Logical Agents (I)

AIMA Chapter 7

Knowledge-Based Agents

- Until now – trying to find an optimal solution via **search**.
- Assignment of values to variables – **CSP.**
- No real model of what the agent **knows**.
- This class: **represent** agent domain **knowledge** using **logical formulas.**

Knowledge Base (KB)

Inference Engine ← Domain-independent algorithms

Knowledge Base ← Domain-specific content

- 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
- Agents can be viewed at the knowledge level

  i.e., specify knowledge and goals, regardless of implementation
- Or at the implementation level
  - i.e., data structures in KB and algorithms that manipulate them

What is the best action at time t?

What did I perceive at time t?

What happened?

What have I done?

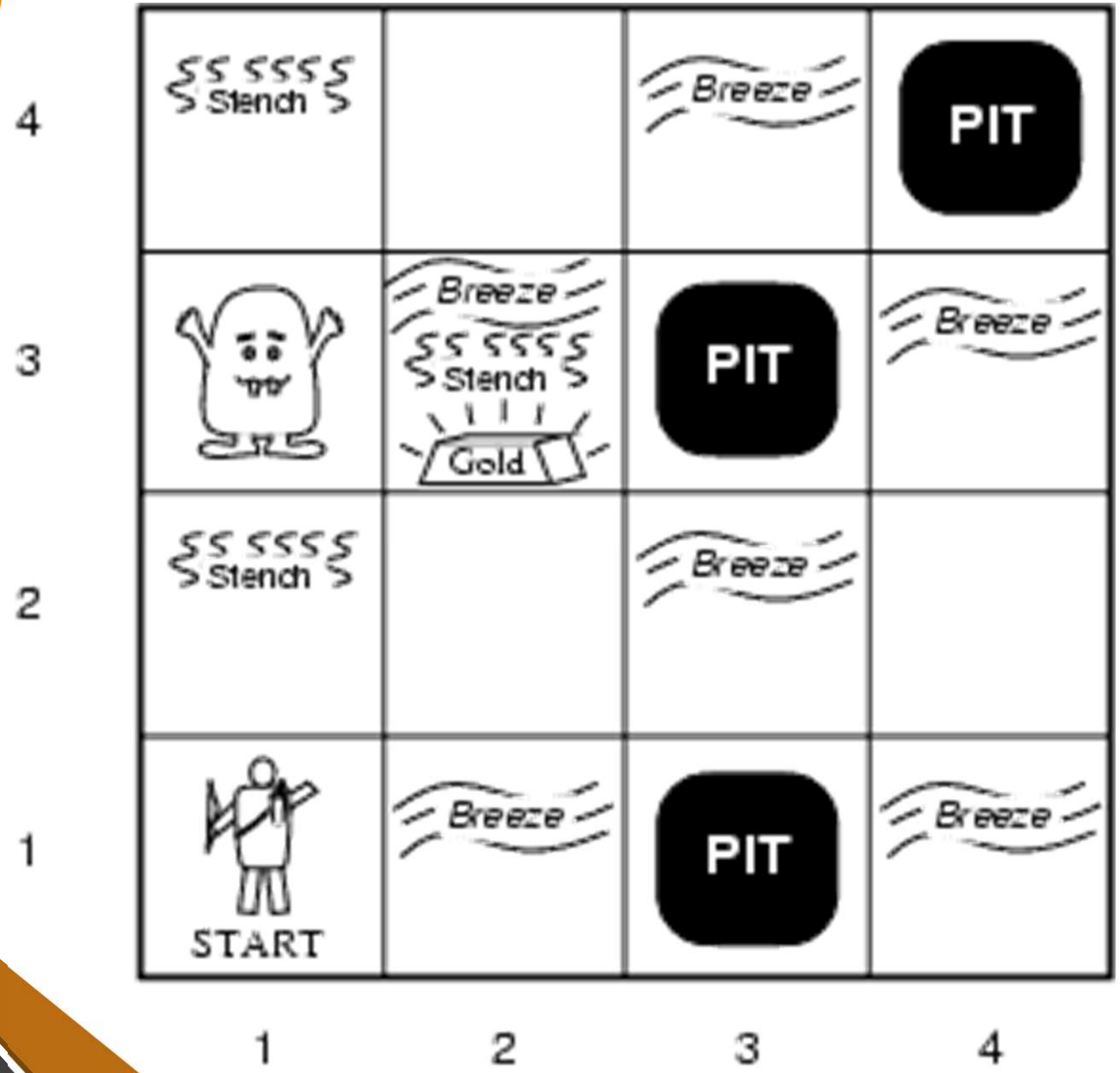
```
function KB-AGENT(percept) returns an action
   persistent: KB, a knowledge base
               t, a counter, initially 0, indicating time

   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 world representations
- Deduce hidden world properties, and deduce actions

Wumpus World



Performance Measure?

Environment?

Actuators?

Sensors?

Performance measure

- gold $+1000$, death -1000
