Lecture 15

Knowledge-based Agents

Lusi Li

Department of Computer Science

ODU

Reading for This Class:

Chapter 7, Russell and Norvig

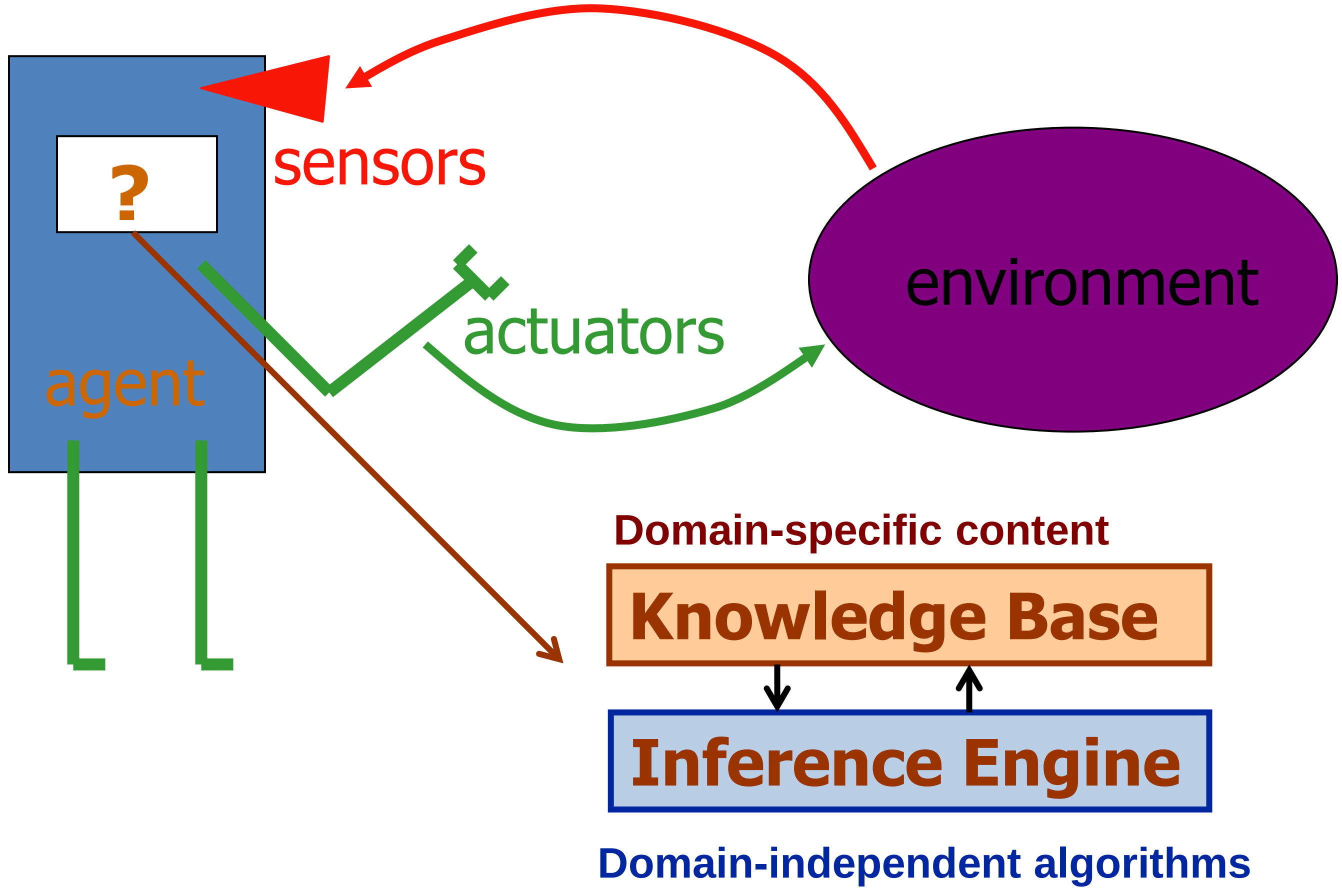
Knowledge-based Agents

- Knowledge-based agents use a process of reasoning over an internal representation of knowledge to decide what actions to take

- A knowledge-based agent includes a knowledge base and an inference engine
- Knowledge base (KB)
  - A set of sentences
    - expressed in a knowledge representation language
    - represent some assertion (or actual facts) about the world
  - Usually starts with some background knowledge
    - can be general (world knowledge) or specific (domain)
- Inference engine
  - Derive new sentences from old

- Actions of an agent
  - TELL
  - ASK

Knowledge-Based Agent



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 TELLS the KB which action is selected and then performs that action.

A Simple Knowledge-Based Agent Program

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

  TELL(KB, $\boxed{\text{MAKE-PERCEPT-SENTENCE}(percept, t)}$)
  $action \leftarrow$ ASK(KB, $\boxed{\text{MAKE-ACTION-QUERY}(t)}$)
  TELL(KB, $\boxed{\text{MAKE-ACTION-SENTENCE}(action, t)}$)
  $t \leftarrow t + 1$
  return *action*

- **Details hidden in three functions**
- **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**

Levels of a Knowledge-based Agent

- **Knowledge Level:**
  - In this level, the behavior of an agent is decided by specifying the following
    - The agent's current knowledge it has perceived.
    - The goal of an agent.

- **Logical Level:**
  - This level is the logical representation of the knowledge level
  - Sentences are encoded in various logics at this level. At the logical level, knowledge is encoded into logical statements.

- **Implementation Level:**
  - This level is the physical representation of logic and knowledge.
  - Here, it is understood that "how the knowledge-based agent actually implements its stored knowledge."

Types of Knowledge

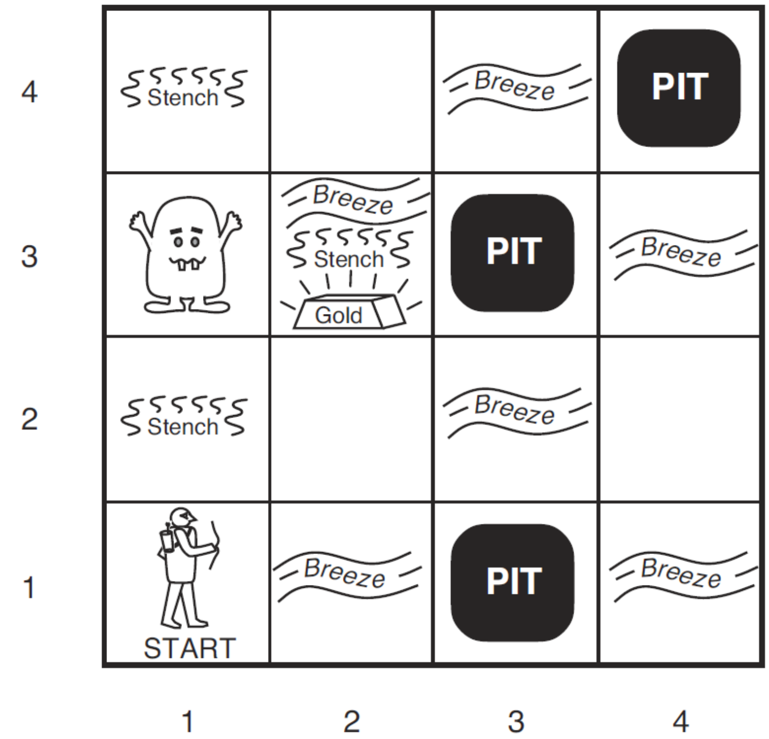
- **Declarative knowledge, e.g.: constraints**
  - **Knowledge about facts and things**
  - **The agent designer TELLS sentences to the empty system one by one until the system becomes knowledgeable enough to deal with the environment.**
  - **It can be used to perform many different sorts of inferences**

