# 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 to this and highly regarded by people
- Logic is also a strong formal system that we can programs computers to use
- We can solve some AI problems by represent-ing them in logic and applying standard proof techniques to generate solutions

# Inference in People

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

# Wason Selection Task

- I have a pack of cards; each has a letter written 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*

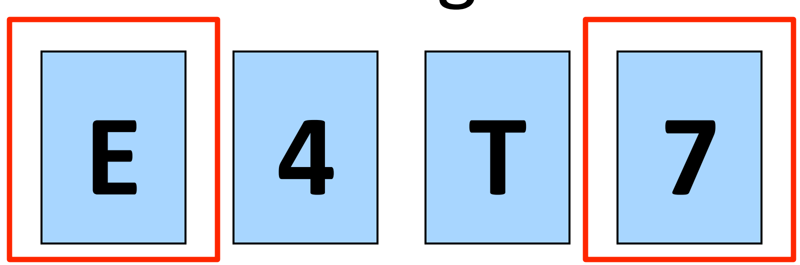
- For these cards, which should you turn over in order to decide whether the rule is true or false?

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

Wikipedia

# Wason Selection Task

- Wason (1966) showed that people are not very good at this task.
- To disprove the 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



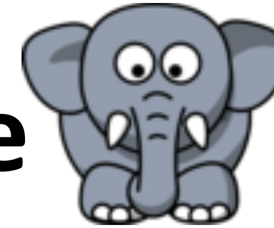
# Wason Selection Task

- This version is easier for people to do, as was shown by Griggs & Cox, 1982
- Your 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**

- May be simpler 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
- This maps these into equivalent statements
  - All elephants are gray
  - No elephant are not gray
- Double negation is common in informal language: th*at won't do you no good*
- But what does this mean: *we cannot underestimate the importance of logic*

# 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 in a formal system can offer other benefits, e.g., letting us prove properties of the approach

• See neats vs. scruffies

# Knowledge-based agents

- Knowledge-based agents have a knowledge base (KB) and an inference system
- A KB is 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 TELLs the KB what it perceives
  2. It ASKs the KB what action it should perform
  3. It performs the chosen action

# Architecture of a KB agent

- **Knowledge Level**
  - The most abstract level: describe agent by saying what it knows
  - Ex: A taxi agent might know that the Golden Gate Bridge connects San Francisco with the Marin County
- **Logical Level**
  - The level at which the knowledge is encoded into *sentences*
  - Ex: links(GoldenGateBridge, SanFrancisco, MarinCounty)
- **Implementation Level**
  - Physical representation of the sentences in the logical level
  - Ex: as a tuple serialized as `(links goldengatebridge sanfrancisco marincounty)`

# Wumpus World environment

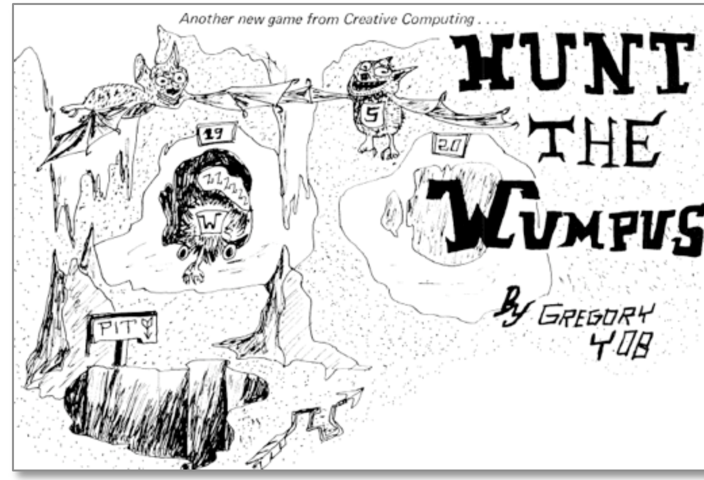
- Based on Hunt the Wumpus computer game
- Agent explores a cave of rooms connected by passageways
- Lurking in a room is the *Wumpus*, a beast that eats any agent that enters its room
- Some rooms have *bottomless pits* that trap any agent that wanders into the room
- Somewhere is a heap of gold in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

# Jargon file on “Hunt the Wumpus”

**WUMPUS** /wuhm'p*s/ n. The central monster (and, in many versions, the name) of a famous family of very early computer games called “Hunt The Wumpus,” dating back at least to 1972 (several years before ADVENT) on the Dartmouth Time-Sharing System. The wumpus lived somewhere in a cave with the topology of a dodecahedron's edge/vertex graph (later versions supported other topologies, including an icosahedron and Mobius strip). The player started somewhere at random in the cave with five “crooked arrows”; these could be shot through up to three connected rooms, and would kill the wumpus on a hit (later versions introduced the wounded wumpus, which got very angry). Unfortunately for players, the movement necessary to map the maze was made hazardous not merely by the wumpus (which would eat you if you stepped on him) but also by bottomless pits and colonies of super bats that would pick you up and drop you at a random location (later versions added “anaerobic termites” that ate arrows, bat migrations, and earthquakes that randomly changed pit locations).

This game appears to have been the first to use a non-random graph-structured map (as opposed to a rectangular grid like the even older Star Trek games). In this respect, as in the dungeon-like setting and its terse, amusing messages, it prefigured ADVENT and Zork and was directly ancestral to both. (Zork acknowledged this heritage by including a super-bat colony.) Today, a port is distributed with SunOS and as freeware for the Mac. A C emulation of the original Basic game is in circulation as freeware on the net.

