Lecture 4b: Knowledge Based Agents

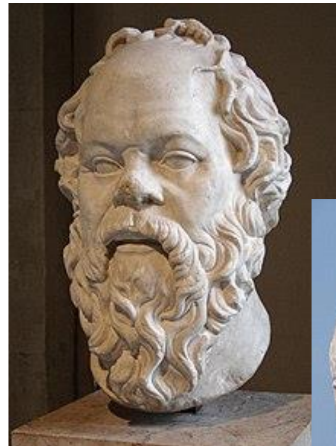
CSCI 360

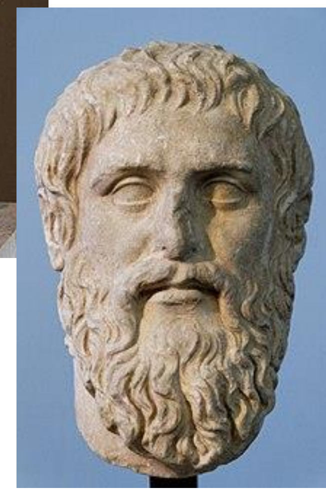
Introduction to Artificial Intelligence

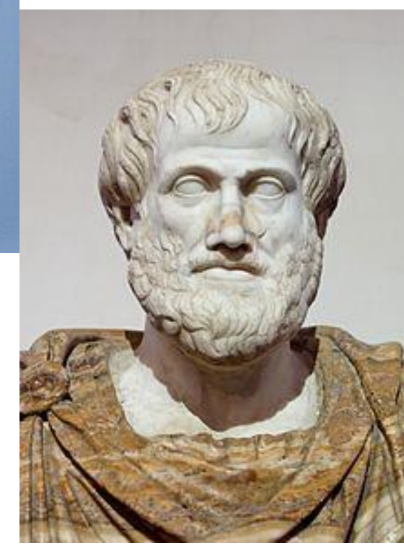
USC

Western philosophy

- **Socrates**







Plato

Aristotle

Western philosophy *(humor)*

- **Socrates** was the greatest of his age, the founding father of Western philosophy, but he didn't publish anything. They killed him.

- His student, **Plato**, published much of what he'd learned from *Socrates*. He became a famous teacher and attracted best students from all over.

- One of his students, **Aristotle**, published even more. So *Alexander the Great* came to sit at his feet to learn to become the most powerful man in the world.

- ***Publish or perish!***

-- by Donald Wunsch

Here is where we are…

Week	30000D	30282R	Topics	Chapters
1	1/7	1/8	Intelligent Agents	[Ch 1.1-1.4 and 2.1-2.4]
	1/9	1/10	Problem Solving and Search	[Ch 3.1-3.3]
2	1/14	1/15	Uninformed Search	[Ch 3.3-3.4]
	1/16	1/17	Heuristic Search (A*)	[Ch 3.5]
3	1/21	1/22	Heuristic Functions	[Ch 3.6]
	1/23	1/24	Local Search	[Ch 4.1-4.2]
	1/25		Project 1 Out	
4	1/28	1/29	Adversarial Search	[Ch 5.1-5.3]
	1/30	1/31	Knowledge Based Agents	[Ch 7.1-7.3]
5	2/4	2/5	Propositional Logic Inference	[Ch 7.4-7.5]
	2/6	2/7	First-Order Logic	[Ch 8.1-8.4]
	2/8		Project 1 Due	
	2/8		Homework 1 Out	
6	2/11	2/12	Rule-Based Systems	[Ch 9.3-9.4]
	2/13	2/14	Search-Based Planning	[Ch 10.1-10.3]
	2/15		Homework 1 Due	
7	2/18	2/19	SAT-Based Planning	[Ch 10.4]
	2/20	2/21	Knowledge Representation	[Ch 12.1-12.5]
8	2/25	2/26	Midterm Review	
	2/27	2/28	**Midterm Exam**	

Outline

- What is AI?
- Problem-solving agent
- Uninformed search
- Informed search (A*)
- Adversarial search
- **Knowledge based agents**

Reasoning

- Agents we've seen so far cannot perform "**reasoning**"
- **Types** of reasoning
  - Deductive reasoning
  - Inductive reasoning
  - Abductive reasoning

Reasoning

- Agents we've seen so far cannot perform "**reasoning**"
- **Types** of reasoning
  - Deductive reasoning
    - Facts + rules ➔ **more facts**
    - *"Socrates is a human" + "All humans are mortal" → "**Socrates is mortal**"*
  - Inductive reasoning
    - Facts + **rules** ➔ more facts
    - *"X/Y/Z is a human" + "X/Y/Z is mortal" → "**All humans must be mortal**"*
  - Abductive reasoning
    - **Facts** + rules ➔ more facts
    - *"All humans are mortal" + "Socrates is a mortal" → "**Socrates may be a human**"*

How to do reasoning?

- Need **representations of knowledge** (either **prior** or **acquired** knowledge)

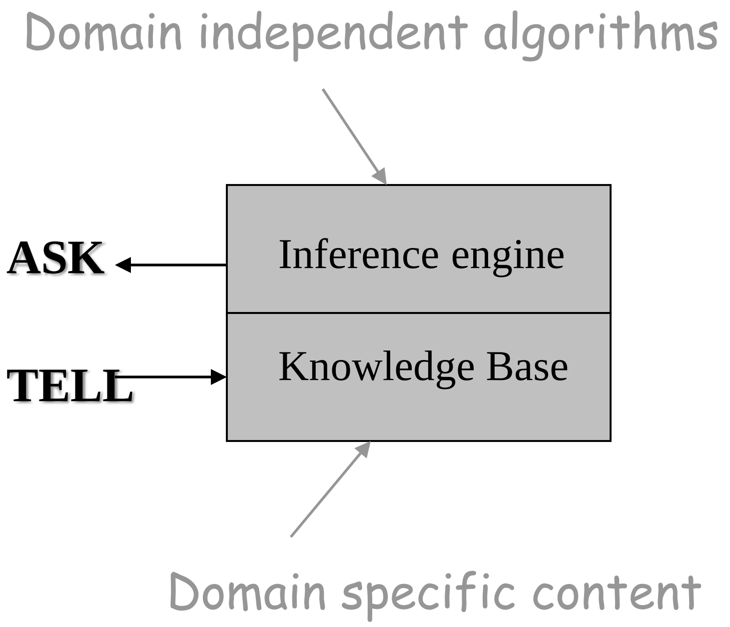
- In "problem-solving" agents, knowledge is baked into the problem statement

  – S ← RESULT(S, A)

  – **Problem**: Limited and inflexible, has to **anticipate all facts** that an agent might need, and bake them into the problem statement

Knowledge representation

- **Logic** as language for knowledge representation
  - Propositional logic (Boolean)
  - First-order logic (FOL)



- Advantage
  - Can combine and recombine information to suit many purposes

Knowledge-based agent

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*))
  $t \leftarrow t + 1$
  return *action*

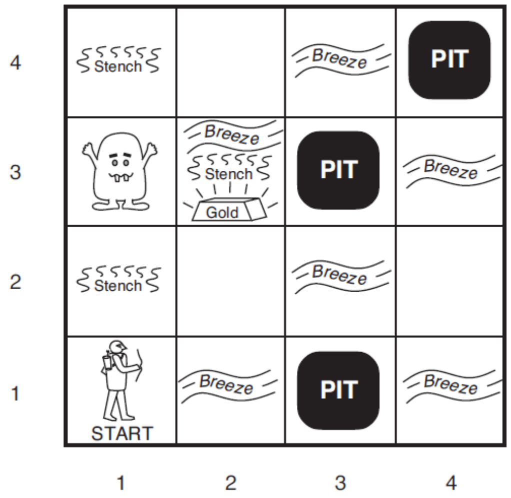
1. TELL KB what was perceived
   - Insert new sentences (representations of facts) into KB
2. ASK KB what to do
   - Use reasoning to examine actions and select the best

Outline for today

- Knowledge Based Agents
- **The Wumpus World**
- Logic
- Propositional Logic

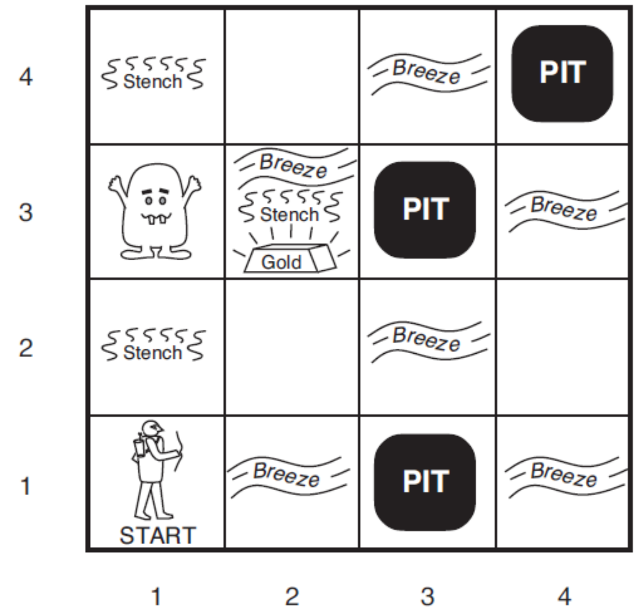
Wumpus world (game)

- Illustrating unique strength of “knowledge-based” agents
  - A cave consisting of dark rooms, on a 4x4 grid
    - Agent can move Forward, Turn Left, or Turn Right
    - Moving Forward into a wall does not change location
  - Beast (named Wumpus) hidden in one room
    - Agent will be eaten if walks into that room
    - Wumpus can be shot by the agent, but the agent has only one arrow
  - Pits hidden in some rooms
    - Agent will die if walks into these rooms



