CSE 473: Logic in AI

Hanna Hajishirzi

(With slides from Luke Zettlemoyer, Dan Weld, Mausam, Stuart Russell, Dieter Fox, Henry Kautz…)

Knowledge Representation

- *Represent knowledge in a manner that facilitates inference (i.e. drawing conclusions) from knowledge.*

- Typically based on
  - Logic
  - Probability
  - Logic and Probability

Propositional Logic: Syntax

- Atoms
  - P, Q, R, …
- Literals
  - P, ¬P
- Sentences
  - Any literal is a sentence
  - If S is a sentence
    - Then (S ∧ S) is a sentence
    - Then (S ∨ S) is a sentence
- Conveniences

$P \rightarrow Q$ same as $\neg P \vee Q$

A Knowledge Base

If the unicorn is mythical, then it is immortal, but if it is not mythical, it is a reptile. If the unicorn is either immortal or a reptile, then it is horned.

$(\neg R \vee H)$

$(\neg I \vee H)$

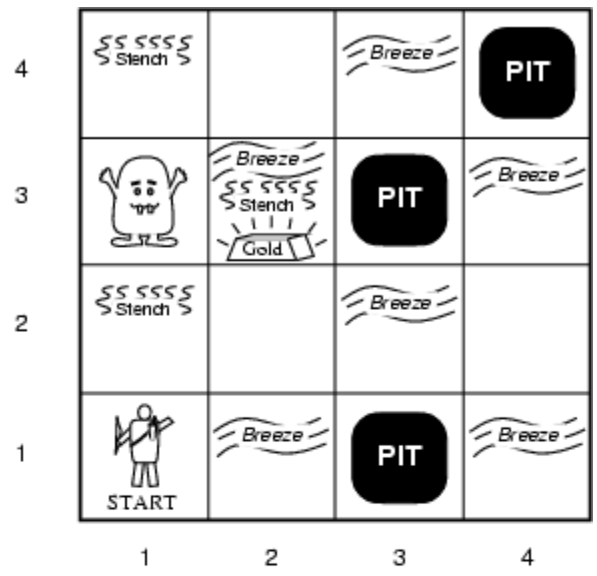
M = mythical
I = immortal
R = reptile
H = horned

$(M \vee R)$

$(\neg M \vee I)$

Wumpus World

- Performance measure
  - Gold: +1000, death: -1000
  - -1 per step, -10 for using the arrow
- Environment
  - Squares adjacent to wumpus are smelly
  - Squares adjacent to pit are breezy
  - Glitter iff gold is in the same square
  - Shooting kills wumpus if you are facing it
  - Shooting uses up the only arrow
  - Grabbing picks up gold if in same square
  - Releasing drops the gold in same square



- Sensors: Stench, Breeze, Glitter, Bump, Scream
- Actuators: Left turn, Right turn, Forward, Grab, Release, Shoot

Wumpus world sentences: KB

Let $P_{i,j}$ be true if there is a pit in [i, j].
Let $B_{i,j}$ be true if there is a breeze in [i, j].

KB:

$\neg P_{1,1}$

$\neg B_{1,1}$

"Pits cause breezes in adjacent squares"

