CS:4420 Artificial Intelligence

Spring 2017

Logical Agents

Cesare Tinelli

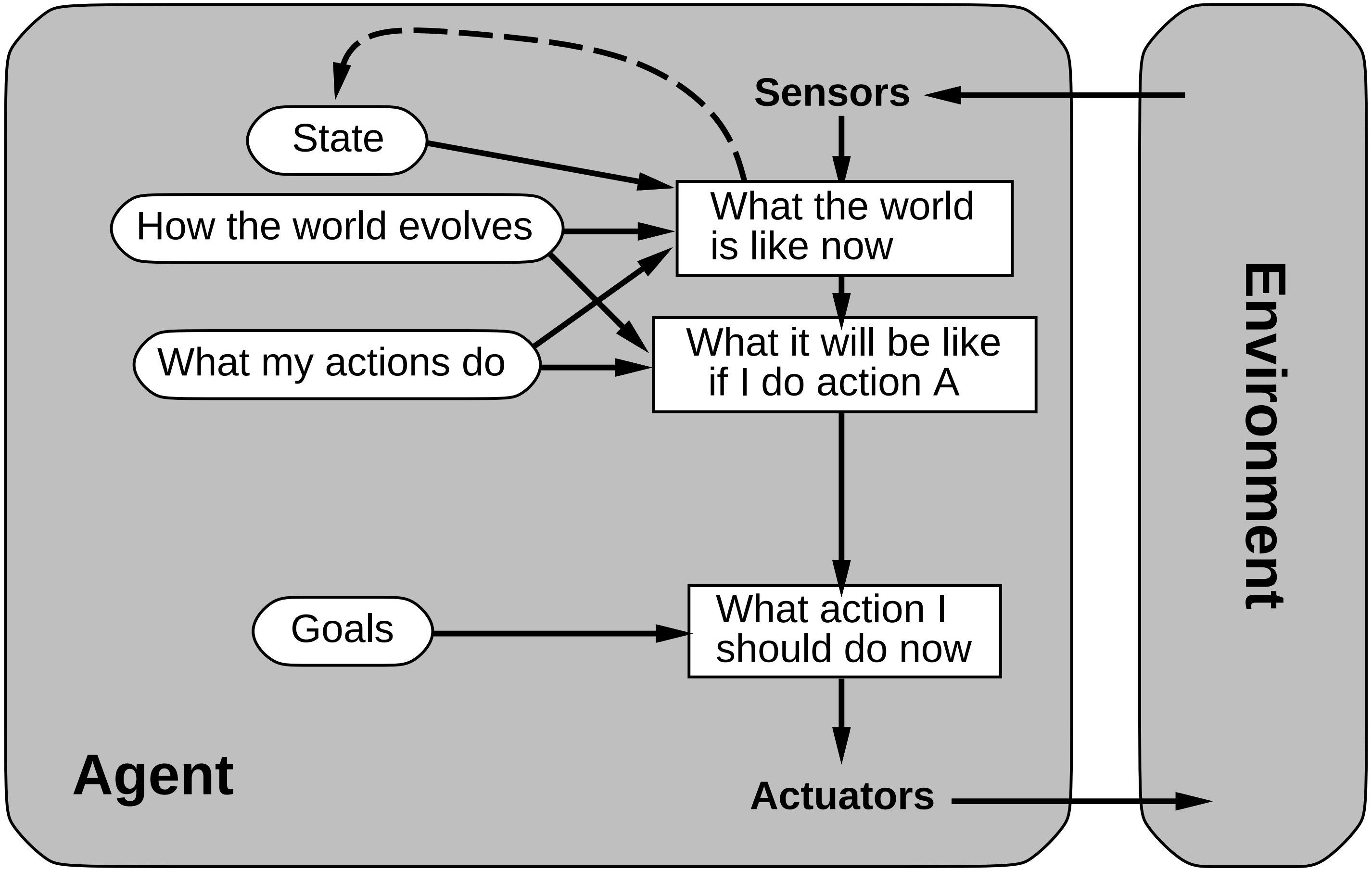
The University of Iowa

Readings

- Chap. 7 of [Russell and Norvig, 2012]

Reasoning Agents

Remember our goal-based agent



(Knowledge-based) Reasoning Agents

Know about the world. They maintain a collection of facts (sentences) about the world, their Knowledge Base, expressed in some formal language

Reason about the world. They are able to derive new facts from those in the KB using some inference mechanism

Act upon the world. They map percepts to actions by querying and updating the KB

Automated Reasoning

Main Assumption (or the "Church Thesis" of AI)

1. Facts about the world can be represented as particular configurations of symbols (*)
2. Reasoning about the world can be achieved by mere symbol manipulation