- -1 per action, -10 for using the arrow

Environment

- 4×4 grid of rooms
- agent, wumpus, gold, pits

Actuators

- Turn left/right, Forward
- Shoot: kills wumpus if facing it; uses up the only arrow
- Grab: picks up gold if in same square
- Climb: get out of cave if in [1,1]

Sensors

- Squares adjacent to wumpus are smelly
- Squares adjacent to pit are breezy
- Glitter iff gold is in the same square
- Gets bumped if agent walks into a wall
- Hears scream if wumpus killed

Properties of Wumpus World

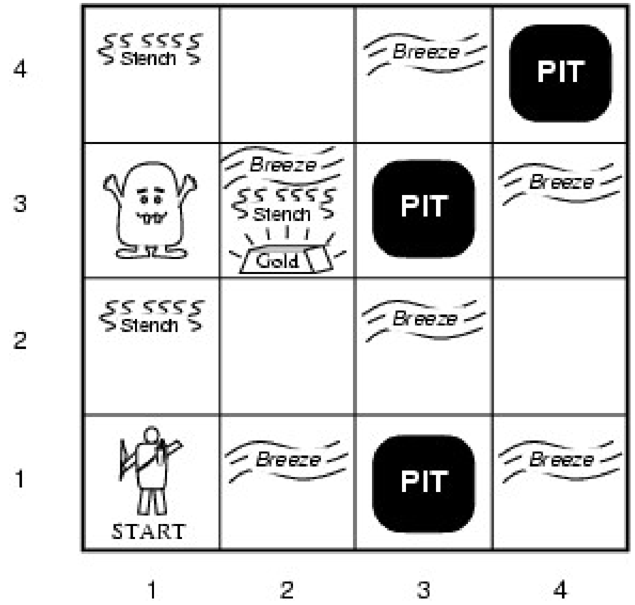
Fully Observable?	No – only local perception
Deterministic?	Yes
Episodic?	No – sequential actions
Static?	Yes – nothing moves
Discrete?	Yes
Single-Agent?	Yes

Exploring a Wumpus World

Agent's view

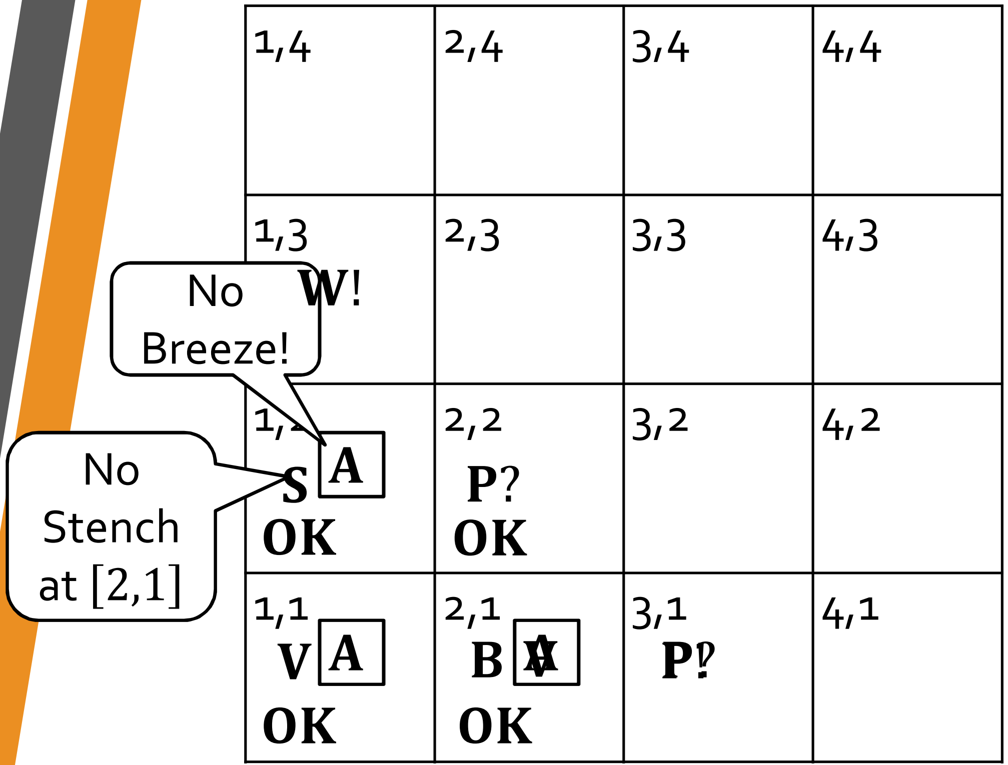
1,4	2,4	3,4	4,4
1,3	2,3	3,3	4,3
1,2 **OK**	2,2 **P?**	3,2	4,2
1,1 **V** **A** **OK**	2,1 **B** **A** **OK**	3,1 **P?**	4,1

- **A** = Agent
- **B** = Breeze
- **G** = Glitter, Gold
- **OK** = Safe Square
- **P** = Pit
- **S** = Stench
- **V** = Visited
- **W** = Wumpus

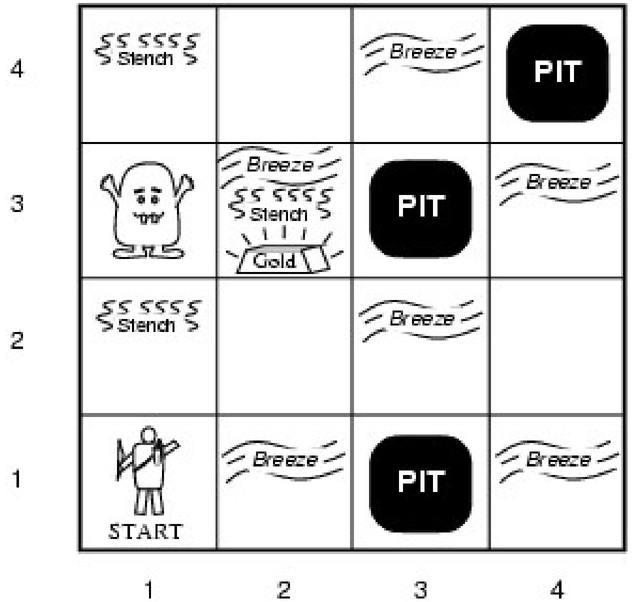


Exploring a Wumpus World

Agent's view



A = Agent
B = Breeze
G = Glitter, Gold
OK = Safe Square
P = Pit
S = Stench