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

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

A Simple Knowledge Based Agent

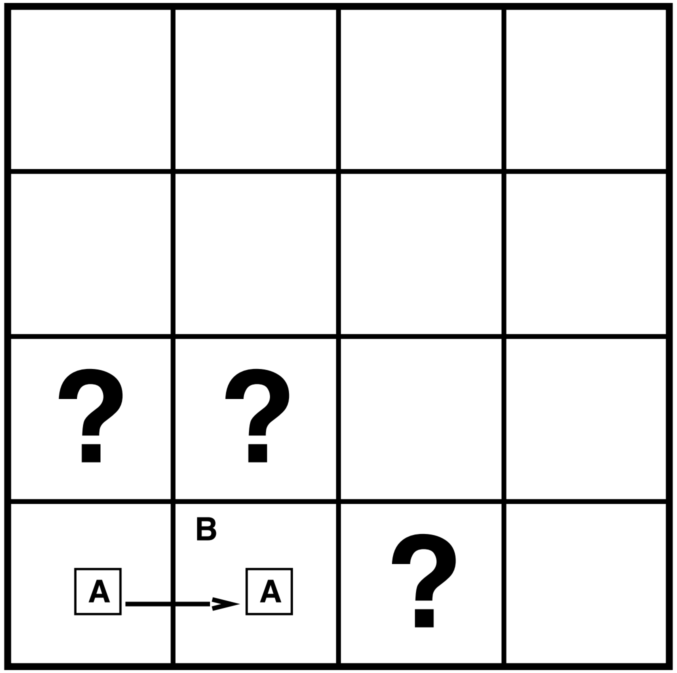
```
function KB-AGENT(percept) returns an action
    static: KB, a knowledge base
            t, a counter, initially 0, indicating time

    TELL(KB, MAKE-PERCEPT-SENTENCE(percept, t))
    action ← ASK(KB, MAKE-ACTION-QUERY(t))
    TELL(KB, MAKE-ACTION-SENTENCE(action, t))
    t ← t + 1
    return action
```

The agent must be able to:

- Represent states, actions, etc.
- Incorporate new percepts
- Update internal representations of the world
- Deduce hidden properties of the world
- Deduce appropriate actions

Entailment in Wumpus World

KB={$\neg P_{1,1}$, $\neg W_{1,1}$, $\neg B_{1,1}$, $\neg G_{1,1}$,
$\neg P_{1,1}$, $\neg W_{1,1}$, $B_{1,1}$, $\neg G_{1,1}$,
…
$B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$
… }



Situation after detecting nothing in [1,1], moving right, breeze in [2,1]

Consider possible models for ?s assuming only pits

3 Boolean choices $\Rightarrow$ 8 possible models

Wumpus Models

Possible assignments for the three locations which we have evidence about:

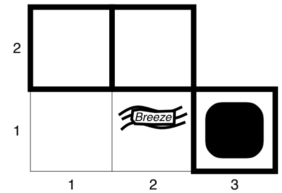
$KB=\{\neg P_{1,1}, \neg W_{1,1}, \neg B_{1,1}, \neg G_{1,1},$

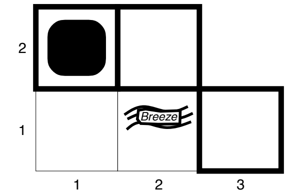
$\neg P_{1,1}, \neg W_{1,1}, B_{1,1}, \neg G_{1,1},$

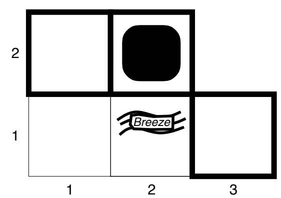
$\ldots$

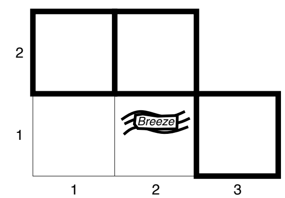
$B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$

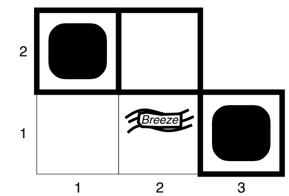
$\ldots \}$



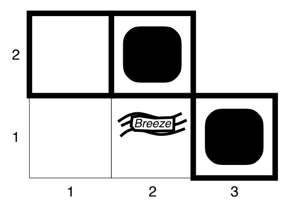


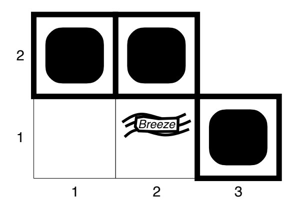


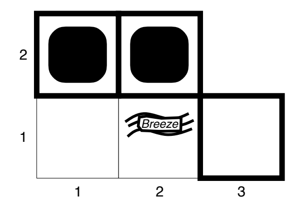




KB is satisfiable!







