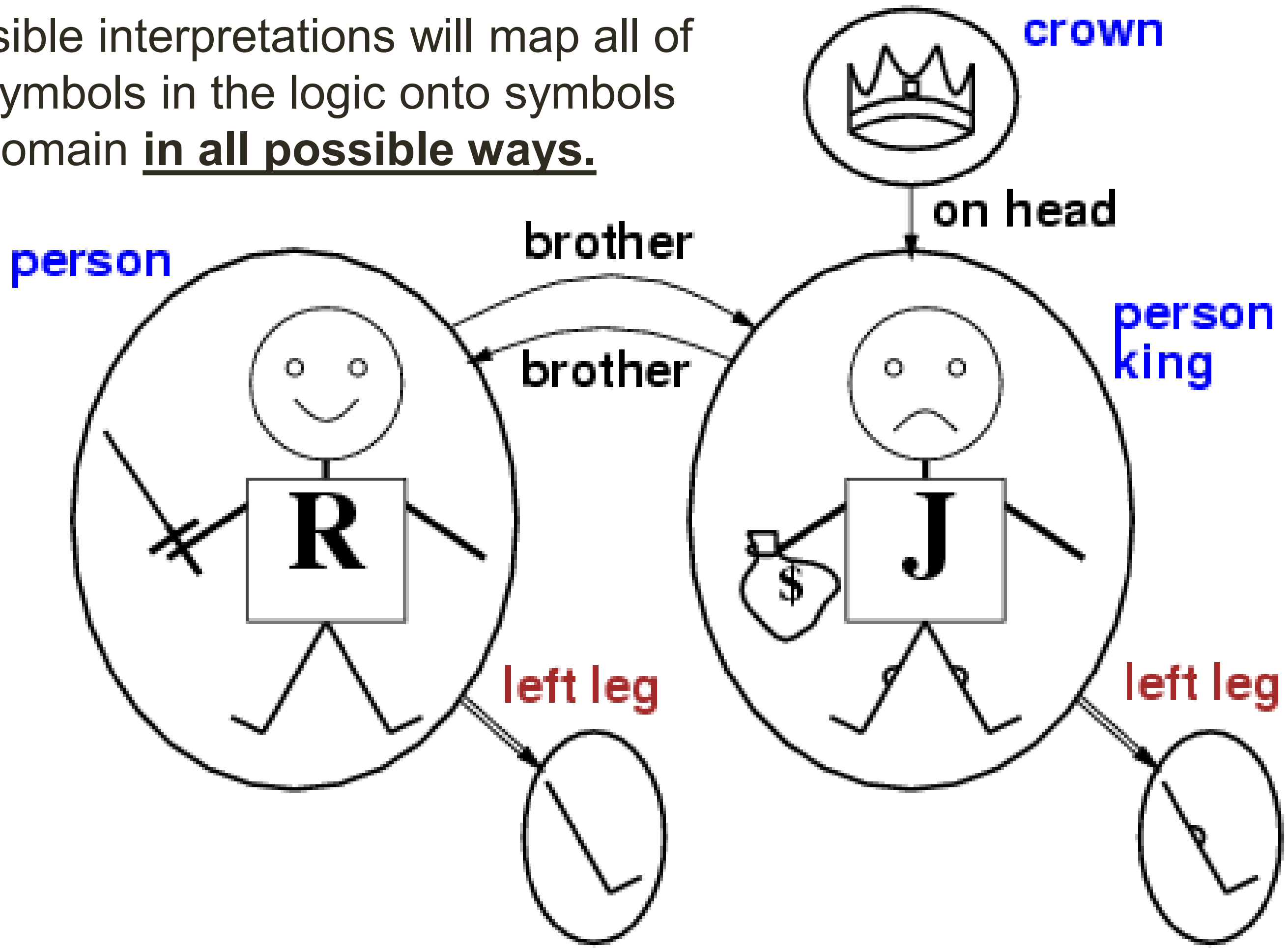
- Think of KB and α as constraints, and models as states.
- M(KB) are the solutions to KB and M(α) the solutions to α.
- Then, KB $\models$ α, i.e., $\models (KB \Rightarrow a)$,
  when all solutions to KB are also solutions to α.

Semantics: Models and Definitions

- An interpretation and possible world **satisfies** a wff (sentence) if the wff has the value "true" under that interpretation in that possible world.
- Model: A domain and an interpretation that satisfies a wff is a **model** of that wff
- Validity: Any wff that has the value "true" in all possible worlds and under all interpretations is **valid.**
- Any wff that does not have a model under any interpretation is inconsistent or **unsatisfiable.**
- Any wff that is true in at least one possible world under at least one interpretation is **satisfiable**.
- If a wff w has a value true under all the models of a set of sentences KB then KB logically **entails** w.

Models for FOL: Example

All possible interpretations will map all of these symbols in the logic onto symbols in the domain **in all possible ways.**



An interpretation maps all symbols in KB onto matching symbols in a possible world. All possible interpretations gives a combinatorial explosion of mappings. Your job, as a Knowledge Engineer, is to write the axioms in KB so **they are satisfied only under the intended interpretation in your own real world.**

Summary of FOL Semantics

- A well-formed formula (“wff”) FOL is true or false with respect to a world and an interpretation (a model).
- The world has objects, relations, functions, and predicates.
- The interpretation maps symbols in the logic to the world.
- The wff is true if and only if (iff) its assertion holds among the objects in the world under the mapping by the interpretation.
- <u>Your job, as a Knowledge Engineer, is to write sufficient KB axioms that ensure that KB is true in your own real world under your own intended interpretation.</u>
  - <u>**The KB axioms must rule out other worlds and interpretations**.</u>

Conversion to CNF

- Everyone who loves all animals is loved by someone:

  $\forall x\ [\forall y\ \mathit{Animal}(y) \Rightarrow \mathit{Loves}(x,y)] \Rightarrow [\exists y\ \mathit{Loves}(y,x)]$

1. Eliminate biconditionals and implications:

   $\forall x\ \neg[\forall y\ \mathit{Animal}(y) \Rightarrow \mathit{Loves}(x,y)] \lor [\exists y\ \mathit{Loves}(y,x)]$
   $\forall x\ \neg[\forall y\ \neg \mathit{Animal}(y) \lor \mathit{Loves}(x,y)] \lor [\exists y\ \mathit{Loves}(y,x)]$

2. Move $\neg$ inwards:
   [Recall: $\neg\forall x\ P(x) \equiv \exists x\ \neg P(x)$; $\neg\ \exists x\ P(x) \equiv \forall x\ \neg P(x)$]

   $\forall x\ [\neg\forall y\ \neg \mathit{Animal}(y) \lor \mathit{Loves}(x,y)] \lor [\exists y\ \mathit{Loves}(y,x)]$
   $\forall x\ [\exists y\ \neg(\neg \mathit{Animal}(y) \lor \mathit{Loves}(x,y))] \lor [\exists y\ \mathit{Loves}(y,x)]$
   $\forall x\ [\exists y\ \neg\neg \mathit{Animal}(y) \land \neg \mathit{Loves}(x,y)] \lor [\exists y\ \mathit{Loves}(y,x)]$
   $\forall x\ [\exists y\ \mathit{Animal}(y) \land \neg \mathit{Loves}(x,y)] \lor [\exists y\ \mathit{Loves}(y,x)]$

Conversion to CNF contd.