Wumpus world: *sensors*

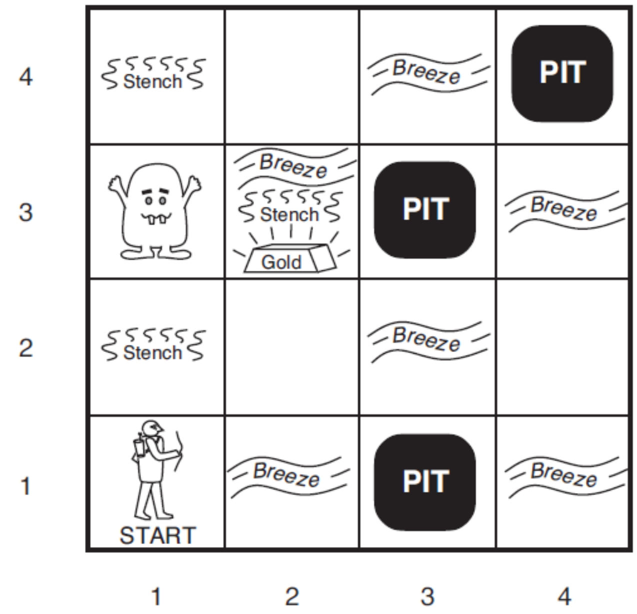
- Sensors:
  - **Stench**: *squares directly (not diagonally) adjacent to Wumpus are smelly*
  - **Breeze**: *squares directly adjacent to pit are breezy*
  - **Glitter**: *in the square where the gold is*
  - **Bump**: *when the agent walks into a wall*
  - **Scream**: *when Wumpus is shot by an arrow, it screams and dies*



(Stench, Breeze, Glitter, Bump, Scream)

Wumpus world: *performance measure*

- Performance measure:
  - **+1000**: *for coming out of the cave with gold*
  - **-1000**: *for dying in a pit or being eaten*
  - **-10**: *for using up the (one and only) arrow*
  - **-1**: *for each action taken*
    - Forward, TurnLeft, TurnRight, Climb
    - Grab,
    - Shoot



Characterization of Wumpus World

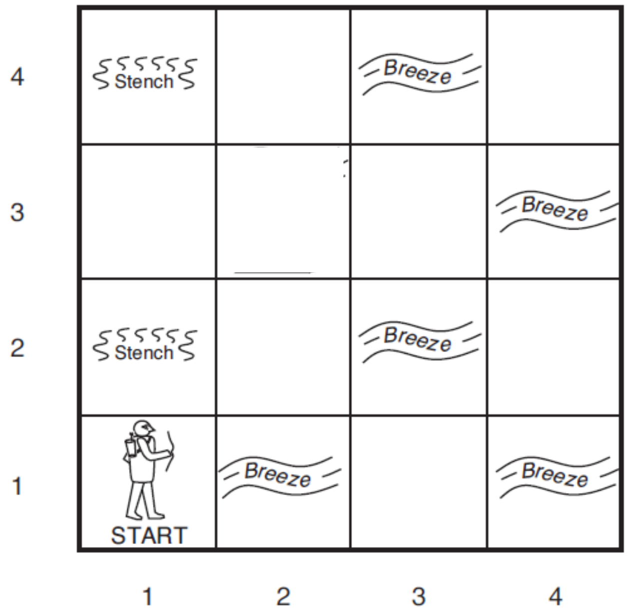
- Deterministic?
- Accessible?
- Static?
- Discrete?
- Episodic?

Characterization of Wumpus World

- Deterministic? Yes – outcome exactly specified
- Accessible? No – only local perception
- Static? Yes – Wumpus and pits do not move
- Discrete? Yes
- Episodic? Yes

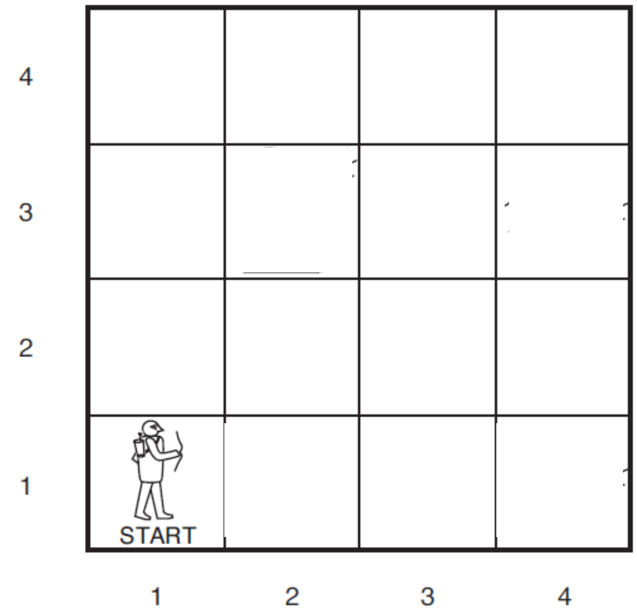
Wumpus world: *how would you play?*

- Recall that "**you don't know what's hidden in each room**"



Wumpus world: *how would you play?*

- Recall that "**you don't know what's hidden in each room**"
- Recall that "**you also don't have percepts until walking into a room**"

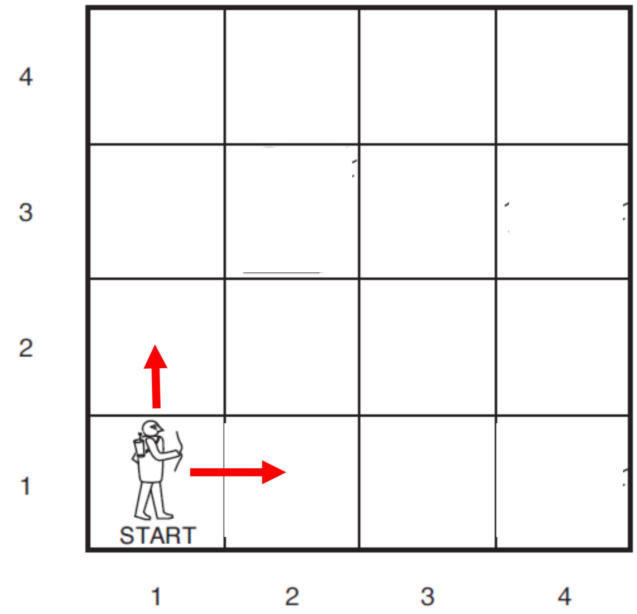


Wumpus world: *how would you play?*

- **What do you know now?**

- Recall that
  - **Stench**: *in the squares directly (not diagonally) adjacent to Wumpus, agent perceives Stench*
  - **Breeze**: *in the squares directly adjacent to a pit, agent perceives Breeze*

- Neither "Stench" nor "Breeze" in [1,1]
  - **Wumpus cannot be in (1,2) or (2,1)**
  - **Pit cannot be in (1,2) or (2,1)**



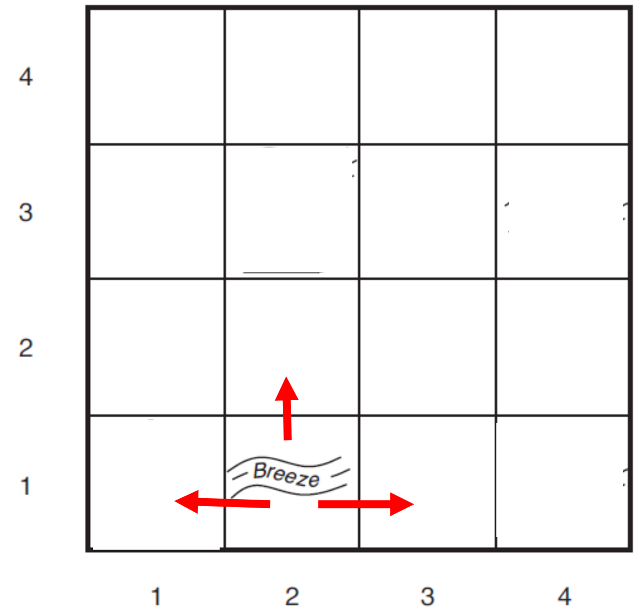
Wumpus world: *how would you play?*

- **What do you know now?**

- Recall that
  - **Breeze**: *in the squares directly adjacent to a pit, the agent perceives a Breeze*

  - Pit may be in (2,2)
  - Pit may be in (3,1)

- *Agent's only choice is going back to (1,1)*



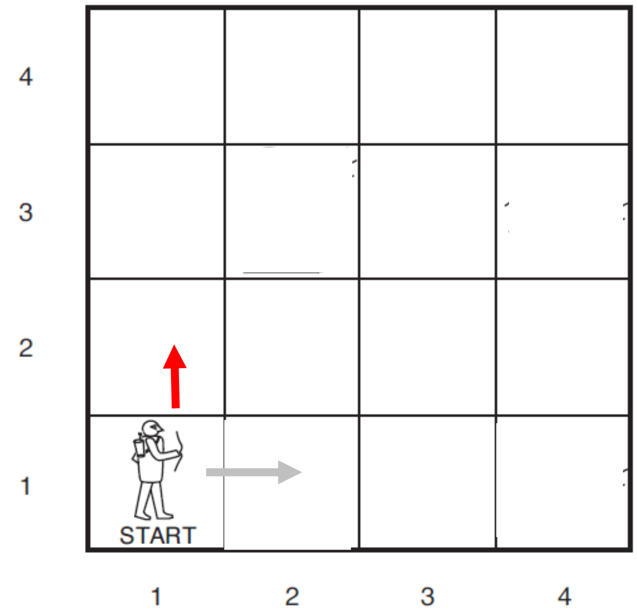
Wumpus world: *how would you play?*

- **What do you know now?**

- Recall that
  - **Breeze**: *in the squares directly adjacent to a pit, the agent perceives a Breeze*

  - Pit may be in (2,2)
  - Pit may be in (3,1)