Wumpus Models

Models that are consistent with our KB:

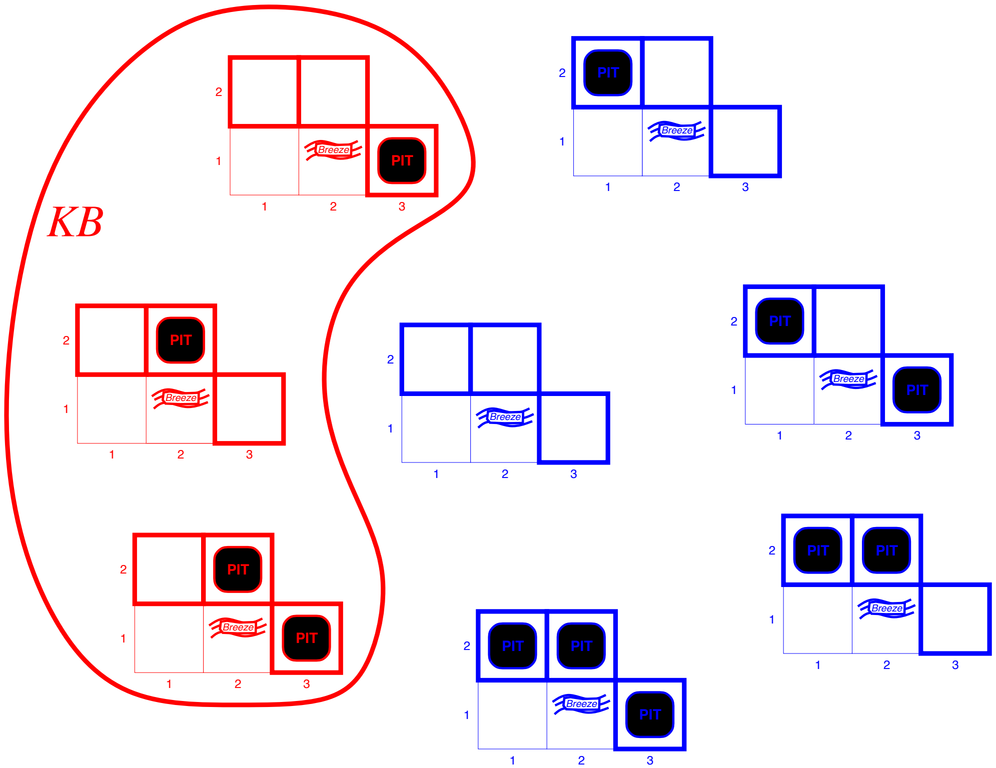
$KB = \{\neg P_{1,1}, \neg W_{1,1}, \neg B_{1,1}, \neg G_{1,1},$

$\neg P_{1,1}, \neg W_{1,1}, B_{1,1}, \neg G_{1,1},$

…

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

… }



KB = wumpus-world rules + observations

Wumpus Models

This KB does entail that [1,2] is safe:

$KB=\{\neg P_{1,1}, \neg W_{1,1}, \neg B_{1,1}, \neg G_{1,1},$

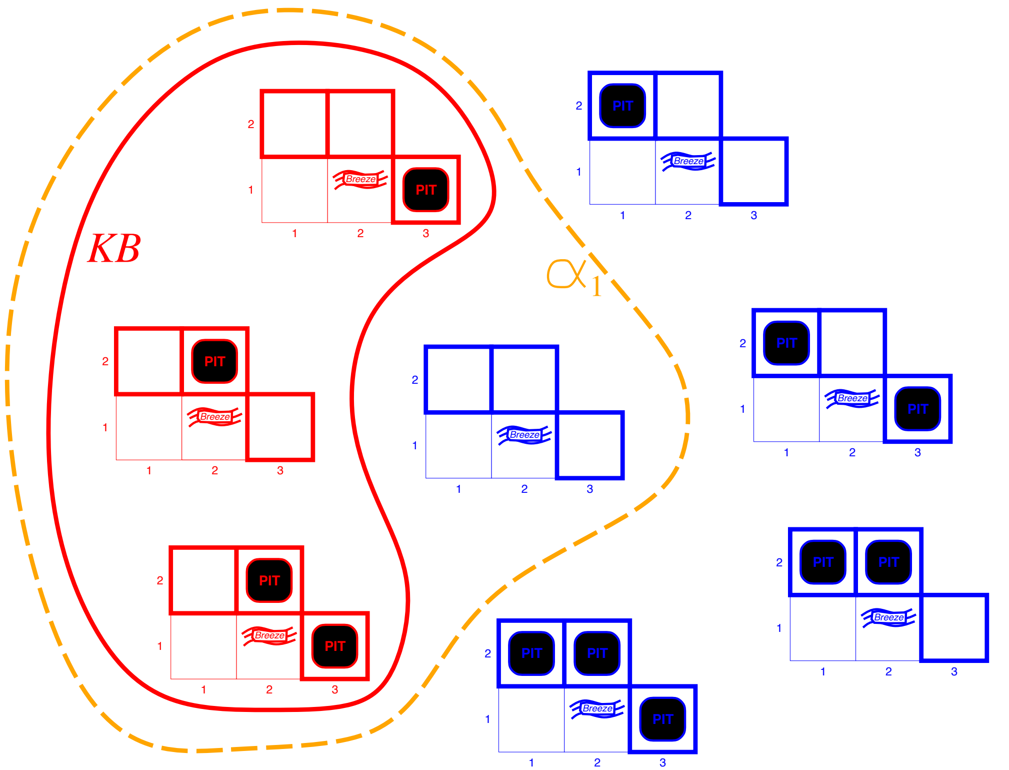
$\neg P_{1,2}, \neg W_{1,2}, B_{1,2}, \neg G_{1,2},$

$\ldots$

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

$\ldots \}$

$\alpha_1 = \neg P_{1,2} \wedge \neg W_{1,2}$



KB = wumpus-world rules + observations

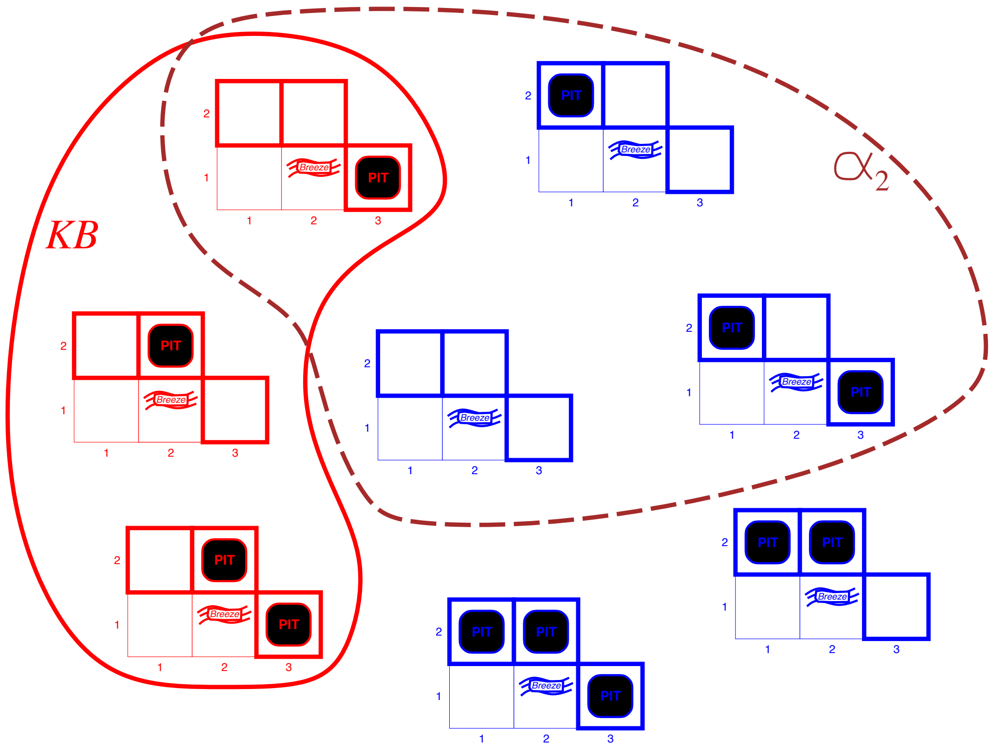
α_1 = "[1,2] is safe", $KB \models \alpha_1$, proved by model checking