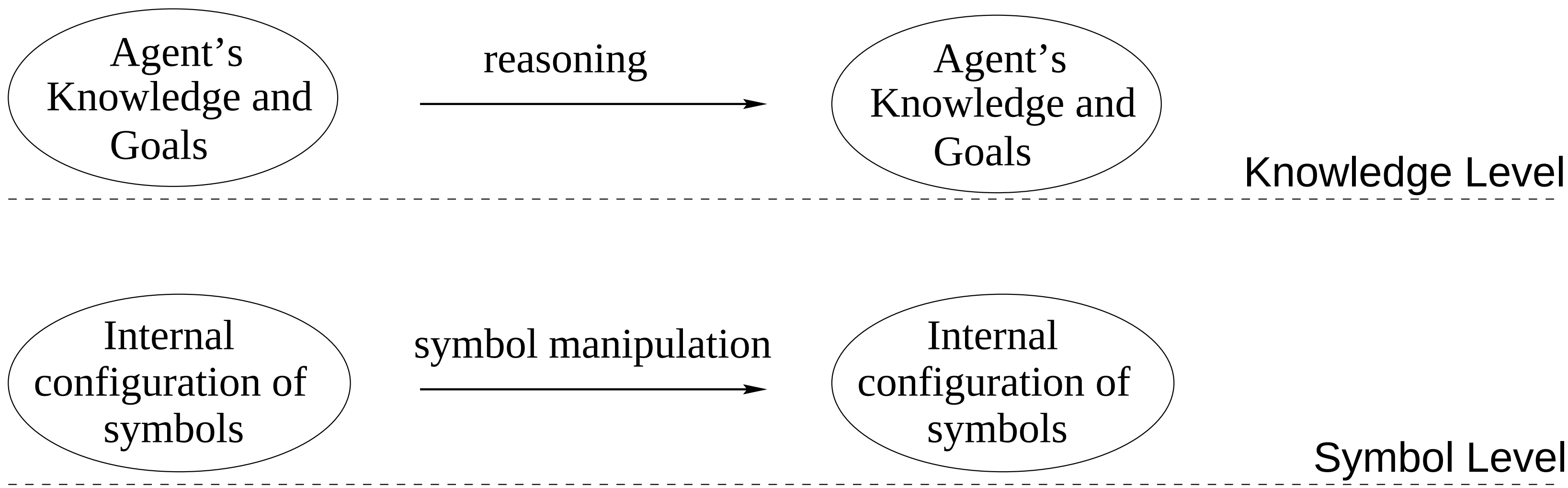
Most AI researchers believe that reasoning is symbol manipulation, nothing else (After all, the human brain is a physical system itself)

(*) I.e., physical entities such as marks on a piece of paper, states in a computer's memory, and so on

Abstraction Levels

We can describe every reasoning agent (natural or not) at two different abstraction levels :

1. Knowledge level: what the agent knows and what the agent's goals are
2. Symbol (or implementation) level: what symbols the agent manipulates and how



Abstraction Levels

We can describe every reasoning agent (natural or not) at two different abstraction levels :

1. Knowledge level: what the agent knows and what the agent's goals are
2. Symbol (or implementation) level: what symbols the agent manipulates and how

At least for artificial agents,

> *the knowledge level is a metaphor for explaining the behavior of the agent, which is really at the symbol level*

Abstraction Levels

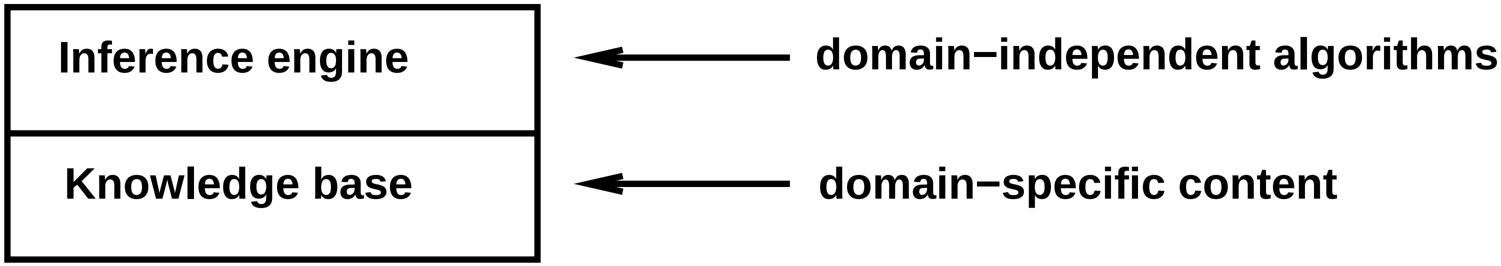
We can describe every reasoning agent (natural or not) at two different abstraction levels :

1. Knowledge level: what the agent knows and what the agent's goals are
2. Symbol (or implementation) level: what symbols the agent manipulates and how

Agents can be viewed at

- the knowledge level
  i.e., what they know and what they can infer, regardless of how implemented
- or at the implementation level
  i.e., data structures to store knowledge and algorithms to manipulate them

Knowledge bases



Knowledge base (KB) = 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 are consequences of the KB

A simple knowledge-based agent

function KB-AGENT(*percept*) **returns** an *action*
  static: 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*

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

An Example: The Wumpus World!

