Knowledge-Based Agents

Chapter 7.1-7.3

Some material adopted from notes by Andreas Geyer-Schulz and Chuck Dyer

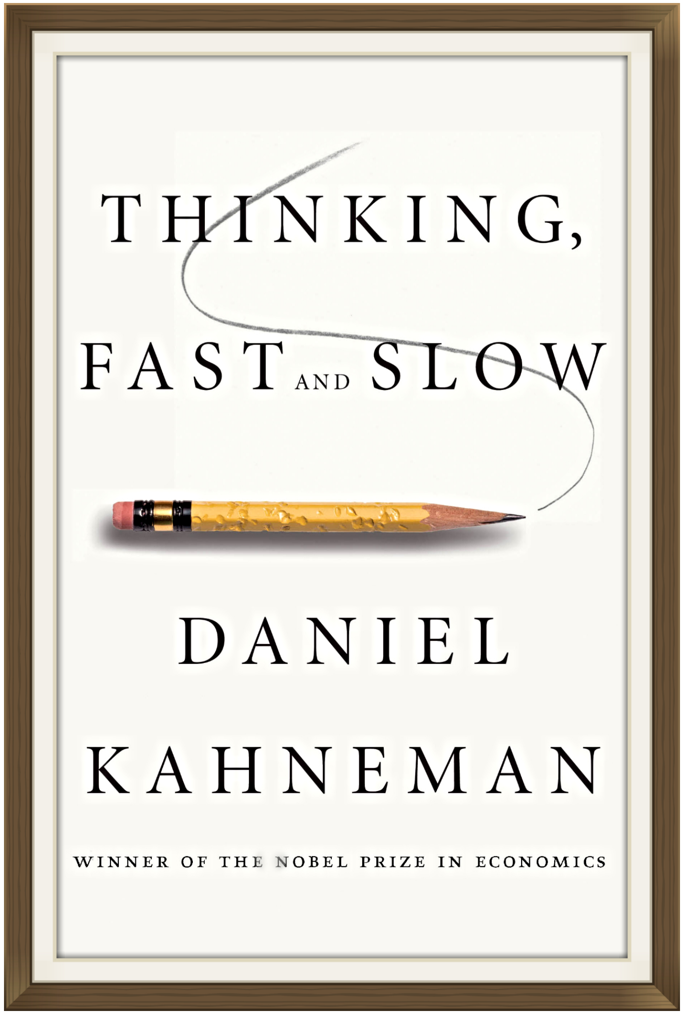
Big Idea

- Drawing reasonable conclusions from a set of data (observations, beliefs, etc.) seems key to intelligence
- Logic is a powerful and well developed approach and highly regarded by people
- Logic is also a strong formal system that computers can use (cf. John McCarthy)
- We can solve some AI problems by representing them in logic and applying standard proof techniques to generate solutions

Inference in People

- People can do logical inference, but are not always very good at it
- Reasoning with negation and disjunction seems particularly difficult
- But, people seem to employ many kinds of reasoning strategies, most of which are neither *complete* nor *sound*

Thinking Fast and Slow



- Popular 2011 book by Nobel prize winning cognitive psychologist
- His model is we have two types of reasoning facilities
- **System 1** operates automatically and quickly, with little or no effort and no sense of voluntary control
- **System 2** allocates attention to the effortful mental activities that demand it, including complex computations

Question #1

Here is a simple puzzle

Don’t try to solve it -- listen to your intuition

Question #1

Here is a simple puzzle

Don’t try to solve it -- listen to your intuition

- A bat and ball cost $1.10
- The bat costs one dollar more than the ball
- How much does the ball cost?

Question #1

Here is a simple puzzle

Don’t try to solve it -- listen to your intuition

- A bat and ball cost $1.10
- The bat costs one dollar more than the ball
- How much does the ball cost?

The ball costs $0.05

Question #2

Determine, as quickly as you can, if the argument is logically valid, i.e. does the conclusion follow the premises?

Question #2

Try to determine, as quickly as you can, if the argument is logically valid. Does the conclusion follow the premises?

- **All roses are flowers**
- **Some flowers fade quickly**

∴ Therefore some roses fade quickly

Question #2

Try to determine, as quickly as you can, if the argument is logically valid. Does the conclusion follow the premises?

- All roses are flowers
- Some flowers fade quickly
- Therefore some roses fade quickly

It is possible that there are no roses among the flowers that fade quickly

Question #3

It takes 5 machines 5 minutes to make 5 widgets

How long would it take 100 machines to make 100 widgets?

Question #3

It takes 5 machines 5 minutes to make 5 widgets

How long would it take 100 machines to make 100 widgets?

- **100 minutes or 5 minutes?**

Question #3

It takes 5 machines 5 minutes to make 5 widgets

How long would it take 100 machines to make 100 widgets?

- 100 minutes or 5 minutes?

5 minutes

Wason Selection Task

- I have a pack of cards; each has a letter on one side and a number on the other
- I claim the following rule is true:

  If a card has a vowel on one side, then it has an even number on the other

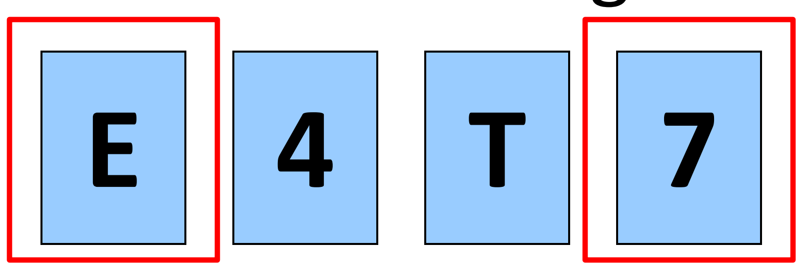
- Given these cards, which should you turn over to decide whether the rule is true or false?