Wumpus Models

This KB does not entail that [2,2] is safe:

KB={$\neg P_{1,1}$, $\neg W_{1,1}$, $\neg B_{1,1}$, $\neg G_{1,1}$,
$\neg P_{1,2}$, $\neg W_{1,2}$, $B_{1,2}$, $\neg G_{1,2}$,
…
$B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$
… }

$\alpha_2 = \neg P_{2,2} \wedge \neg W_{2,2}$



KB = wumpus-world rules + observations

α_2 = "[2,2] is safe", $KB \not\models \alpha_2$

Summary: Models

- Logicians often think in terms of *models*, which are formally structured worlds with respect to which truth can be evaluated
  - In propositional case, each model = truth assignment
  - Set of models can be enumerated in a truth table

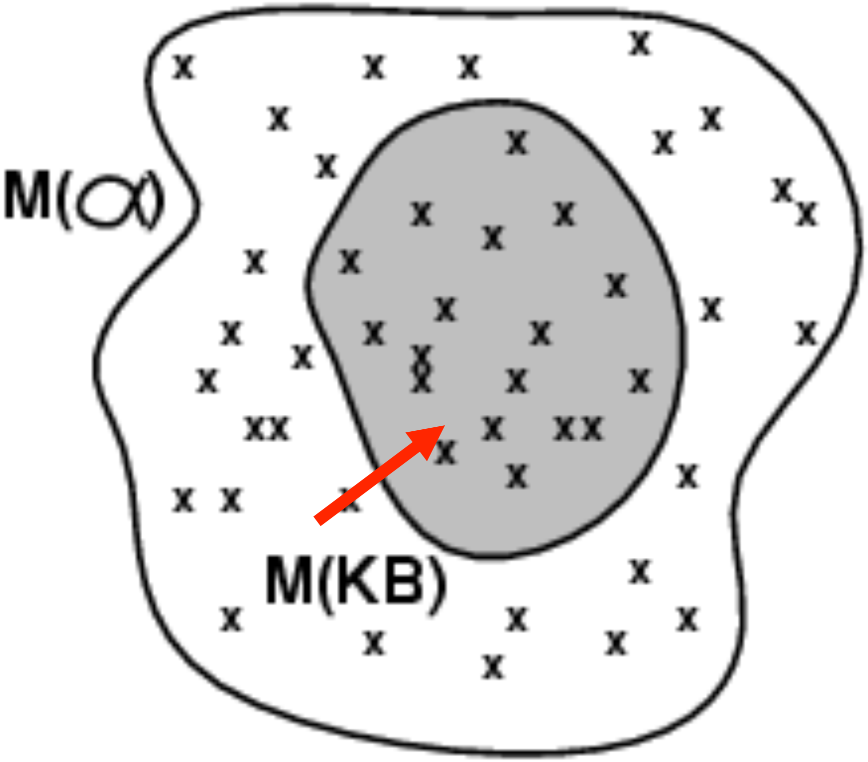
- We say *m* is a model ***of*** a sentence α if α is true in *m*

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

- Entailment: KB $\models$ α iff *M(KB)* $\subseteq$ *M*(α)

  - E.g. *KB* = (P∨Q) ∧(¬P ∨ R)
    α = (P∨R)