Performance measure:

gold +1000, death -1000,

-1 per step, -10 for using the arrow

Environment:

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

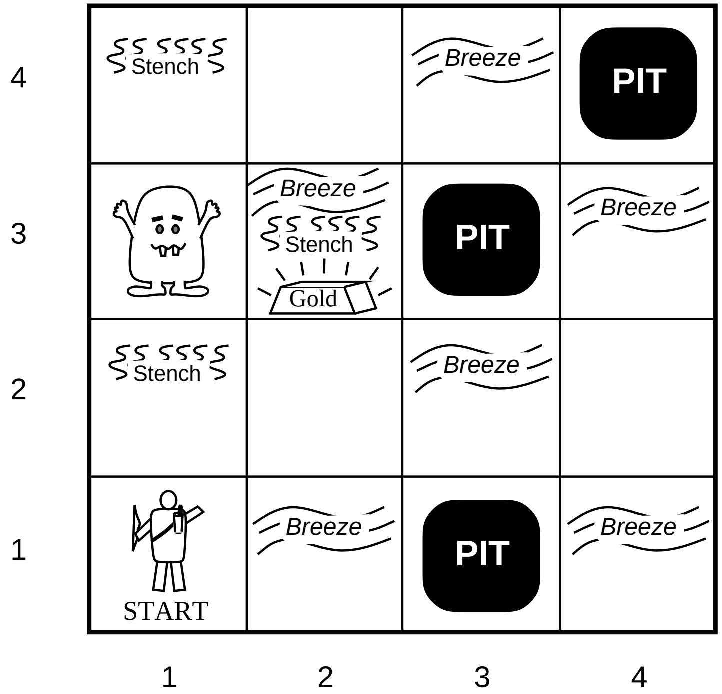
Shooting uses up the only arrow

Grabbing picks up gold if in same square

Releasing drops the gold in same square

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

Sensors: Breeze, Glitter, Smell



Wumpus World Characterization

Observable?

Deterministic?

Episodic?

Static?

Discrete?

Single-agent?

Wumpus World Characterization

Observable? Partially—only local perception

Deterministic?

Episodic?

Static?

Discrete?

Single-agent?

Wumpus World Characterization

Observable? Partially—only local perception

Deterministic? Yes—outcomes exactly specified

Episodic?

Static?

Discrete?

Single-agent?

Wumpus World Characterization

Observable? Partially—only local perception

Deterministic? Yes—outcomes exactly specified

Episodic? No—sequential at the level of actions

Static?

Discrete?

Single-agent?

Wumpus World Characterization

Observable? Partially—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?

Single-agent?

Wumpus World Characterization

Observable? Partially—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?

Wumpus World Characterization

Observable? Partially—only local perception

Deterministic? Yes—outcomes exactly specified

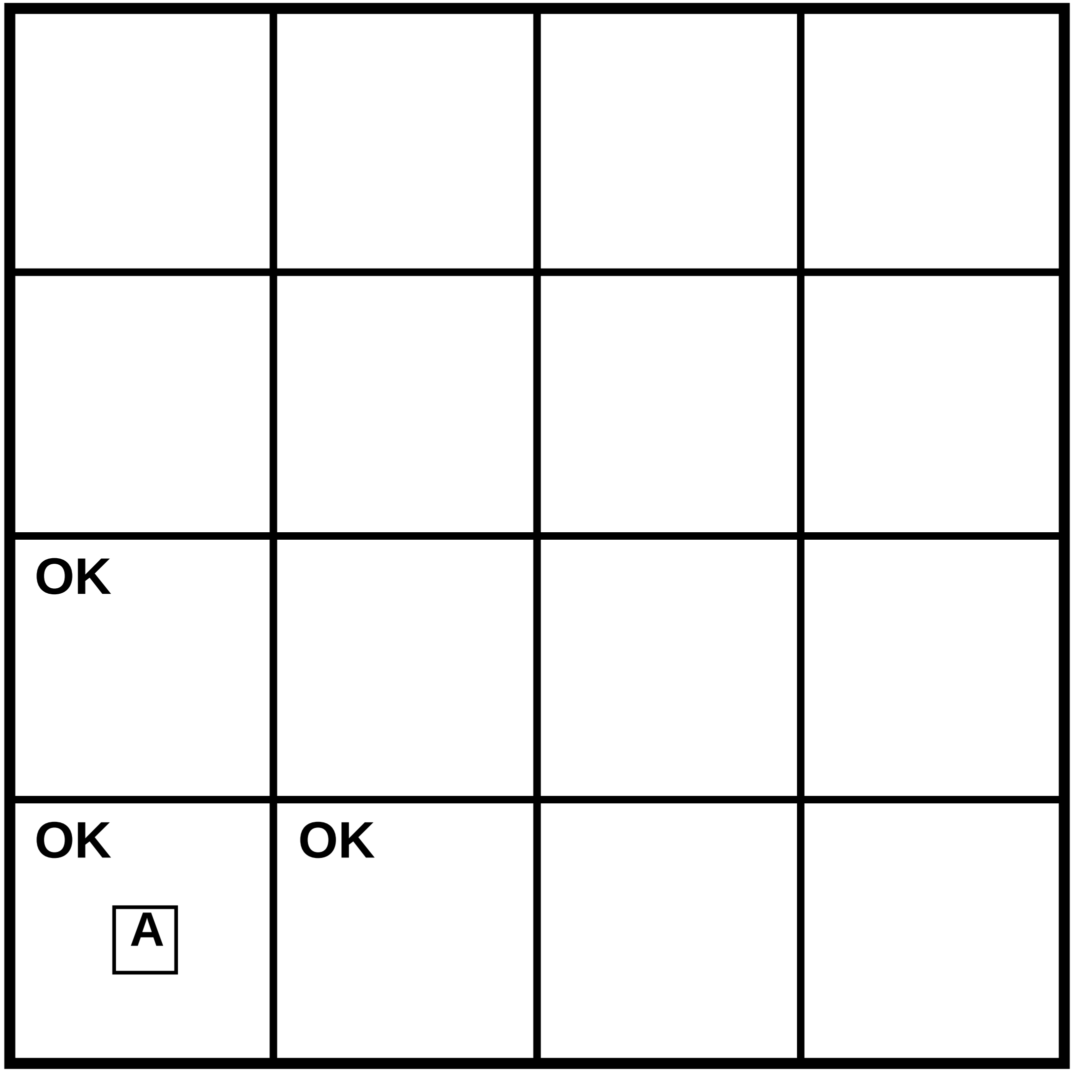
Episodic? No—sequential at the level of actions

