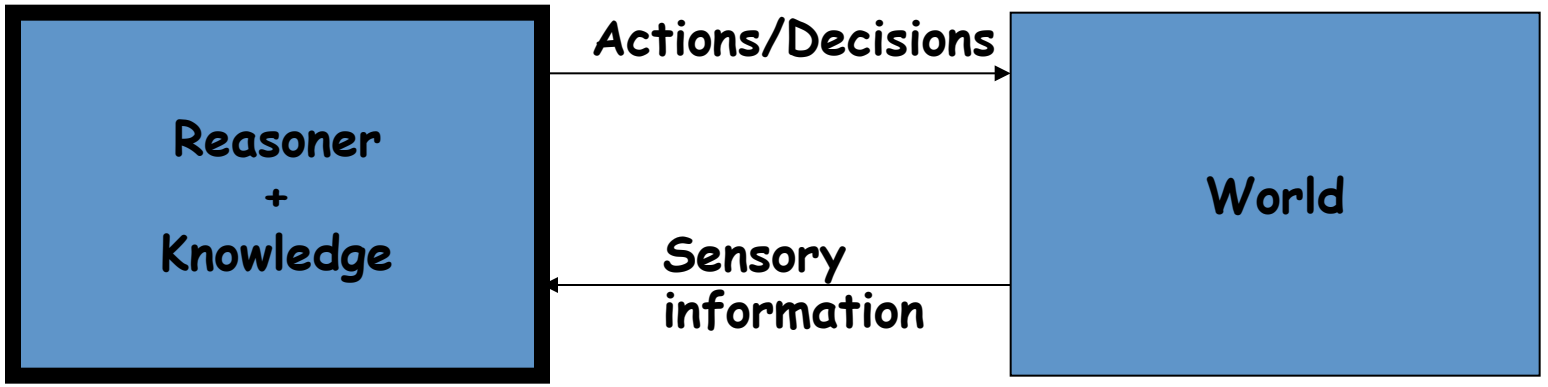
CSE 473: Logic in AI

Hanna Hajishirzi

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

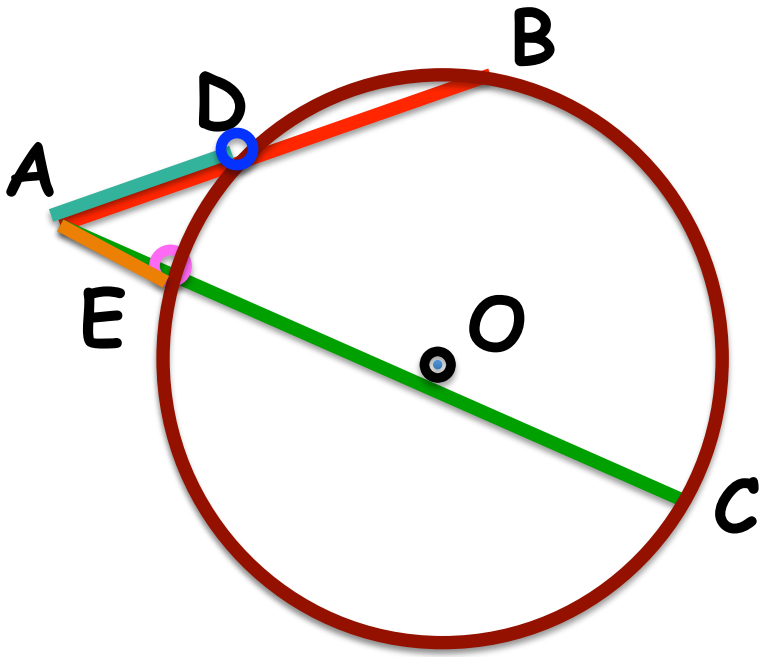
Knowledge Representation and Reasoning (KR&R)

- Advice taker: a paradigm for KR&R
  - Represent knowledge (with statements)
  - Add statements when you want to give advice (control knowledge = statements)
  - World vs Reasoner (Decision Maker)



Solving Geometry Questions

In the diagram, secant AB intersects circle O at D, secant AC intersects circle O at E, AE = 4, AC = 24, and AB = 16. Find AD.