- **Procedural, e.g.: functions**
  - **Knowledge about how to do something**
    - such as reasoning, decision making, and problem solving
  - **Such knowledge can only be used in one way -- by executing it**

- **Approaches to designing a knowledge-based agent:**
  - **Declarative approach**
  - **Procedural approach**

Example: A Wumpus World

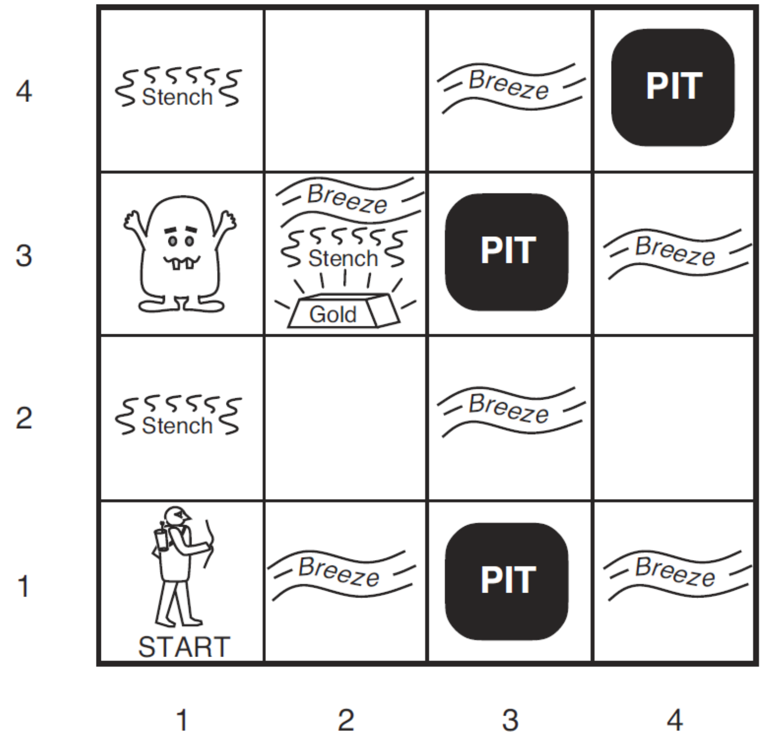
- **A Wumpus World**
  - **Cave of 4×4 grid of rooms**
  - **Wumpus: A deadly beast who kills anyone entering his room.**
  - **Pits: Bottomless pits that will trap you forever.**
  - **Gold**
  - **Agent always starts in [1,1]**
  - **The task of the agent is to find the gold, return to [1,1] and climb out of the cave.**



A Wumpus World PEAS Description

- **Performance measure**
  - +1000 for picking up gold
  - -1000 get falling into pit or eaten by the wumpus
  - -1 per step
  - -10 for using the (only) arrow

- **Environment: 4 x 4 grid of rooms**
  - 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 (shot Wumpus)]**

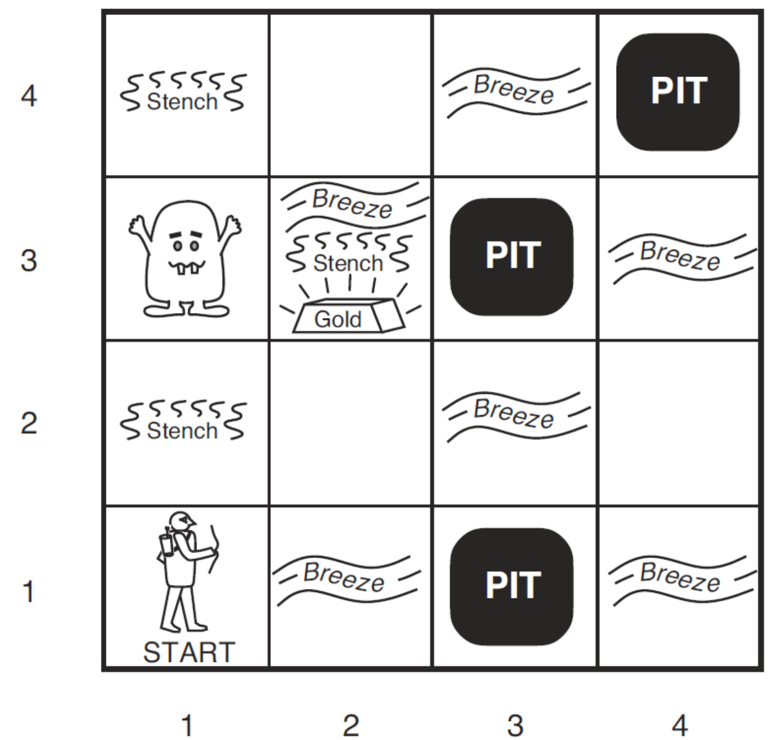
- **Actuators: Left turn, Right turn, Forward, Grab, Release, Shoot**

Reasoning in the Wumpus World

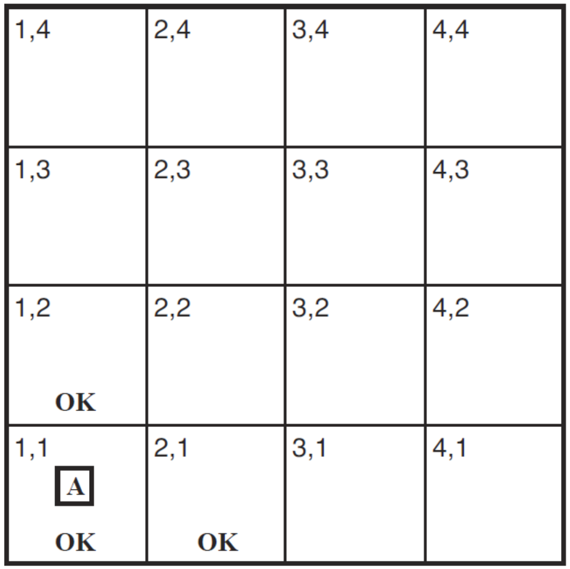
- **Agent has initial ignorance about the configuration**
  - **Agent knows his/her initial location, i.e., [1,1]**
  - **It's safe at [1,1]**
  - **Agent knows the rules of the environment**

- **Goal is to explore environment, make inferences (reasoning) to try to find the gold.**

- **Random instantiations of this problem used to test agent reasoning and decision algorithms.**