- How to check?
  - One way is to enumerate all elements in the truth table – slow ☹



Pros and Cons of Propositional Logic

- Propositional logic is *declarative*: pieces of syntax correspond to facts
- Propositional logic allows *partial/disjunctive/negated* information (unlike most data structures and databases)
- Propositional logic is *compositional*:
  - meaning of $B_{1,1} \wedge P_{1,2}$ derived from meanings of $B_{1,1}$ and $P_{1,2}$
- Propositional logic has very limited expressive power (unlike natural language)
  - E.g., cannot say "pits cause breezes in adjacent squares" except by writing one sentence for each square

Why First Order Logic

Propositional logic: Deals with facts and propositions (can be true or false):

- $P_{1,1}$ -- “there is a pit in (1,1)”
- George_Monkey -- “George is a monkey”
- George_Curious -- “George is curious”
- Jack_Monkey – “Jack is a monkey’’
- 473student1_curious – “student 1 is a curious’’
- 473student2_curious – “student 2 is a curious
- (George_Monkey ∧ ¬473student1_Monkey) ∨ …

FOL Definitions

Constants: Name a specific object.

George, Monkey2, Larry, Hanna…

Variables: Refer to an object without naming it.

X, Y, …

Relations (predicates): Properties of or relationships between objects.

Curious(.), PokesInTheEyes(.,.), SmarterThan(.,.)…

Functions: Mapping from objects to objects.

banana-of(.), grade-of(.), child-of(.,.)

Syntax of First Order Logic

Constants	$KingJohn,\ 2,\ UCB,..$
Predicates	$Brother,\ >, \ldots$
Functions	$Sqrt,\ LeftLegOf, \ldots$
Variables	$x,\ y,\ a,\ b, \ldots$
Connectives	$\wedge\ \vee\ \neg\ \Rightarrow\ \Leftrightarrow$
Equality	$=$
Quantifiers	$\forall\ \exists$

Atomic sentence = $predicate(term_1, \ldots, term_n)$ or $term_1 = term_2$

Term = $function(term_1, \ldots, term_n)$ or $constant$ or $variable$

Atomic Sentences:

E.g., $Brother(KingJohn, RichardTheLionheart)$

$> (Length(LeftLegOf(Richard)), Length(LeftLegOf(KingJohn)))$

Complex Sentences:

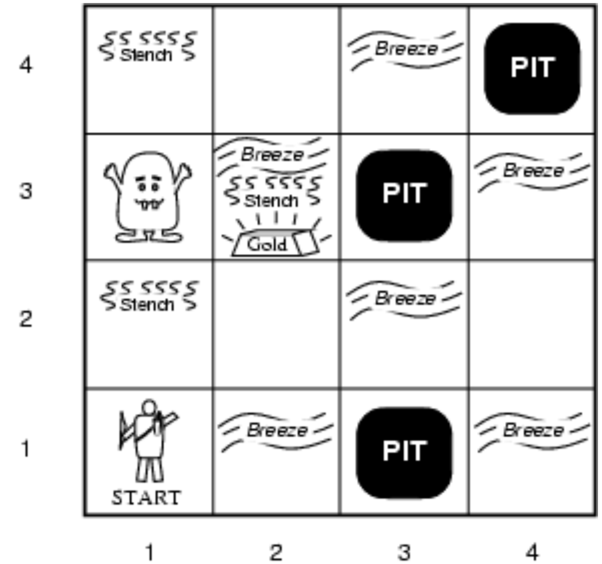
E.g. $Sibling(KingJohn, Richard) \Rightarrow Sibling(Richard, KingJohn)$