E **4** **T** **7**

Wikipedia

Wason Selection Task

- Wason (1966) showed people are bad at this task
- To disprove rule P=>Q, find a situation in which P is true but Q is false, i.e., show P^~Q
- To disprove **vowel => even**, find a card with a vowel and an odd number
- Thus, turn over the cards showing **vowels** and turn over cards showing **odd numbers**



Wason Selection Task

- This version is easier for people, as shown by Griggs & Cox, 1982
- You are the bouncer in a bar; which of these people do you card given the rule: *You must be 21 or older to drink beer.*

beer	**coke**	**22**	**20**

Perhaps easier because it's more familiar or because people have special strategies to reason about certain situations, such as cheating in a social situation

Negation in Natural Language

- We often model the meaning of natural language sentences as a logic statements
- Logic maps these into equivalent statements
  - All elephants are gray
  - No elephant are not gray
- Double negation is common in informal language: *that won't do you no good*
- But what does this mean: *we cannot underestimate the importance of logic*

Misnegation

we cannot underestimate the importance of logic

Does it mean:

- Logic is very important
- Logic is not very important

Language Log has many posts withexamples of the phenomenon

Logic as a Methodology

Even if people don’t use formal logical reasoning for solving a problem, logic might be a good approach for AI for a number of reasons

- Airplanes don’t need to flap their wings
- Logic may be a good implementation strategy
- Solution by a formal system can offer other benefits, e.g., letting us prove properties of the approach (e.g., complexity)

• See neats vs. scruffies

Knowledge-based agents

- Knowledge-based agents have a **knowledge base** (KB) and an **inference system**
- KB: a set of representations of facts believed true
- Each individual representation is called a **sentence**
- Sentences are expressed in a **knowledge representation language**
- The agent operates as follows:
  1. It **TELL**s the KB facts it perceives
  2. It **ASK**s the KB what action it should perform
  3. It performs the chosen action

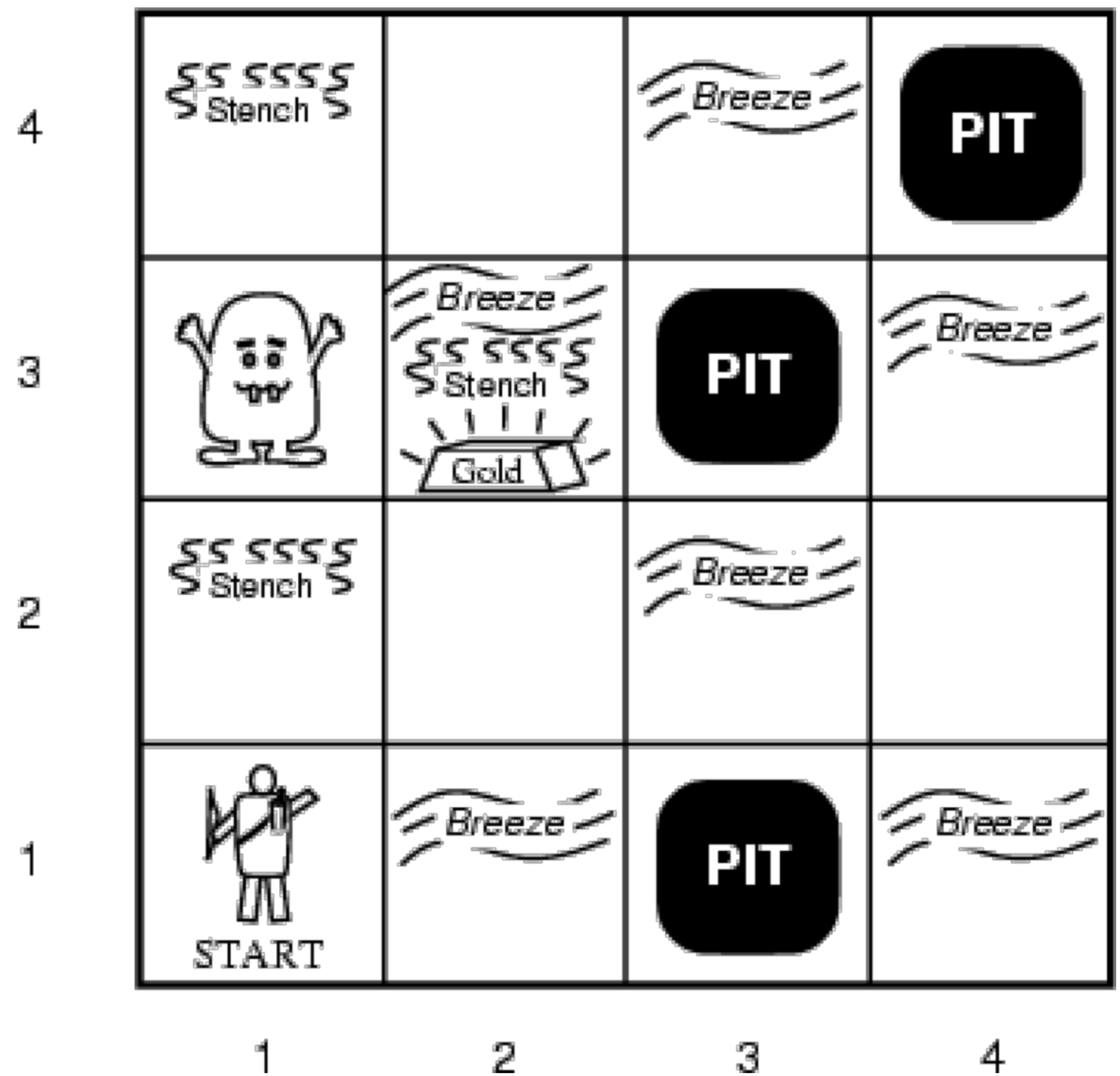
Wumpus World environment

- Cf. 1972 Hunt the Wumpus computer game
- Agent explores cave of rooms connected by passageways
- Lurking in a room is the *Wumpus*, a beast that eats any agent that enters its room
- Some rooms have *bottomless pits* that trap any agent that wanders into the room
- Somewhere is a heap of gold in a room
- Goal: collect gold & exit w/o being eaten

AIMA’s Wumpus World

The agent always starts in [1,1]

Agent’s task: find the gold, return to [1,1], and exit cave



Agent in a Wumpus world: Percepts

- The agent perceives
  - **stench** in square containing Wumpus and adjacent squares (not diagonally)
  - **breeze** in squares adjacent to a pit
  - **glitter** in the square where the gold is
  - **bump**, if it walks into a wall
  - Woeful **scream** everywhere if Wumpus killed
- Percepts given as five-tuple, e.g., if stench and breeze, but no glitter, bump or scream:

  `(Stench, Breeze, None, None, None)`
- Agent cannot perceive its location, e.g., (2,2)

Wumpus World Actions

- **go forward**
- **turn right** 90 degrees
- **turn left** 90 degrees
- **grab**: Pick up object in same square as agent
- **shoot**: Fire arrow in direction agent faces. It continues until it hits & kills Wumpus or hits outer wall. Agent has one arrow, so only first shoot action has effect
- **Climb:** leave cave, only effective in start square
- **die**: automatically and irretrievably happens if agent enters square with pit or living Wumpus

Wumpus World Goal

Agent's goal: find the gold and bring it back to the start square as quickly as possible, without getting killed

Reward function:

- +1,000 points for exiting cave with gold
- -1 point for every action taken
- -10,000 points for getting killed