- *Agent's only choice is going to (1,2)*

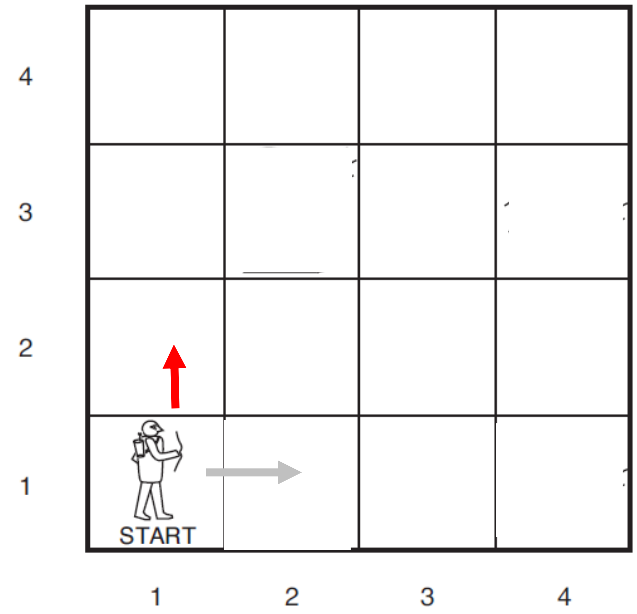


Wumpus world: *how would you play?*

- **What do you know now?**

- Recall that
  - **Breeze**: *in the squares directly adjacent to a pit, the agent perceives a Breeze*

  - Pit may be in (2,2)
  - Pit may be in (3,1)
  - Pit cannot be in (1,2)
  - Pit cannot be in (2,1)



- *Agent may go to (1,2)*

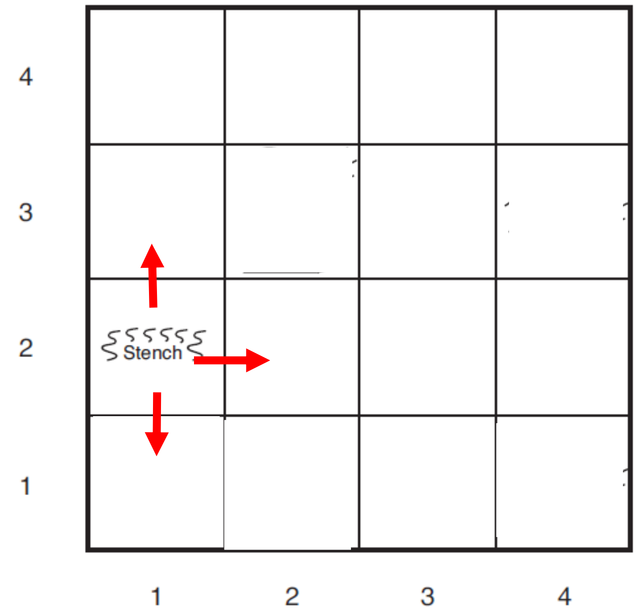
Wumpus world: *how would you play?*

- **What do you know now?**

- Recall that
  - **Stench**: *in the square containing Wumpus, and in the directly (not diagonally) adjacent squares, the agent perceives a Stench*

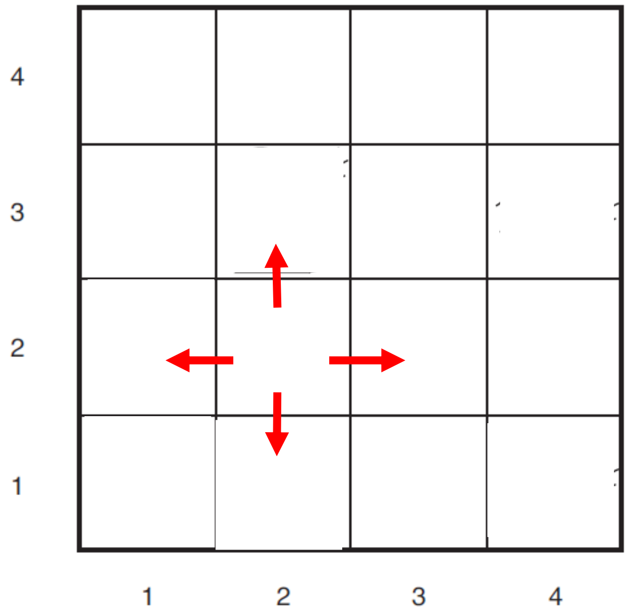
  - ~~Pit may be in (2,2)~~
  - Pit may be in (3,1)
  - Pit cannot be in (1,2)
  - Pit cannot be in (2,1)
  - Pit cannot be in (2,2)
  - Wumpus cannot be in (2,2)
  - Wumpus MUST be in (1,3)

- *Agent may go to (2,2)*



Wumpus world: *how would you play?*

- **What do you know now?**
- Recall that
  - ~~Pit may be in (2,2)~~
  - Pit may be in (3,1)
  - Pit cannot be in (1,2)
  - Pit cannot be in (2,1)
  - Pit cannot be in (2,2)
  - Wumpus cannot be in (2,2)
  - Wumpus Must be in (1,3)
  - Pit must be in (3,1)
  - Pit cannot be in (2,3)



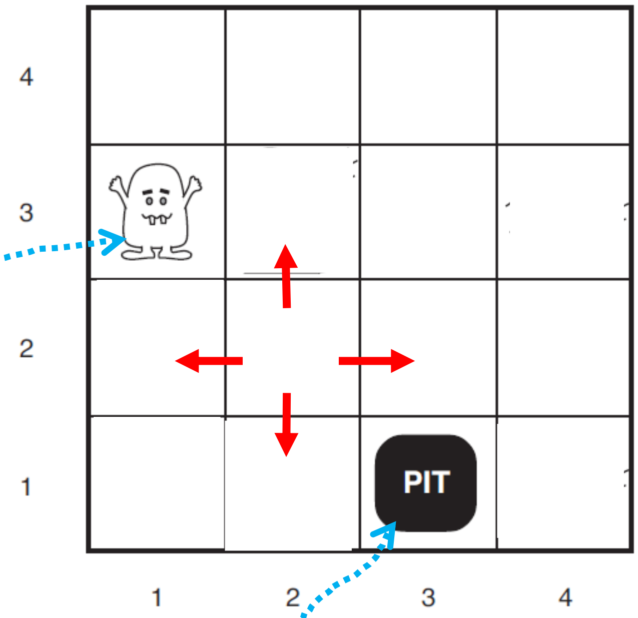
- ***Agent may go to (2,3)*** *or (3,2) or (2,1) or (1,2)*

Wumpus world: *how would you play?*

- **What do you know now?**

- Recall that

  - ~~Pit may be in (2,2)~~
  - Pit may be in (3,1)
  - Pit cannot be in (1,2)
  - Pit cannot be in (2,1)
  - Pit cannot be in (2,2)
  - **Wumpus Must be in (1,3)**
  - Wumpus cannot be in (2,2)
  - **Pit must be in (3,1)**
  - Pit cannot be in (2,3)



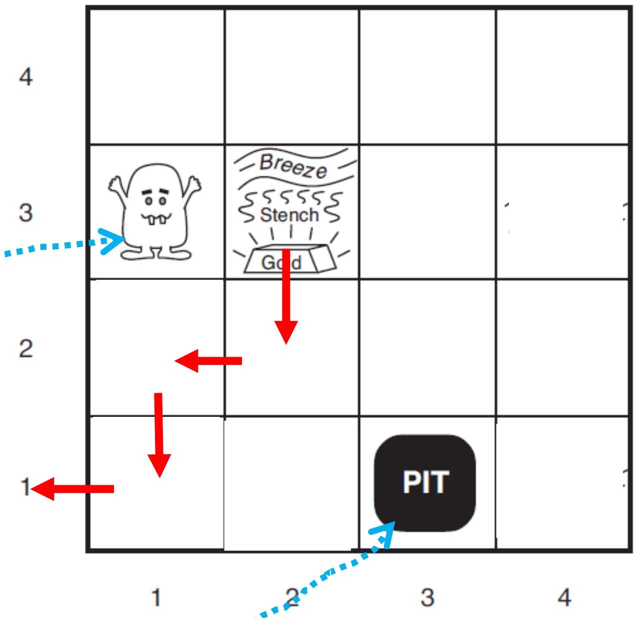
- ***Agent may go to (2,3)*** *or (3,2) or (2,1) or (1,2)*