3. Standardize variables: each quantifier should use a different variable

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

4. Skolemize: a more general form of existential instantiation. Each existential variable is replaced by a Skolem function of the enclosing universally quantified variables:

   $\forall x\ [\mathit{Animal}(F(x)) \wedge \neg \mathit{Loves}(x,F(x))] \vee \mathit{Loves}(G(x),x)$

5. Drop universal quantifiers:

   $[\mathit{Animal}(F(x)) \wedge \neg \mathit{Loves}(x,F(x))] \vee \mathit{Loves}(G(x),x)$

6. Distribute $\vee$ over $\wedge$:

   $[\mathit{Animal}(F(x)) \vee \mathit{Loves}(G(x),x)] \wedge [\neg \mathit{Loves}(x,F(x)) \vee \mathit{Loves}(G(x),x)]$

A note on Skolem functions

Consider the statement: $\forall x \, \exists y \, P(x, y)$

The statement asserts that, for all x, there is (at least) one y such that P(x,y). Recall that each x may have a different y, and so y depends on x.

So, at least abstractly, there is a list that pairs each x to a y that satisfies P(x,y):

{ (x1, y1), (x2, y2), (x3, y3), (x4, y4) … }

where P(x1, y1) = TRUE; P(x2, y2) = TRUE; P(x3, y3) = TRUE; and so on.

So, at least abstractly, there is a function that maps xi to yi. Call that function F(), where F(x1) = y1; F(x2) = y2; F(x3) = y3; and so on. (We don't know what that function is, but we do know that it must exist --- even if we can't write it down.)

So P(x1, F(x1)) = TRUE; P(x2, F(x2)) = TRUE; P(x3, F(x3)) = TRUE; and so on.

In other words, $\forall x \, \exists y \, P(x, y) \equiv \forall x \, P(x, F(x))$, where F() is as described above.

Simple FOL Resolution Example

- ∀ x Person(x) => HasHead(x)   "Every person has a head."
- Person(John)   "John is a person."
- **Query Sentence:** HasHead(John)   "John has a head."
- Resulting KB plus negated goal in CNF:
  - (¬Person(x) ∨ HasHead(x))
  - Person(John)
  - ¬ HasHead(John)
- Resolve (¬Person(x) ∨ HasHead(x)) with Person(John) and substitution {x/John} to yield HasHead(John)
  - Note that after the substitution, the first clause becomes

    (¬Person(x) ∨ HasHead(x))
- Resolve HasHead(John) with ¬ HasHead(John) to yield ( )

Unification

- Recall: Subst(θ, p) = result of substituting θ into sentence p

- Unify algorithm: takes 2 sentences p and q and returns a unifier if one exists

  Unify(p,q) = θ   where Subst(θ, p) = Subst(θ, q)

  where θ is a list of variable/substitution pairs
  that will make p and q syntactically identical

- Example:
  p = Knows(John,x)
  q = Knows(John, Jane)

  Unify(p,q) = {x/Jane}

Unification examples

- simple example: query = Knows(John,x), i.e., who does John know?

p	q	θ
Knows(John,x)	Knows(John,Jane)	{x/Jane}
Knows(John,x)	Knows(y,OJ)	{x/OJ,y/John}
Knows(John,x)	Knows(y,Mother(y))	{y/John,x/Mother(John)}
Knows(John,x)	Knows(x,OJ)	{fail}

- Last unification fails: only because x can't take values John and OJ at the same time
  - But we know that if John knows x, and everyone (x) knows OJ, we should be able to infer that John knows OJ
- Problem is due to use of same variable x in both sentences
- Simple solution: Standardizing apart eliminates overlap of variables, e.g., Knows(z,OJ)

Unification examples

1) UNIFY(Knows(John, x), Knows(John, Jane))	{ x / Jane }
2) UNIFY(Knows(John, x), Knows(y, Jane))	{ x / Jane, y / John }
3) UNIFY(Knows(y, x), Knows(John, Jane))	{ x / Jane, y / John }
4) UNIFY(Knows(John, x), Knows(y, Father (y)))	{ y / John, x / Father (John) }
5) UNIFY(Knows(John, F(x)), Knows(y, F(F(z))))	{ y / John, x / F (z) }
6) UNIFY(Knows(John, F(x)), Knows(y, G(z)))	None
7) UNIFY(Knows(John, F(x)), Knows(y, F(G(y))))	{ y / John, x / G (John) }

Unification

- To unify *Knows(John,x)* and *Knows(y,z)*,

  θ = {y/John, x/z } or θ = {y/John, x/John, z/John}

- The first unifier is more general than the second.

- There is a single most general unifier (MGU) that is unique up to renaming of variables.

  MGU = { y/John, x/z }

- General algorithm in Figure 9.1 in the text

Unification Algorithm

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

  if θ = failure **then return** failure
  else if $x = y$ **then return** θ
  else if VARIABLE?(x) **then return** UNIFY-VAR(x, y, θ)
  else if VARIABLE?(y) **then return** UNIFY-VAR(y, x, θ)
  else if COMPOUND?(x) **and** COMPOUND?(y) **then**
    return UNIFY(x.ARGS, y.ARGS, UNIFY(x.OP, y.OP, θ))
  else if LIST?(x) **and** LIST?(y) **then**
    return UNIFY(x.REST, y.REST, UNIFY(x.FIRST, y.FIRST, θ))
  else return failure

---

function UNIFY-VAR(var, x, θ) **returns** a substitution

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

Figure 9.1 The unification algorithm. The algorithm works by comparing the structures of the inputs, element by element. The substitution θ that is the argument to UNIFY is built up along the way and is used to make sure that later comparisons are consistent with bindings that were established earlier. In a compound expression such as $F(A, B)$, the OP field picks out the function symbol F and the ARGS field picks out the argument list (A, B).

Unification Algorithm

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

  if θ = failure **then return** failure
  else if $x = y$ **then return** θ
  else if VARIABLE?(x) **then return** UNIFY-VAR(x, y, θ)
  else if VARIABLE?(y) **then return** UNIFY-VAR(y, x, θ)
  else if COMPOUND?(x) **and** COMPOUND?(y) **then**
    return UNIFY(x.ARGS, y.ARGS, UNIFY(x.OP, y.OP, θ))
  else if LIST?(x) **and** LIST?(y) **then**
    return UNIFY(x.REST, y.REST, UNIFY(x.FIRST, y.FIRST, θ))
  else return failure

If we have failed or succeeded, then fail or succeed.

---

function UNIFY-VAR(var, x, θ) **returns** a substitution

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

Figure 9.1 The unification algorithm. The algorithm works by comparing the structures of the inputs, element by element. The substitution θ that is the argument to UNIFY is built up along the way and is used to make sure that later comparisons are consistent with bindings that were established earlier. In a compound expression such as $F(A, B)$, the OP field picks out the function symbol F and the ARGS field picks out the argument list (A, B).

Unification Algorithm

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

  if θ = failure **then return** failure
  else if $x = y$ **then return** θ
  else if VARIABLE?(x) **then return** UNIFY-VAR(x, y, θ)
  else if VARIABLE?(y) **then return** UNIFY-VAR(y, x, θ)
  else if COMPOUND?(x) **and** COMPOUND?(y) **then**
    return UNIFY(x.ARGS, y.ARGS, UNIFY(x.OP, y.OP, θ))
  else if LIST?(x) **and** LIST?(y) **then**
    return UNIFY(x.REST, y.REST, UNIFY(x.FIRST, y.FIRST, θ))
  else return failure

If we can unify a variable then do so.

---

function UNIFY-VAR(var, x, θ) **returns** a substitution

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

Figure 9.1 The unification algorithm. The algorithm works by comparing the structures of the inputs, element by element. The substitution θ that is the argument to UNIFY is built up along the way and is used to make sure that later comparisons are consistent with bindings that were established earlier. In a compound expression such as $F(A, B)$, the OP field picks out the function symbol F and the ARGS field picks out the argument list (A, B).