V = Visited
W = Wumpus

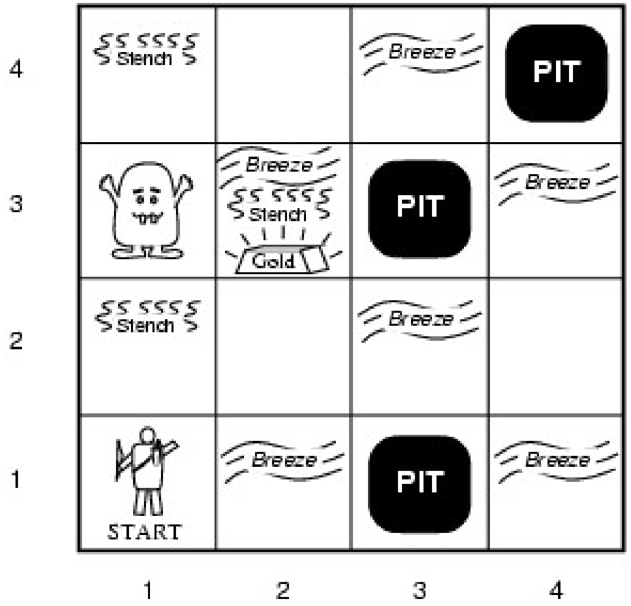


Exploring a Wumpus World

Agent's view

1,4	2,4	3,4	4,4
1,3 W!	2,3 A OK	3,3	4,3
1,2 S V OK	2,2 V OK	3,2 OK	4,2
1,1 V OK	2,1 B V OK	3,1 P!	4,1

A = Agent
B = Breeze
G = Glitter, Gold
OK = Safe Square
P = Pit
S = Stench
V = Visited
W = Wumpus

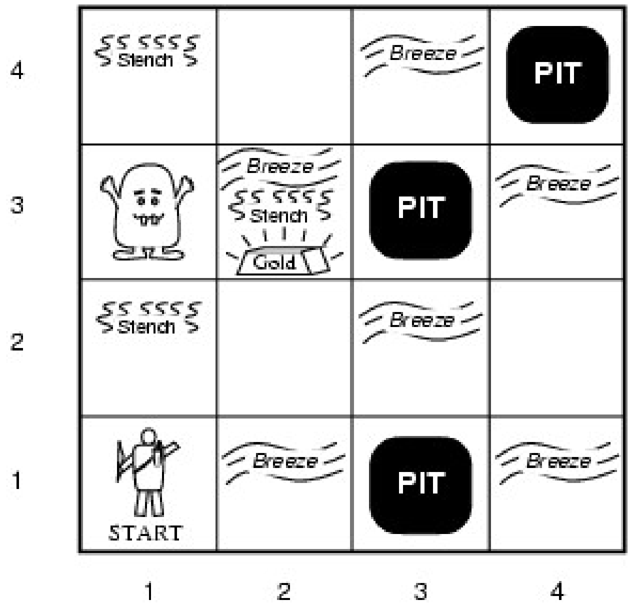


Exploring a Wumpus World

Agent's view

1,4	2,4	3,4	4,4
1,3 **W!**	2,3 $\boxed{A}$ **S B G OK**	3,3	4,3
1,2 **S V OK**	2,2 **V OK**	3,2 **OK**	4,2
1,1 **V OK**	2,1 **B V OK**	3,1 **P!**	4,1

A = Agent
B = Breeze
G = Glitter, Gold
OK = Safe Square
P = Pit
S = Stench
V = Visited
W = Wumpus



Logic in General

- Logic: formal language for KR, infer conclusions
- Syntax: defines the sentences in the language
- Semantics: define the "meaning" of sentences;
  - i.e., define truth of a sentence in a world
- E.g., language of arithmetic
  - $x + 2 \geq y$ is a sentence; $x2y +>$ is not a sentence
  - $x + 2 \geq y$ is true in a world where $x = 7, y = 1$
  - $x + 2 \geq y$ is false in a world where $x = 0, y = 6$

Entailment

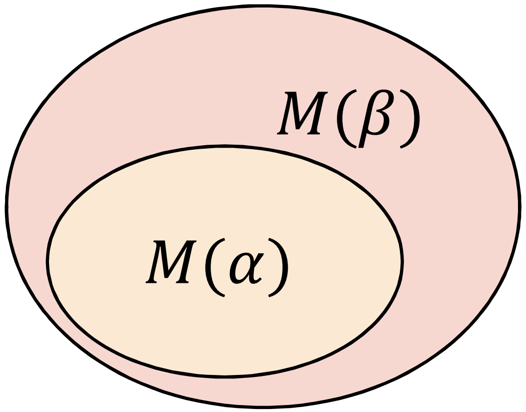
- Modeling: m models α if α is true under m.
  For example, what are models for the following?