# Wumpus History

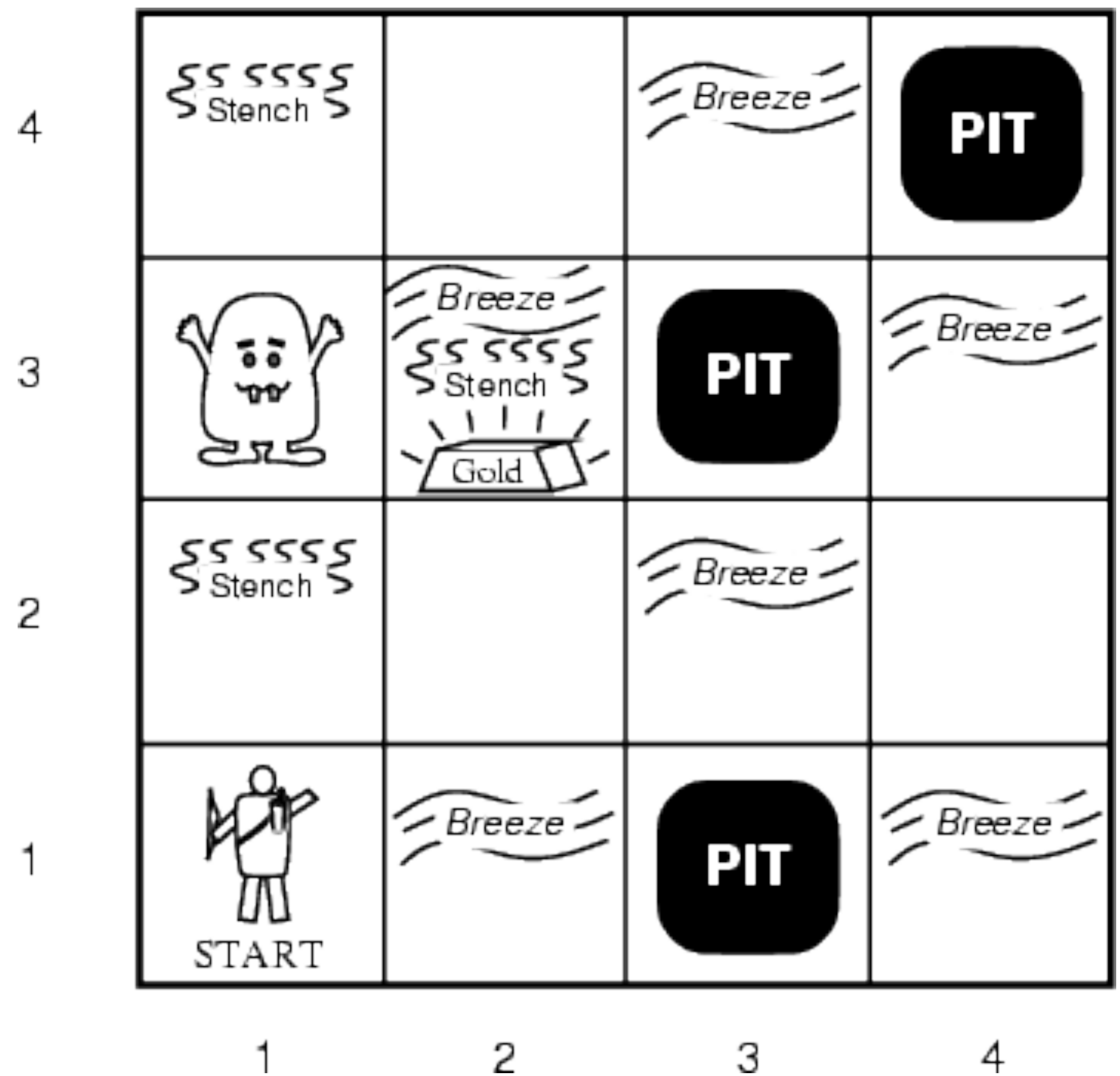



- See [Hunt the Wumpus] for details
- Early (c. 1972) text-based game written in BASIC written by Gregory Yob, a student at UMASS, Dartmouth
- Defined a genre of games including adventure, zork, and nethack
- Eventually commercialized (c. 1980) for early personal computers
- The Hunt the Wumpus basic code is available in a 1976 article in Creative Computing by Yob!

# AIMA's Wumpus World

The agent always starts in the field [1,1]

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



# Agent in a Wumpus world: Percepts

- The agent perceives
  - **stench** in the square containing the Wumpus and in the adjacent squares (not diagonally)
  - **breeze** in the squares adjacent to a pit
  - **glitter** in the square where the gold is
  - **bump**, if it walks into a wall
  - awoeful **scream** everywhere in the cave, if Wumpus is 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 own location

# Wumpus World Actions

- **go forward**
- **turn right** 90 degrees
- **turn left** 90 degrees
- **grab**: Pick up object in the same square as the agent
- **shoot**: Fire arrow in straight line in direction agent is facing. It continues until it hits and kills Wumpus or hits outer wall. Agent has only one arrow, so only first shoot action has an effect
- **climb** is used to leave cave, only effective in start square
- **die**: This action automatically and irretrievably happens if agent enters square with pit or live Wumpus

# Wumpus World Goal

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

- –1,000 point reward for climbing out of the cave with the gold
- –1 point deducted for every action taken
- –10,000 point penalty for getting killed