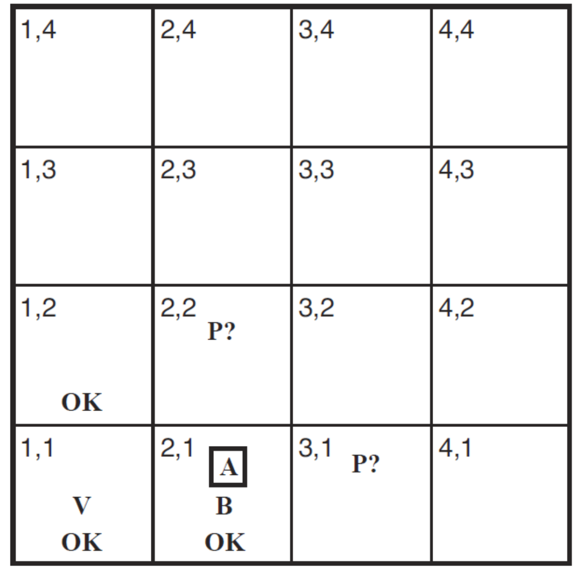


Exploring the Wumpus World



(a)

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



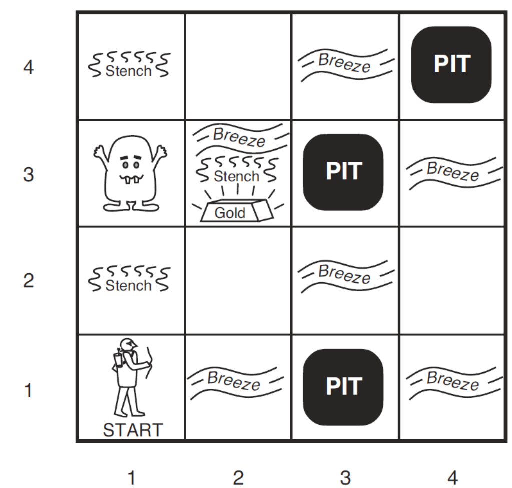
(b)

[1,1] The KB initially contains the rules of the environment.

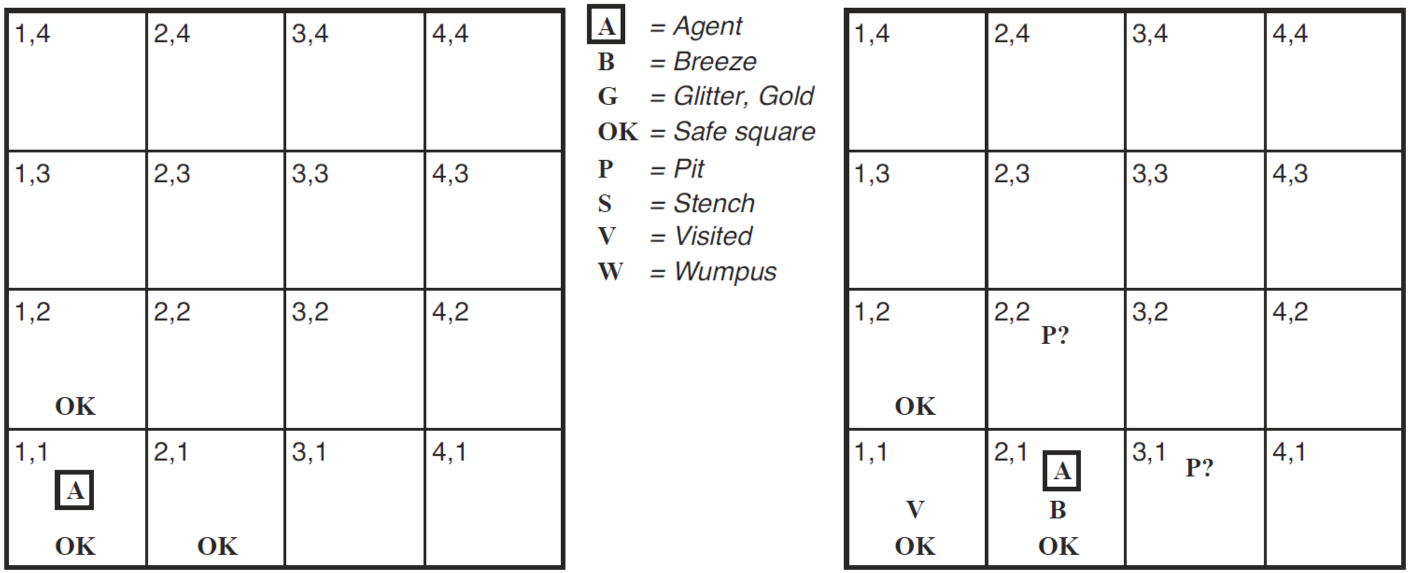
[Stench, Breeze, Glitter, Bump, Scream]

The first percept is [*none, none, none, none, none*],

move to safe cell e.g. [2,1]



Exploring the Wumpus World



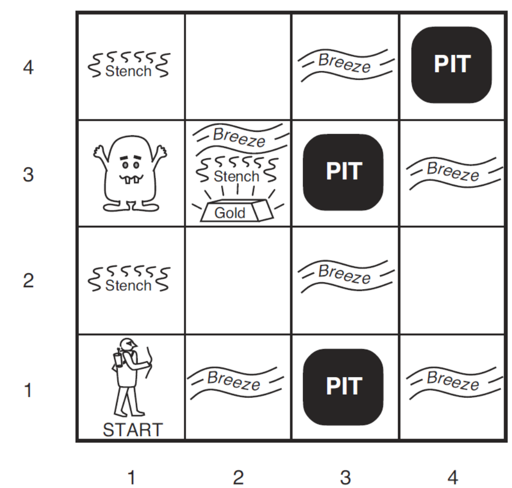
(a) (b)

[2,1] = breeze

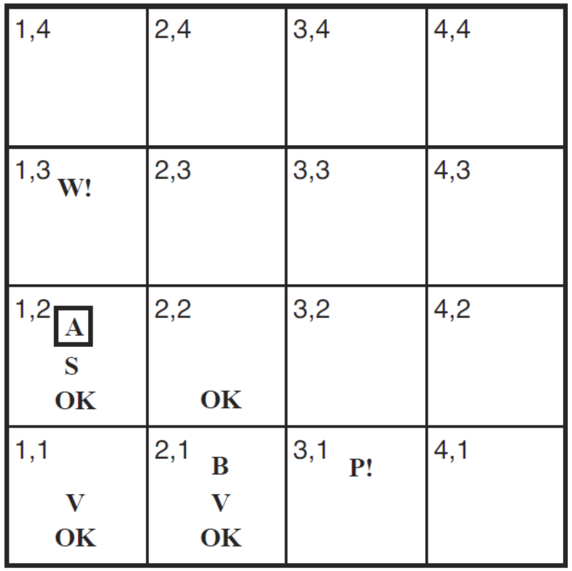
The second percept is [*none, breeze, none, none, none*]

indicate that there is a pit in [2,2] or [3,1] or both

return to [1,1] to try next safe cell [1,2]