Wumpus world characterization

Recall environment characteristics from ch. 2

- **Fully Observable?**
- **Deterministic?**
- **Episodic?**
- **Static**?
- **Discrete**?
- **Single-agent?**

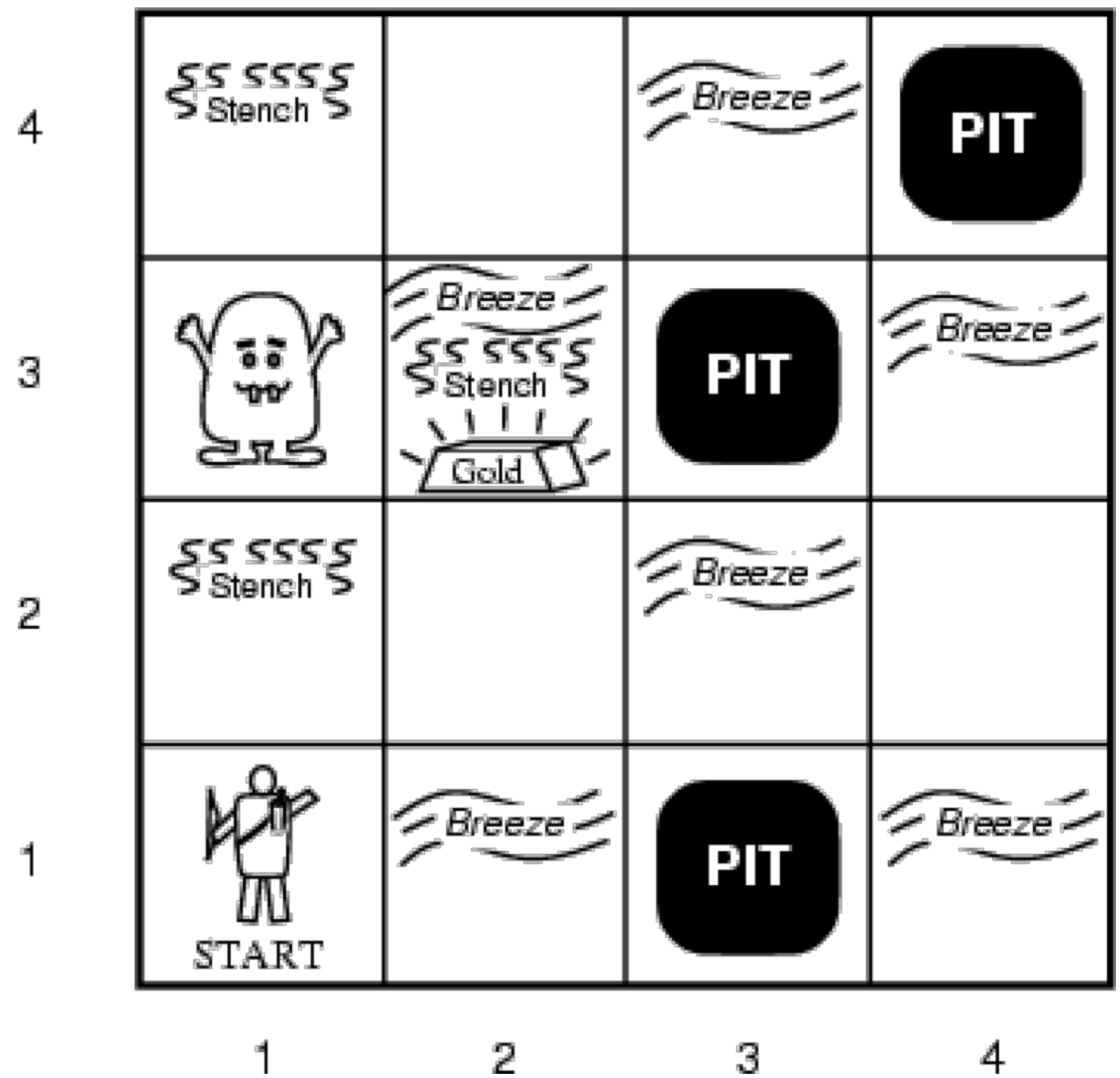
Wumpus world characterization

- **Fully Observable** No, only local perception
- **Deterministic** Yes, outcomes exactly specified
- **Episodic** No, sequential at level of actions
- **Static** Yes, Wumpus and Pits do not move
- **Discrete** Yes
- **Single-agent?** Yes, Wumpus essentially a natural feature