# Wumpus world characterization

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

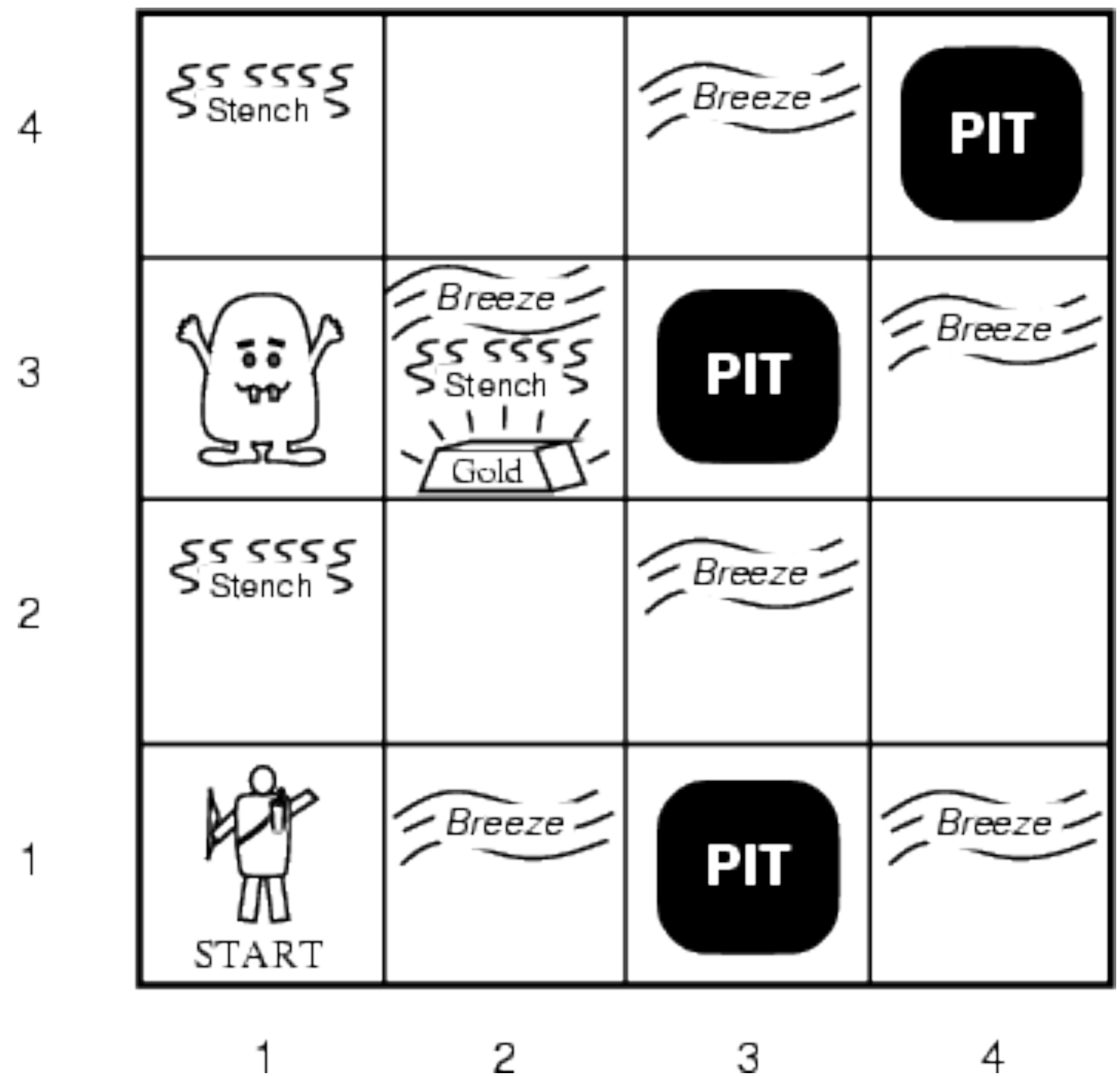
# Wumpus world characterization

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

# AIMA's Wumpus World

The agent always starts in the field [1,1]

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



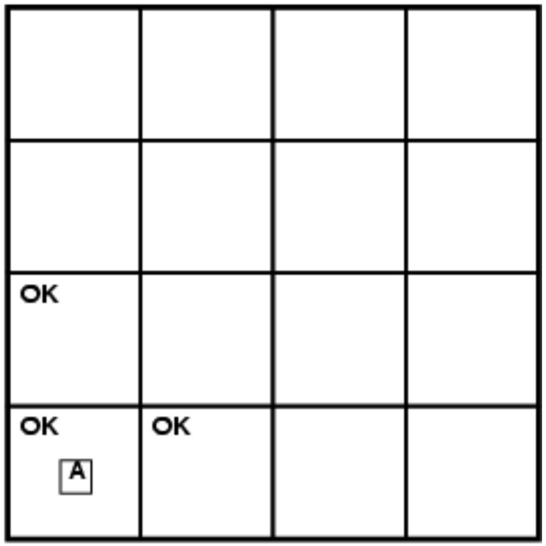
# The Hunter’s first step


(a) (b)

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

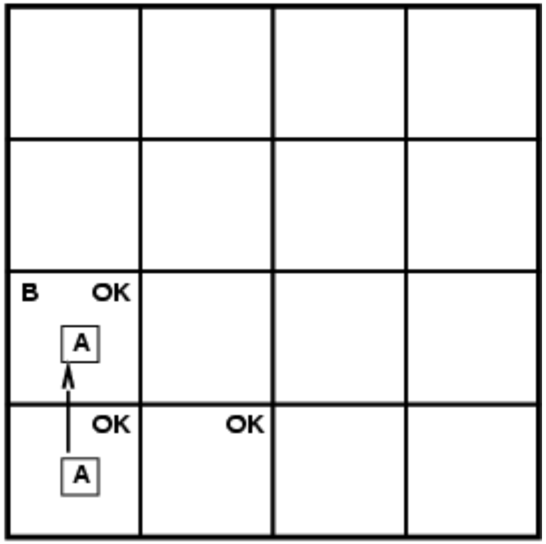
Moving to [2,1] is a safe move that reveals a breeze but no stench, implying that the Wumpus is not adjacent but that one or more pits are

# Exploring a wumpus world



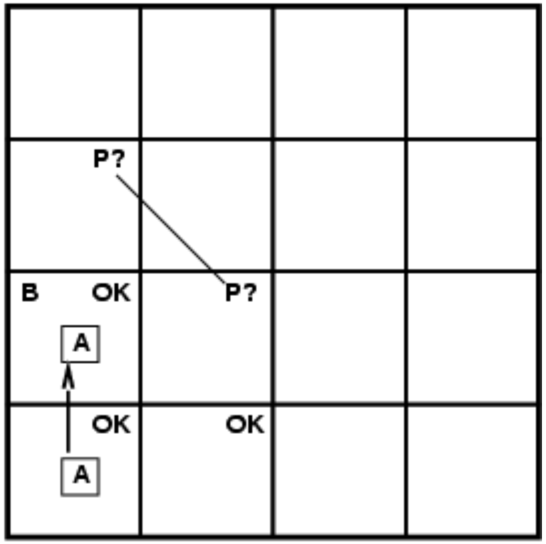

| | |
|---|---|
| A | agent |
| B | breeze |
| G | glitter |
| OK | safe cell |
| P | pit |
| S | stench |
| W | wumpus |