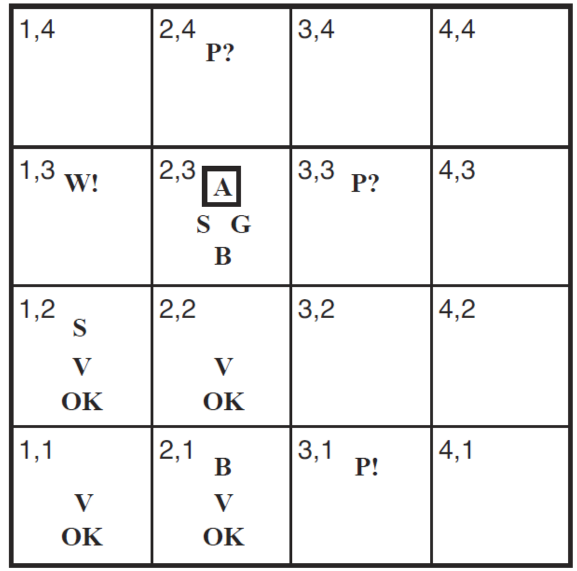


Exploring the Wumpus World



(a)

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



(b)

[1,2] = stench

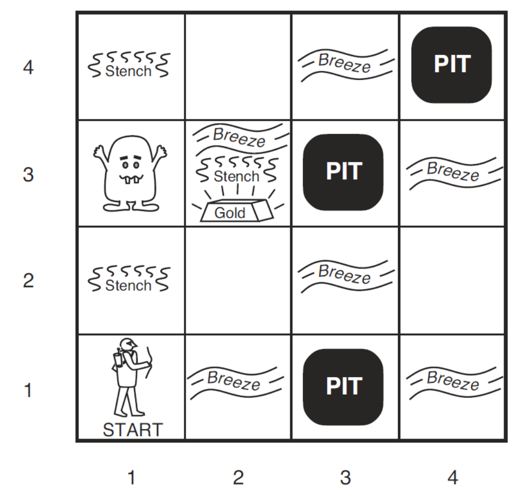
The third percept is [stench, none, none, none, none]

indicate that Wumpus is in [2,2] or [1,3] or both

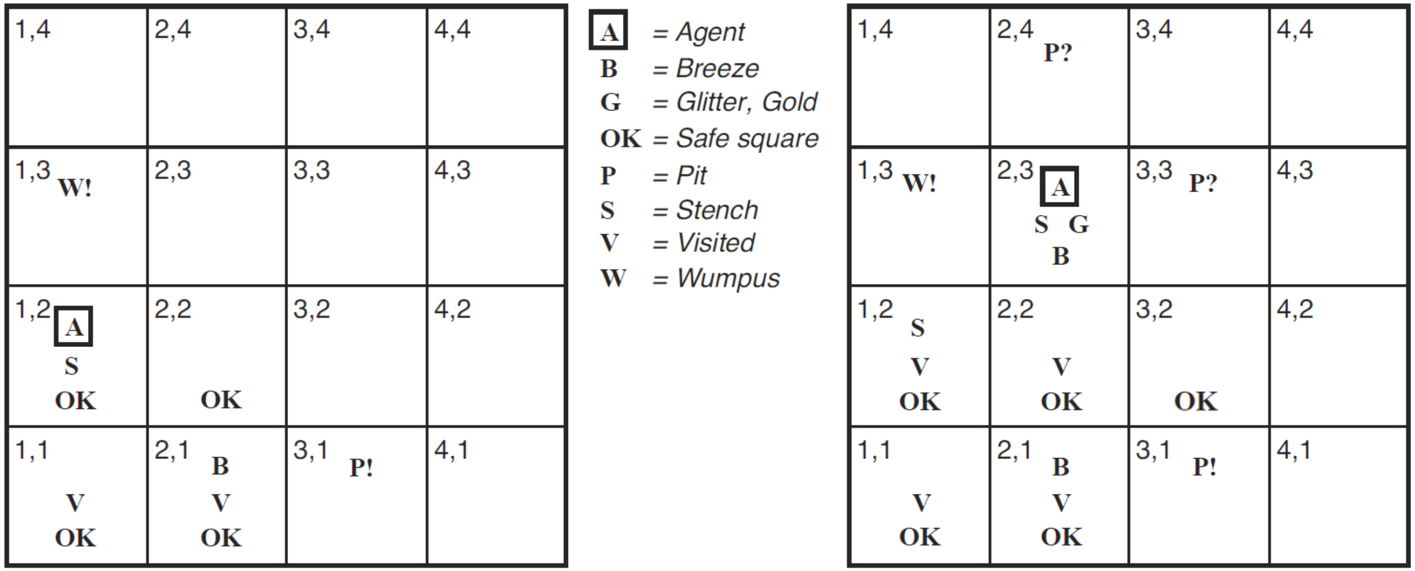
Yet not in [2,2] otherwise stench would have been detected in [2,1]

THUS wumpus is in [1,3]
THUS [2,2] is safe because of lack of breeze in [1,2]
THUS pit in [3,1] (again a clever inference)

move to next safe cell [2,2]



Exploring the Wumpus World

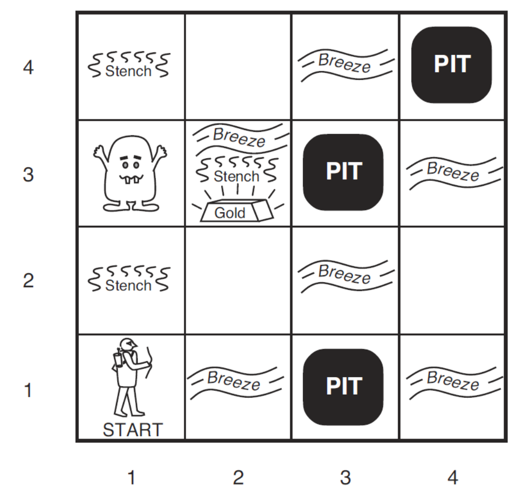


(a) (b)

[2,2] = none, we assume the agent moves to [2,3]

The fifth percept is [stench, breeze, glitter, none, none]

detect a glitter, so it should grab the gold and then return home.



What our example has shown us

- Can represent general knowledge about an environment by a set of rules and facts

- Can gather evidence (percept) and then infer new facts by combining evidence with the rules (KB)

- The conclusions are guaranteed to be correct if
  - The evidence is correct
  - The rules are correct
  - The inference procedure is correct
    → logical reasoning

- The inference may be quite complex
  - E.g., evidence at different times, combined with different rules, etc.

Logic in general

- Logics are formal languages for representing information such that conclusions can be drawn

- Syntax defines what sentences are legal (well-formed)
  - E.g., arithmetic
  - $x+2 \geq y$ is a sentence; $x2+y > ()$ is not a sentence

- Semantics define the "meaning" of sentences
  - i.e., define truth of each sentence w.r.t to each possible world
  - E.g.,
  - $x+2 \geq y$ is true in a world where $x = 7$, $y = 1$
  - $x+2 \geq y$ is false in a world where $x = 0$, $y = 6$

- In standard logics, every sentence must be either true or false in each possible world – there is no in between.