Static? Yes—Wumpus and Pits do not move

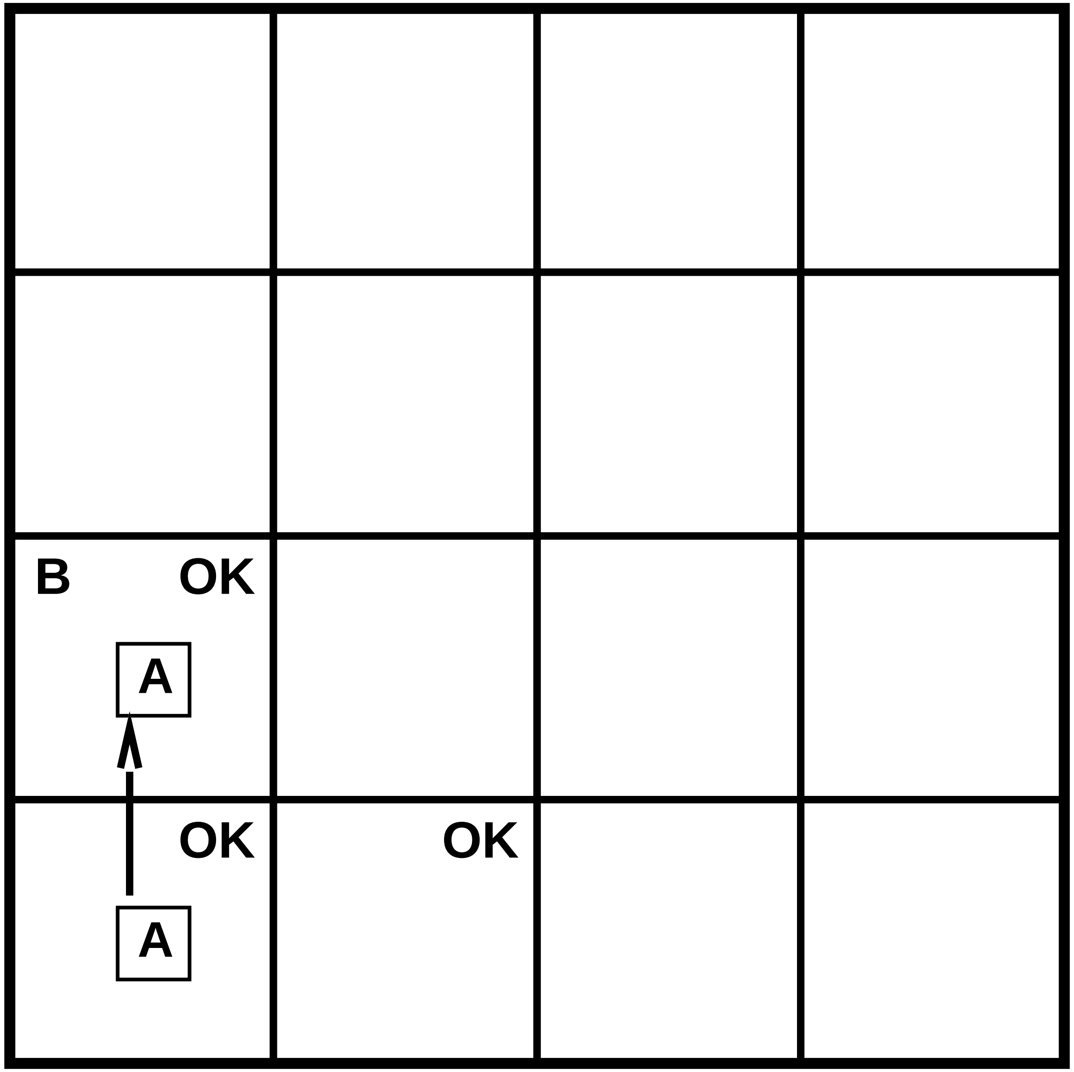
Discrete? Yes