# Exploring a wumpus world



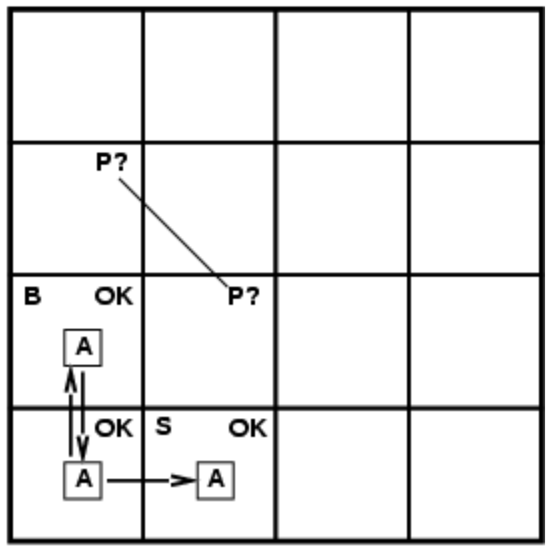

A agent
B breeze
G glitter
OK safe cell
P pit
S stench
W wumpus

# Exploring a wumpus world



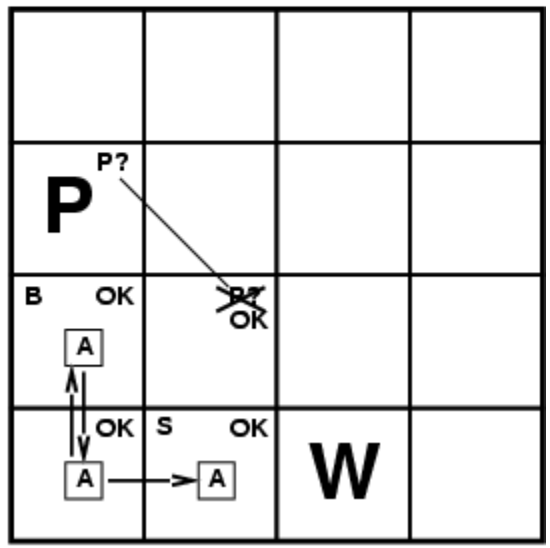

| | |
|---|---|
| A | agent |
| B | breeze |
| G | glitter |
| OK | safe cell |
| P | pit |
| S | stench |
| W | wumpus |

# Exploring a wumpus world



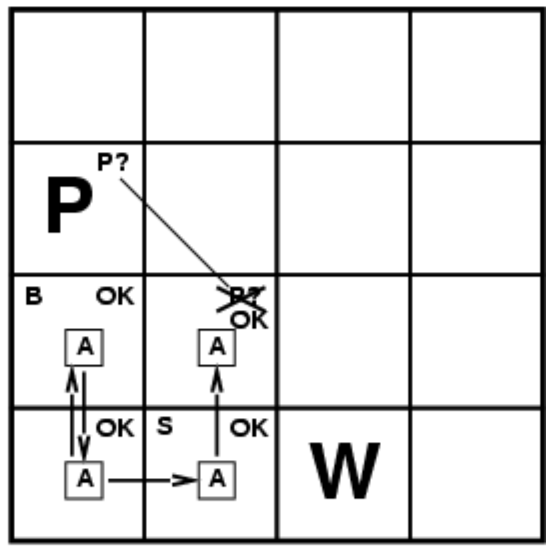

| | |
|---|---|
| A | agent |
| B | breeze |
| G | glitter |
| OK | safe cell |
| P | pit |
| S | stench |
| W | wumpus |

# Exploring a wumpus world



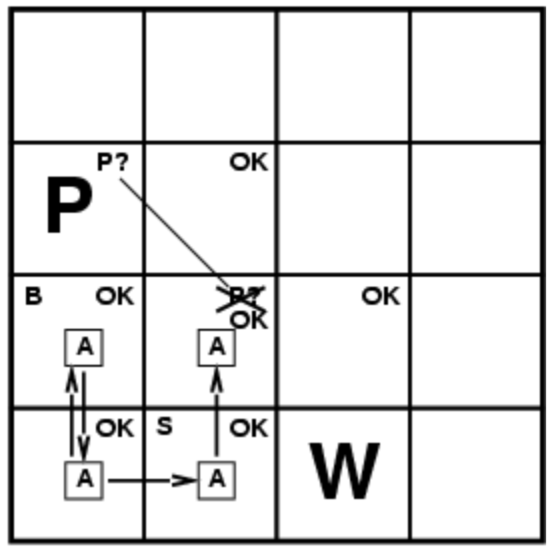

| | |
|---|---|
| A | agent |
| B | breeze |
| G | glitter |
| OK | safe cell |
| P | pit |
| S | stench |
| W | wumpus |

# Exploring a wumpus world



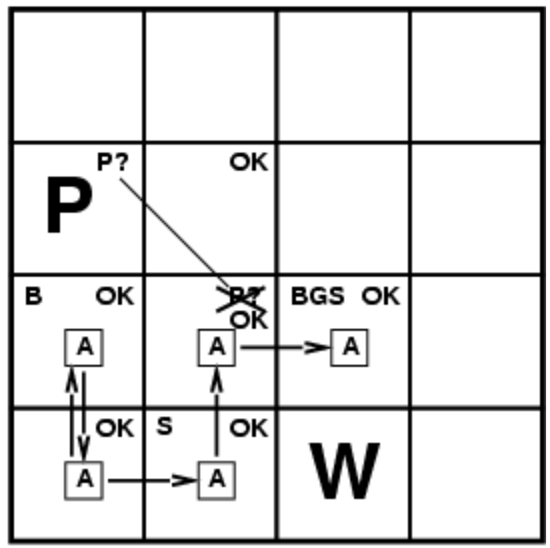

| | |
|---|---|
| A | agent |
| B | breeze |
| G | glitter |
| OK | safe cell |
| P | pit |
| S | stench |
| W | wumpus |

# Exploring a wumpus world



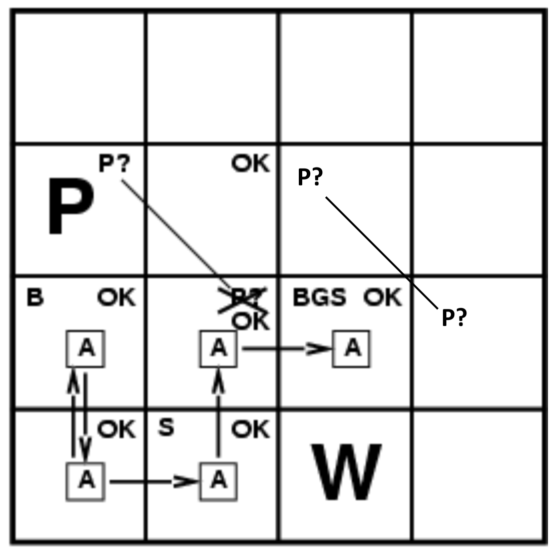

| | |
|---|---|
| A | agent |
| B | breeze |
| G | glitter |
| OK | safe cell |
| P | pit |
| S | stench |
| W | wumpus |