Wumpus world: *how would you play?*

- **What do you know now?**

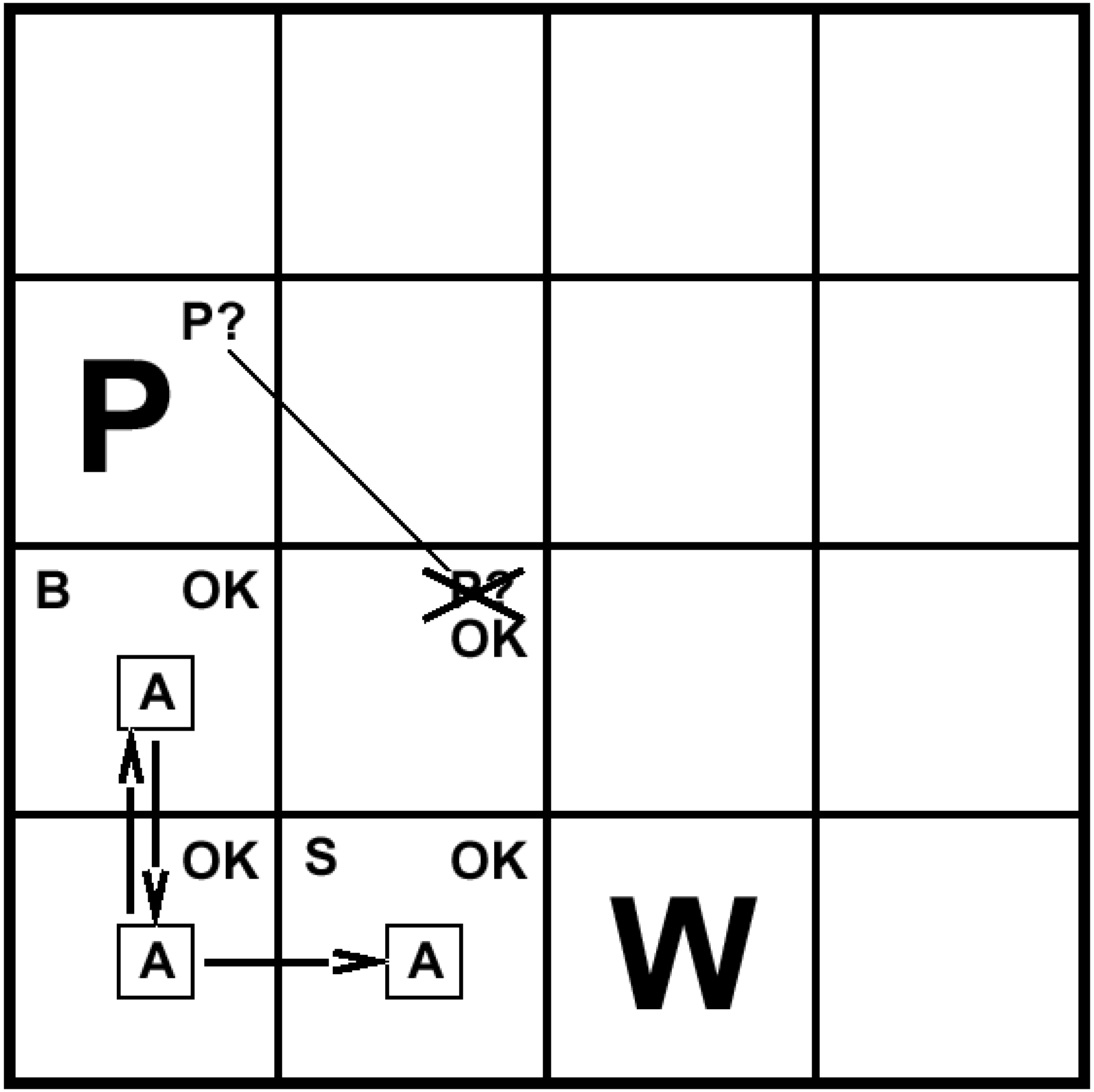
- Recall that

  - Pit may be in (3,1)
  - Pit cannot be in (1,2)
  - Pit cannot be in (2,1)
  - Pit cannot be in (2,2)
  - Wumpus Must be in (1,3)
  - Wumpus cannot be in (2,2)
  - Pit must be in (3,1)
  - Pit cannot be in (2,3)



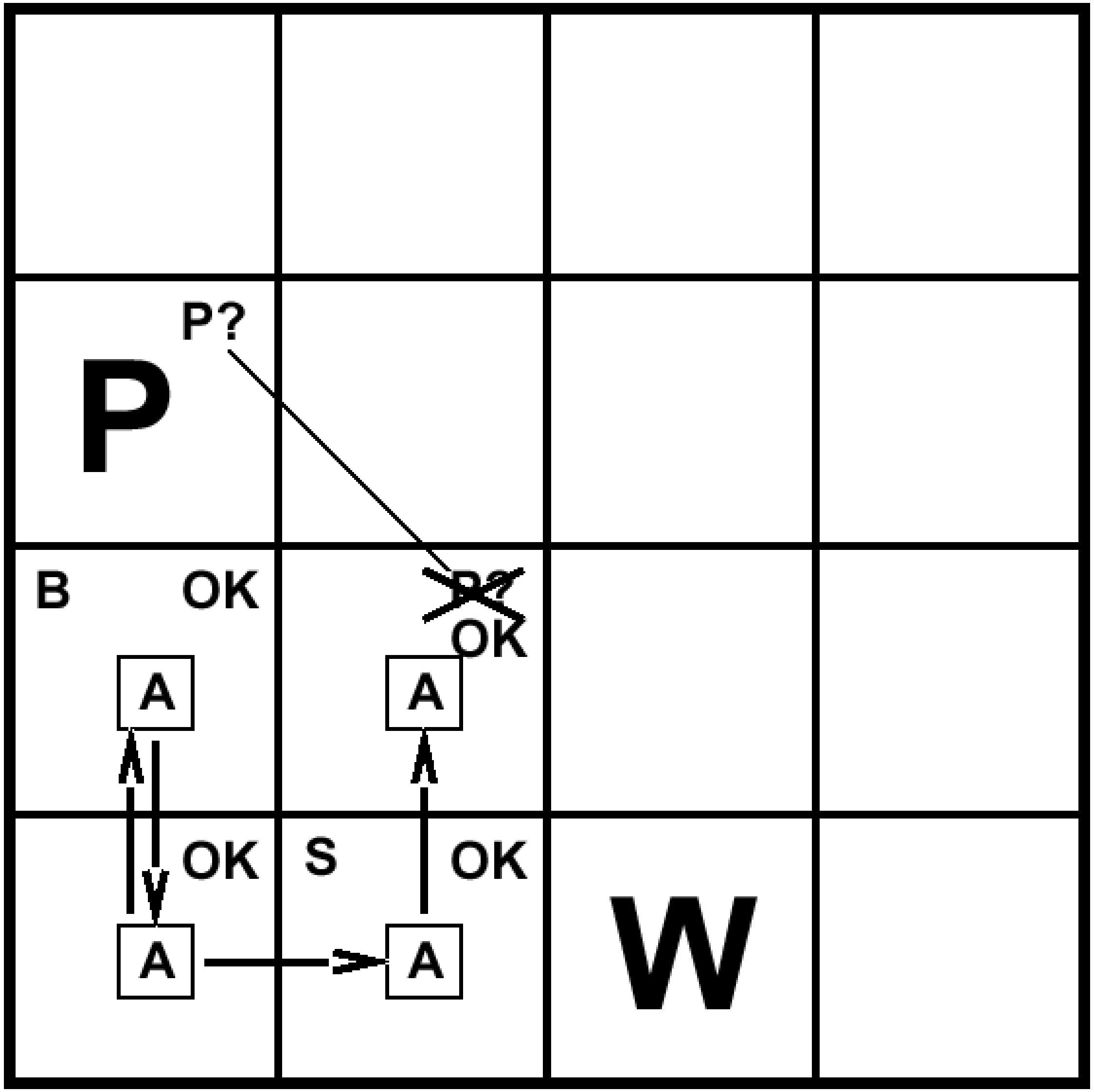
- *Agent finds gold, and backtracks to (1,1)*

Exploring*... learning a transition model*



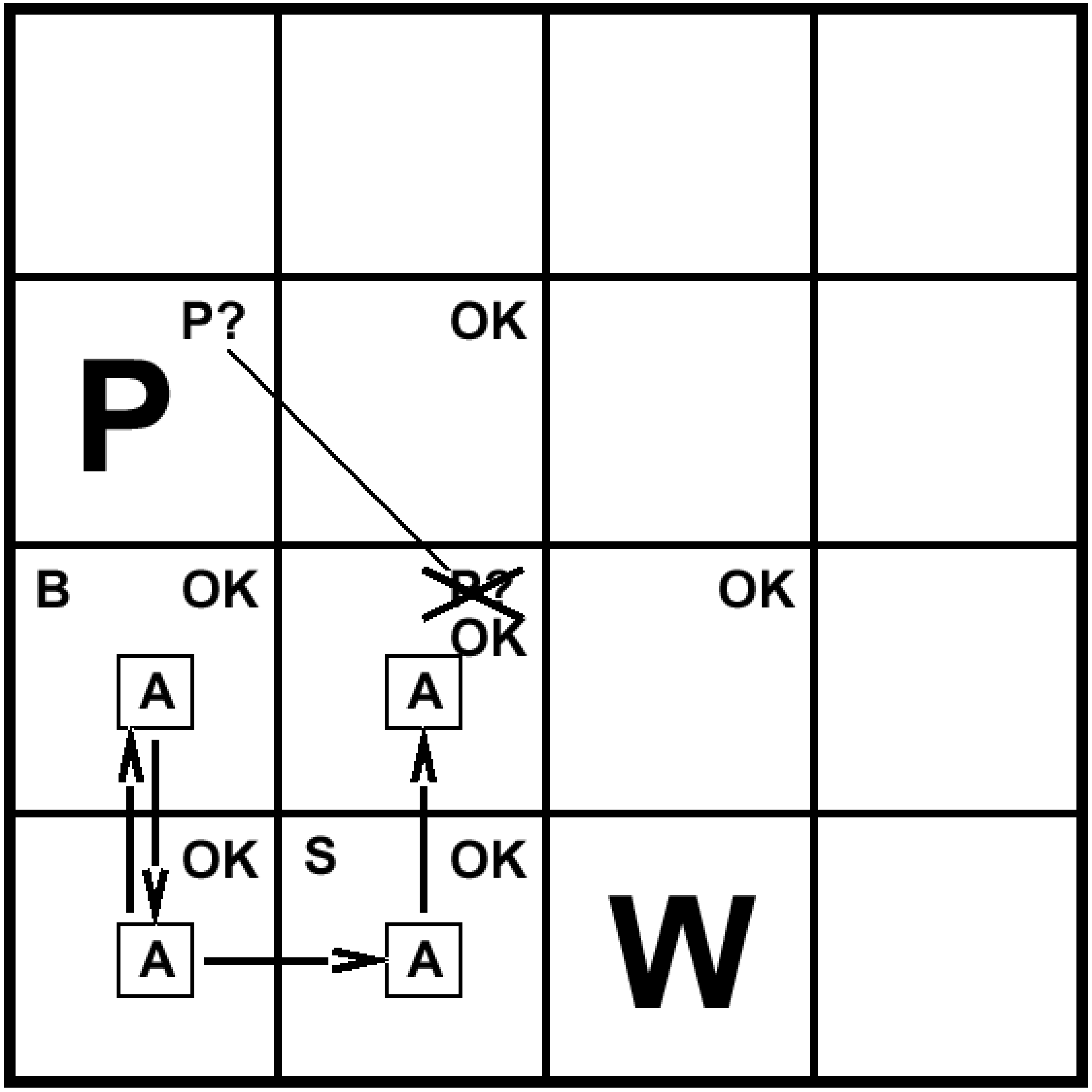
A= Agent
B= Breeze
S= Smell
P= Pit
W= Wumpus
OK = Safe
V = Visited
G = Glitter