$$\alpha = (q \in \mathbb{Z}_+) \wedge (\forall n, m \in \mathbb{Z}_+ : q = nm \Rightarrow n \vee m = 1)$$

- We let $M(\alpha)$ be the set of all models for α
- Entailment means that one thing follows from another:

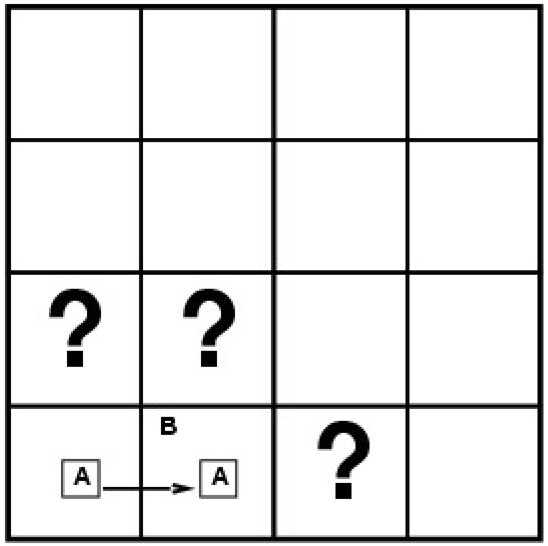
$$\alpha \vDash \beta \text{ or equivalently } M(\alpha) \subseteq M(\beta)$$

- For example:
  $\alpha = (q \text{ is prime})$ entails
  $\beta = (q \text{ is odd}) \vee (q = 2)$.

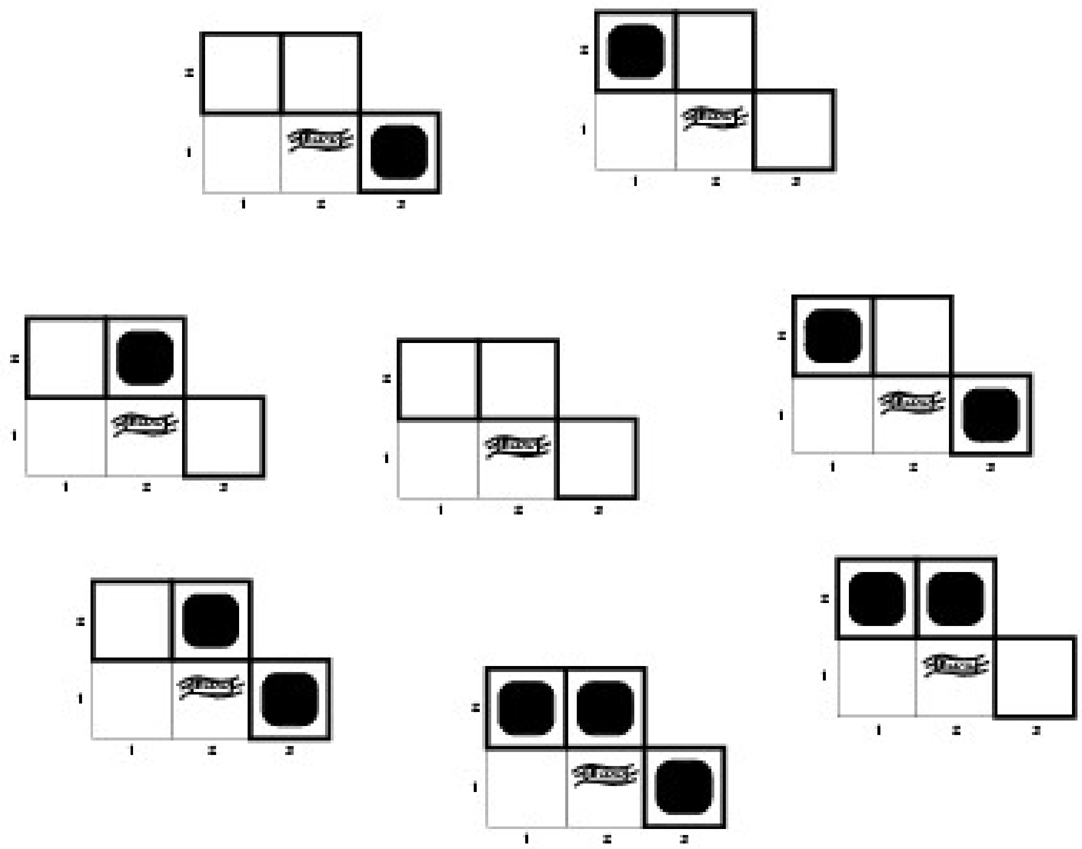


Entailment in the Wumpus World

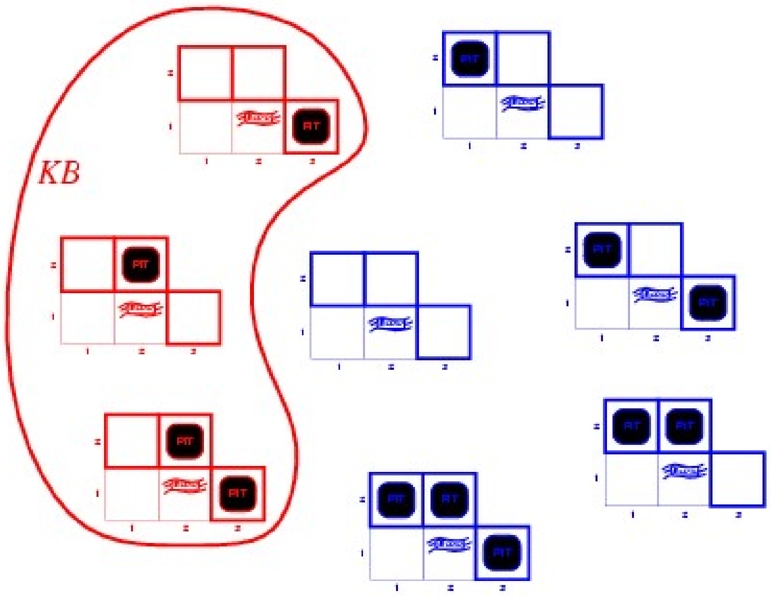
- Situation after detecting nothing in [1,1], moving right, breeze in [2,1]
- Consider possible models for KB assuming only pits
- 3 Boolean choices ⇒ 8 possible models



Wumpus Models

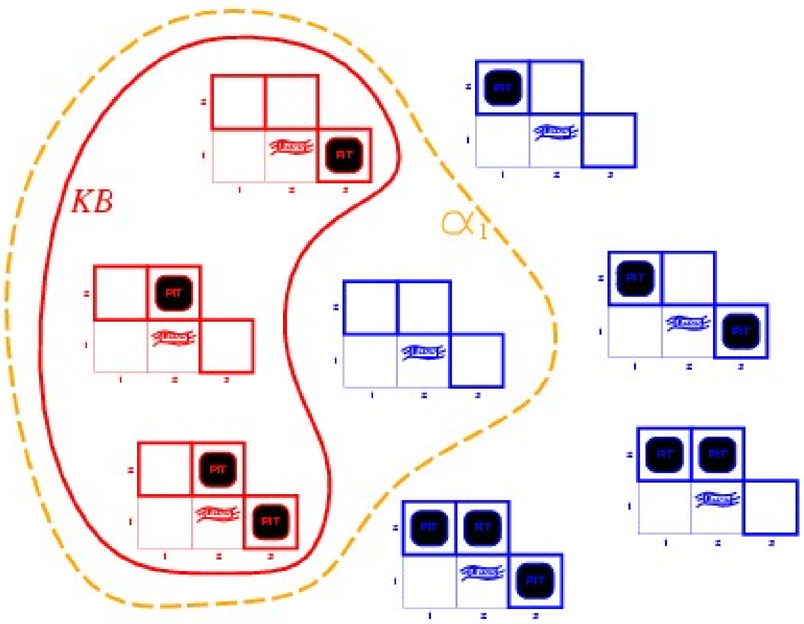


Wumpus Models



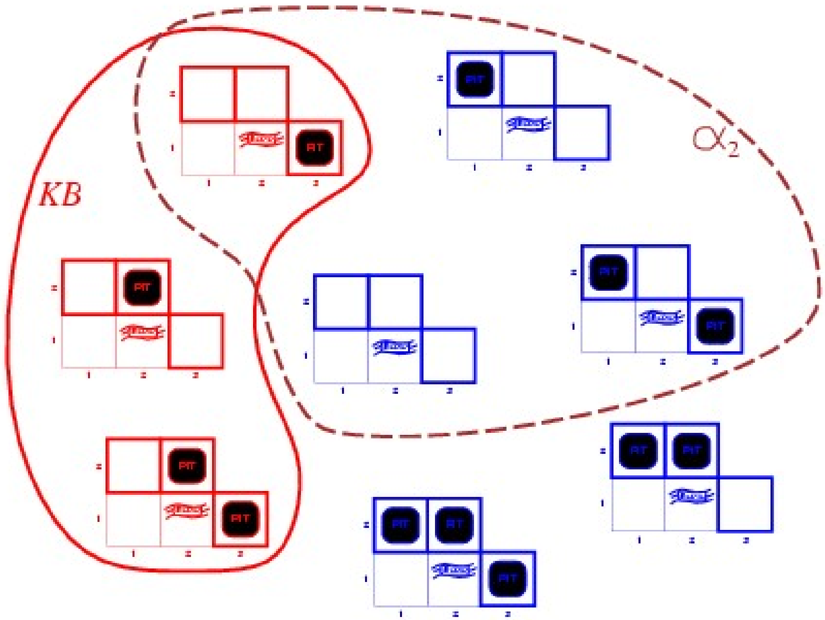
- KB = wumpus-world rules + percepts

Wumpus Models



- KB = wumpus-world rules + percepts
- α_1 = "[1,2] is safe", $KB \vDash \alpha_1$, proved by model checking
- The agent can infer that [1,2] is safe

Wumpus Models



- KB = wumpus-world rules + percepts
- α_2 = “[2,2] is safe”, $KB \not\models \alpha_2$
- The agent cannot infer that [2,2] is safe (or unsafe)!