# 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 |

# Wumpus World games online

- AIMA code
  - Python
  - Lisp
- http://scv.bu.edu/cgi-bin/wcl – Web-based version you can play
- http://codenautics.com/wumpus/ - Mac version

# Logic in general

- **Logics** are formal languages for representing information such 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

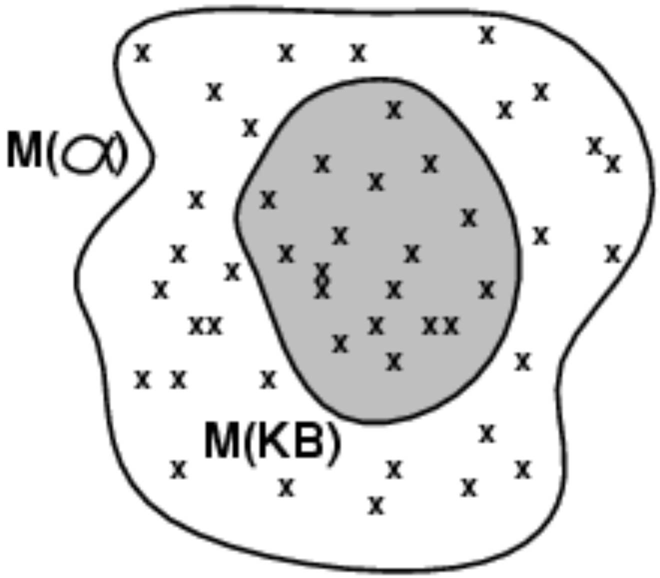
- $x+2 \geq y$ is a sentence; $x2+y > \{\}$ is not a sentence
- $x+2 \geq y$ is true iff the number $x+2$ is no less than the number $y$
- $x+2 \geq y$ is true in a world where $x = 7, y = 1$
- $x+2 \geq y$ is false in a world where $x = 0, y = 6$
- $x+1 > x$ is true for all numbers $x$

# Entailment

- Entailment means that one thing follows from another:
- KB $\models$ α
- 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 sentences (i.e., syntax) that is based on **semantics**

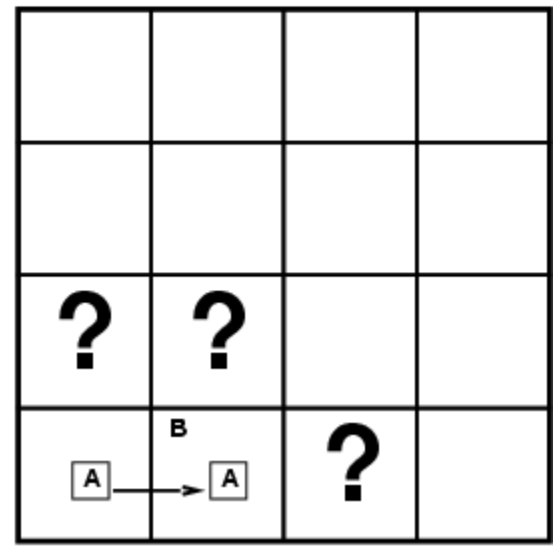
# Models

- Logicians typically think in terms 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*
- *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