Exploring*... learning a transition model*



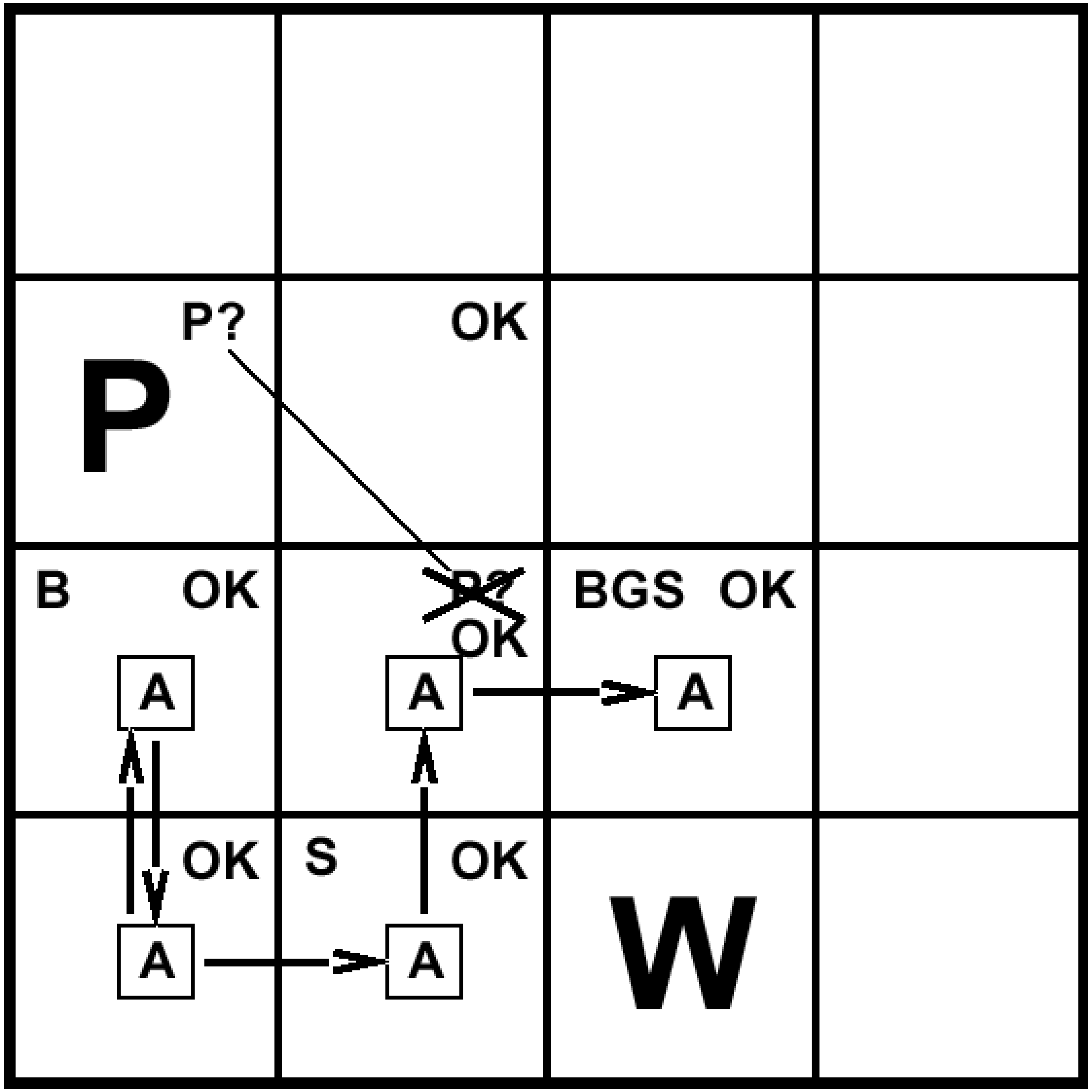
A= Agent
B= Breeze
S= Smell
P= Pit
W= Wumpus
OK = Safe
V = Visited
G = Glitter

Exploring*... learning a transition model*



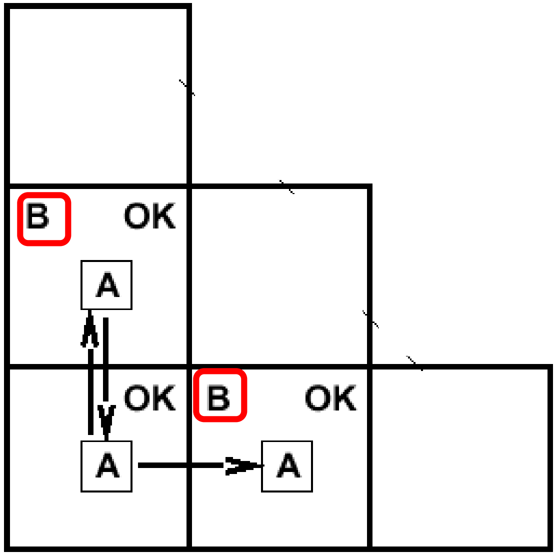
A= Agent
B= Breeze
S= Smell
P= Pit
W= Wumpus
OK = Safe
V = Visited
G = Glitter

Exploring... *learning a transition model*



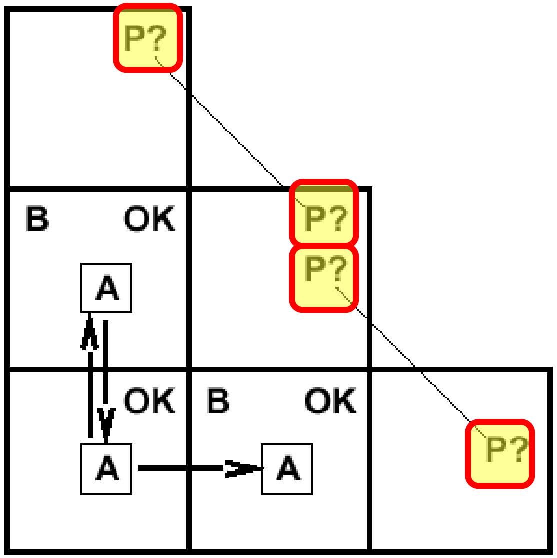
A= Agent
B= Breeze
S= Smell
P= Pit
W= Wumpus
OK = Safe
V = Visited
G = Glitter

Other tight spots



Breeze in (1,2) and (2,1)

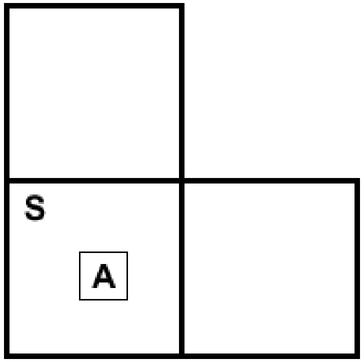
Other tight spots



Breeze in (1,2) and (2,1)
$\Rightarrow$ no safe actions

Assuming pits uniformly distributed, (2,2) is most likely to have a pit

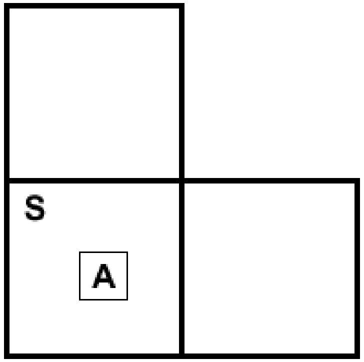
Other tight spots



Smell in (1,1)

$\Rightarrow$ cannot move

Other tight spots



Smell in (1,1)

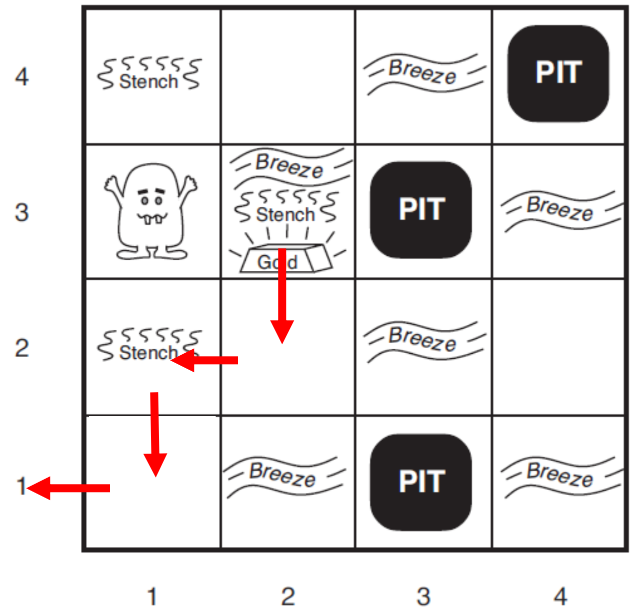
$\Rightarrow$ cannot move

Can use a strategy of <u>coercion</u>:

shoot straight ahead
wumpus was there $\Rightarrow$ dead $\Rightarrow$ safe
wumpus wasn't there $\Rightarrow$ safe

Question: Can you play the game ***using search*** alone?

- **No,** unless you want to risk "being eaten" or "dying in a pit" **multiple times**, before learning the **transition model** of the environment

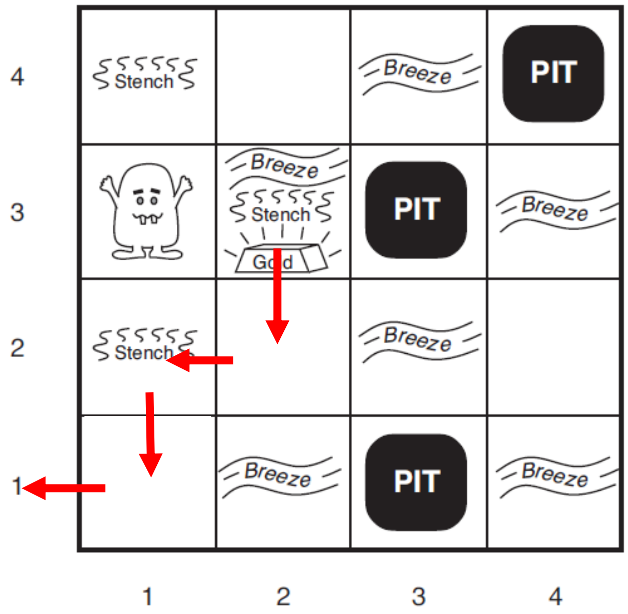


Question: Can you play the game ***using search*** alone?