$>(1,2) \vee \leq(1,2)$

$>(1,2) \wedge \neg>(1,2)$

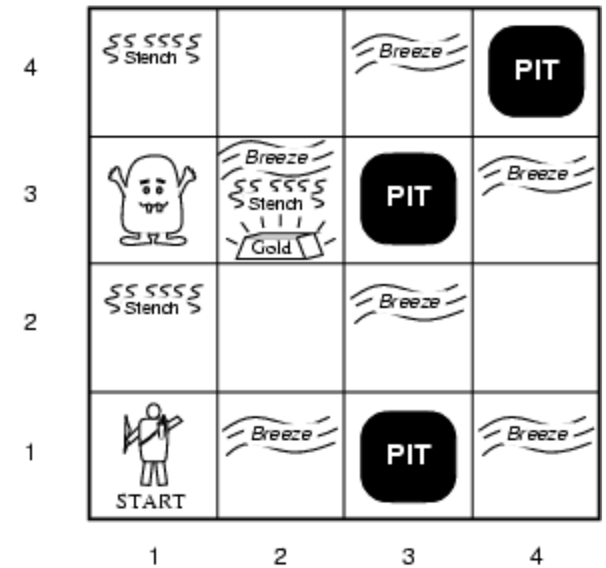
Wumpus World

- Performance measure
  - Gold: +1000, death: -1000
  - -1 per step, -10 for using the arrow
- Environment
  - Squares adjacent to wumpus are smelly
  - Squares adjacent to pit are breezy
  - Glitter iff gold is in the same square
  - Shooting kills wumpus if you are facing it
  - Shooting uses up the only arrow
  - Grabbing picks up gold if in same square
  - Releasing drops the gold in same square



- Sensors: Stench, Breeze, Glitter, Bump, Scream
- Actuators: Left turn, Right turn, Forward, Grab, Release, Shoot

Wumpus World



Properties of locations:

$\forall x, t \;\; At(Agent, x, t) \land Smelt(t) \;\Rightarrow\; Smelly(x)$

$\forall x, t \;\; At(Agent, x, t) \land Breeze(t) \;\Rightarrow\; Breezy(x)$

Diagnostic rule—infer cause from effect

$$\forall y \;\; Breezy(y) \;\Rightarrow\; \exists x \;\; Pit(x) \land Adjacent(x, y)$$

Causal rule—infer effect from cause

$$\forall x, y \;\; Pit(x) \land Adjacent(x, y) \;\Rightarrow\; Breezy(y)$$

Neither of these is complete—e.g., the causal rule doesn't say whether squares far away from pits can be breezy

Definition for the *Breezy* predicate:

$$\forall y \;\; Breezy(y) \;\Leftrightarrow\; [\exists x \;\; Pit(x) \land Adjacent(x, y)]$$

First Order Models

Sentences are true with respect to a model and an interpretation

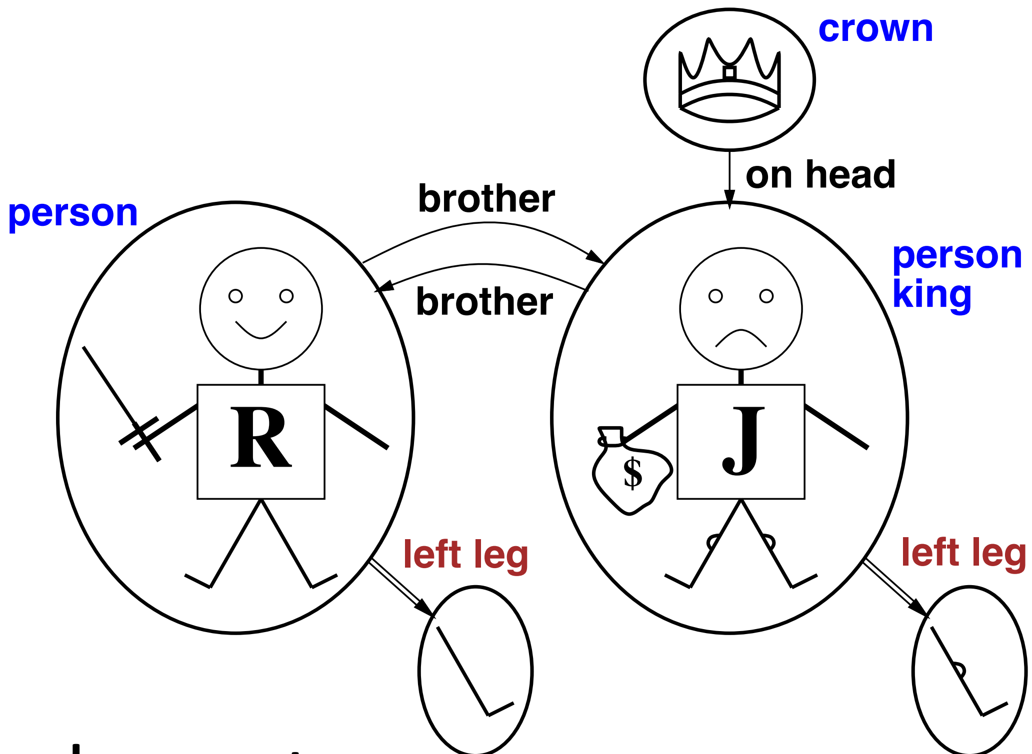
Model contains ≥ 1 objects (domain elements) and relations among them