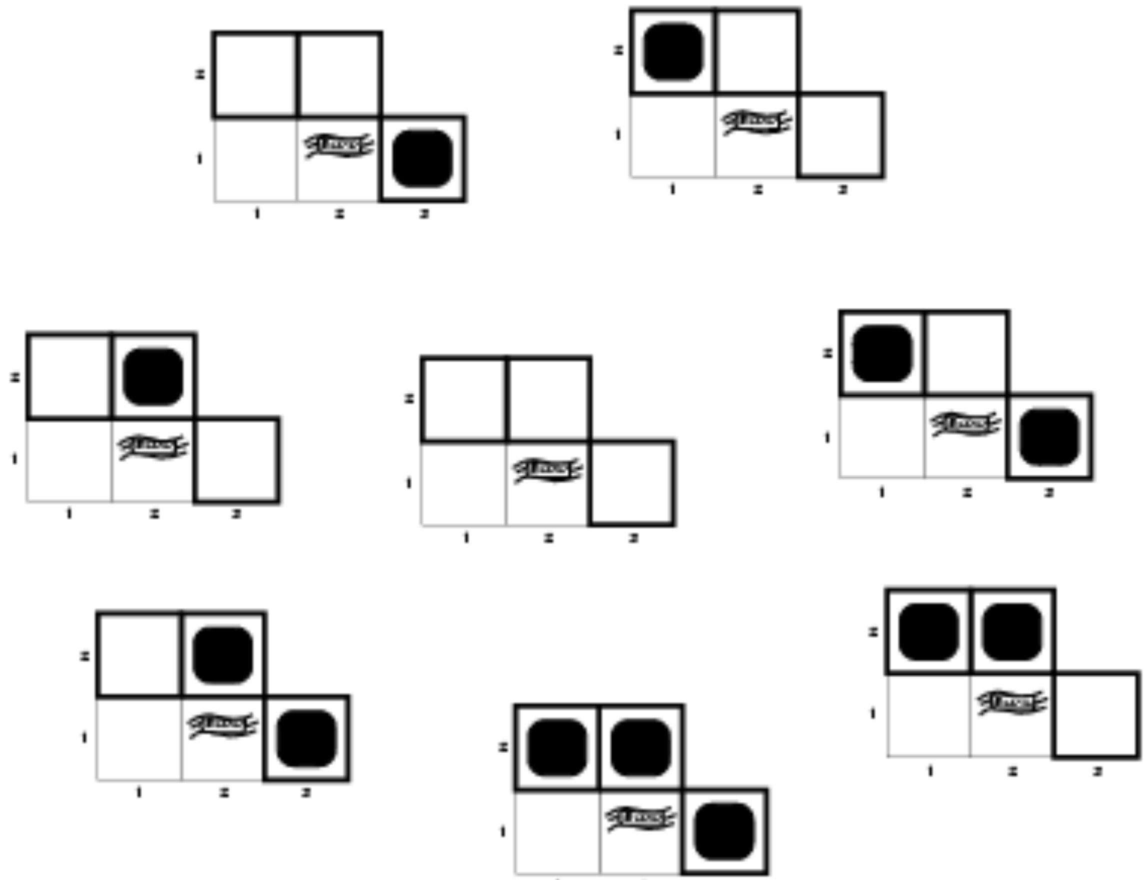
# Entailment in the wumpus world

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



# 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 |



# 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

…

*these also follow*

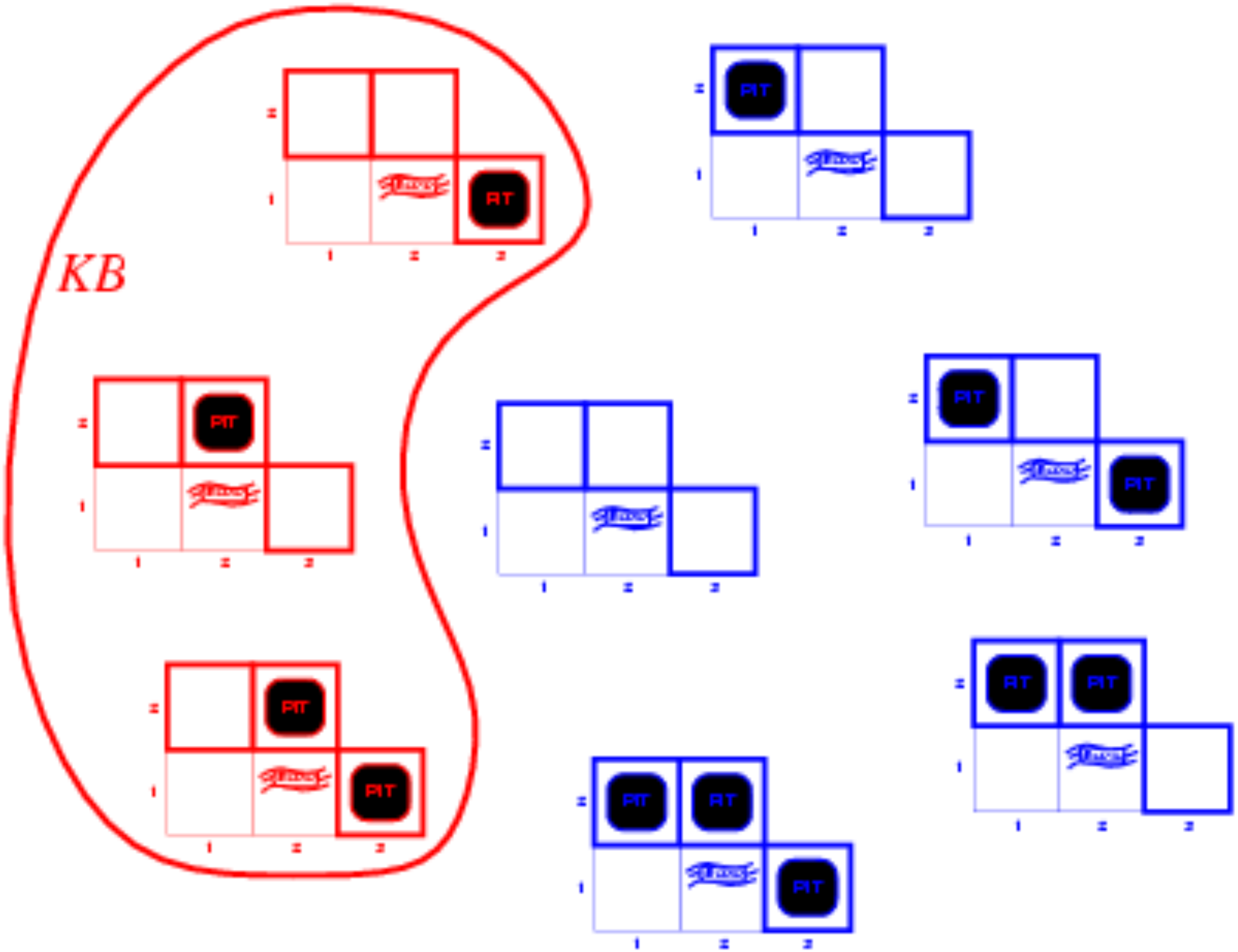
~B21 => ~P11

~B12 => ~P11

~B11 => ~P21

~B22 => ~P21

…




*KB* = wumpus-world rules + observations

# Wumpus World Rules (2)

- Cell safe if it has neither a pit or wumpus

  OK11 => ~P11 ∧ ~W11

  OK12 => ~P12 ∧ ~W12

  …

- From which we can derive

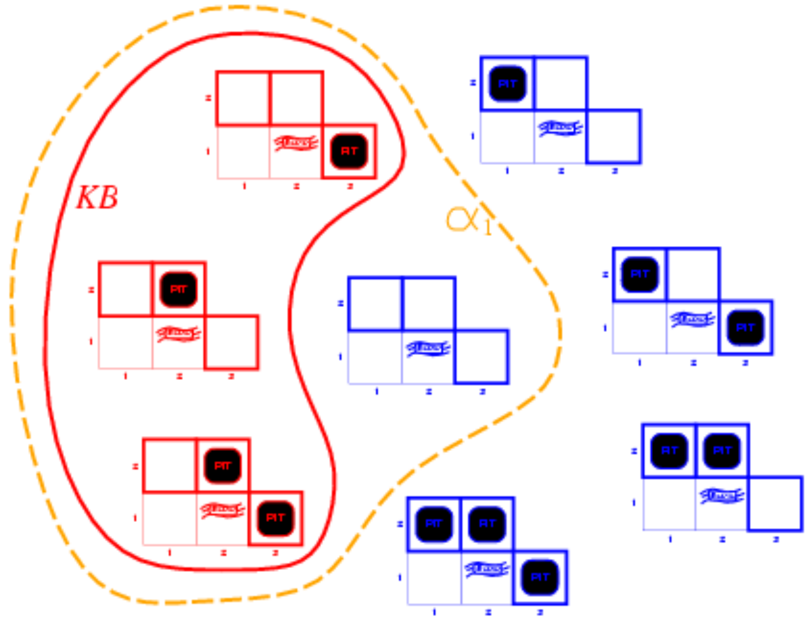
  P11 ∨ W11 => ~OK11

  P11 => ~OK11

  W11 => ~OK11

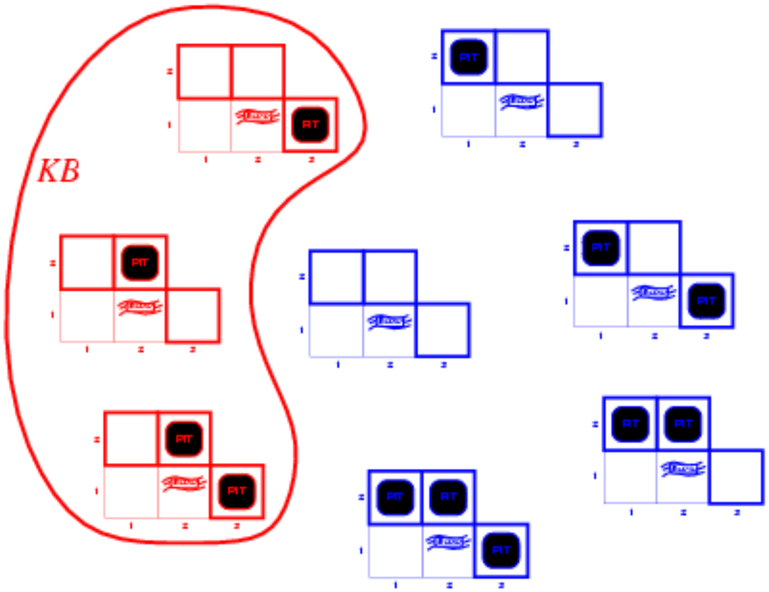
  …

# Wumpus models



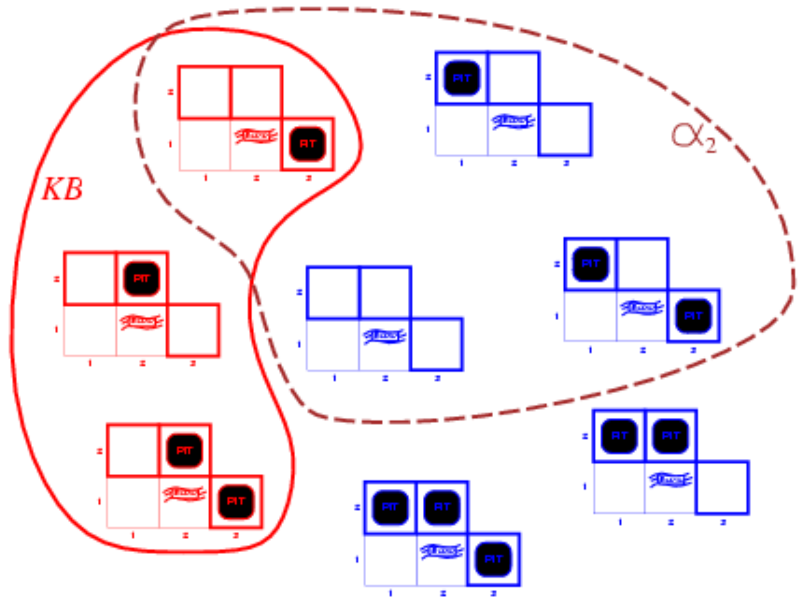

- *KB* = wumpus-world rules + observations
- $\alpha_1$ = "[1,2] is safe"
- *Since all models include* $\alpha_1$
- *KB* $\models \alpha_1$, proved by **model checking**

# Wumpus models




- *KB* = wumpus-world rules + observations