More on Possible Worlds

- m is a model of a sentence α if α is true in m
- M(α) is the set of all models of α
- Possible worlds ~ models
  - Possible worlds: potentially real environments
  - Models: mathematical abstractions that establish the truth or falsity of every sentence
- Example:
  - x + y = 4, where x = # men, y = # women
  - Possible models = all possible assignments of integers to x and y.
  - For CSPs, possible model = complete assignment of values to variables.

Entailment

- **Entailment means that a sentence follows logically from another:**

$$\alpha \models \beta \text{ iff } M(\alpha) \subseteq M(\beta)$$

- **A sentence α entails another sentence β in every model α is true, β must also be true**
  - **Note the direction of the ⊆ here: if α |= β, then α is a stronger assertion than β: it rules out more possible worlds**
  - **What if α is false? since no model makes "false" true, M(α) is ∅ (∅ is a sub set of any set) , then β can be either true or false.**

- **Examples:**
  - **(x = 0) |= (xy = 0) ?**
  - **(p = True) |= (p ∨ q) ?**
  - **(p ∧ q) |= (p ∨ q) ?**
  - **False |= True ?**
  - **True |= False ?**

Entailment

- Entailment means that a sentence follows logically from another:

$$\alpha \models \beta \text{ iff } M(\alpha) \subseteq M(\beta)$$

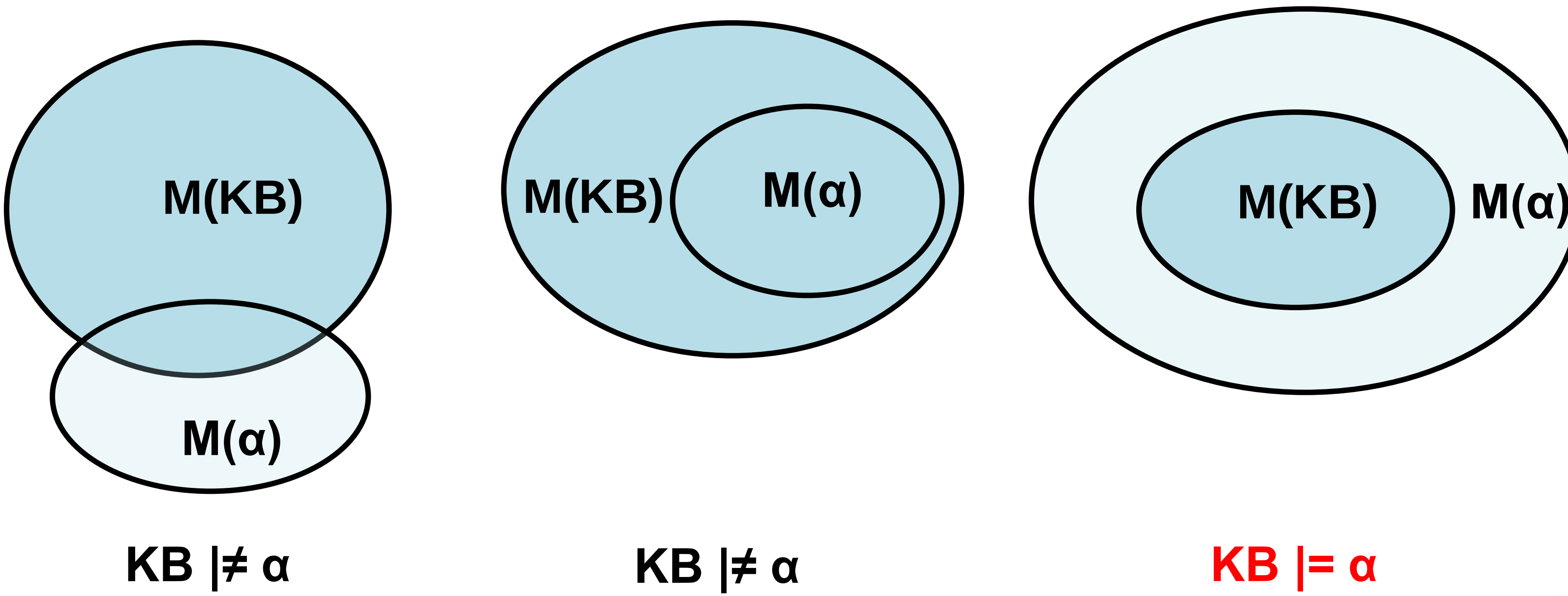
- A sentence α entails another sentence β in every model α is true, β must also be true
  - Note the direction of the ⊆ here: if α |= β, then α is a stronger assertion than β: it rules out more possible worlds
  - What if α is false? since no model makes "false" true, M(α) is ∅ (∅ is a sub set of any set) , then β can be either true or false.

- Examples:
  - (x = 0) |= (xy = 0) ? Yes
  - (p = True) |= (p ∨ q) ? Yes
  - (p ∧ q) |= (p ∨ q) ? Yes
  - False |= True ? Yes
  - True |= False ? No

Entailment

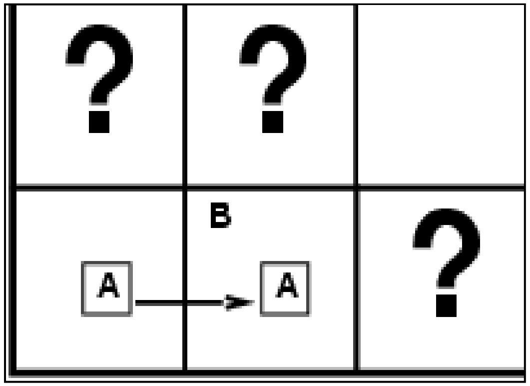
$KB \models \alpha$ *iff* $M(KB) \subseteq M(\alpha)$

- KB entails sentence α iff α is true in all worlds where KB is true
- E.g., KB = “the Phillies won” and “the Reds won” entails α = “Either the Phillies won or the Reds won”