Interpretation specifies referents for

constant symbols $\rightarrow$ objects
predicate symbols $\rightarrow$ relations
function symbols $\rightarrow$ functional relations

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 $predicate$

Example: A World of Kings and Legs



- Syntactic elements:

Constants:

Richard, John, RsLeftLeg, …

Functions:

leftleg(.), onheadof(.), …

Relations:

On(.,.) IsKing(.), IsPerson(.), …

Semantics

- Logical connectives: and, or, not, $\Rightarrow$, $\Leftrightarrow$
- Quantifiers:
  - $\forall$ For all (Universal quantifier)
  - $\exists$ There exists (Existential quantifier)
- Examples
  - George is a monkey and he is curious

    Monkey(George) $\wedge$ Curious(George)
  - All monkeys are curious

    $\forall$m: Monkey(m) $\Rightarrow$ Curious(m)
  - There is a curious monkey

    $\exists$m: Monkey(m) $\wedge$ Curious(m)

Quantifier / Connective Interaction

$M(x)$ == "x is a monkey"
$C(x)$ == "x is curious"

$$\forall x: \; M(x) \wedge C(x)$$

"Everything is a curious monkey"

$$\forall x: \; M(x) \Rightarrow C(x)$$

"All monkeys are curious"

$$\exists x: \; M(x) \wedge C(x)$$

"There exists a curious monkey"

$$\exists x: \; M(x) \Rightarrow C(x)$$

"There exists an object that is *either* a curious monkey, *or* not a monkey at all"

Fun With Sentences

- Brothers are siblings.

$$\forall x, y \;\; Brother(x, y) \;\Rightarrow\; Sibling(x, y).$$

- “Sibling” is symmetric.

$$\forall x, y \;\; Sibling(x, y) \;\Leftrightarrow\; Sibling(y, x).$$

- One’s mother is one’s female parent.

$$\forall x, y \;\; Mother(x, y) \;\Leftrightarrow\; (Female(x) \land Parent(x, y)).$$

- A first cousin is a child of a parent’s sibling.

$$\forall x, y \;\; FirstCousin(x, y) \;\Leftrightarrow\; \exists p, ps \;\; Parent(p, x) \land Sibling(ps, p) \land Parent(ps, y)$$

Propositional. Logic *vs.* First Order

Ontology	Facts (P, Q,…)	Objects, Properties, Relations
Syntax	Atomic sentences Connectives	Variables & quantification Sentences have structure: terms father-of(mother-of(X)))
Semantics	Truth Tables	Interpretations & Models (Much more complicated)
Inference Algorithm	DPLL, WalkSAT Fast in practice	Unification Forward, Backward chaining Prolog, theorem proving
Complexity	NP-Complete	Semi-decidable May run forever if $KB \not\models \alpha$