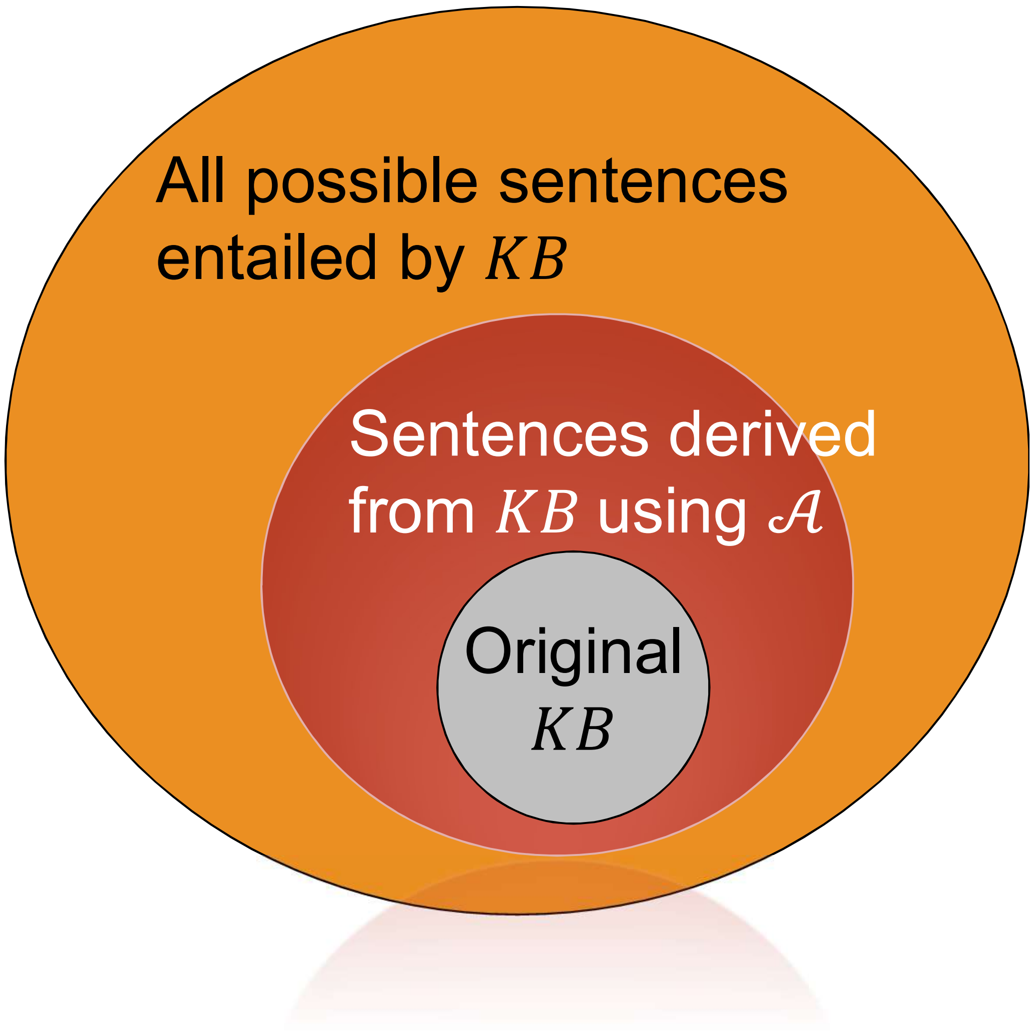
Inference algorithm: is a sentence α is derived from KB?

- Define $KB \vdash_{\mathcal{A}} \alpha$ to be "sentence α is derived from KB by inference algorithm $\mathcal{A}$"
  - $\mathcal{A}$ is **sound** if $KB \vdash_{\mathcal{A}} \alpha$ implies $KB \vDash \alpha$.
    "don't infer nonsense"
  - $\mathcal{A}$ is **complete** if $KB \vDash \alpha$, implies $KB \vdash_{\mathcal{A}} \alpha$.
    "If it's implied, it can be inferred"

Is an inference algorithm complete and sound?

Completeness: $\mathcal{A}$ **is complete if whenever** $KB \vDash \alpha$, **it is also true that** $KB \vdash_{\mathcal{A}} \alpha$

- An incomplete inference algorithm cannot reach all possible conclusions
- Equivalent to completeness in search (chapter 3)



Propositional Logic: Syntax

- A simple logic – illustrates basic ideas
- Defines allowable sentences
- Sentences are represented by symbols e.g. S_1,S_2
- Logical connectives for constructing complex sentences from simpler ones:
  - If S is a sentence, $\neg S$ is a sentence (negation)
  - If S_1 and S_2 are sentences:
    - $S_1 \wedge S_2$ is a sentence (conjunction)
    - $S_1 \vee S_2$ is a sentence (disjunction)