Knowledge about geometry axioms and theorems

KR Hypothesis

Any ***intelligent process*** will have ingredients that

1) We as external observers interpret as knowledge
2) This knowledge plays a formal, causal & essential role in guiding the behavior

- Brian Smith (paraphrased)

Some KR Languages

- Propositional Logic
- Predicate Calculus
- Frame Systems
- Rules with Certainty Factors
- Bayesian Belief Networks
- Influence Diagrams
- Semantic Networks
- Concept Description Languages
- Non-monotonic Logic

Knowledge Representation

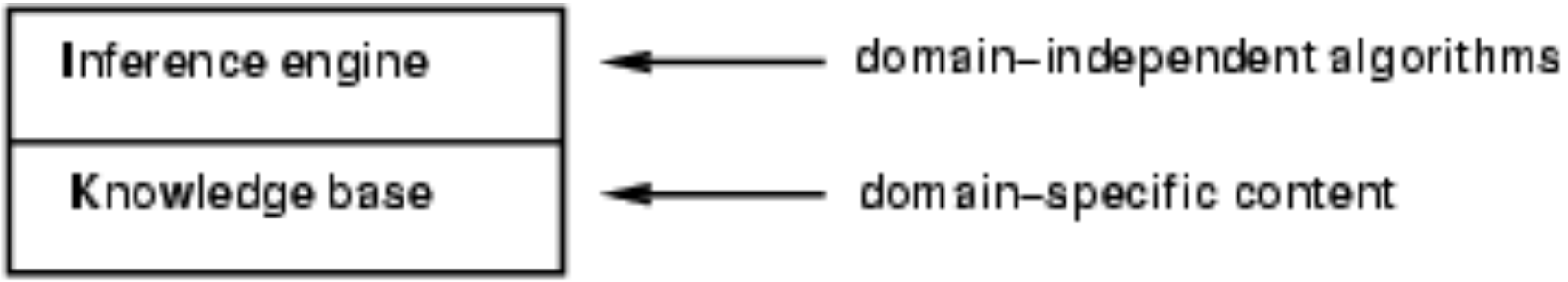
- *Represent knowledge in a manner that facilitates inference (i.e. drawing conclusions) from knowledge.*

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

Basic Idea of Logic

- By starting with true assumptions, you can deduce true conclusions.

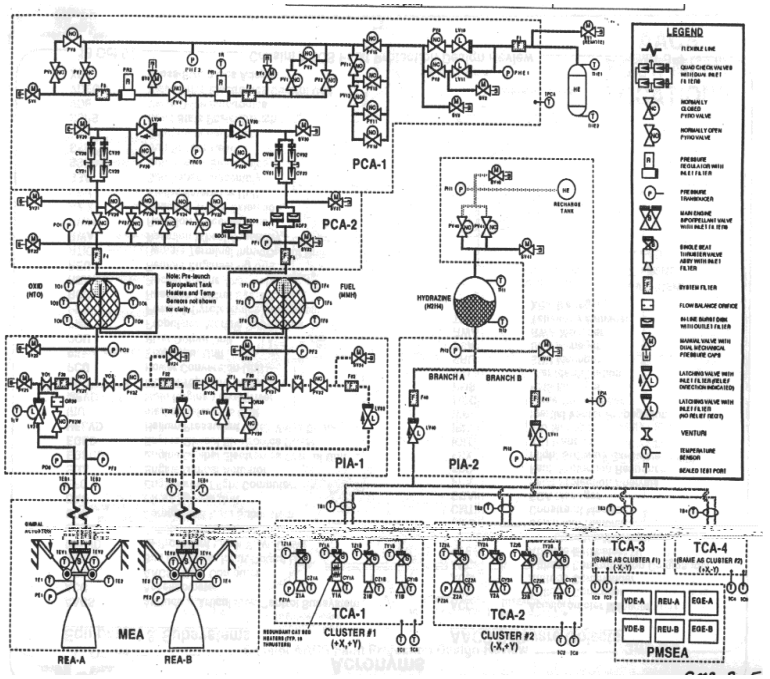
Knowledge bases



- Knowledge base = set of sentences in a formal language
- Declarative approach to building an agent (or other system):
  - `Tell` it what it needs to know
- Then it can `Ask` itself what to do - answers should follow from the KB