- **No,** unless you want to risk "being eaten" or "dying in a pit" **multiple times**, before learning the **transition model** of the environment

- With a "**Knowledge Base (KB)**", the agent can infer facts such as
  - [2,2] cannot have Pit
  - [2,2] cannot have Wumpus
  - [1,3] must have Wumpus
  - [3,1] must have Pit

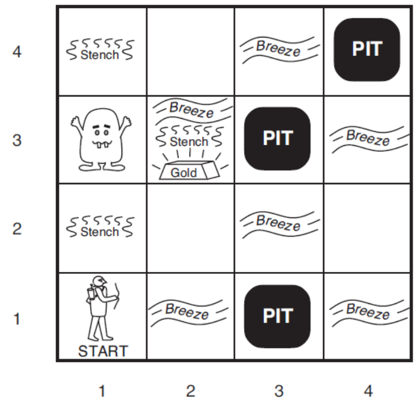
- ***Correctness is guaranteed***
  - As long as KB is correct



What’s a Knowledge Base (KB)

- Consists of a set of **sentences**, each about something of the environment that the agent knew

- **Symbols:**



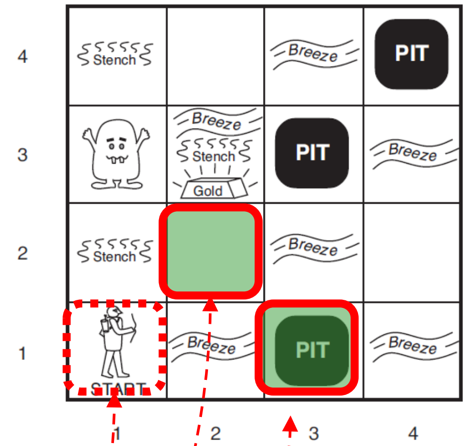
$P_{x,y}$ is true if there is a pit in $[x, y]$.
$W_{x,y}$ is true if there is a wumpus in $[x, y]$, dead or alive.
$B_{x,y}$ is true if the agent perceives a breeze in $[x, y]$.
$S_{x,y}$ is true if the agent perceives a stench in $[x, y]$.

- **Sentence#1:** There is no pit in [1,1] $\boxed{\neg P_{1,1}\ .}$
- **Sentence#2:** There is breeze in [2,1] $\boxed{B_{2,1}\ .}$
- **Sentence#3:** A square is breezy IFF pit is in a neighboring square

$$B_{1,1} \Leftrightarrow (P_{1,2} \lor P_{2,1})\ .$$
$$B_{2,1} \Leftrightarrow (P_{1,1} \lor P_{2,2} \lor P_{3,1})\ .$$
$$B_{3,1} \ldots$$
$$B_{4,1} \ldots$$

What's a Knowledge Base (KB)

- Consists of a set of **sentences**, each about something of the environment that the agent knew



- **Symbols:**

$P_{x,y}$ is true if there is a pit in $[x, y]$.
$W_{x,y}$ is true if there is a wumpus in $[x, y]$, dead or alive.
$B_{x,y}$ is true if the agent perceives a breeze in $[x, y]$.
$S_{x,y}$ is true if the agent perceives a stench in $[x, y]$.

- **Sentence#1:** There is no pit in [1,1] $\quad \boxed{\neg P_{1,1}\,.}$
- **Sentence#2:** There is breeze in [2,1] $\quad \boxed{B_{2,1}\,.}$
- **Sentence#3:** A square is breezy IFF pit is in a neighboring square

$$B_{1,1} \;\Leftrightarrow\; (P_{1,2} \lor P_{2,1})\,.$$
$$\boxed{B_{2,1} \;\Leftrightarrow\; (P_{1,1} \lor P_{2,2} \lor P_{3,1})\,.}$$
$$B_{3,1} \ldots$$
$$B_{4,1} \ldots$$

Outline for today

- Knowledge Based Agents
- The Wumpus World
- **Logic**
- Propositional Logic

Syntax vs. Semantics *(of ordinary arithmetic)*

- **Syntax** specifies all **well formed** sentences (in KB)
  - $(x + y = 4)$ is a **well-formed** sentence
  - $(x\ 4\ y + =)$ is **not** a well-formed sentence

Syntax vs. Semantics *(of ordinary arithmetic)*

- **Syntax** specifies all **well formed** sentences (in KB)
  - ***(x + y = 4)*** is a **well-formed** sentence
  - ***(x 4 y + =)*** is **not** a well-formed sentence

- **Semantics** defines the **meaning** of the sentences
  - ***(x + y = 4) is true*** in worlds where
    - …
    - $x = 0, y = 4$
    - $x = 1, y = 3$
    - $x = 2, y = 2$
    - $x = 3, y = 1$
    - $x = 4, y = 0$
    - …

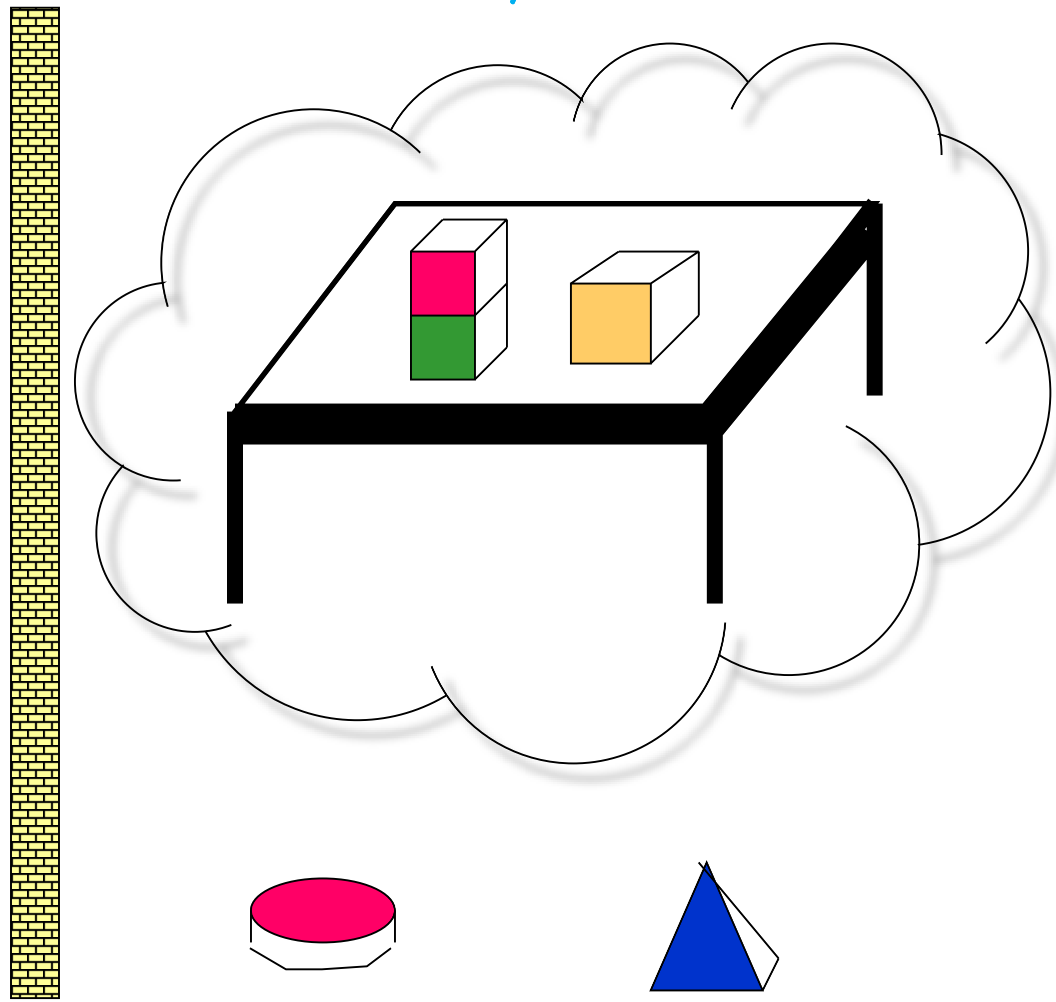
Sentence (α)

Models of α

The Semantic Wall

Symbol System

Physical World



+BLOCKA+

+BLOCKB+

+BLOCKC+

P_1: *(IS_ON +BLOCKA+ +BLOCKB+)*
P_2: *(IS_RED +BLOCKA+)*

Truth depends on Interpretation

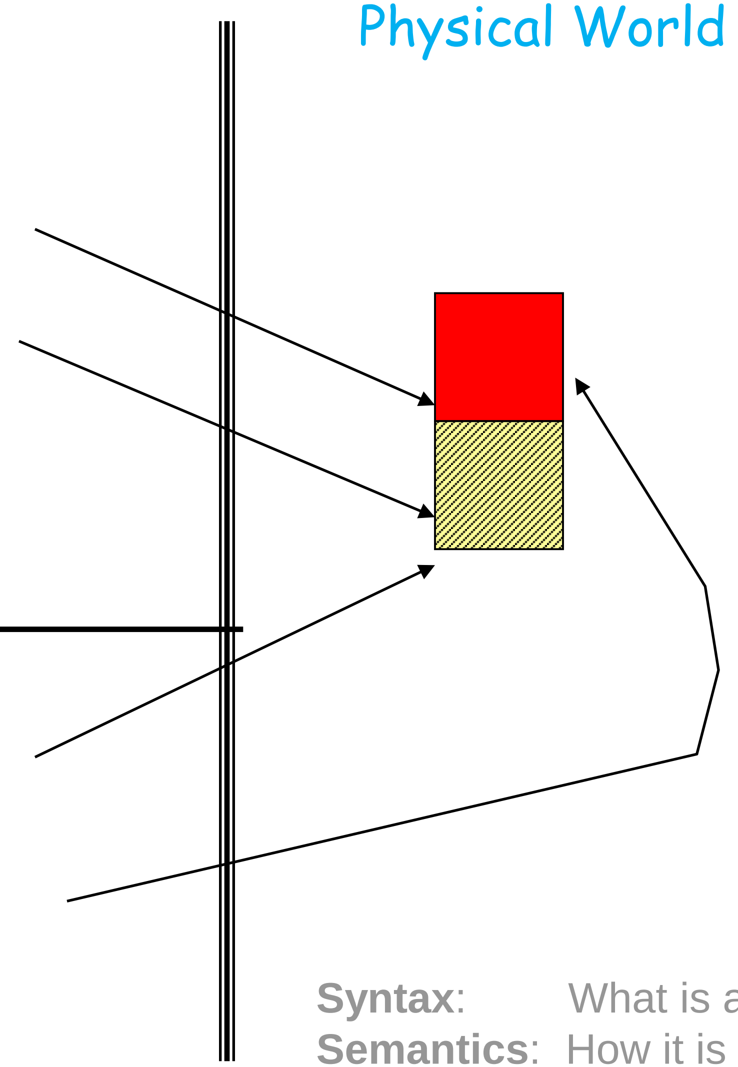
Representation 1

A

B

ON(A,B) T
ON(B,A) F

Physical World



Representation 2

ON(A,B) F   A

ON(B,A) T   B

Syntax: What is allowed on the LHS
Semantics: How it is related to the RHS

Entailment

$\boxed{\alpha_1 = \text{“There is no pit in [1,2].”}}$

in every model in which KB is true, α_1 is also true.

$$\boxed{KB \models \alpha_1}$$

Based on the **facts put into KB so far**, can we make this claim, with certainty?