Unification Algorithm

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

  if θ = failure **then return** failure
  else if $x = y$ **then return** θ
  else if VARIABLE?(x) **then return** UNIFY-VAR(x, y, θ)
  else if VARIABLE?(y) **then return** UNIFY-VAR(y, x, θ)
  else if COMPOUND?(x) **and** COMPOUND?(y) **then**
    return UNIFY(x.ARGS, y.ARGS, UNIFY(x.OP, y.OP, θ))
  else if LIST?(x) **and** LIST?(y) **then**
    return UNIFY(x.REST, y.REST, UNIFY(x.FIRST, y.FIRST, θ))
  else return failure

---

function UNIFY-VAR(var, x, θ) **returns** a substitution

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

If we already have bound variable *var* to a value, try to continue on that basis.

Figur
of the
up alo

There is an implicit assumption that "{var/val} ∈ θ", if it succeeds, binds val to the value that allowed it to succeed,

that were established earlier. In a compound expression such as $F(A, B)$, the OP field picks out the function symbol F and the ARGS field picks out the argument list (A, B).

Unification Algorithm

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

  if θ = failure **then return** failure
  else if $x = y$ **then return** θ
  else if VARIABLE?(x) **then return** UNIFY-VAR(x, y, θ)
  else if VARIABLE?(y) **then return** UNIFY-VAR(y, x, θ)
  else if COMPOUND?(x) **and** COMPOUND?(y) **then**
    return UNIFY(x.ARGS, y.ARGS, UNIFY(x.OP, y.OP, θ))
  else if LIST?(x) **and** LIST?(y) **then**
    return UNIFY(x.REST, y.REST, UNIFY(x.FIRST, y.FIRST, θ))
  else return failure

---

function UNIFY-VAR(var, x, θ) **returns** a substitution

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

If we already have bound *x* to a value, try to continue on that basis.

Figure 9.1 The unification algorithm. The algorithm works by comparing the structures of the inputs, element by element. The substitution θ that is the argument to UNIFY is built up along the way and is used to make sure that later comparisons are consistent with bindings that were established earlier. In a compound expression such as $F(A, B)$, the OP field picks out the function symbol F and the ARGS field picks out the argument list (A, B).

Unification Algorithm

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

  if θ = failure **then return** failure
  else if $x = y$ **then return** θ
  else if VARIABLE?(x) **then return** UNIFY-VAR(x, y, θ)
  else if VARIABLE?(y) **then return** UNIFY-VAR(y, x, θ)
  else if COMPOUND?(x) **and** COMPOUND?(y) **then**
    return UNIFY(x.ARGS, y.ARGS, UNIFY(x.OP, y.OP, θ))
  else if LIST?(x) **and** LIST?(y) **then**
    return UNIFY(x.REST, y.REST, UNIFY(x.FIRST, y.FIRST, θ))
  else return failure

---

function UNIFY-VAR(var, x, θ) **returns** a substitution

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

If *var* occurs anywhere within *x*, then no substitution will succeed.

Figure 9.1 The unification algorithm. The algorithm works by comparing the structures of the inputs, element by element. The substitution θ that is the argument to UNIFY is built up along the way and is used to make sure that later comparisons are consistent with bindings that were established earlier. In a compound expression such as $F(A, B)$, the OP field picks out the function symbol F and the ARGS field picks out the argument list (A, B).

Unification Algorithm

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

  if θ = failure **then return** failure
  else if $x = y$ **then return** θ
  else if VARIABLE?(x) **then return** UNIFY-VAR(x, y, θ)
  else if VARIABLE?(y) **then return** UNIFY-VAR(y, x, θ)
  else if COMPOUND?(x) **and** COMPOUND?(y) **then**
    return UNIFY(x.ARGS, y.ARGS, UNIFY(x.OP, y.OP, θ))
  else if LIST?(x) **and** LIST?(y) **then**
    return UNIFY(x.REST, y.REST, UNIFY(x.FIRST, y.FIRST, θ))
  else return failure

---

function UNIFY-VAR(var, x, θ) **returns** a substitution

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

Else, try to bind *var* to *x*, and recurse.

Figure 9.1 The unification algorithm. The algorithm works by comparing the structures of the inputs, element by element. The substitution θ that is the argument to UNIFY is built up along the way and is used to make sure that later comparisons are consistent with bindings that were established earlier. In a compound expression such as $F(A, B)$, the OP field picks out the function symbol F and the ARGS field picks out the argument list (A, B).

Unification Algorithm

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

  if θ = failure **then return** failure
  else if $x = y$ **then return** θ
  else if VARIABLE?(x) **then return** UNIFY-VAR(x, y, θ)
  else if VARIABLE?(y) **then return** UNIFY-VAR(y, x, θ)
  else if COMPOUND?(x) **and** COMPOUND?(y) **then**
    return UNIFY(x.ARGS, y.ARGS, UNIFY(x.OP, y.OP, θ))
  else if LIST?(x) **and** LIST?(y) **then**
    return UNIFY(x.REST, y.REST, UNIFY(x.FIRST, y.FIRST, θ))
  else return failure

If a predicate/function, unify the arguments.

---

function UNIFY-VAR(var, x, θ) **returns** a substitution

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

Figure 9.1 The unification algorithm. The algorithm works by comparing the structures of the inputs, element by element. The substitution θ that is the argument to UNIFY is built up along the way and is used to make sure that later comparisons are consistent with bindings that were established earlier. In a compound expression such as $F(A, B)$, the OP field picks out the function symbol F and the ARGS field picks out the argument list (A, B).

Unification Algorithm

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

  if θ = failure **then return** failure
  else if $x = y$ **then return** θ
  else if VARIABLE?(x) **then return** UNIFY-VAR(x, y, θ)
  else if VARIABLE?(y) **then return** UNIFY-VAR(y, x, θ)
  else if COMPOUND?(x) **and** COMPOUND?(y) **then**
    return UNIFY(x.ARGS, y.ARGS, UNIFY(x.OP, y.OP, θ))
  else if LIST?(x) **and** LIST?(y) **then**
    return UNIFY(x.REST, y.REST, UNIFY(x.FIRST, y.FIRST, θ))
  else return failure

If unifying arguments, unify the remaining arguments.

---

function UNIFY-VAR(var, x, θ) **returns** a substitution

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

Figure 9.1 The unification algorithm. The algorithm works by comparing the structures of the inputs, element by element. The substitution θ that is the argument to UNIFY is built up along the way and is used to make sure that later comparisons are consistent with bindings that were established earlier. In a compound expression such as $F(A, B)$, the OP field picks out the function symbol F and the ARGS field picks out the argument list (A, B).

Unification Algorithm

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

  if θ = failure **then return** failure
  else if $x = y$ **then return** θ
  else if VARIABLE?(x) **then return** UNIFY-VAR(x, y, θ)
  else if VARIABLE?(y) **then return** UNIFY-VAR(y, x, θ)
  else if COMPOUND?(x) **and** COMPOUND?(y) **then**
    return UNIFY(x.ARGS, y.ARGS, UNIFY(x.OP, y.OP, θ))
  else if LIST?(x) **and** LIST?(y) **then**
    return UNIFY(x.REST, y.REST, UNIFY(x.FIRST, y.FIRST, θ))
  else return failure Otherwise, fail.

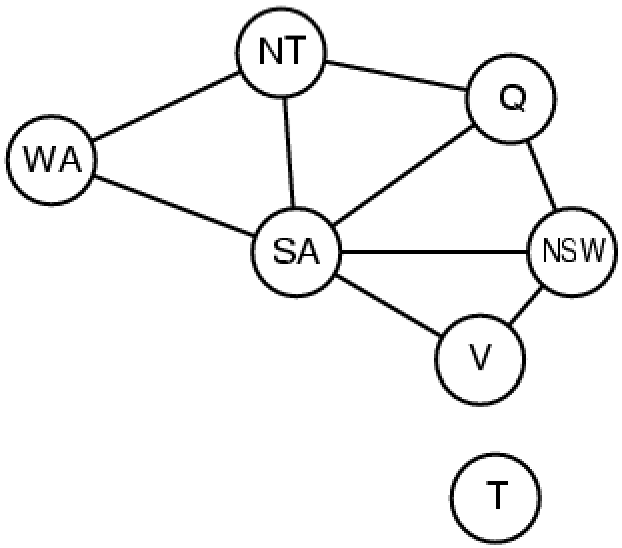
---

function UNIFY-VAR(var, x, θ) **returns** a substitution

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

Figure 9.1 The unification algorithm. The algorithm works by comparing the structures of the inputs, element by element. The substitution θ that is the argument to UNIFY is built up along the way and is used to make sure that later comparisons are consistent with bindings that were established earlier. In a compound expression such as $F(A, B)$, the OP field picks out the function symbol F and the ARGS field picks out the argument list (A, B).