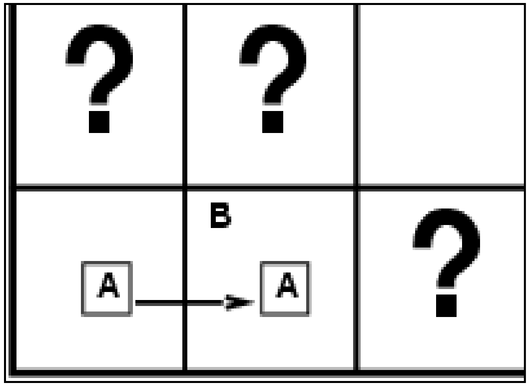
Wumpus World

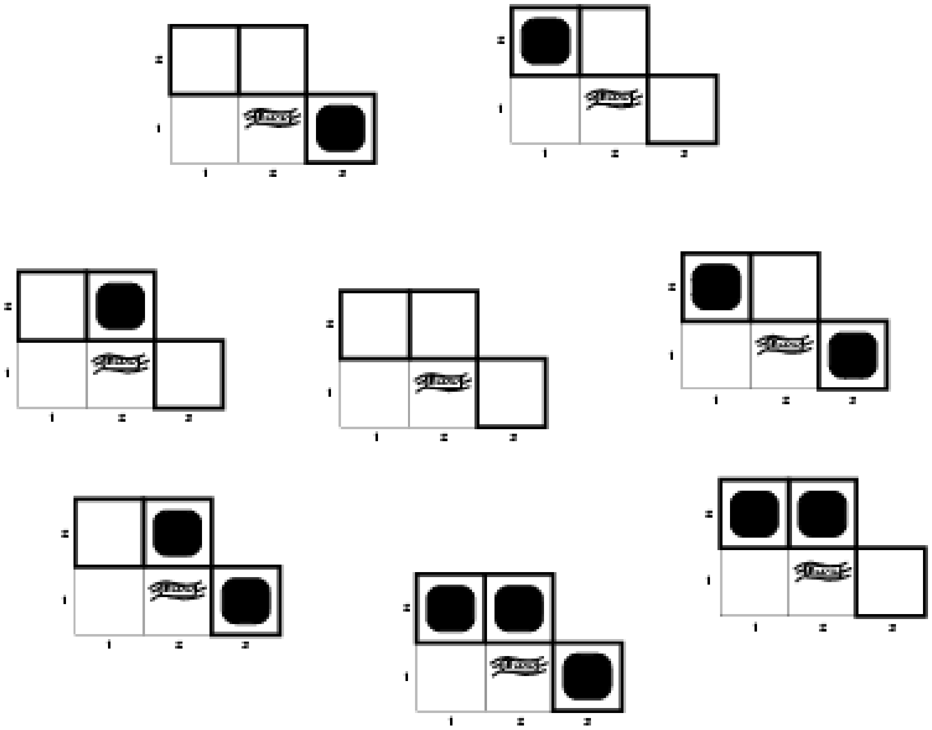
- Consider possible models for *KB* assuming only pits in a reduced Wumpus world
- Situation after detecting nothing in [1,1], moving right, detecting breeze in [2,1]
- The agent wants to know if the adjacent squares contain pits.



Wumpus World

- **All possible models in this reduced Wumpus world.**





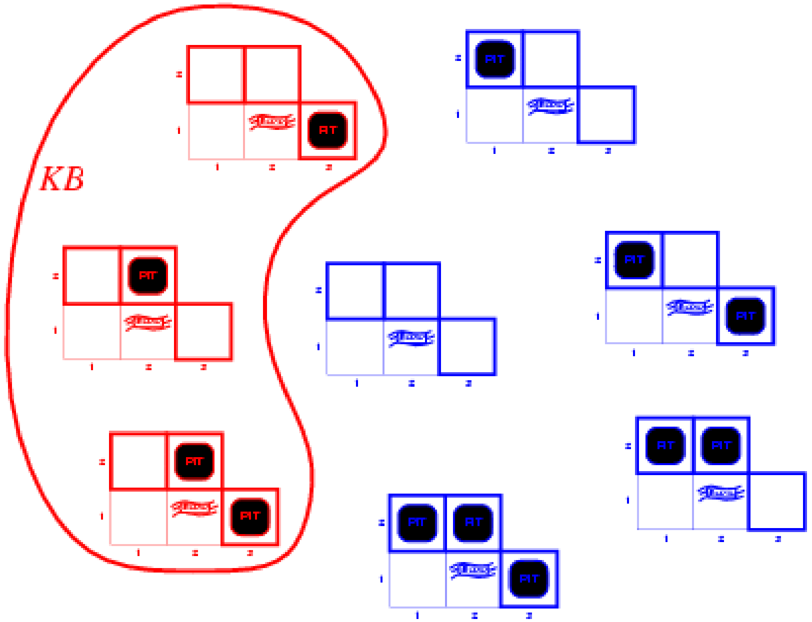
Inferring Conclusions

- Consider 2 possible conclusions given a KB
  - $\alpha 1$ = "[1,2] is safe"
  - $\alpha 2$ = "[2,2] is safe"

- One possible inference procedure
  - Start with KB
  - Model-checking
    - Check if KB $\models \alpha$ by checking if in all possible models where KB is true that α is also true

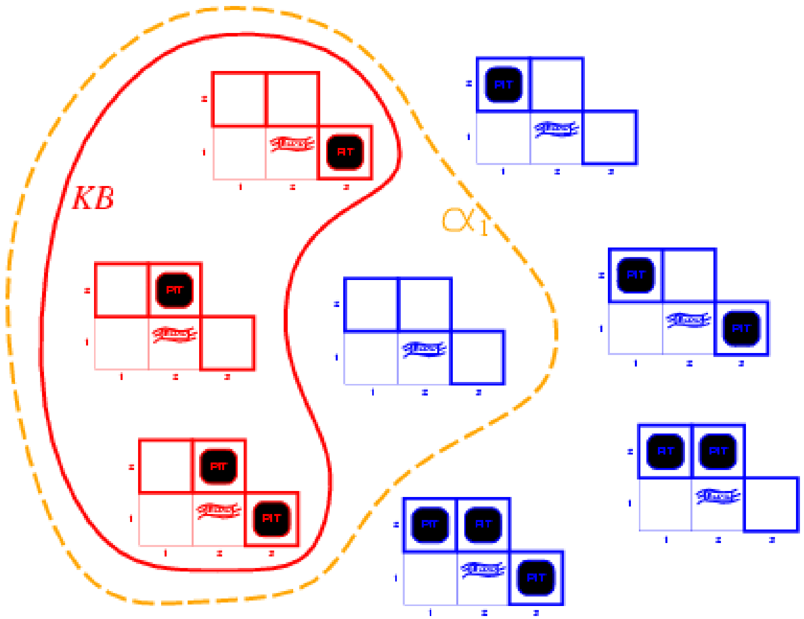
- Comments:
  - Model-checking enumerates all possible worlds
    - Only works on finite domains, will suffer from exponential growth of possible models

Wumpus World



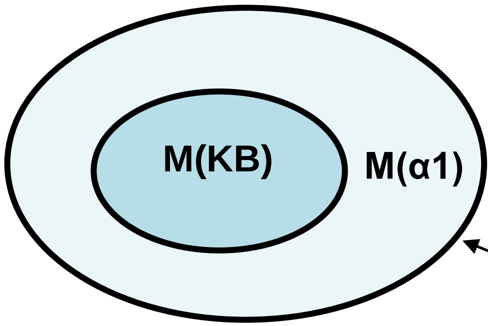
- **KB = wumpus-world rules + observations**