Deep Space One

- Autonomous diagnosis & repair “Remote Agent”
- Compiled schematic to 7,000 var SAT problem



Muddy Children Problem

- Mom to N children “Don’t get dirty”
- While playing, K≥1 get mud on forehead
- Father: “Some of you are dirty!”
- Father: “Raise your hand if you are dirty”
  - No one raises hand
- Father: “Raise your hand if you are dirty”
  - No one raises hand

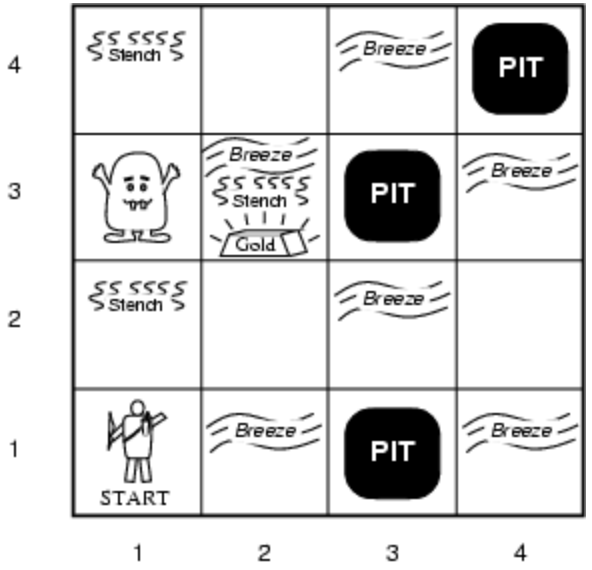
…

- Father: “Raise your hand if you are dirty”
  - All dirty children raise hand

} K times

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

Components of KR

- Syntax: defines the sentences in the language
- Semantics: defines the "meaning" of sentences
- Inference Procedure
  - Algorithm
  - Sound?
  - Complete?
  - Complexity
- Knowledge Base

Propositional Logic

- Syntax
  - Atomic sentences: P, Q, …
  - Connectives: $\wedge$, $\vee$, $\neg$, $\Rightarrow$
- Semantics
  - Truth Tables
- Inference
  - Modus Ponens
  - Resolution
  - DPLL
  - GSAT

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 → Q same as ¬ P ∨ Q

Truth tables for connectives

P	Q	$\neg P$	$P \wedge Q$	$P \vee Q$	$P \Rightarrow Q$	$P \Leftrightarrow Q$
$false$	$false$	$true$	$false$	$false$	$true$	$true$
$false$	$true$	$true$	$false$	$true$	$true$	$false$
$true$	$false$	$false$	$false$	$true$	$false$	$false$
$true$	$true$	$false$	$true$	$true$	$true$	$true$

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.

(¬ R ∨ H) **(¬I ∨ H)**

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

(M ∨ R)

(¬ M ∨ I)

Prop. Logic: **Knowledge Engr**

1) One of the women is a biology major
2) Lisa is not next to Dave in the ranking
3) Dave is immediately ahead of Jim
4) Jim is immediately ahead of a bio major
5) Mary or Lisa is ranked first