Hard matching example



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()*

Diff(Red,Blue) *Diff (Red,Green)*
Diff(Green,Red) *Diff(Green,Blue)*
Diff(Blue,Red) *Diff(Blue,Green)*

- To unify the grounded propositions with premises of the implication you need to solve a CSP!
- *Colorable*() is inferred iff the CSP has a solution
- CSPs include 3SAT as a special case, hence matching is NP-hard

Resolution: brief summary

- Full first-order version:

$$\frac{\ell_1 \vee \cdots \vee \ell_k, \qquad m_1 \vee \cdots \vee m_n}{(\ell_1 \vee \cdots \vee \ell_{i-1} \vee \ell_{i+1} \vee \cdots \vee \ell_k \vee m_1 \vee \cdots \vee m_{j-1} \vee m_{j+1} \vee \cdots \vee m_n)\theta}$$

  where $\mathtt{Unify}(\ell_i, \neg m_j) = \theta$.

- The two clauses are assumed to be standardized apart so that they share no variables.
- For example,

$$\frac{\begin{array}{l}\neg Rich(x) \vee Unhappy(x) \\ Rich(Ken)\end{array}}{Unhappy(Ken)}$$

  with $\theta = \{x/Ken\}$

- Apply resolution steps to CNF(KB $\wedge \neg\alpha$); complete for FOL

Examples of Sound Inference Patterns

Classical Syllogism (due to Aristotle)	
All Ps are Qs	All Men are Mortal
X is a P	Socrates is a Man
Therefore, X is a Q	Therefore, Socrates is Mortal
Implication (Modus Ponens)	
P implies Q	Smoke implies Fire
P	Smoke
Therefore, Q	Therefore, Fire
Contrapositive (Modus Tollens)	
P implies Q	Smoke implies Fire
Not Q	Not Fire
Therefore, Not P	Therefore, not Smoke
Law of the Excluded Middle (due to Aristotle)	
A Or B	Alice is a Democrat or a Republican
Not A	Alice is not a Democrat
Therefore, B	Therefore, Alice is a Republican

Examples of Sound Inference Patterns

<u>Classical Syllogism (due to Aristotle)</u>

All Ps are Qs	All Men are Mortal
X is a P	Socrates is a Man
Therefore, X is a Q	Therefore, Socrates is Mortal

$\forall$ x Man(x) $\Rightarrow$ Mortal(x)
Man(Socrates)
Therefore, Mortal(Socrates)

($\neg$Man(x) Mortal(x))
(Man(Socrates))

Examples of Sound Inference Patterns

<u>Classical Syllogism (due to Aristotle)</u>

All Ps are Qs
X is a P
Therefore, X is a Q

All Men are Mortal
Socrates is a Man
Therefore, Socrates is Mortal

$\forall$ x Man(x) $\Rightarrow$ Mortal(x)
Man(Socrates)
Therefore, Mortal(Socrates)

($\boxed{\neg\text{Man}(x)}$ Mortal(x))
($\boxed{\text{Man(Socrates)}}$)

Mortal(Socrates), with substitution θ = { x/Socrates }

<u>The classical syllogism is proven sound by Resolution!</u>

Examples of Sound Inference Patterns

Implication (Modus Ponens)

P implies Q
P
Therefore, Q

Smoke implies Fire
Smoke
Therefore, Fire

Smoke $\Rightarrow$ Fire
Smoke
Therefore, Fire

($\neg$Smoke Fire)
(Smoke)

Examples of Sound Inference Patterns

Implication (Modus Ponens)

P implies Q	Smoke implies Fire
P	Smoke
Therefore, Q	Therefore, Fire

Smoke $\Rightarrow$ Fire
Smoke
Therefore, Fire

($\boxed{\neg\text{Smoke}}$ Fire)
($\boxed{\text{Smoke}}$)

(Fire)

Implication (Modus Ponens) is proven sound by Resolution!

Examples of Sound Inference Patterns

Contrapositive (Modus Tollens)

P implies Q	Smoke implies Fire
Not Q	Not Fire
Therefore, Not P	Therefore, not Smoke

Smoke $\Rightarrow$ Fire
$\neg$Fire
Therefore, $\neg$Smoke

($\neg$Smoke Fire)
($\neg$Fire)

Examples of Sound Inference Patterns

<u>Contrapositive (Modus Tollens)</u>

P implies Q	Smoke implies Fire
Not Q	Not Fire
Therefore, Not P	Therefore, not Smoke

Smoke $\Rightarrow$ Fire
$\neg$Fire
Therefore, $\neg$Smoke

($\neg$Smoke $\boxed{\text{Fire}}$)
($\boxed{\neg\text{Fire}}$)

($\neg$Smoke)

<u>The Contrapositive (Modus Tollens) is proven sound by Resolution!</u>

Examples of Sound Inference Patterns

Law of the Excluded Middle (due to Aristotle)

A Or B	Alice is a Democrat or a Republican
Not A	Alice is not a Democrat
Therefore, B	Therefore, Alice is a Republican

$(\text{Democrat(Alice)} \lor \text{Republican(Alice)})$
$\neg\text{Democrat(Alice)}$
Therefore, Republican(Alice)

$(\text{Democrat(Alice)} \lor \text{Republican(Alice)})$
$(\neg\text{Democrat(Alice)})$

Examples of Sound Inference Patterns

<u>Law of the Excluded Middle (due to Aristotle)</u>

A Or B	Alice is a Democrat or a Republican
Not A	Alice is not a Democrat
Therefore, B	Therefore, Alice is a Republican

(Democrat(Alice) $\lor$ Republican(Alice))
$\neg$Democrat(Alice)
Therefore, Republican(Alice)

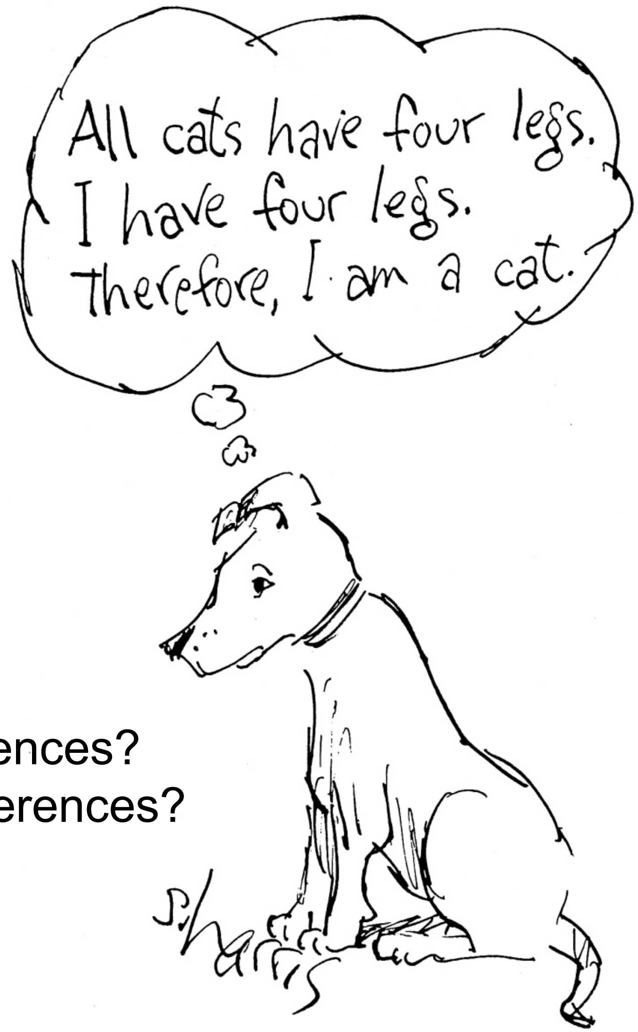
$$\frac{\begin{array}{l}(\boxed{\text{Democrat(Alice)}} \lor \text{Republican(Alice)}) \\ (\boxed{\neg\text{Democrat(Alice)}})\end{array}}{(\text{Republican(Alice)})}$$

<u>The Law of the Excluded Middle is proven sound by Resolution!</u>

So --- how do we keep it from "Just making things up." ?