Single-agent? Yes—Wumpus is essentially a natural feature

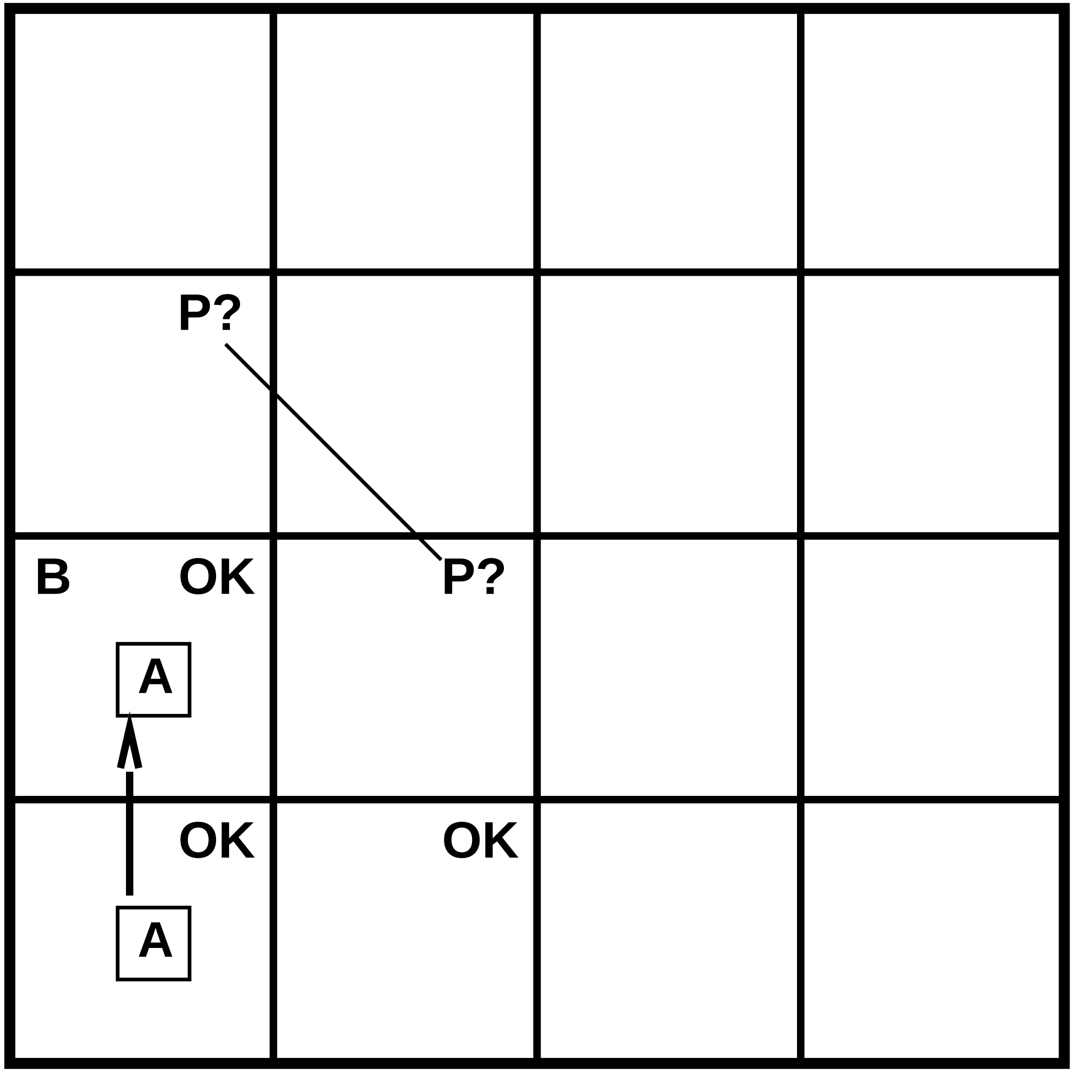