# Wumpus models




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

# Inference, Soundness, Completeness

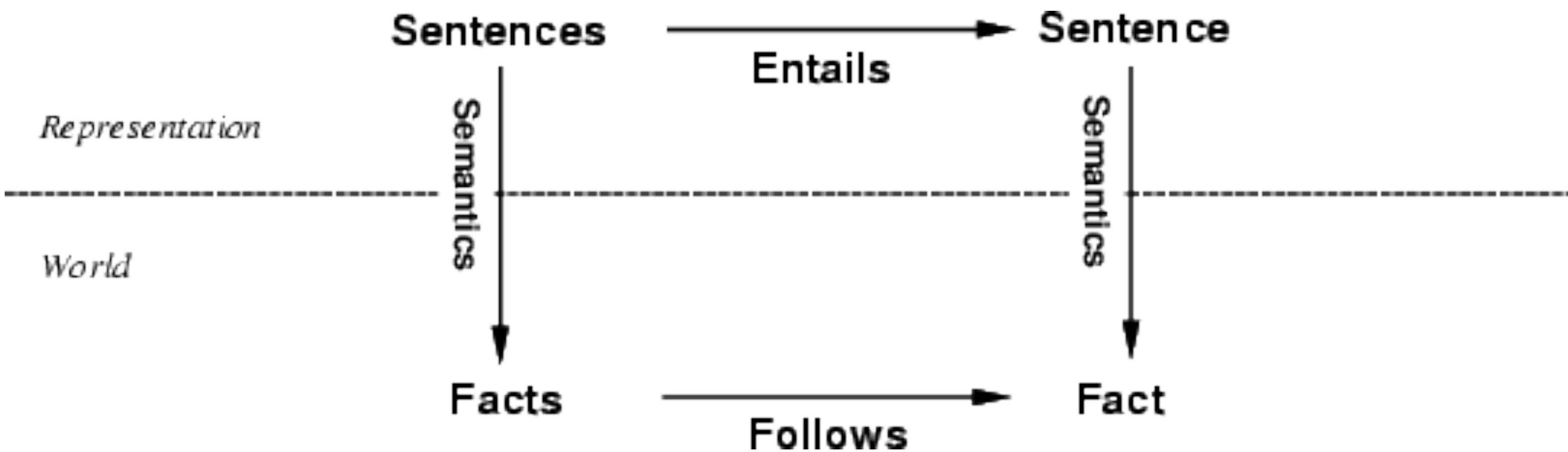
- $KB \vdash_i \alpha$ = sentence α can be derived from *KB* by procedure *i*
- **Soundness:** *i* is sound if whenever $KB \vdash_i \alpha$, it is also true that $KB \models \alpha$
- **Completeness:** *i* is complete if whenever $KB \models \alpha$, it is also true that $KB \vdash_i \alpha$

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

# Representation, reasoning, and logic

- Object of knowledge representation (KR): express knowledge in a **computer-tractable** form, so that agents can perform well
- A KR language is defined by:
  - **Syntax:** defines all possible sequences of symbols that constitute sentences of the language
    - Ex: Sentences in a book, bit patterns in computer memory
  - **Semantics**: determines facts in the world to which the sentences refer
    - Each sentence makes a claim about the world.
    - An agent is said to believe a sentence about the world.