Entailment *(another example)*

- ***Let*** α **=** ***(x > 10)***
- ***Let*** β **=** ***(x > 0)***

$$\boxed{\alpha \models \beta}$$

- ***α*** entails ***β***
- ***β*** follows logically from ***α***

- In every model in which ***α*** is satisfied, ***β*** is also satisfied

Entailment *(another example: the models…)*

- ***Let α = (x > 10)***
- ***Let β = (x > 0)***

$$\boxed{\alpha \models \beta}$$

- ***α*** entails ***β***
- ***β*** follows logically from ***α***

- In every model in which ***α*** is satisfied, ***β*** is also satisfied

$$\boxed{\alpha \models \beta \text{ if and only if } M(\alpha) \subseteq M(\beta)}$$

Entailment *(another example: Venn diagram…)*

- ***Let*** α **= *(x > 10)***
- ***Let*** β **= *(x > 0)***

$$\boxed{\alpha \models \beta}$$

- ***α*** entails ***β***
- ***β*** follows logically from ***α***

- In every model in which ***α*** is satisfied, ***β*** is also satisfied

Entailment *(another example: algebra…)*

- ***Let*** α **=** ***(x > 10)***
- ***Let*** β **=** ***(x > 0)***

$$\boxed{\alpha \models \beta}$$

- ***α*** entails ***β***
- ***β*** follows logically from ***α***

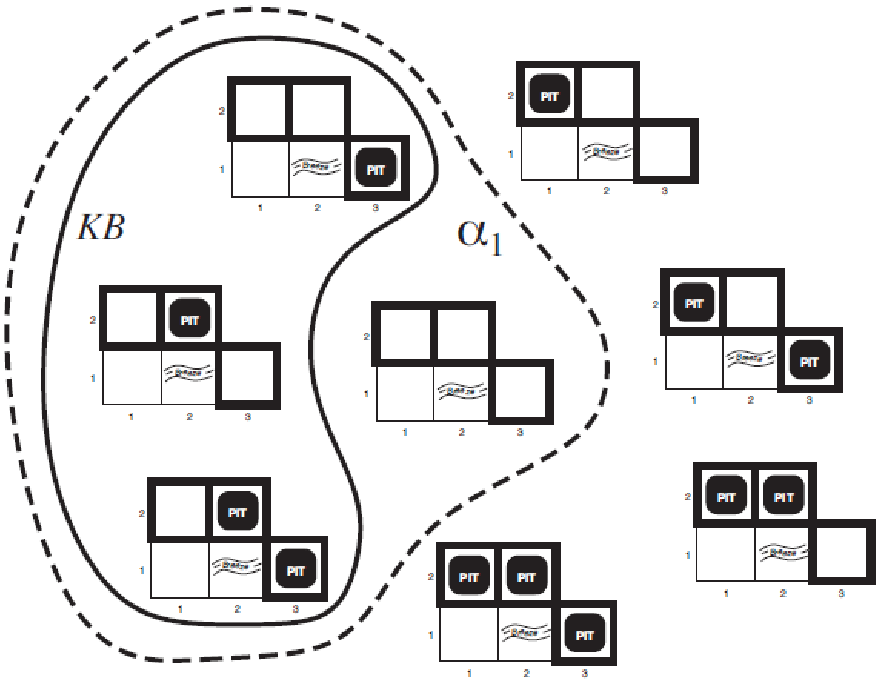
- In every model in which ***α*** is satisfied, ***β*** is also satisfied

Entailment

α_1 = "There is no pit in [1,2]."

in every model in which KB is true, α_1 is also true.

$$\boxed{KB \models \alpha_1}$$



First, find all models in KB

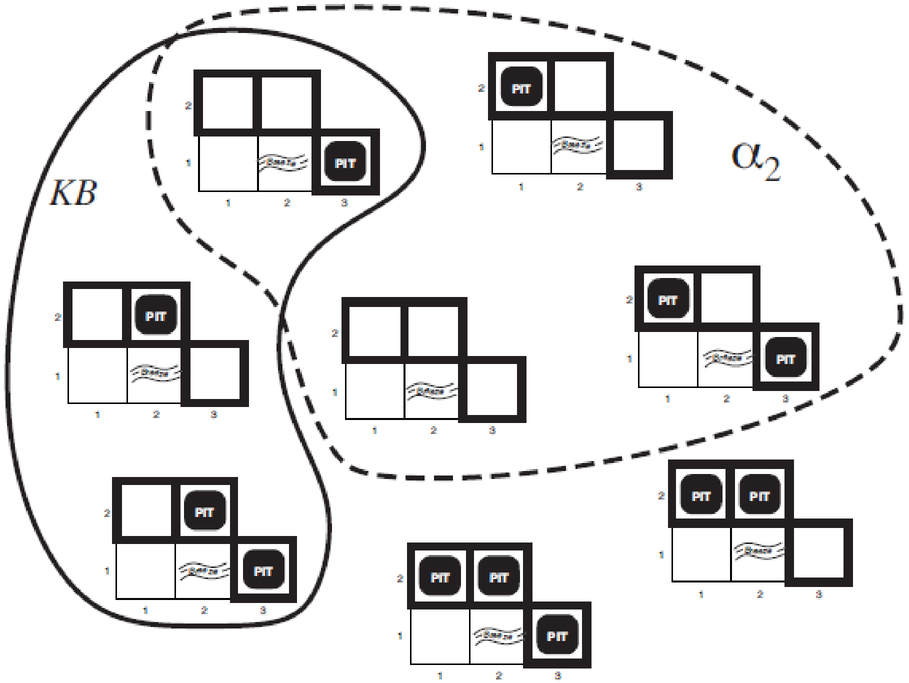
Then, for each model, check if the sentence is true

Entailment

α_2 = "There is no pit in [2,2]."

in some models in which KB is true, α_2 is false.

$\boxed{KB \not\models \alpha_2}$



First, find all models in KB

Then, for each model, check if the sentence is true

Logical inference

- Two methods
  - Method#1: Based on entailment (model checking)
  - Method#2: Based on inference rules (theorem proving)

- Enumerate all models to check if "α is true in all models in which KB is true"

$$\boxed{M(KB) \subseteq M(\alpha).}$$

Soundness and Completeness

- **Logical inference** based on “model checking” is **sound**
  - Truth-preserving: only entailed sentences will be derived

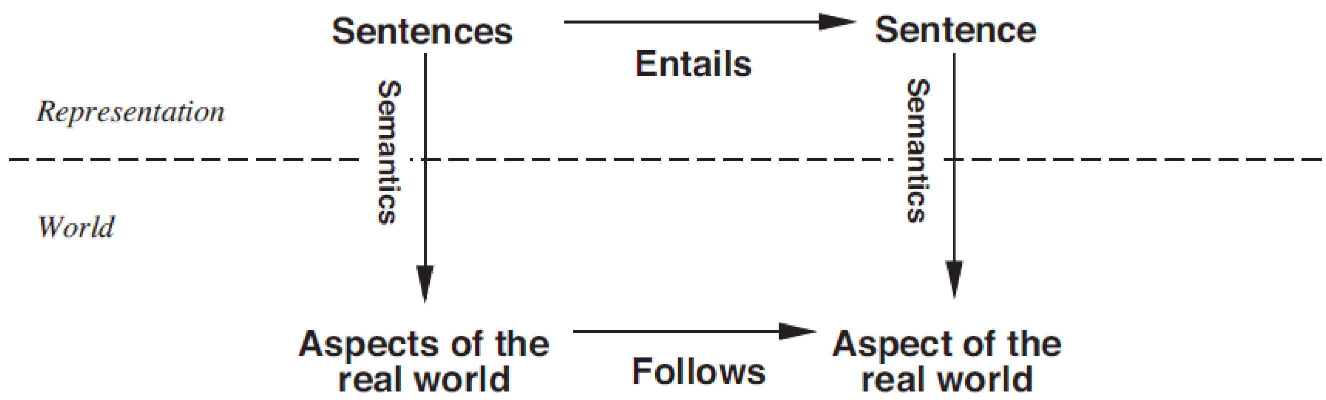
- It is also **complete**
  - Can derive any sentence that is entailed

When playing the **Wumpus World** game, for example, the agent is guaranteed to be as good as anyone (or anything)…

Logical sentences vs. Physical configurations

$(x > 10)$ $\qquad$ $(x > 0)$

... $\qquad$ ...



... $\qquad$ ...