Is this inference correct?

How do you know?
How can you tell?



How can we **make correct** inferences?
How can we **avoid incorrect** inferences?

"Einstein Simplified: Cartoons on Science" by Sydney Harris, 1992, Rutgers University Press

Example of an Unsound Inference Pattern

An Unsound Inference Pattern

All Ps are Qs	All Cats are FourFootedAnimals
X is a Q	I am a FourFootedAnimal
Therefore, X is a P	Therefore, I am a Cat

$\forall$ x Cat(x) $\Rightarrow$ FourFootedAnimal(x)
FourFootedAnimal(Me)
Therefore, Cat(Me)

($\neg$Cat(x) FourFootedAnimal(x))
(FourFootedAnimal(Me))

Example of an Unsound Inference Pattern

An Unsound Inference Pattern

All Ps are Qs	All Cats are FourFootedAnimals
X is a Q	I am a FourFootedAnimal
Therefore, X is a P	Therefore, I am a Cat

$\forall$ x Cat(x) $\Rightarrow$ FourFootedAnimal(x)
FourFootedAnimal(Me)
Therefore, Cat(Me)

($\neg$Cat(x) FourFootedAnimal(x))
(FourFootedAnimal(Me))

No Resolution is possible! No pair of complementary literals!

<u>Resolution shows that the premises do not entail the conclusion!</u>

Example knowledge base

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

- Prove that Col. West is a criminal

Example knowledge base (Horn clauses)

... it is a crime for an American to sell weapons to hostile nations:

American(x) $\wedge$ Weapon(y) $\wedge$ Sells(x,y,z) $\wedge$ Hostile(z) $\Rightarrow$ Criminal(x)

Nono … has some missiles, i.e., $\exists$x Owns(Nono,x) $\wedge$ Missile(x):

Owns(Nono,M_1) $\wedge$ Missile(M_1)

… all of its missiles were sold to it by Colonel West

Missile(x) $\wedge$ Owns(Nono,x) $\Rightarrow$ Sells(West,x,Nono)

Missiles are weapons:

Missile(x) $\Rightarrow$ Weapon(x)

An enemy of America counts as "hostile“:

Enemy(x,America) $\Rightarrow$ Hostile(x)

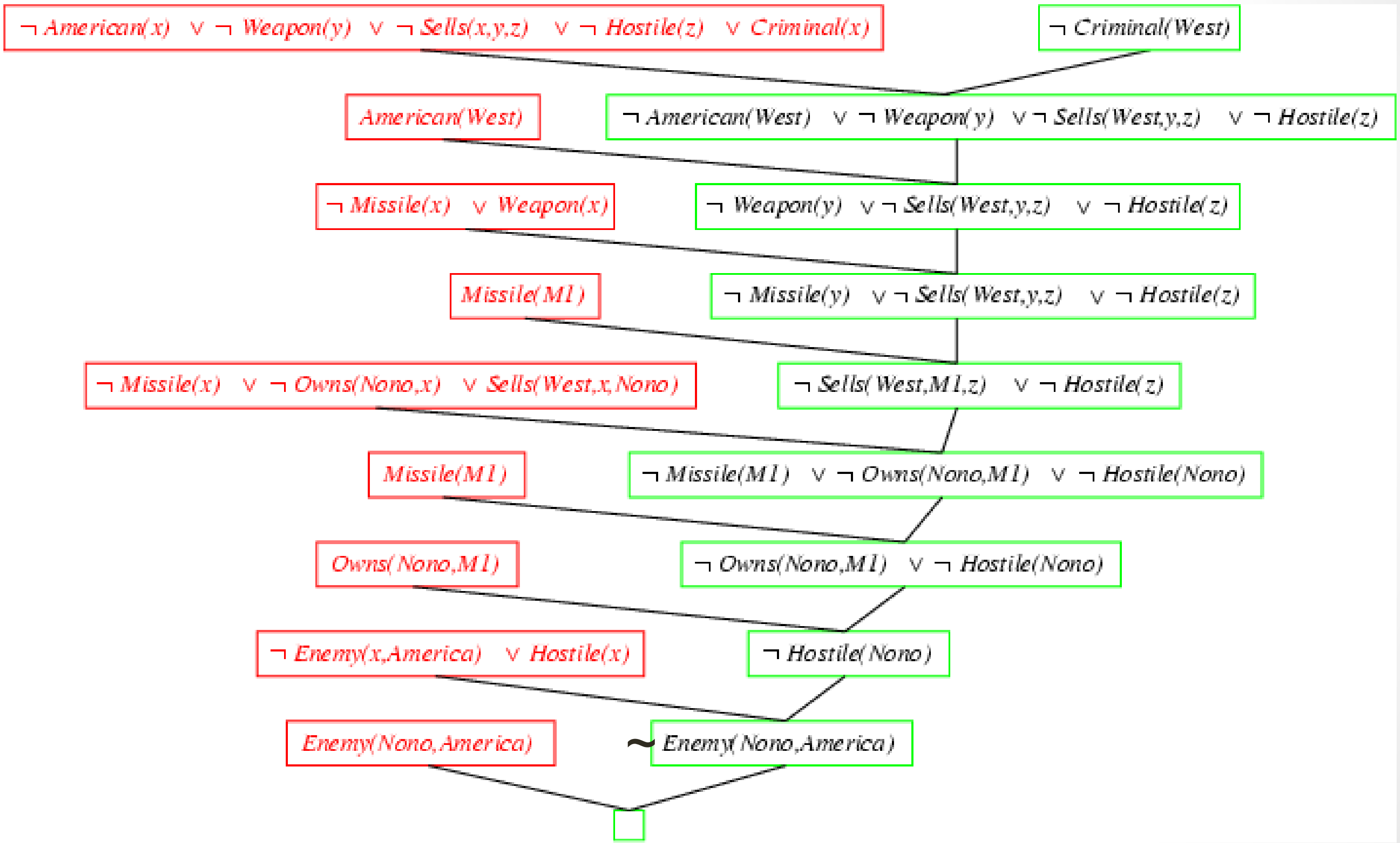
West, who is American …

American(West)