    - $S_1 \Rightarrow S_2$ is a sentence (implication)
    - $S_1 \Leftrightarrow S_2$ is a sentence (biconditional)

Propositional Logic: Semantics

A model is then just a **truth assignment to the basic variables**.

If a model has n variables, how many truth assignments are there?

All other sentences' truth value is derived according to logical rules.

$$x_1 = T; x_2 = F; x_3 = T$$

$$(x_1 \wedge \neg x_2) \Rightarrow \neg\big(x_3 \vee (\neg x_1 \wedge x_2)\big) = ?$$

Knowledge Base for Wumpus World

- $P_{ij} = \text{True} \Leftrightarrow$ there is a pit in $[i,j]$.
- $B_{ij} = \text{True} \Leftrightarrow$ there is breeze in $[i,j]$
- Rules:
  - $R_1: \neg P_{1,1}$
  - $R_4: \neg B_{1,1}$
  - $R_5: P_{2,1}$

KB is true iff $\bigwedge_{k=1,\ldots,5} R_k$ is true

- "Pits cause breezes in adjacent squares"
  - $R_2: B_{1,1} \Leftrightarrow \left(P_{1,2} \vee P_{2,1}\right)$
  - $R_3: B_{2,1} \Leftrightarrow \left(P_{1,1} \vee P_{2,2} \vee P_{3,1}\right)$

Inference

- Given a knowledge base, infer something non-obvious about the world.
- Mimic logical human reasoning
- After exploring 3 squares, we have some understanding of the Wumpus world
- Inference ⇨ Deriving knowledge out of percepts

Given KB and α, we want to know if $KB \vdash \alpha$