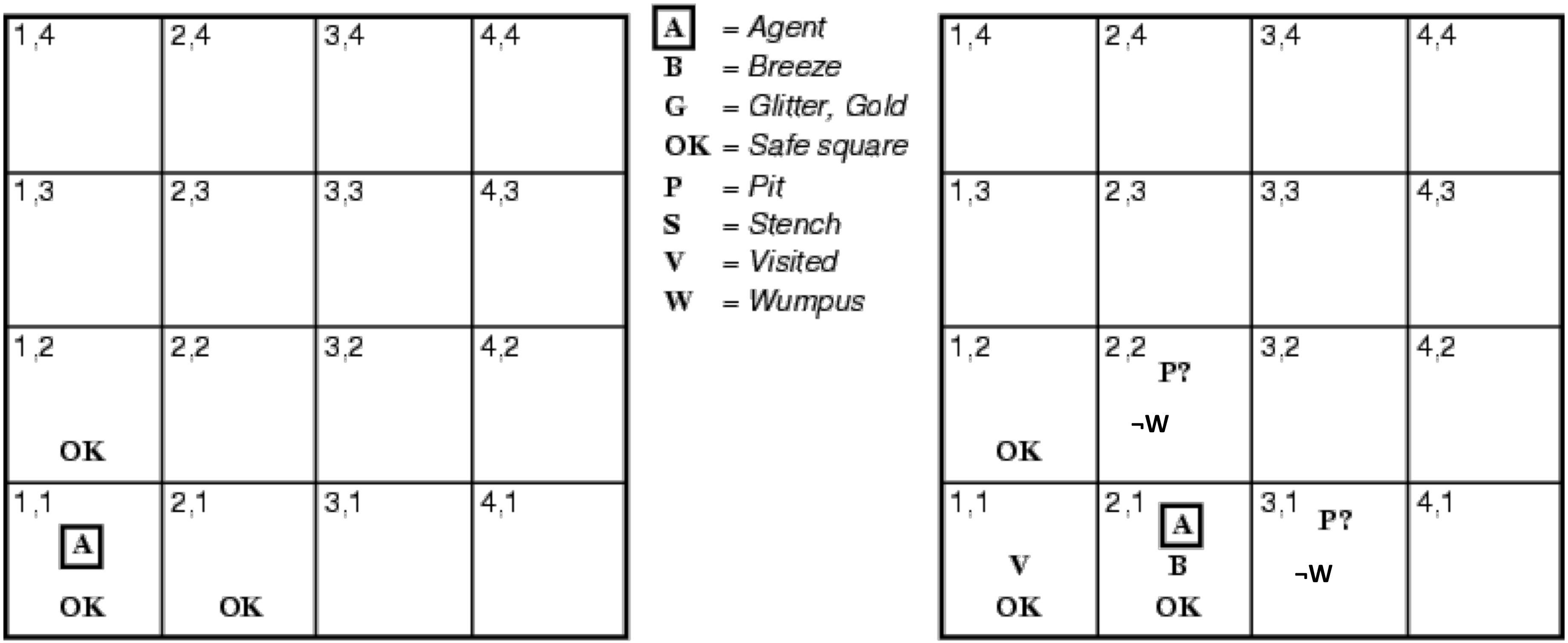
AIMA's Wumpus World

The agent always starts in [1,1]

Agent's task: find gold, return [1,1], and climb out of the cave



The Hunter's first step

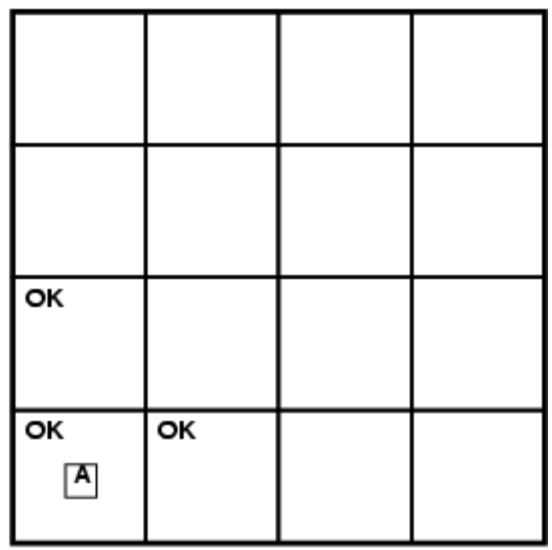


(a) (b)

Since agent is alive and perceives neither breeze nor stench at [1,1], it **knows** [1,1] and its neighbors are OK

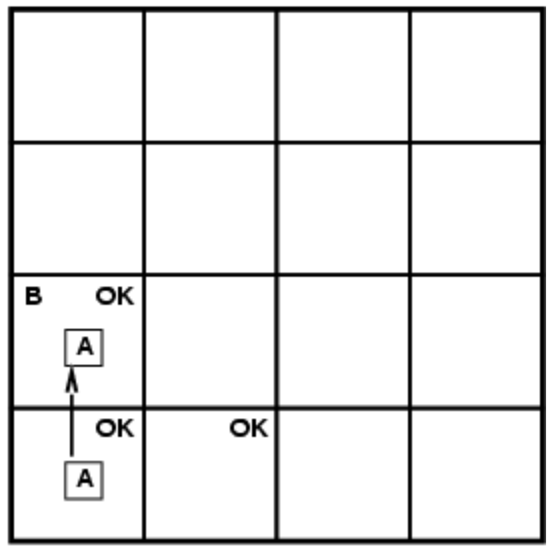
Moving to [2,1] is a **safe move** that reveals a breeze but no stench, **implying** that Wumpus isn't adjacent but one or more pits are

Exploring a wumpus world



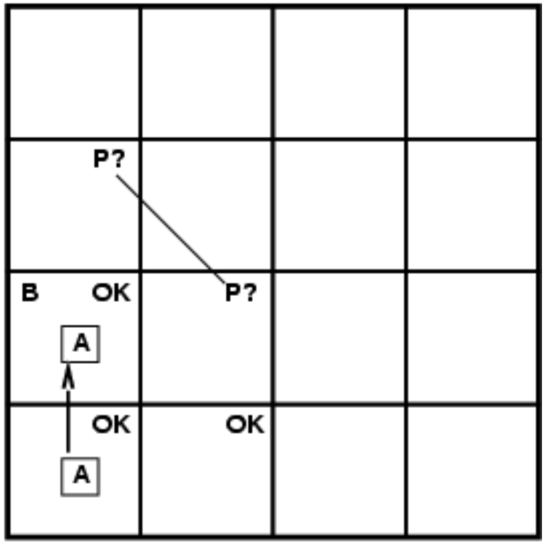
A	agent
B	breeze
G	glitter
OK	safe cell
P	pit
S	stench
W	wumpus

Exploring a wumpus world



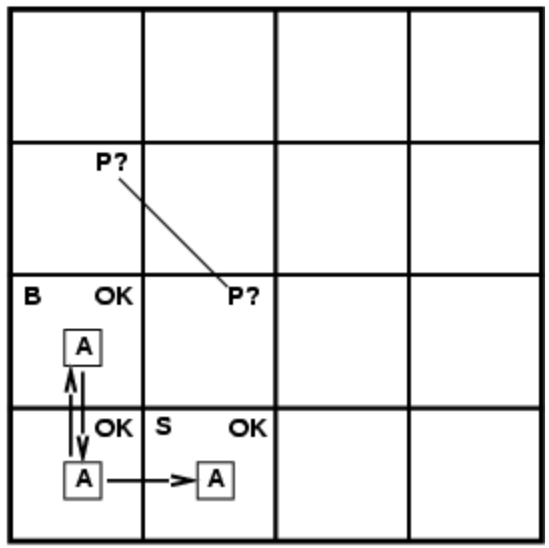
A	agent
B	breeze
G	glitter
OK	safe cell
P	pit
S	stench
W	wumpus

Exploring a wumpus world



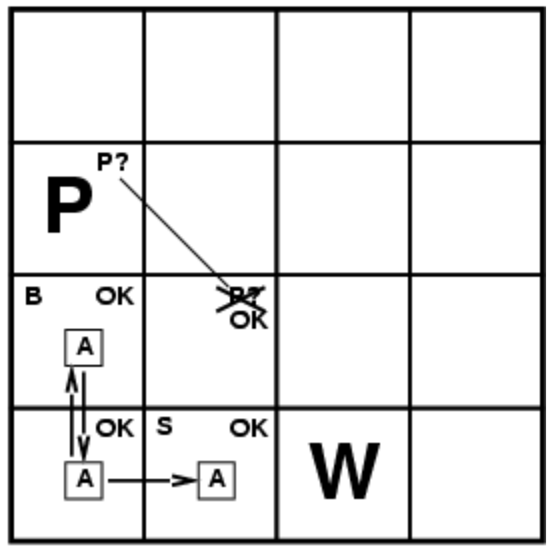
A	agent
B	breeze
G	glitter
OK	safe cell
P	pit
S	stench
W	wumpus