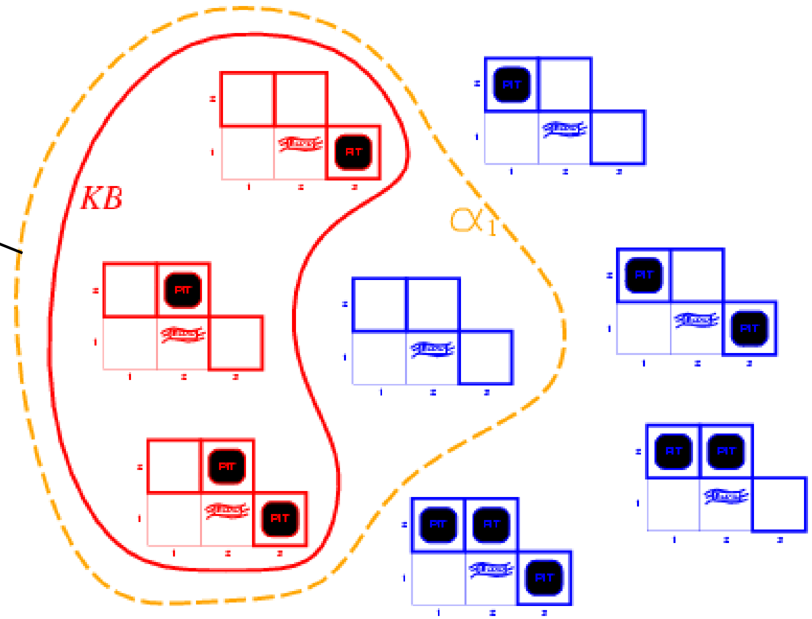
Entailment in the Wumpus World



- KB = wumpus-world rules + observations
- α1 = "[1,2] is safe"

Entailment in the Wumpus World

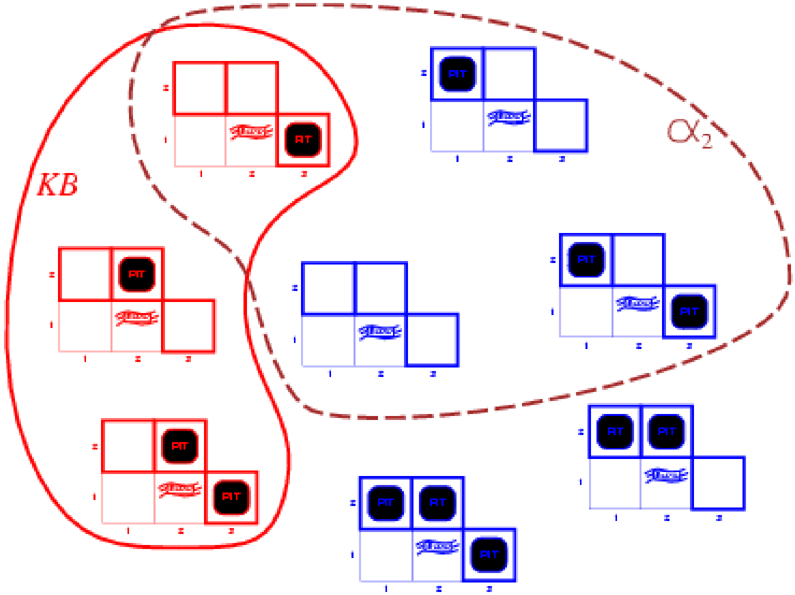




KB |= α1

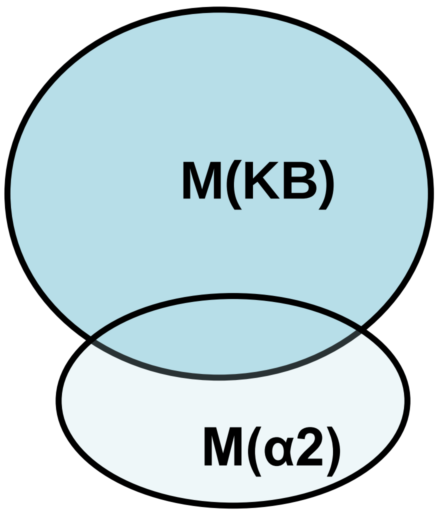
- **KB = wumpus-world rules + observations**
- **α1 = "[1,2] is safe”**
- **KB ╞ α1, proved by model checking**

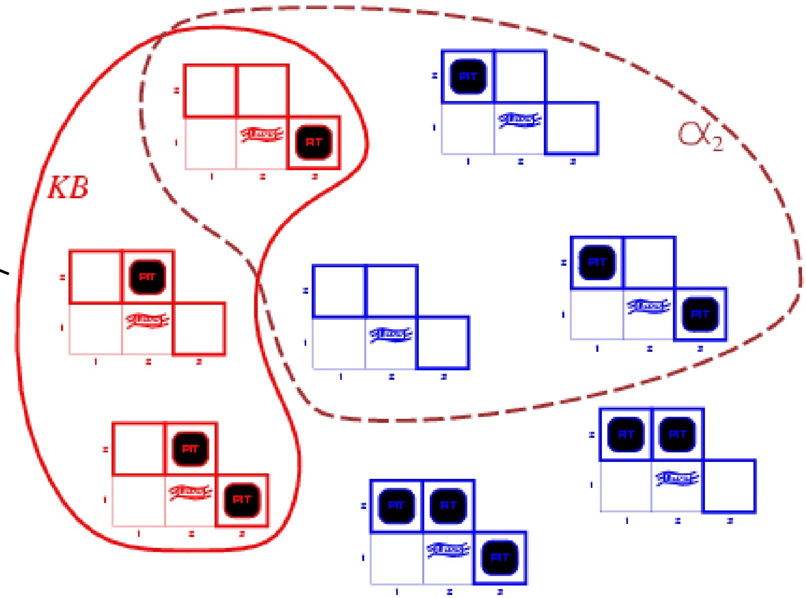
Entailment in the Wumpus World



- KB = wumpus-world rules + observations
- α2 = "[2,2] is safe"

Entailment in the Wumpus World





KB |≠ α2

- **KB = wumpus-world rules + observations**
- **α2 = "[2,2] is safe", KB |≠ α2**
  - **There are some models entailed by KB where α2 is false.**
- **The examples not only show entailment, but also how entailment can be used to derive conclusions.**

Logical Inference

- Inference is a procedure that allows new sentences to be derived from a knowledge base.