Exploring a Wumpus World



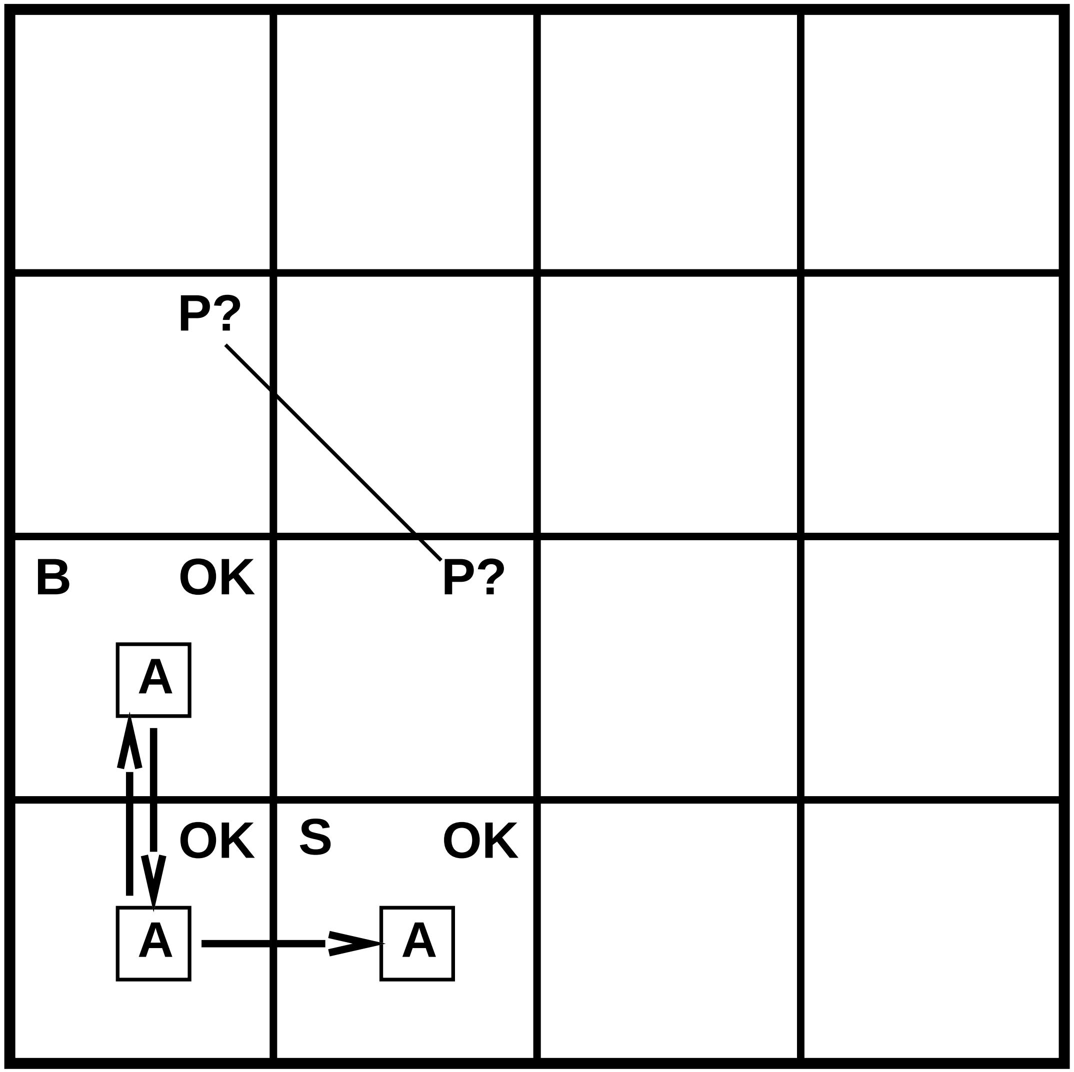
Exploring a Wumpus World