The country Nono, an enemy of America …

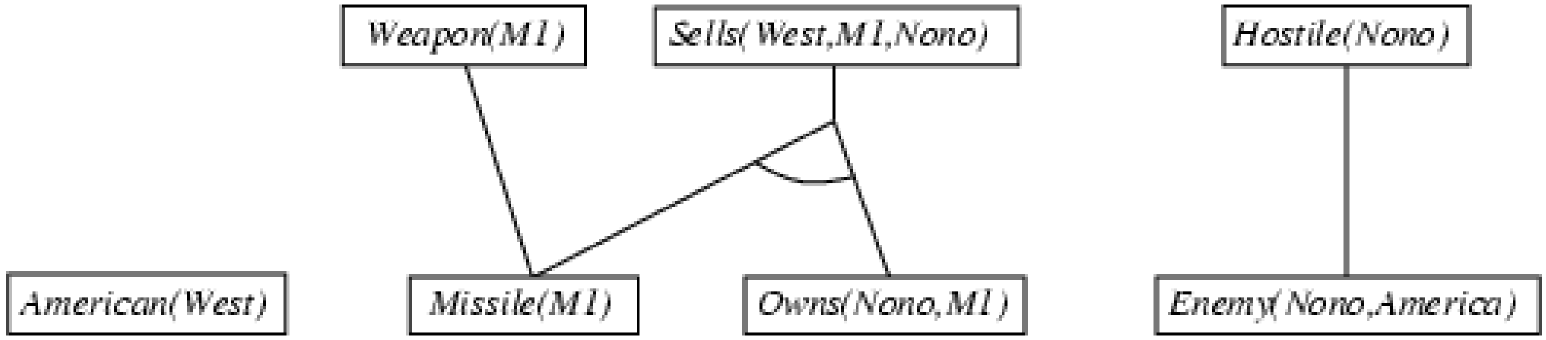
Enemy(Nono,America)

Resolution proof:



Forward chaining proof: (Horn clauses)

Forward chaining proof (Horn clauses)

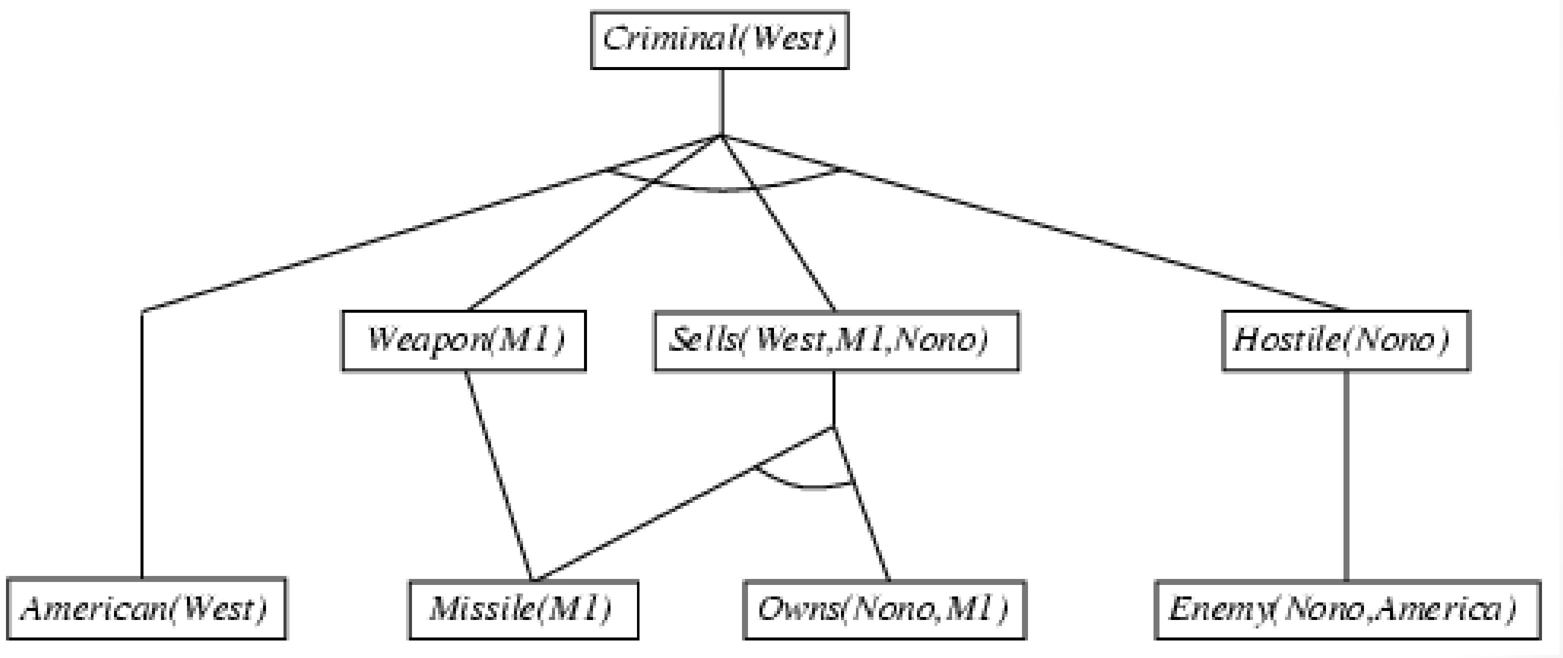


$Enemy(x,America) \Rightarrow Hostile(x)$

$Missile(x) \wedge Owns(Nono,x) \Rightarrow Sells(West,x,Nono)$

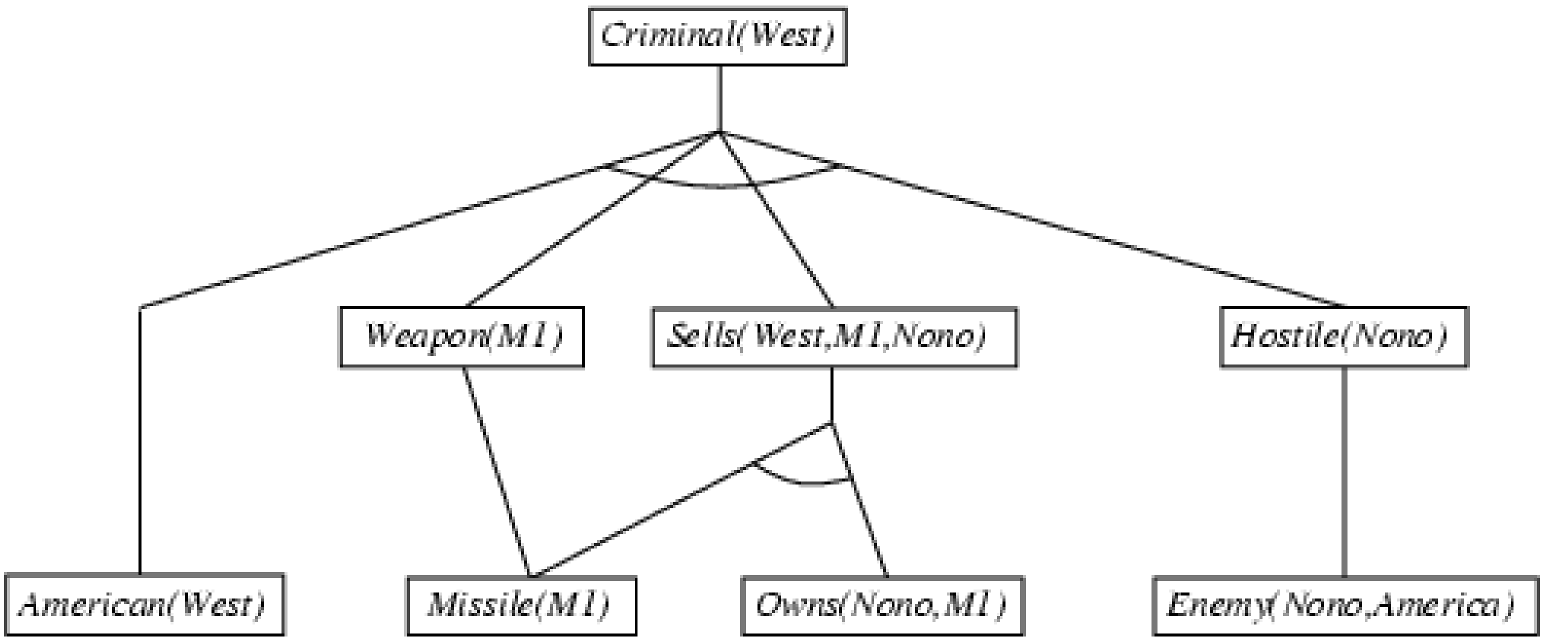
$Missile(x) \Rightarrow Weapon(x)$

Forward chaining proof (Horn clauses)