Exploring a wumpus world



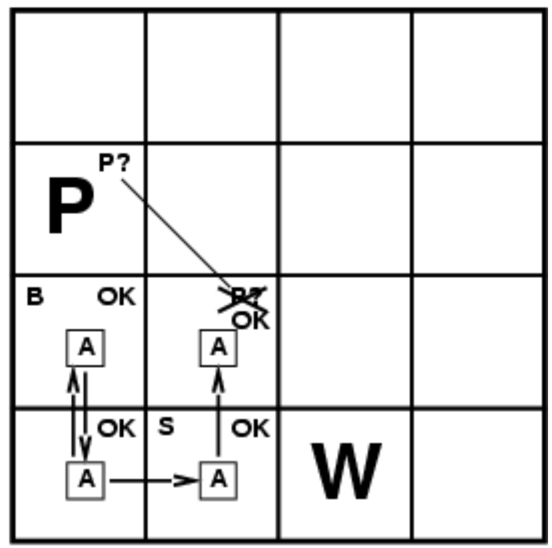
A	agent
B	breeze
G	glitter
OK	safe cell
P	pit
S	stench
W	wumpus

Exploring a wumpus world



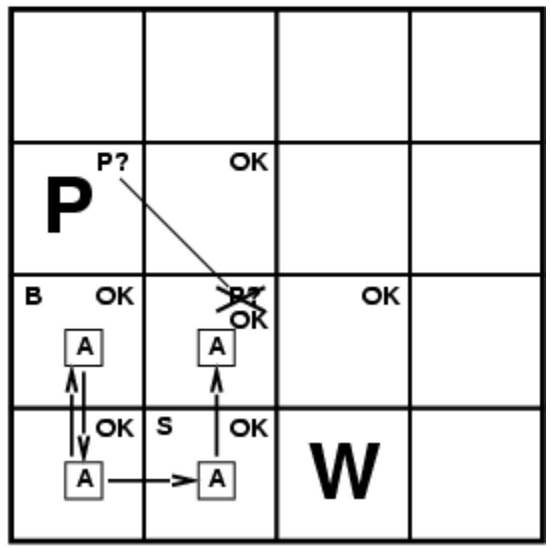
A	agent
B	breeze
G	glitter
OK	safe cell
P	pit
S	stench
W	wumpus

Exploring a wumpus world



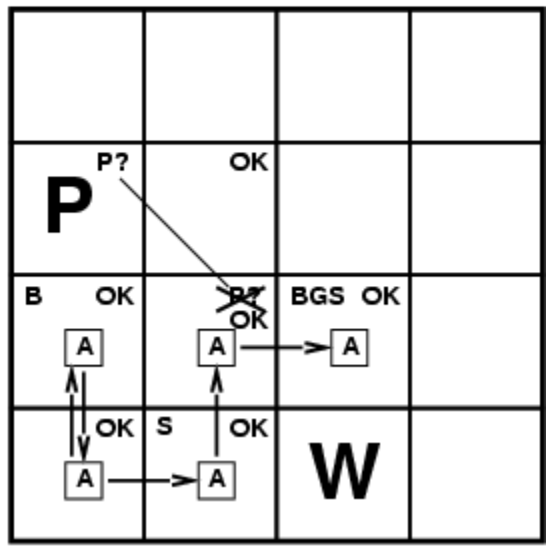
A	agent
B	breeze
G	glitter
OK	safe cell
P	pit
S	stench
W	wumpus

Exploring a wumpus world



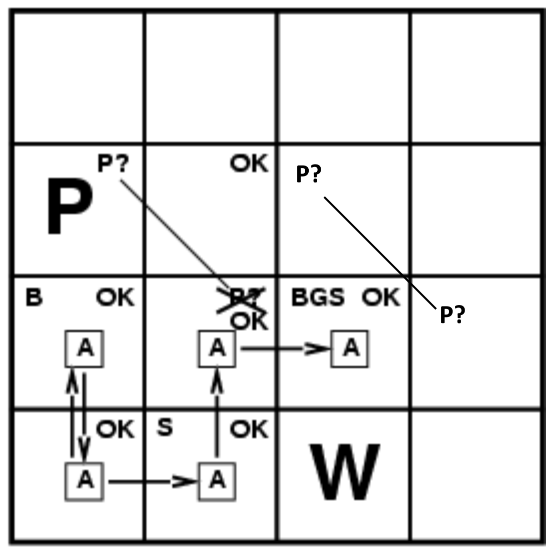
A	agent
B	breeze
G	glitter
OK	safe cell
P	pit
S	stench
W	wumpus

Exploring a wumpus world



A	agent
B	breeze
G	glitter
OK	safe cell
P	pit
S	stench
W	wumpus

Exploring a wumpus world



A	agent
B	breeze
G	glitter
OK	safe cell
P	pit
S	stench
W	wumpus

Logic in general

- **Logics** are formal languages for representing information so that conclusions can be drawn
- **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., the language of arithmetic

- $x+2 \geq y$ is a sentence; $x2+y > \{\}$ is not a sentence
- $x+2 \geq y$ is true iff the number $x+2$ is no less than the number y
- $x+2 \geq y$ is true in a world where $x = 7$, $y = 1$
- $x+2 \geq y$ is false in a world where $x = 0$, $y = 6$
- $x+1 > x$ is true for all numbers x

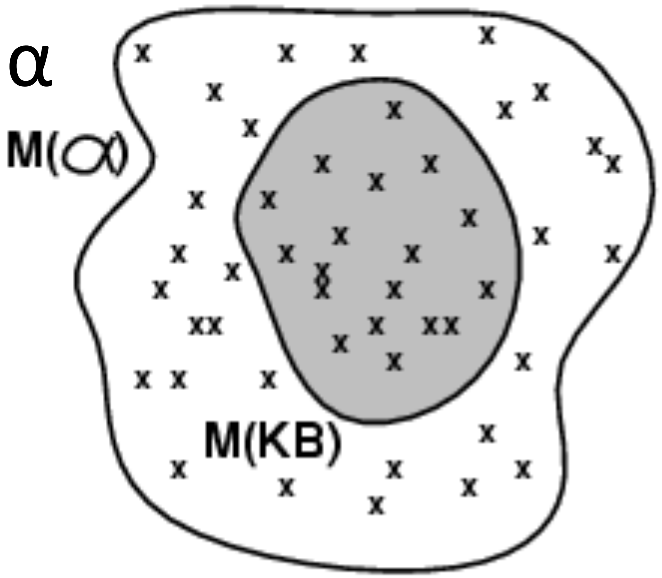
Entailment

- Entailment: one thing follows from another
- KB $\models \alpha$
- Knowledge base *KB* entails sentence α iff α is true in *all possible worlds* where *KB* is true
  - E.g., the KB containing "UMBC won" and "JHU won" entails "Either UMBC won or JHU won"
  - E.g., x+y = 4 entails 4 = x+y
  - Entailment is a relationship between (sets of) sentences (i.e., syntax) that is based on **semantics**