Truth Table for Inference

Is α_1 true whenever KB is true?

		$_{,1}$	$P_{1,2}$	$P_{2,1}$	$P_{2,2}$	$P_{3,1}$	KB	α_1
		se	$false$	$false$	$false$	$false$	$false$	$true$
$alse$	$false$	$false$	$false$	$false$	$false$	$true$	$false$	$true$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$false$	$true$	$false$	$false$	$false$	$false$	$false$	$false$	$true$
$false$	$true$	$false$	$false$	$false$	$false$	$true$	$\underline{true}$	$\underline{true}$
$false$	$true$	$false$	$false$	$false$	$true$	$false$	$\underline{true}$	$\underline{true}$
$false$	$true$	$false$	$false$	$false$	$true$	$true$	$\underline{true}$	$\underline{true}$
$false$	$true$	$false$	$false$	$true$	$false$	$false$	$false$	$true$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$true$	$true$	$true$	$true$	$true$	$true$	$true$	$false$	$false$

$R_1: \neg P_{1,1}$

$R_4: \neg B_{1,1}$

$R_5: B_{2,1}$

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

Does KB entail α_1?

Can we infer that [1,2] is safe from pits?

Inference by Truth-Table Enumeration

- Depth-first enumeration of all models is sound and complete
- For n symbols, time complexity is $\mathcal{O}(2^n)$, space complexity is $\mathcal{O}(n)$