American(x) $\wedge$ *Weapon(y)* $\wedge$ *Sells(x,y,z)* $\wedge$ *Hostile(z)* $\Rightarrow$ *Criminal(x)*

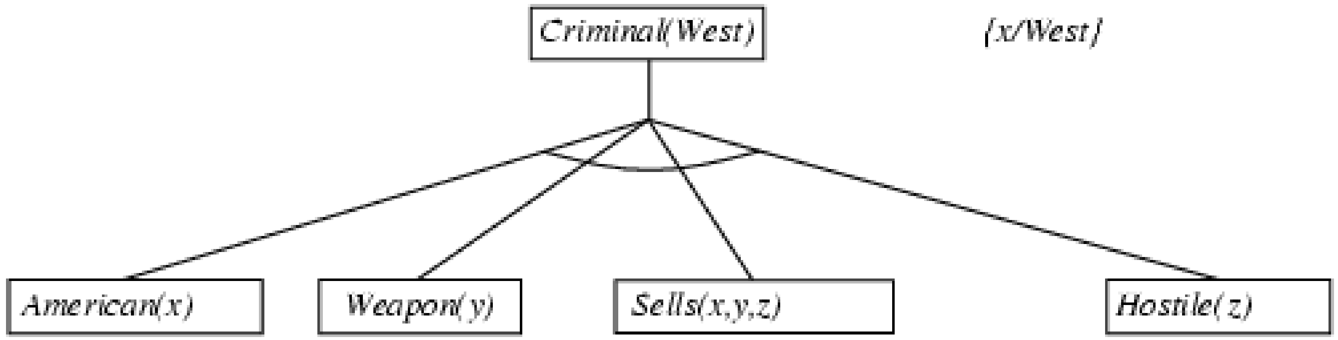
Forward chaining proof (Horn clauses)



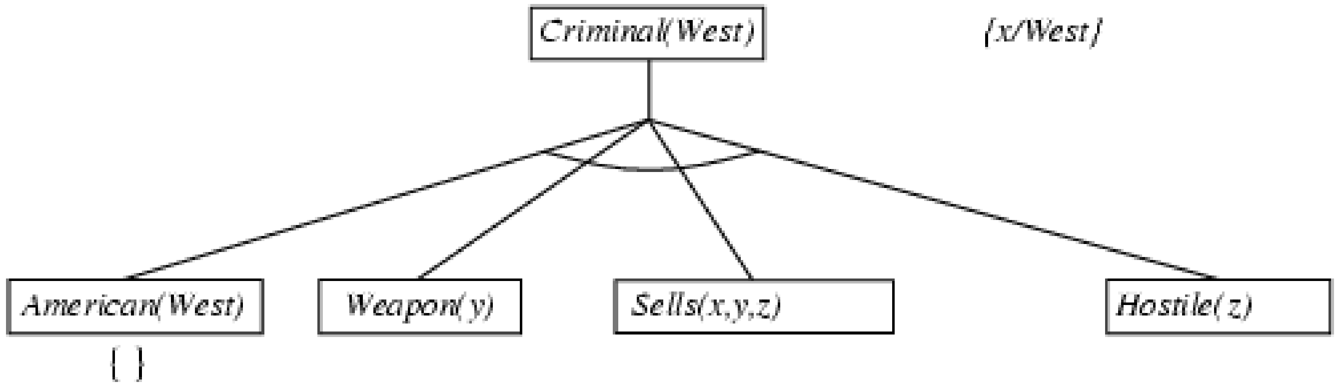
*American(x) $\wedge$ Weapon(y) $\wedge$ Sells(x,y,z) $\wedge$ Hostile(z) $\Rightarrow$ Criminal(x)
*Owns(Nono,M1) and Missile(M1)
*Missile(x) $\wedge$ Owns(Nono,x) $\Rightarrow$ Sells(West,x,Nono)
*Missile(x) $\Rightarrow$ Weapon(x)
*Enemy(x,America) $\Rightarrow$ Hostile(x)
*American(West)
*Enemy(Nono,America)

Backward chaining example (Horn clauses)

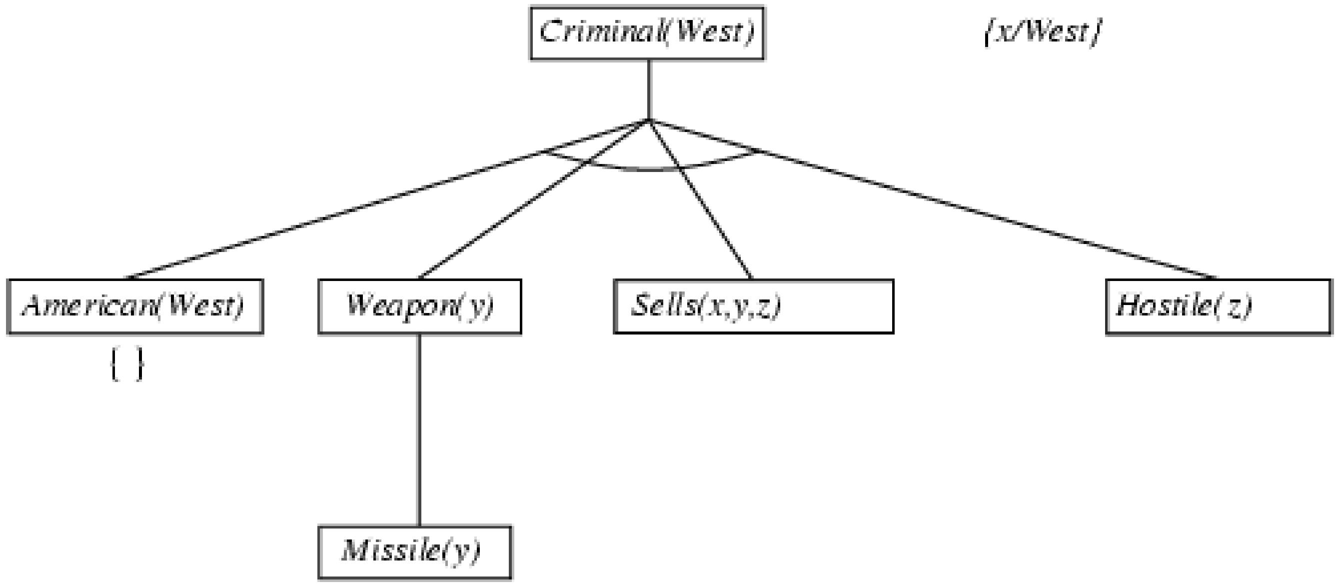
Backward chaining example (Horn clauses)