Models

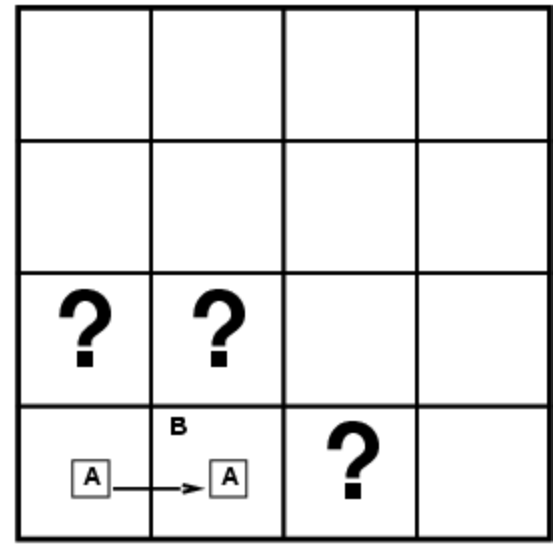
- Logicians talk of models: formally structured worlds w.r.t which truth can be evaluated
- *m* is a model of sentence α if α is true in *m*

  Lots of other things might or might not be true or might be unknown in *m*
- *M(α)* is the set of all models of α
- Then KB $\models$ α iff *M(KB)* $\subseteq$ *M*(α)
  - *KB* = UMBC and JHU won
  - α = UMBC won
  - Then KB $\models$ α



Entailment in the Wumpus World

- Situation after detecting nothing in [1,1], moving right, breeze in [1,2]
- Possible models for *KB* assuming only pits and restricting cells to {(1,3)(2,1)(2,2)}
- Two observations: ~B11, B12
- Three propositional variables variables: P13, P21, P22
- ⇒ 8 possible models

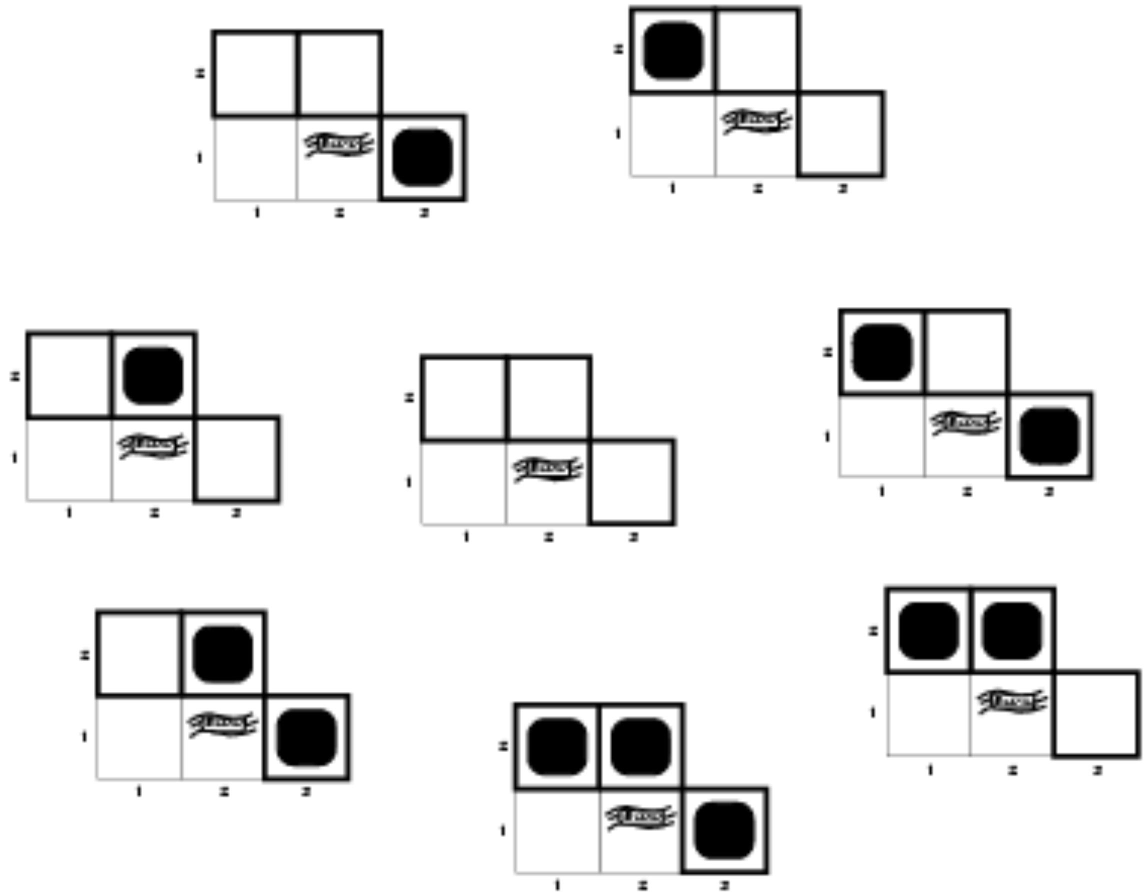


B11: breeze in (1,1)
P13: pit in (1,3)

Wumpus models

P13	P21	P22
F	F	F
F	F	T
F	T	F
F	T	T
T	F	F
T	F	T
T	T	F
T	T	T

Each row is a possible world



Wumpus World Rules (1)

- **If a cell has a pit, then a breeze is observable in every adjacent cell**
- In propositional calculus we can not have rules with variables (e.g., forall X…)