```
function TT-ENTAILS?(KB, α) returns true or false
  inputs: KB, the knowledge base, a sentence in propositional logic
          α, the query, a sentence in propositional logic

  symbols ← a list of the proposition symbols in KB and α
  return TT-CHECK-ALL(KB, α, symbols, { })
```

```
function TT-CHECK-ALL(KB, α, symbols, model) returns true or false
  if EMPTY?(symbols) then
     if PL-TRUE?(KB, model) then return PL-TRUE?(α, model)
     else return true // when KB is false, always return true
  else do
     P ← FIRST(symbols)
     rest ← REST(symbols)
     return (TT-CHECK-ALL(KB, α, rest, model ∪ {P = true})
             and
             TT-CHECK-ALL(KB, α, rest, model ∪ {P = false }))
```

Check all possible truth assignments

Validity and Satisfiability

A sentence is valid if it is true in all models,

e.g., $True$, $A \vee \neg A$, $A \Rightarrow A$, $\big(A \wedge (A \Rightarrow B)\big) \Rightarrow B$

Validity is connected to entailment via the Deduction Theorem:

$KB \vDash \alpha$ iff $(KB \Rightarrow \alpha)$ is valid

A sentence is satisfiable if it is true in some model

e.g., $A \vee B$, C

A sentence is unsatisfiable if it is true in no models

e.g., $A \wedge \neg A$

Satisfiability is connected to entailment via the following:

$KB \vDash \alpha$ if and only if $(KB \wedge \neg\alpha)$ is unsatisfiable