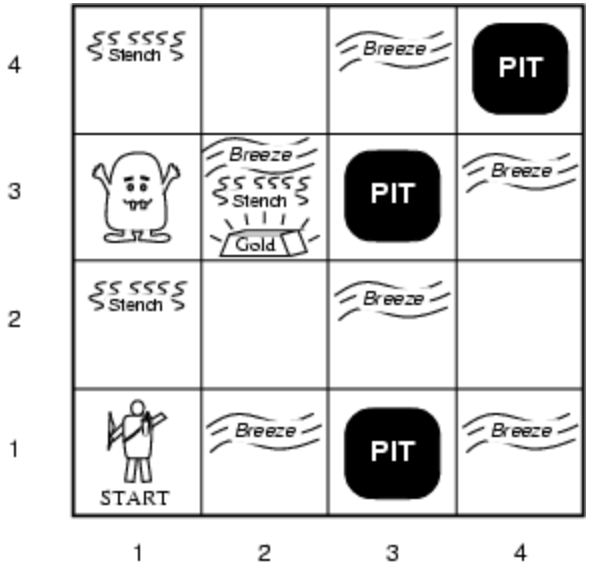
1. Choose Vocabulary

Universe: Lisa, Dave, Jim, Mary
LD = "Lisa is immediately ahead of Dave"
D = "Dave is a Bio Major"

2. Choose initial sentences

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"

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

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

Full Encoding of Wumpus World

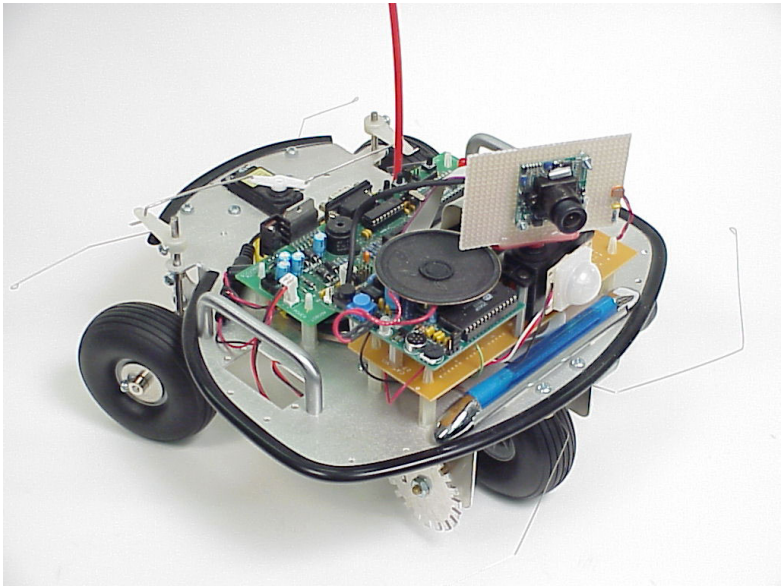
In propositional logic:

$\neg P_{1,1}$
$\neg W_{1,1}$
$B_{x,y} \Leftrightarrow (P_{x,y+1} \vee P_{x,y-1} \vee P_{x+1,y} \vee P_{x-1,y})$
$S_{x,y} \Leftrightarrow (W_{x,y+1} \vee W_{x,y-1} \vee W_{x+1,y} \vee W_{x-1,y})$
$W_{1,1} \vee W_{1,2} \vee \ldots \vee W_{4,4}$
$\neg W_{1,1} \vee \neg W_{1,2}$
$\neg W_{1,1} \vee \neg W_{1,3}$
…

$\Rightarrow$ 64 distinct proposition symbols, 155 sentences

State Estimation

- Maintaining a KB which records what you know about the (partially observed) world state
  - Prop logic
  - First order logic
  - Probabilistic encodings