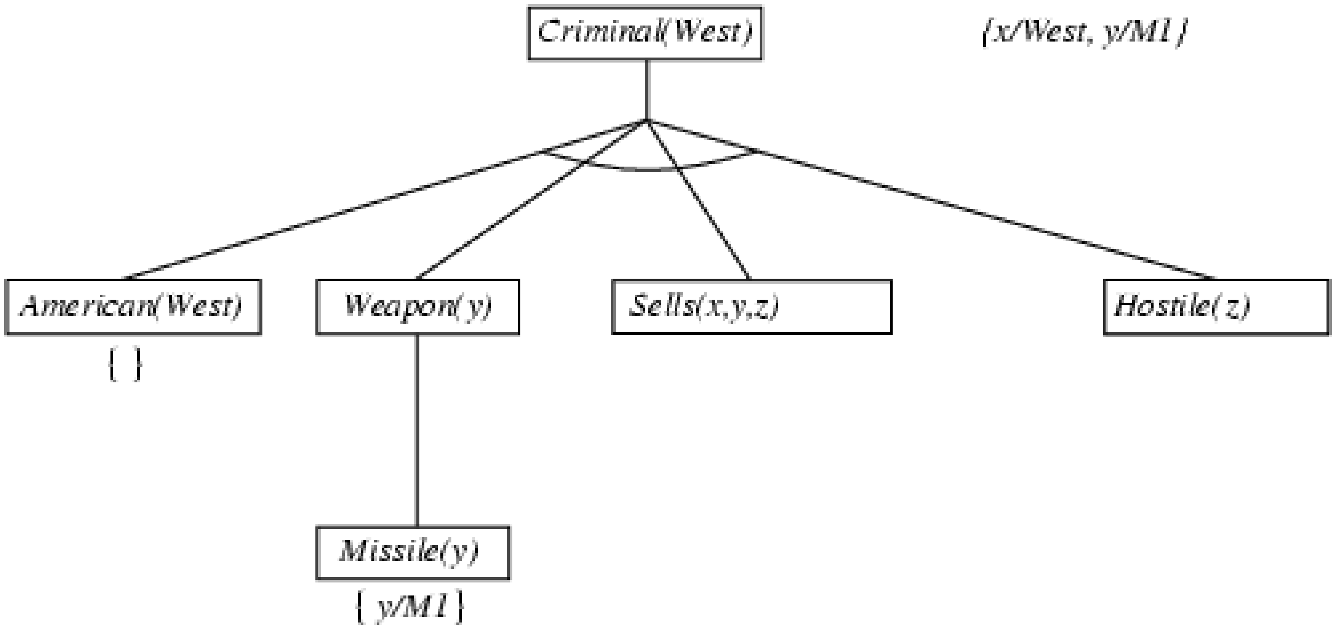
Backward chaining example (Horn clauses)



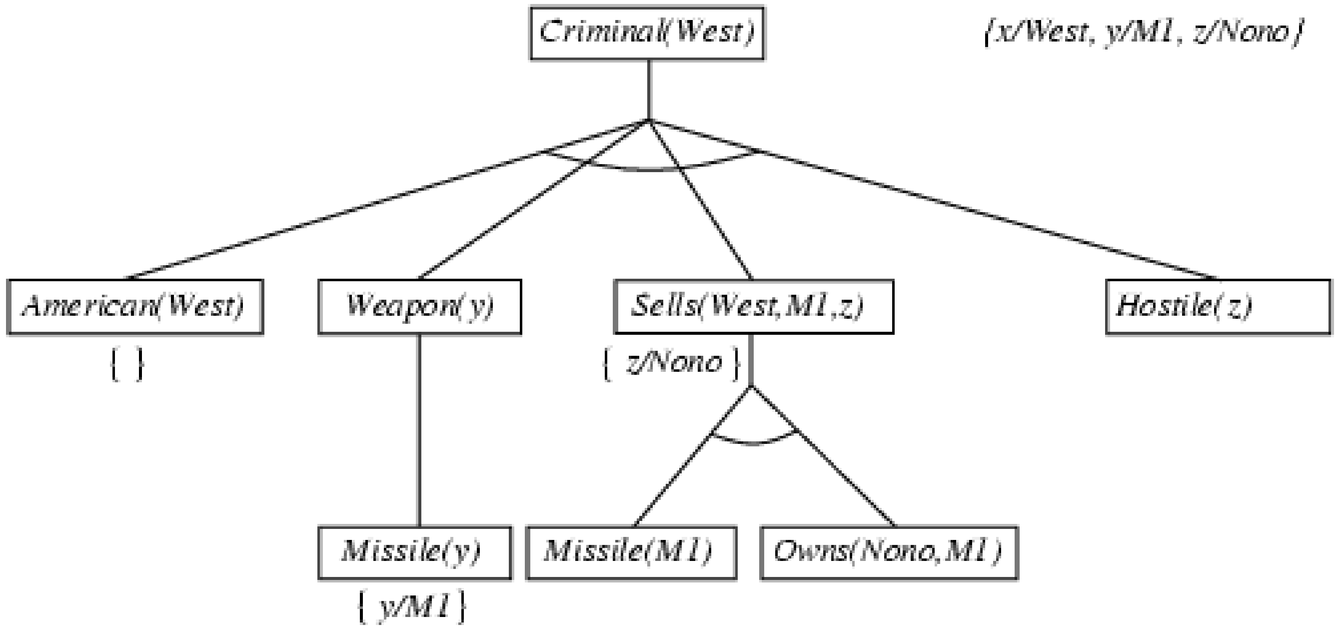
Backward chaining example (Horn clauses)



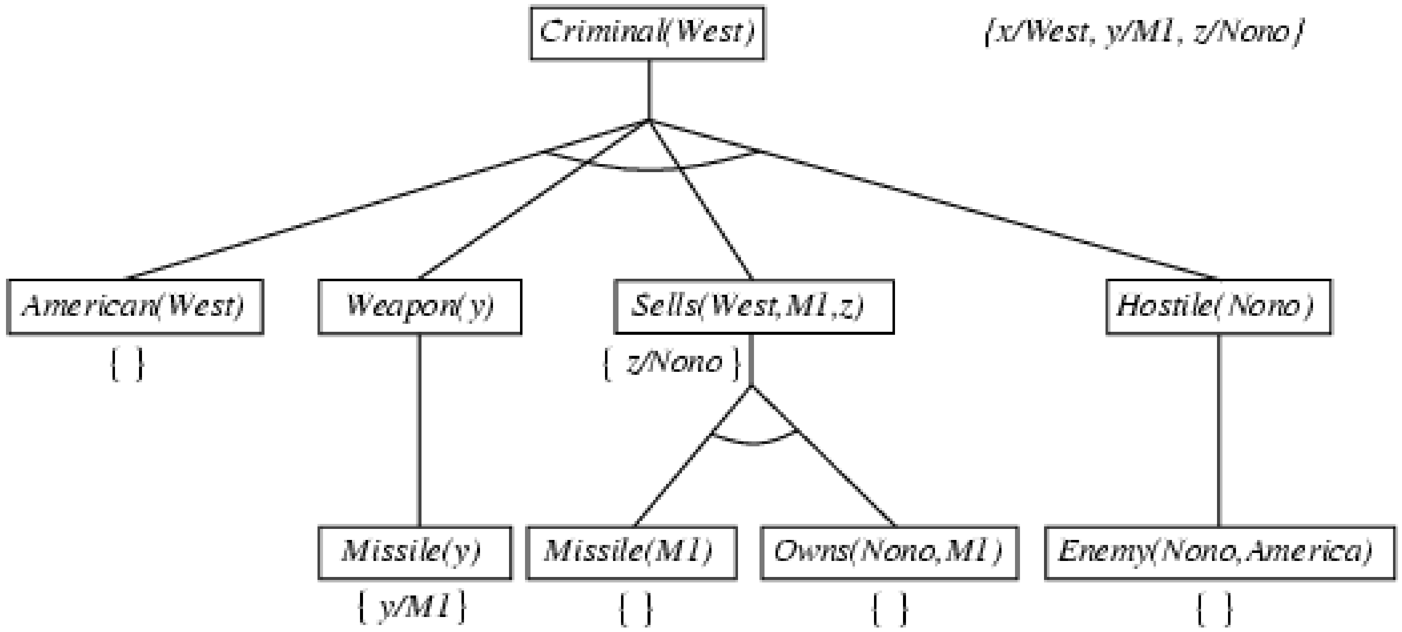
Backward chaining example (Horn clauses)



Backward chaining example (Horn clauses)



Backward chaining example (Horn clauses)



Summary

- First-order logic:
  - Much more expressive than propositional logic
  - Allows objects and relations as semantic primitives
  - Universal and existential quantifiers
- Syntax: constants, functions, predicates, equality, quantifiers
- Nested quantifiers
- Translate simple English sentences to FOPC and back
- Semantics: correct under any interpretation and in any world
- Unification: Making terms identical by substitution
  - The terms are universally quantified, so substitutions are justified.