# The connection between sentences and facts

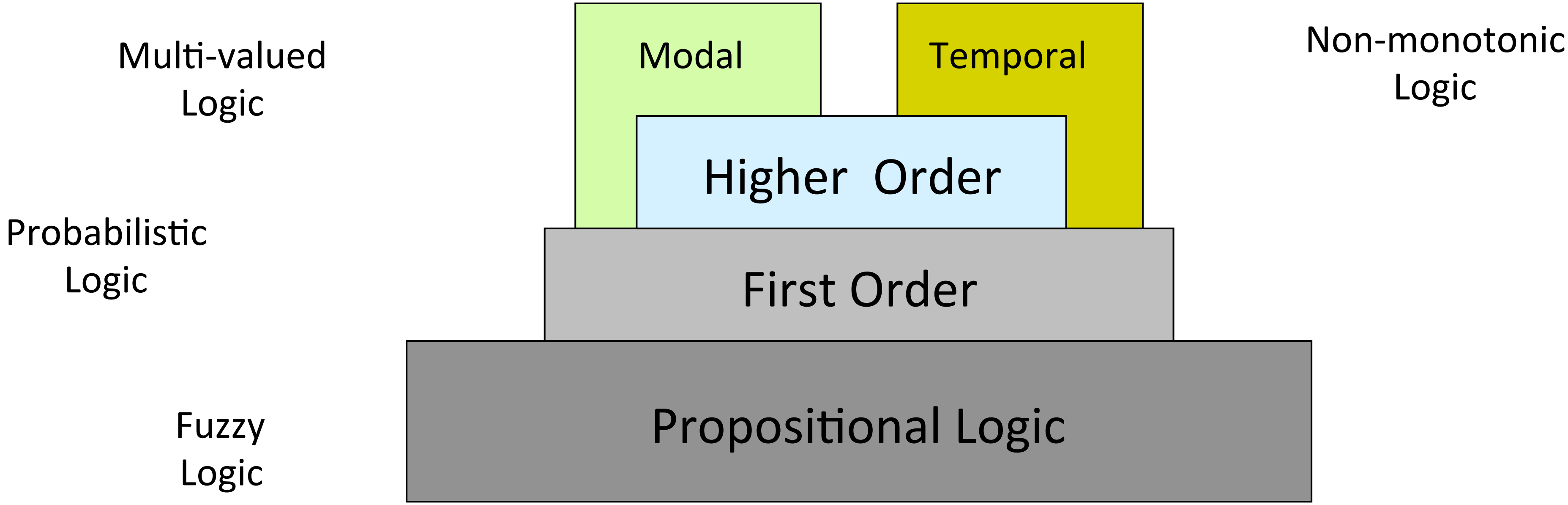



*Semantics* maps sentences in logic to facts in the world. The property of one fact following from another is mirrored by the property of one sentence being entailed by another.

# 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

# Logic as a KR language



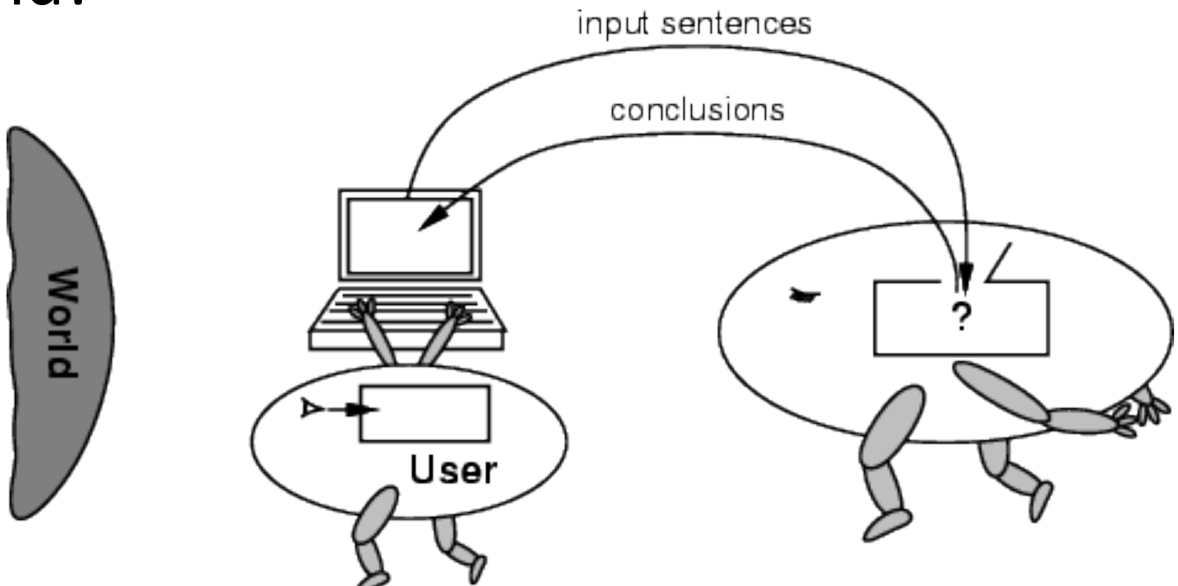
# Ontology and epistemology

- **Ontology** is the study of what there is—an inventory of what exists. An ontological commitment is a commitment to an existence claim.
- **Epistemology** is a major branch of philosophy that concerns the forms, nature, and preconditions of knowledge.

| Language | Ontological Commitment (What exists in the world) | Epistemological Commitment (What an agent believes about facts) |
|---|---|---|
| Propositional logic | facts | true/false/unknown |
| First-order logic | facts, objects, relations | true/false/unknown |
| Temporal logic | facts, objects, relations, times | true/false/unknown |
| Probability theory | facts | degree of belief 0…1 |
| Fuzzy logic | degree of truth | degree of belief 0…1 |

# 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 the world for making good decisions
- Agent's knowledge stored in a knowledge base (KB) as **sentences** in a knowledge representation (KR) language
- A knowledge-based agent needs a **KB** and an **inference mechanism**. It operates by storing sentences in its KB, inferring new sentences and using them to deduce which actions to take
- A **representation language** is defined by its syntax and semantics, which specify structure of sentences and how they relate to facts of the world
- The **interpretation** of a sentence is fact to which it refers. If the fact is part of the actual world, then the sentence is true