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
```

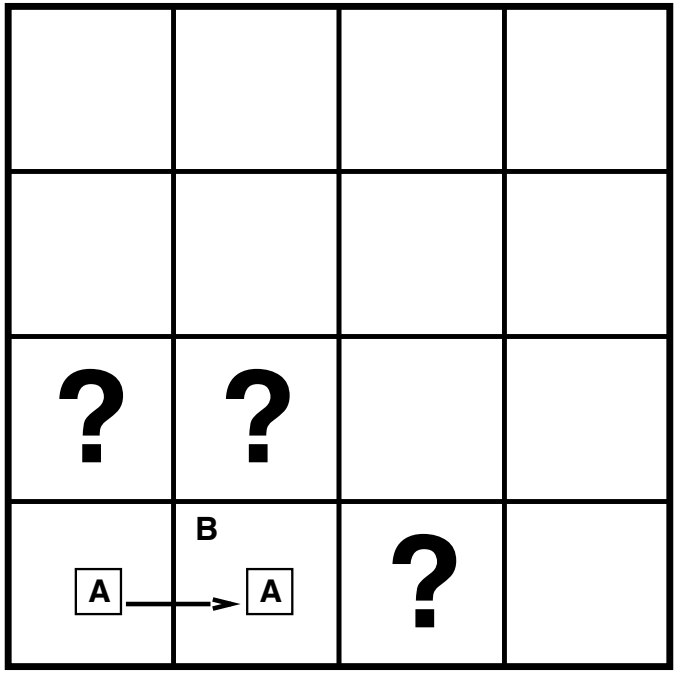
The agent must be able to:

- 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]



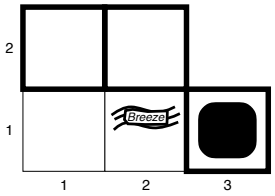
Consider possible models for ?s assuming only pits

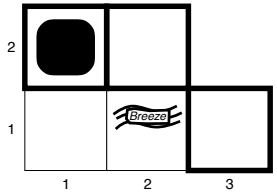
3 Boolean choices $\Rightarrow$ 8 possible models

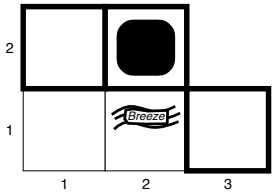
Wumpus Models

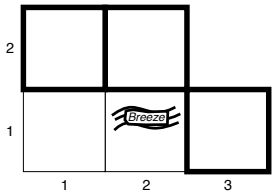
Possible assignments for the three locations which we have evidence about:

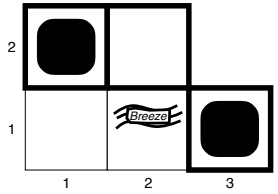
KB={$\neg P_{1,1}$, $\neg W_{1,1}$, $\neg B_{1,1}$, $\neg G_{1,1}$,
$\neg P_{1,1}$, $\neg W_{1,1}$, $B_{1,1}$, $\neg G_{1,1}$,
…