P11 => B21

P11 => B12

P21 => B11

P21 => B22 …

If a pit in (1,1) then a breeze in (2,1), …

these also follow

~B21 => ~P11

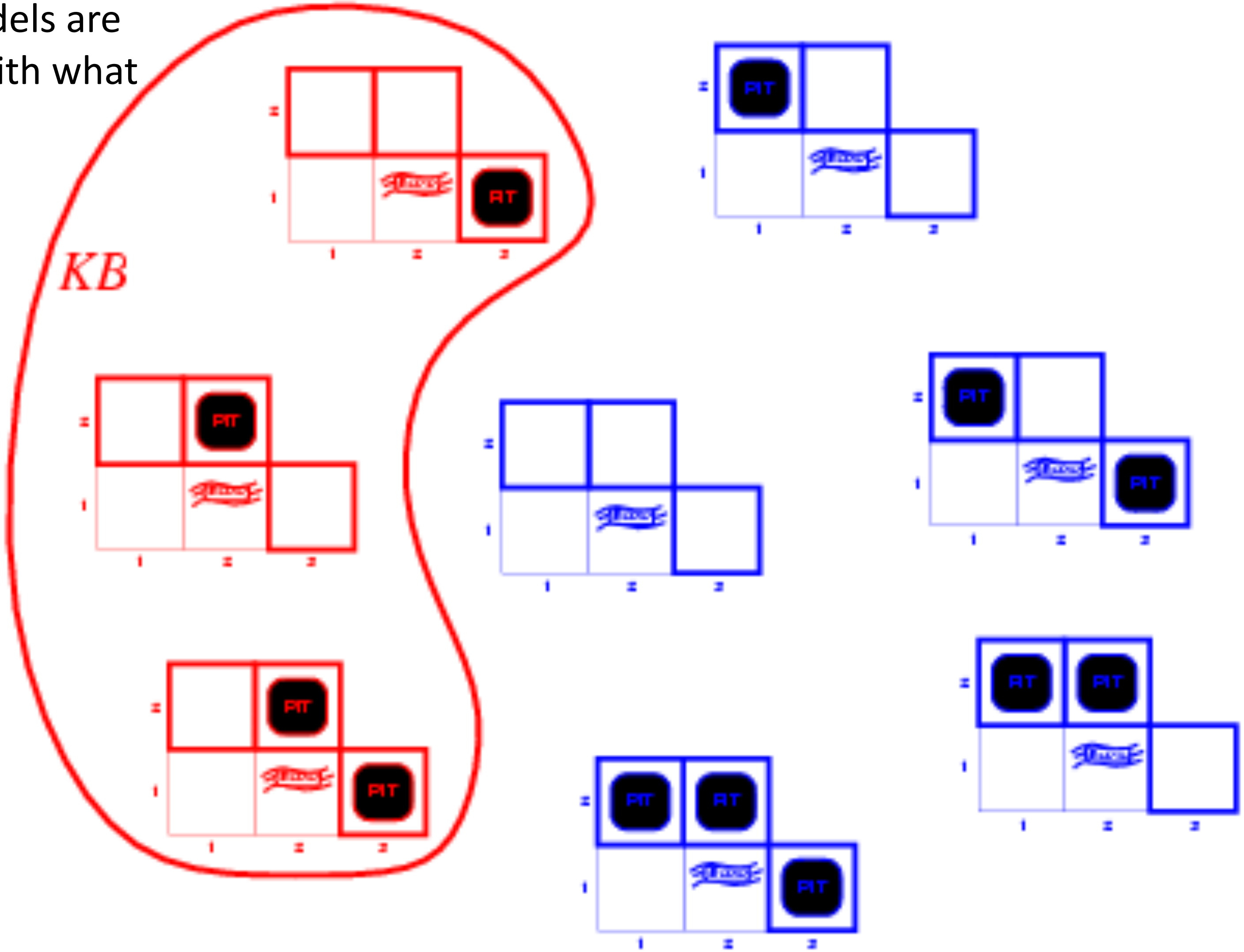
~B12 => ~P11

~B11 => ~P21

~B22 => ~P21

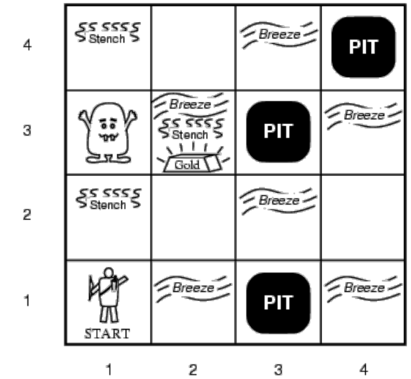
…

Only three of the possible models are consistent with what we know



KB = wumpus-world rules + observations

Wumpus World Rules (2)



- **Cell safe if it has neither a pit nor wumpus**

  OK11 => ~P11 ∧ ~W11

  OK12 => ~P12 ∧ ~W12 …

- From which we can derive

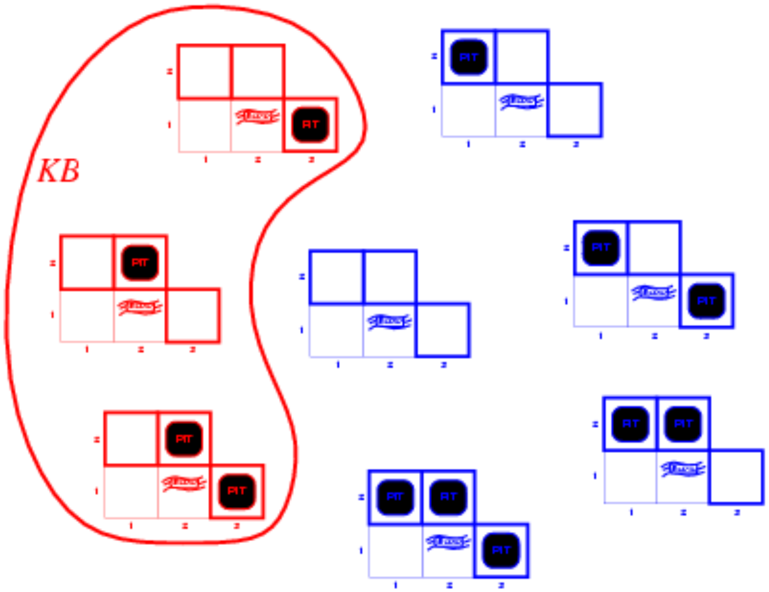
  P11 ∨ W11 => ~OK11

  P11 => ~OK11

  W11 => ~OK11 …

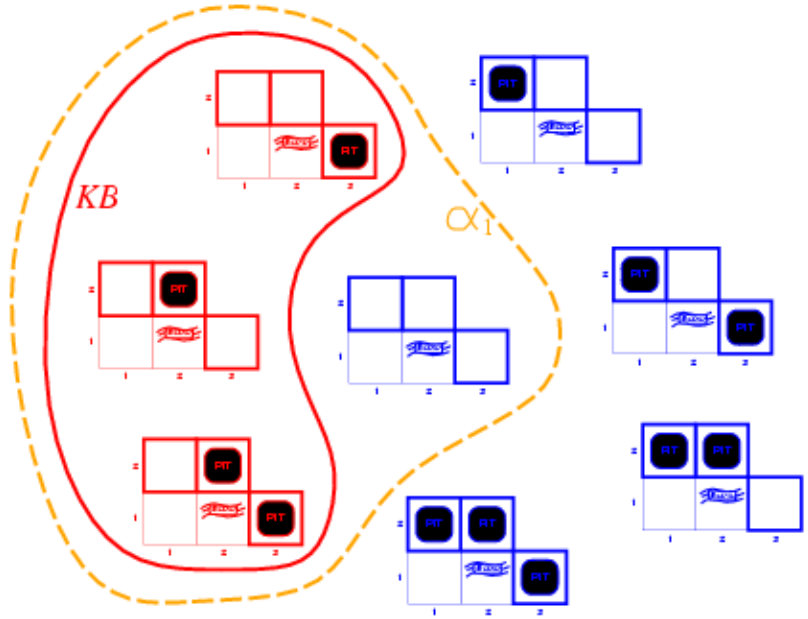
OK11: (1,1) is safe
W11: Wumpus in (1,1)