Is in this set $= \{11, 12, \ldots, +\infty\}$ $\qquad$ *Is in this set* $= \{1, 2, \ldots, 11, 12, \ldots +\infty\}$

Outline for today

- Knowledge Based Agents
- The Wumpus World
- Logic
- **Propositional Logic**

Syntax of Propositional Logic

- Propositional symbols
  - *P, Q, R, etc.*
  - Boolean variables: either “***true***” or “***false***”

- Logical connectives
  - NOT ($\neg$)
  - AND ($\wedge$)
  - OR ($\vee$)
  - IMPLIES ($\rightarrow$)
  - EQUIVALENT ($\leftrightarrow$)

Gramma of sentences

- Recursive definition

$$
\begin{array}{rcl}
\mathit{Sentence} & \rightarrow & \mathit{AtomicSentence} \mid \mathit{ComplexSentence} \\
\mathit{AtomicSentence} & \rightarrow & \mathit{True} \mid \mathit{False} \mid P \mid Q \mid R \mid \ldots \\
\mathit{ComplexSentence} & \rightarrow & (\ \mathit{Sentence}\ ) \mid [\ \mathit{Sentence}\ ] \\
& \mid & \neg\ \mathit{Sentence} \\
& \mid & \mathit{Sentence} \land \mathit{Sentence} \\
& \mid & \mathit{Sentence} \lor \mathit{Sentence} \\
& \mid & \mathit{Sentence} \Rightarrow \mathit{Sentence} \\
& \mid & \mathit{Sentence} \Leftrightarrow \mathit{Sentence} \\
\\
\text{OPERATOR PRECEDENCE} & : & \neg, \land, \lor, \Rightarrow, \Leftrightarrow
\end{array}
$$

Semantics

- *Truth table*

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

Semantics

- The truth table

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

Semantics

- The truth table

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

Semantics

- The truth table

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*
		2	1	3	3	2

Truth table can tell you, for each formula, the
number of models (satisfying assignments)

Checking entailment: *enumeration method*

Let $\boxed{\alpha = A \vee B}$ and $\boxed{KB = (A \vee C) \wedge (B \vee \neg C)}$

Is it the case that $KB \models \alpha$?
Check all possible models—α must be true wherever KB is true

A	B	C	$A \vee C$	$B \vee \neg C$	KB	α
$False$	$False$	$False$				
$False$	$False$	$True$				
$False$	$True$	$False$				
$False$	$True$	$True$				
$True$	$False$	$False$				
$True$	$False$	$True$				
$True$	$True$	$False$				
$True$	$True$	$True$				

Checking entailment: *enumeration method*

Let $\boxed{\alpha = A \vee B}$ and $\boxed{KB = (A \vee C) \wedge (B \vee \neg C)}$

Is it the case that $KB \models \alpha$?

Check all possible models—α must be true wherever KB is true

A	B	C	$A \vee C$	$B \vee \neg C$	KB	α
$False$	$False$	$False$	$False$	$True$	$False$	$False$
$False$	$False$	$True$	$True$	$False$	$False$	$False$
$False$	$True$	$False$	$False$	$True$	$False$	$True$
$False$	$True$	$True$	$True$	$True$	$True$	$True$
$True$	$False$	$False$	$True$	$True$	$True$	$True$
$True$	$False$	$True$	$True$	$False$	$False$	$True$
$True$	$True$	$False$	$True$	$True$	$True$	$True$
$True$	$True$	$True$	$True$	$True$	$True$	$True$

Checking entailment: *enumeration method*

Let $\boxed{\alpha = A \vee B}$ and $\boxed{KB = (A \vee C) \wedge (B \vee \neg C)}$

Is it the case that $KB \models \alpha$?
Check all possible models—α must be true wherever KB is true

A	B	C	$A \vee C$	$B \vee \neg C$	KB	α
$False$	$False$	$False$	$False$	$True$	$False$	$False$
$False$	$False$	$True$	$True$	$False$	$False$	$False$
$False$	$True$	$False$	$False$	$True$	$False$	$True$
$False$	$True$	$True$	$True$	$True$	$True$	$True$
$True$	$False$	$False$	$True$	$True$	$True$	$True$
$True$	$False$	$True$	$True$	$False$	$False$	$True$
$True$	$True$	$False$	$True$	$True$	$True$	$True$
$True$	$True$	$True$	$True$	$True$	$True$	$True$

Checking entailment: *another example*

- *Based on "model checking"*

$$\left.\begin{array}{lllll} R_1: & \neg P_{1,1}\,. & & \\ R_2: & B_{1,1} & \Leftrightarrow & (P_{1,2} \vee P_{2,1})\,. \\ R_3: & B_{2,1} & \Leftrightarrow & (P_{1,1} \vee P_{2,2} \vee P_{3,1})\,. \\ R_4: & \neg B_{1,1}\,. & & \\ R_5: & B_{2,1}\,. & & \end{array}\right\}\ KB$$

$B_{1,1}$	$B_{2,1}$	$P_{1,1}$	$P_{1,2}$	$P_{2,1}$	$P_{2,2}$	$P_{3,1}$	R_1	R_2	R_3	R_4	R_5	KB
false	*false*	*false*	*false*	*false*	*false*	*false*	*true*	*true*	*true*	*true*	*false*	*false*
false	*false*	*false*	*false*	*false*	*false*	*true*	*true*	*true*	*false*	*true*	*false*	*false*
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
false	*true*	*false*	*false*	*false*	*false*	*false*	*true*	*true*	*false*	*true*	*true*	*false*
false	*true*	*false*	*false*	*false*	*false*	*true*	*true*	*true*	*true*	*true*	*true*	<u>*true*</u>
false	*true*	*false*	*false*	*false*	*true*	*false*	*true*	*true*	*true*	*true*	*true*	<u>*true*</u>
false	*true*	*false*	*false*	*false*	*true*	*true*	*true*	*true*	*true*	*true*	*true*	<u>*true*</u>
false	*true*	*false*	*false*	*true*	*false*	*false*	*true*	*false*	*false*	*true*	*true*	*false*
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
true	*true*	*true*	*true*	*true*	*true*	*true*	*false*	*true*	*true*	*false*	*true*	*false*

Logical inference

- Two methods
  - Method#1: Based on entailment (model checking)
  - Method#2: Based on inference rules (theorem proving)

- Standard logical equivalences

$$
\begin{aligned}
(\alpha \land \beta) &\equiv (\beta \land \alpha) \quad \text{commutativity of } \land \\
(\alpha \lor \beta) &\equiv (\beta \lor \alpha) \quad \text{commutativity of } \lor \\
((\alpha \land \beta) \land \gamma) &\equiv (\alpha \land (\beta \land \gamma)) \quad \text{associativity of } \land \\
((\alpha \lor \beta) \lor \gamma) &\equiv (\alpha \lor (\beta \lor \gamma)) \quad \text{associativity of } \lor \\
\neg(\neg\alpha) &\equiv \alpha \quad \text{double-negation elimination} \\
(\alpha \Rightarrow \beta) &\equiv (\neg\beta \Rightarrow \neg\alpha) \quad \text{contraposition} \\
(\alpha \Rightarrow \beta) &\equiv (\neg\alpha \lor \beta) \quad \text{implication elimination} \\
(\alpha \Leftrightarrow \beta) &\equiv ((\alpha \Rightarrow \beta) \land (\beta \Rightarrow \alpha)) \quad \text{biconditional elimination} \\
\neg(\alpha \land \beta) &\equiv (\neg\alpha \lor \neg\beta) \quad \text{De Morgan} \\
\neg(\alpha \lor \beta) &\equiv (\neg\alpha \land \neg\beta) \quad \text{De Morgan} \\
(\alpha \land (\beta \lor \gamma)) &\equiv ((\alpha \land \beta) \lor (\alpha \land \gamma)) \quad \text{distributivity of } \land \text{ over } \lor \\
(\alpha \lor (\beta \land \gamma)) &\equiv ((\alpha \lor \beta) \land (\alpha \lor \gamma)) \quad \text{distributivity of } \lor \text{ over } \land
\end{aligned}
$$

Validity

- A sentence is “**valid**” if it is true in “**all**” models (i.e., in each and every possible model)
  - Valid sentences are also called “**tautologies**”

- Can you think of an example?

Validity

- A sentence is "**valid**" if it is true in "**all**" models (i.e., in each and every possible model)
  - Valid sentences are also called "**tautologies**"

- Can you think of an example?
  - $True \vee A$
  - $A \vee (\neg A)$
  - $\neg(A \wedge (\neg A))$
  - $A \Leftrightarrow A$
  - $((P \vee Q) \Leftrightarrow P) \vee (\neg P \wedge Q)$
  - $(P \Leftrightarrow Q) \Rightarrow (P \Rightarrow Q)$

Validity

- A sentence is "**valid**" if it is true in "**all**" models (i.e., in each and every possible model)
  - Valid sentences are also called "**tautologies**"

- More examples

$$\alpha \text{ is valid if and only if } True \models \alpha.$$

$$\text{For any } \alpha,\ False \models \alpha.$$

Satisfiability

- A sentence is "**satisfiable**" if it is true in "**some**" models (i.e., in *at least one* model)

- Some examples:

$$(A \lor B) \land \neg(A \Rightarrow B) \text{ is satisfiable.}$$

$$(A \Leftrightarrow B) \land (\neg A \lor B) \text{ is satisfiable.}$$

Connection between Validity and Entailment

- "Checking entailment" can be done by "checking validity"

$$\boxed{\textit{For any sentences } \alpha \textit{ and } \beta,\ \alpha \models \beta \textit{ if and only if the sentence } (\alpha \Rightarrow \beta) \textit{ is valid.}}$$

Connection between Validity and Satisfiability

- A sentence is "**valid**" if and only if its negation is "**unsatisfiable**"

$$\alpha \models \beta \textit{ if and only if the sentence } (\alpha \wedge \neg\beta) \textit{ is unsatisfiable.}$$

This is the intuition behind "***Proof by contradiction***"

Assume that "alpha" does not entail "beta"
Then, there must be a case where "alpha" is ture but "beta" is false…
…
Since the case ("alpha" is true but "beta" is false) is unsatisfiable, our assumption doesn't hold.
Thus, the original statement ("alpha" entails "beta") is correct.

QED

More examples

- Is the sentence valid, satisfiable, or neither?

$$[(Food \Rightarrow Party) \vee (Drinks \Rightarrow Party)] \Rightarrow [(Food \wedge Drinks) \Rightarrow Party]$$

Outline

- What is AI?
- Problem-solving agent
  - Uninformed search
  - Informed search (A*)
  - Local search
  - Adversarial search
- **Knowledge-based agent**
  - The Wumpus World
  - Propositional Logic