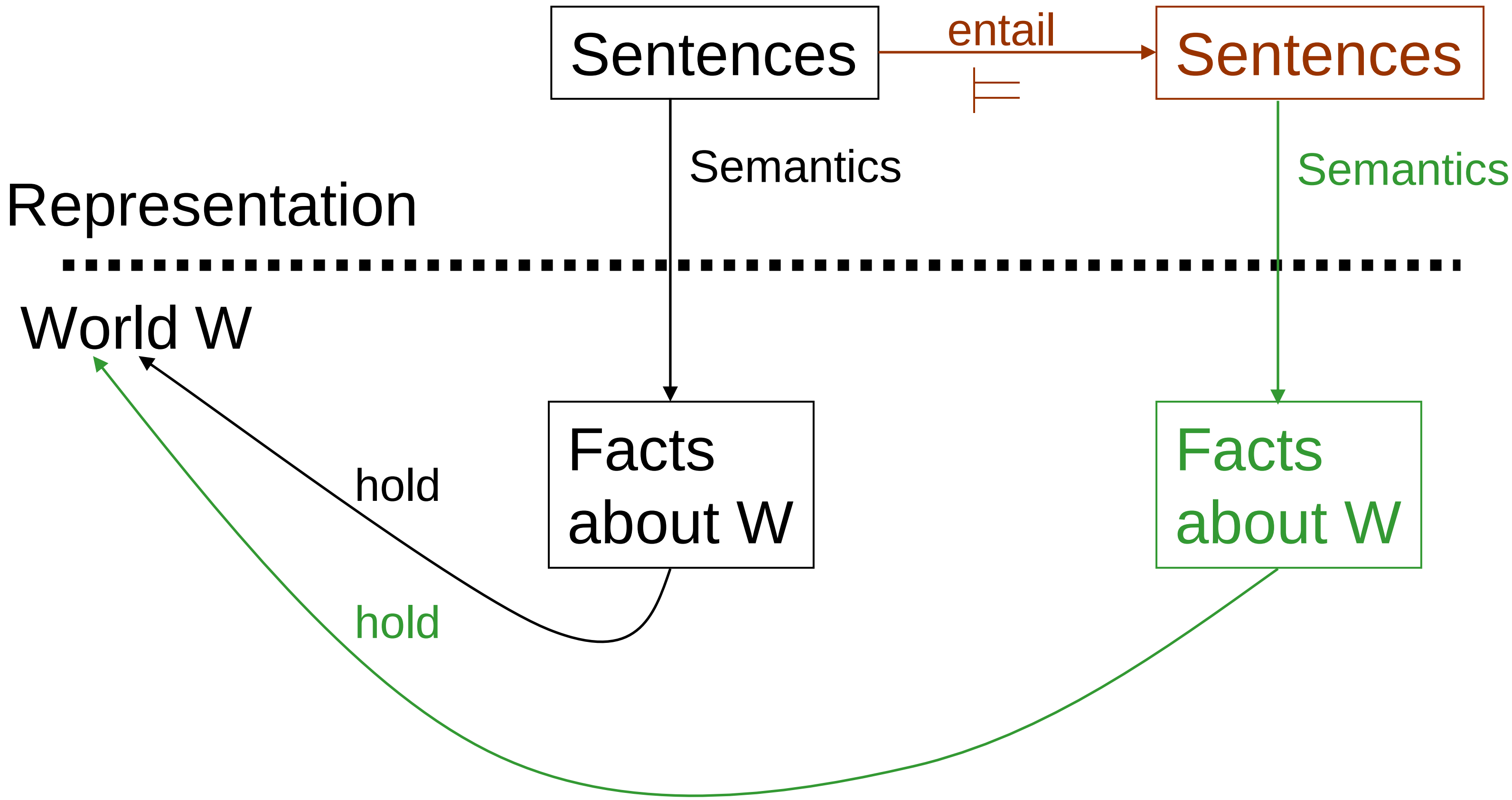
- KB $|\!-_i$ α
  - sentence α can be derived from KB by the inference procedure i

- An inference procedure is sound or truth preserving if it only derives entailed sentences.

- An inference procedure is complete if it can derive any sentence that is entailed.

- E.g., model-checking is sound and complete
  - enumerate all possible models and check whether α is true.

Connection World-Representation



If KB is true in the real world, then any sentence α derived from KB by a sound inference procedure is also true in the real world.

Propositional logic: Syntax

- **Propositional logic is the simplest logic**
- **Syntax defines allowable sentences.**

Sentences:

- **Atomic sentences = single proposition symbols**
  - **E.g., P, Q, R etc are sentences**
  - **Special cases: True = always true, False = always false**

- **Complex sentences are combined by connectives (operators) :**
  - **If P is a sentence, $\neg$P is a sentence (negation)**
  - **If P and Q are sentences, P $\wedge$ Q is a sentence (conjunction)**
  - **If P and Q are sentences, P $\vee$ Q is a sentence (disjunction)**
  - **If P and Q are sentences, P $\Rightarrow$ Q is a sentence (implication)**
  - **If P and Q are sentences, P $\Leftrightarrow$ Q is a sentence (biconditional)**

- **Literal: atomic sentence or negated atomic sentence**

Syntax Summary

A BNF (Backus–Naur Form) grammar of sentences in propositional logic, along with operator precedence, from highest to lowest.

$$
\begin{aligned}
Sentence &\rightarrow AtomicSentence \mid ComplexSentence \\
AtomicSentence &\rightarrow True \mid False \mid P \mid Q \mid R \mid \ldots \\
ComplexSentence &\rightarrow (\ Sentence\ ) \mid [\ Sentence\ ] \\
&\mid \neg\ Sentence \\
&\mid Sentence \land Sentence \\
&\mid Sentence \lor Sentence \\
&\mid Sentence \Rightarrow Sentence \\
&\mid Sentence \Leftrightarrow Sentence \\
\\
\textsc{Operator Precedence} &: \neg, \land, \lor, \Rightarrow, \Leftrightarrow
\end{aligned}
$$

- Counter examples:
  - $(A \land \Rightarrow R)$
  - $(A\ B) \lor (\neg C)$
  - $A \Rightarrow B \Rightarrow C$

Correct examples:

$(A \Rightarrow B) \Rightarrow C$

$A \Rightarrow (B \Rightarrow C)$

Propositional logic: Semantics

- The semantics define the rules for determining the truth of a sentence with respect to a particular model – a mapping of the atomic symbols into true and false.

- Each model specifies true/false for each proposition symbol
  E.g. m1 = $\{P_{1,2} = \text{false}, P_{2,2} = \text{true}, P_{3,1} = \text{false}\}$
  - With these symbols, 8 possible models, can be enumerated automatically.

- Rules for evaluating truth with respect to a model *m*:
  1) $\neg P$ is true iff P is false
  2) $P \wedge Q$ is true iff P is true and Q is true
  3) $P \vee Q$ is true iff P is true or Q is true
  4) $P \Rightarrow Q$ is true unless P is true and Q is false
  5) $P \Leftrightarrow Q$ is true iff $P \Rightarrow Q$ is true and $Q \Rightarrow P$ is true

- The following sentence is evaluated in m1:

$$\neg P_{1,2} \wedge (P_{2,2} \vee P_{3,1}) = \mathit{true} \wedge (\mathit{true} \vee \mathit{false}) = \mathit{true} \wedge \mathit{true} = \mathit{true}$$

Model of Propositional Logic

- Assignment of a truth value – true or false – to every atomic sentence
- Examples:
  - Let A, B, C, and D be the propositional symbols
  - m = {A=true, B=false, C=false, D=true} is a model
  - m′ = {A=true, B=false, C=false} is not a model

- With n propositional symbols, one can define 2^n possible models

- A model for a KB is a “possible world” (assignment of truth values to propositional symbols) in which each sentence in the KB is True.

Summary

- **Logic**
  - **Knowledge-based Agent**
  - **Entailment**
  - **Propositional logic**

What I want you to do

- **Review Chapter 7**