Wumpus models



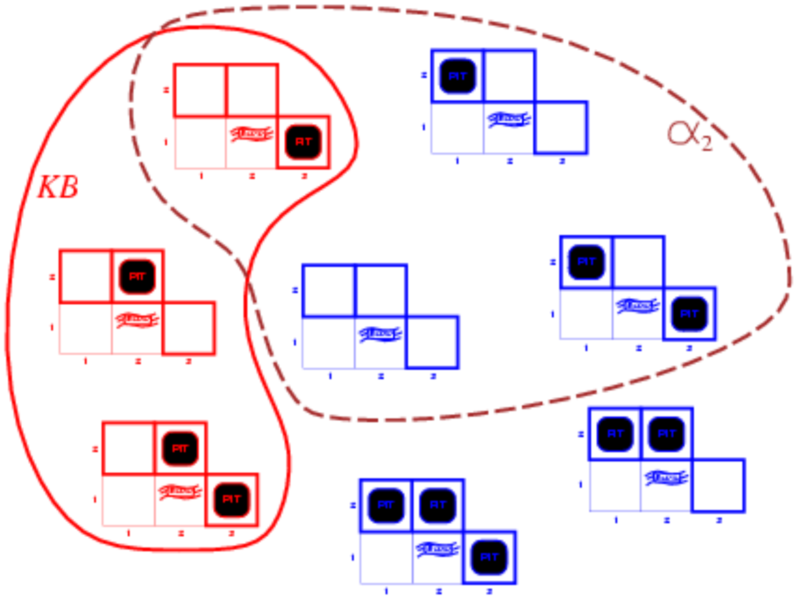
- *KB* = wumpus-world rules + observations

Wumpus models



- *KB* = wumpus-world rules + observations
- α_1 = "[1,2] is safe"
- *Since all models include* α_1
- *KB* $\models \alpha_1$, proved by **model checking**

Is (2,2) Safe?



- *KB* = wumpus-world rules + observations
- α_2 = "[2,2] is safe"
- Since some models don't include α_2, $KB \not\models \alpha_2$
- We cannot prove OK22; it might be true or false

Inference, Soundness, Completeness

- *KB* $\vdash_i$ α = sentence α can be derived from *KB* by procedure *i*
- **Soundness:** *i* is sound if whenever *KB* $\vdash_i$ α, it is also true that *KB* ╞ α
- **Completeness:** *i* is complete if whenever *KB* ╞ α, it is also true that *KB* $\vdash_i$ α

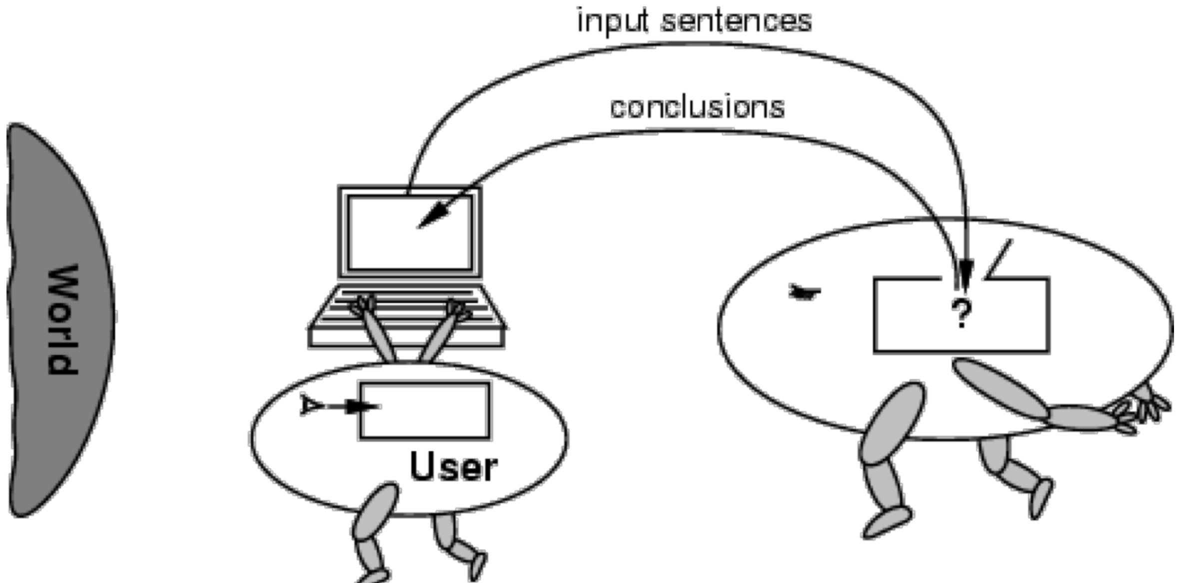
- Preview: **first-order logic** is expressive enough to say almost anything of interest and has a **sound** and **complete** inference procedure

Soundness and completeness

- A *sound* inference method derives only entailed sentences
- Analogous to the property of *completeness* in search, a *complete* inference method can derive any sentence that is entailed

No independent access to the world

- Reasoning agents often gets knowledge about facts of the world as a sequence of logical sentences and must draw conclusions only from them w/o independent access to world
- Thus, it is very important that the agents' reasoning is sound!



Summary

- Intelligent agents need knowledge about world for good decisions
- Agent's knowledge stored in a knowledge base (KB) as **sentences** in a knowledge representation (KR) language
- Knowledge-based agents needs a **KB** & **inference mechanism**. They store sentences in KB, infer new sentences & use them to **deduce** which actions to take
- A **representation language** defined by its syntax & semantics, which specify structure of sentences & how they relate to facts of the world
- **Interpretation** of a sentence is fact to which it refers. If fact is part of the actual world, then the sentence is true