$B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$
… }



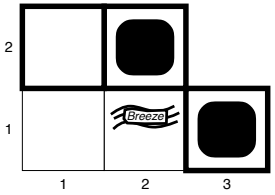


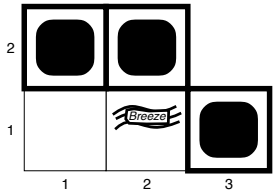


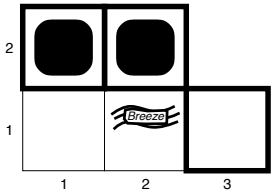




KB is satisfiable!



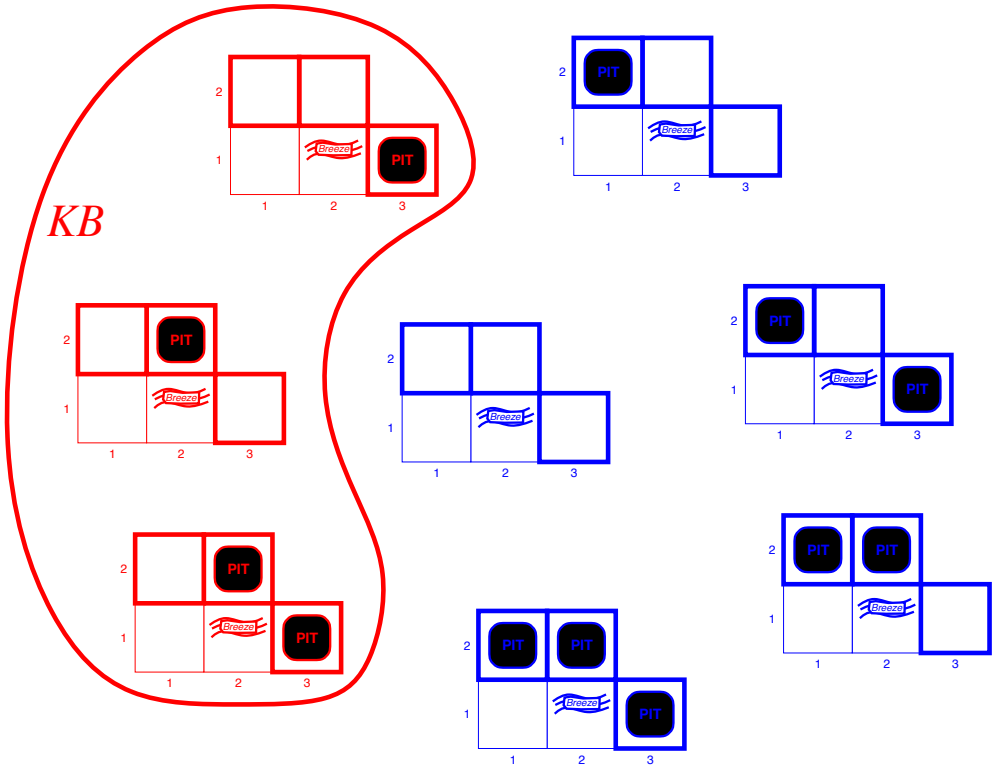




Wumpus Models

Models that are consistent with our KB:

KB={¬$P_{1,1}$, ¬$W_{1,1}$, ¬$B_{1,1}$, ¬$G_{1,1}$,
¬$P_{1,1}$, ¬$W_{1,1}$, $B_{1,1}$, ¬$G_{1,1}$,
…
$B_{1,1}$ ⇔ ($P_{1,2}$ ∨ $P_{2,1}$)
… }



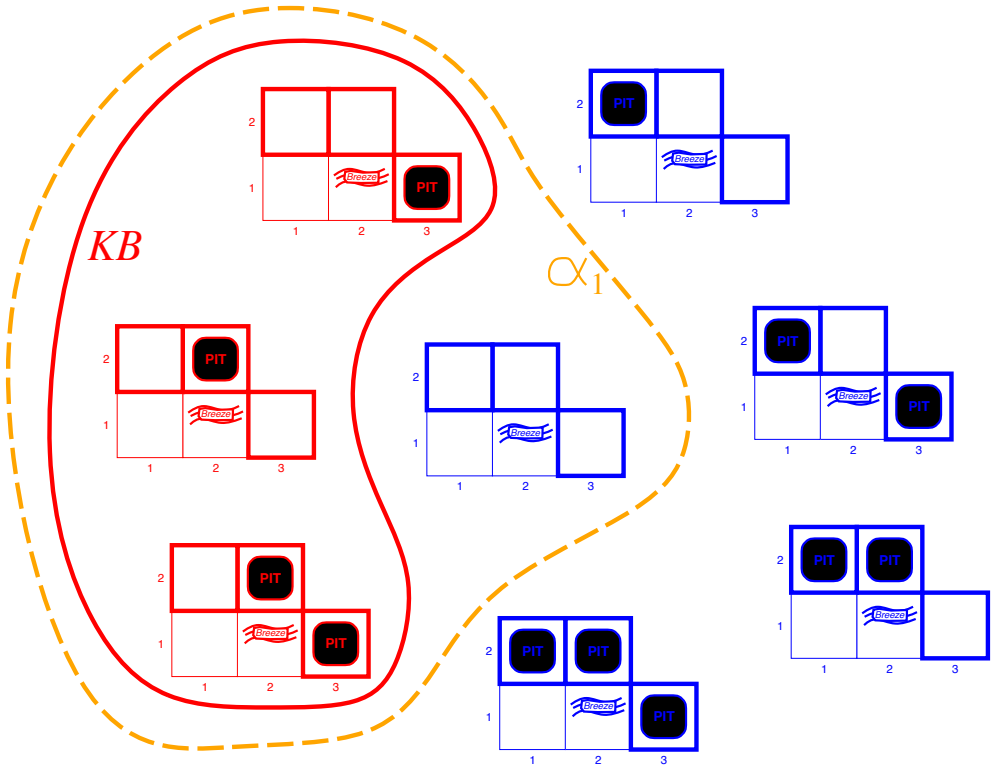
KB = wumpus-world rules + observations

Wumpus Models

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

KB={¬$P_{1,1}$, ¬$W_{1,1}$, ¬$B_{1,1}$, ¬$G_{1,1}$,
¬$P_{1,2}$, ¬$W_{1,2}$, $B_{1,2}$, ¬$G_{1,2}$,
…