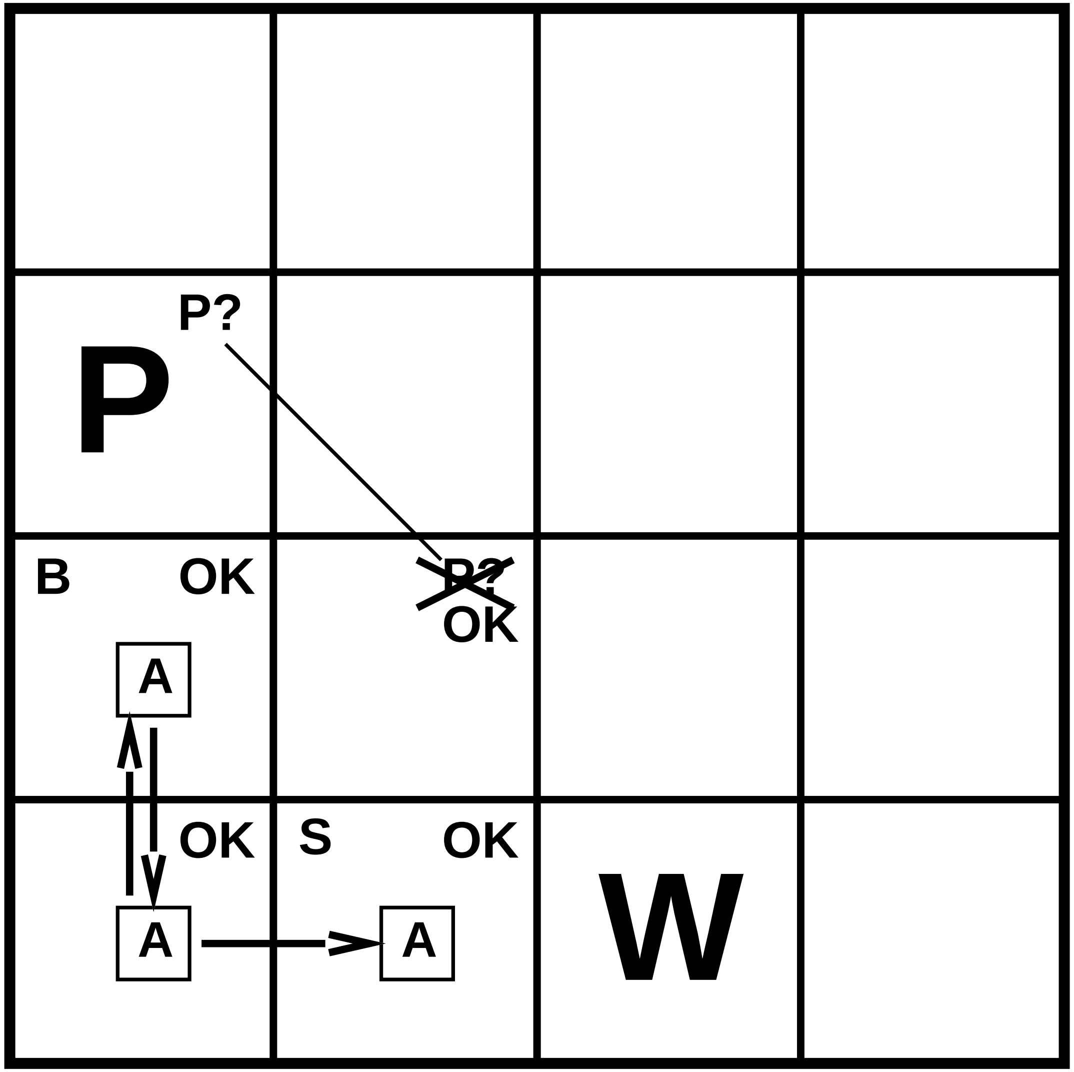
Exploring a Wumpus World



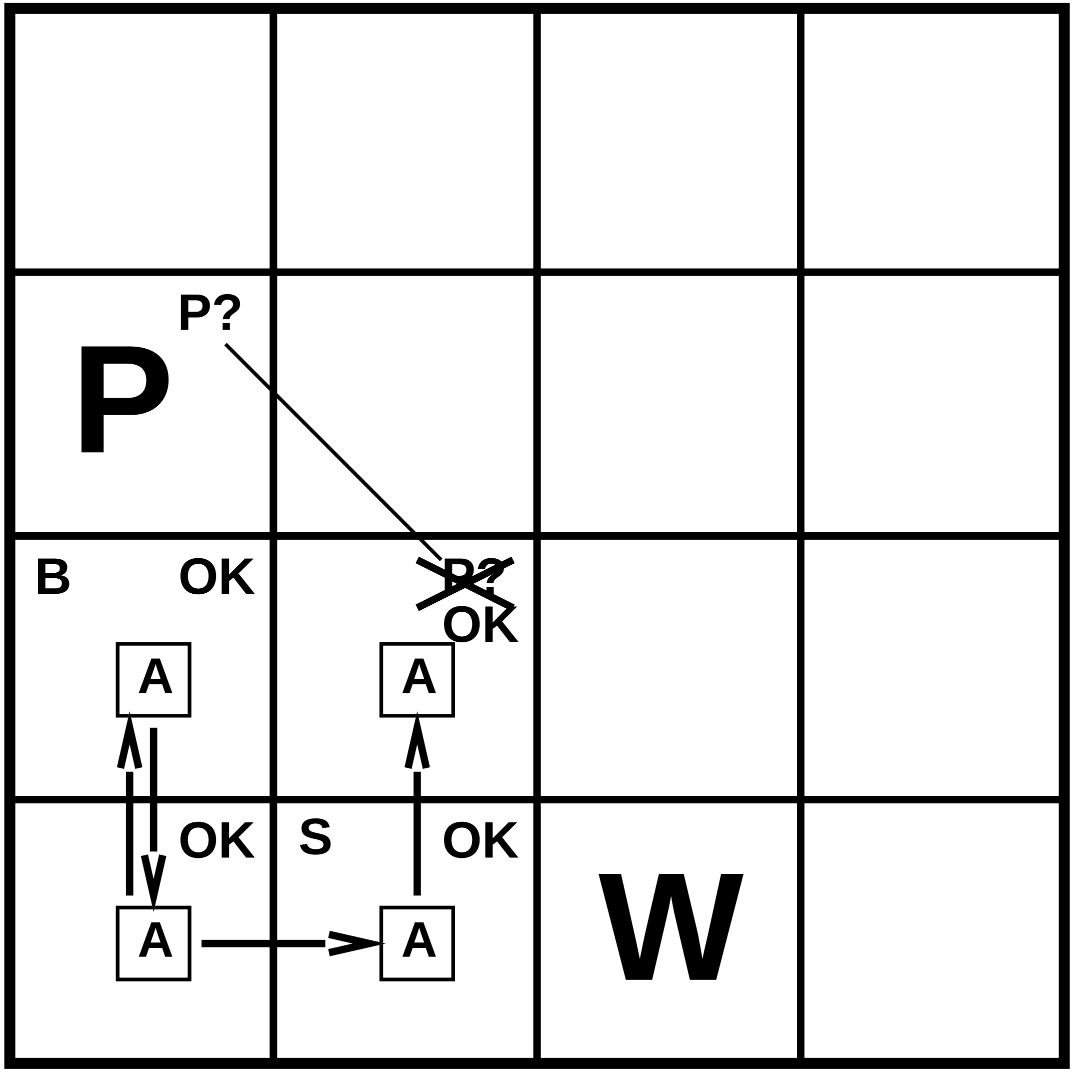