$B_{1,1}$ ⇔ ($P_{1,2}$ ∨ $P_{2,1}$)
… }

α_1= ¬$P_{1,2}$ ∧ ¬$W_{1,2}$



KB = wumpus-world rules + observations

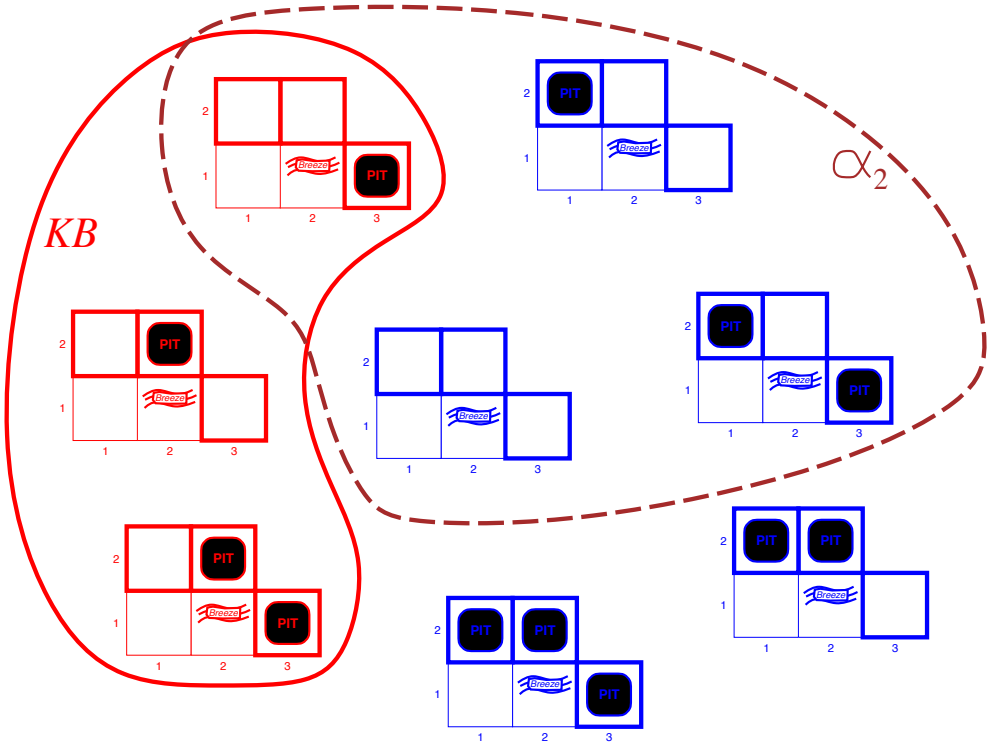
α_1 = "[1,2] is safe", $KB \models \alpha_1$, proved by model checking

Wumpus Models

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

KB={¬$P_{1,1}$, ¬$W_{1,1}$, ¬$B_{1,1}$, ¬$G_{1,1}$,
¬$P_{1,2}$, ¬$W_{1,2}$, $B_{1,2}$, ¬$G_{1,2}$,
…
$B_{1,1}$ ⇔ ($P_{1,2}$ ∨ $P_{2,1}$)
… }

α_2= ¬$P_{2,2}$ ∧ ¬$W_{2,2}$



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 $\models \alpha$ iff $M(KB) \subseteq M(\alpha)$
  - E.g. *KB* = $(P \vee Q) \wedge (\neg P \vee R)$
    $\alpha = (P \vee R)$
- How to check?
  - One way is to enumerate all elements in the truth table – slow ☹