Exploring a Wumpus World



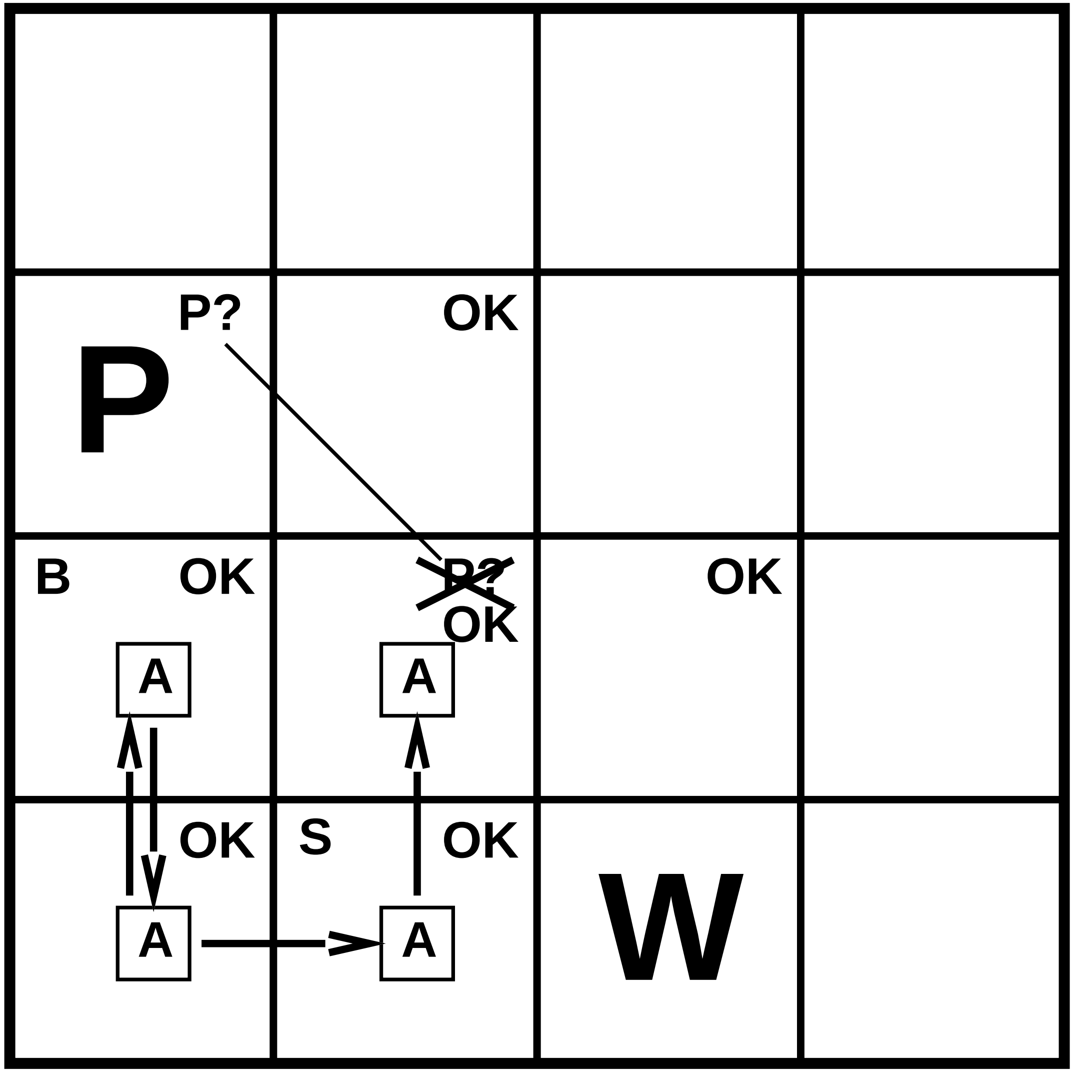
Exploring a Wumpus World



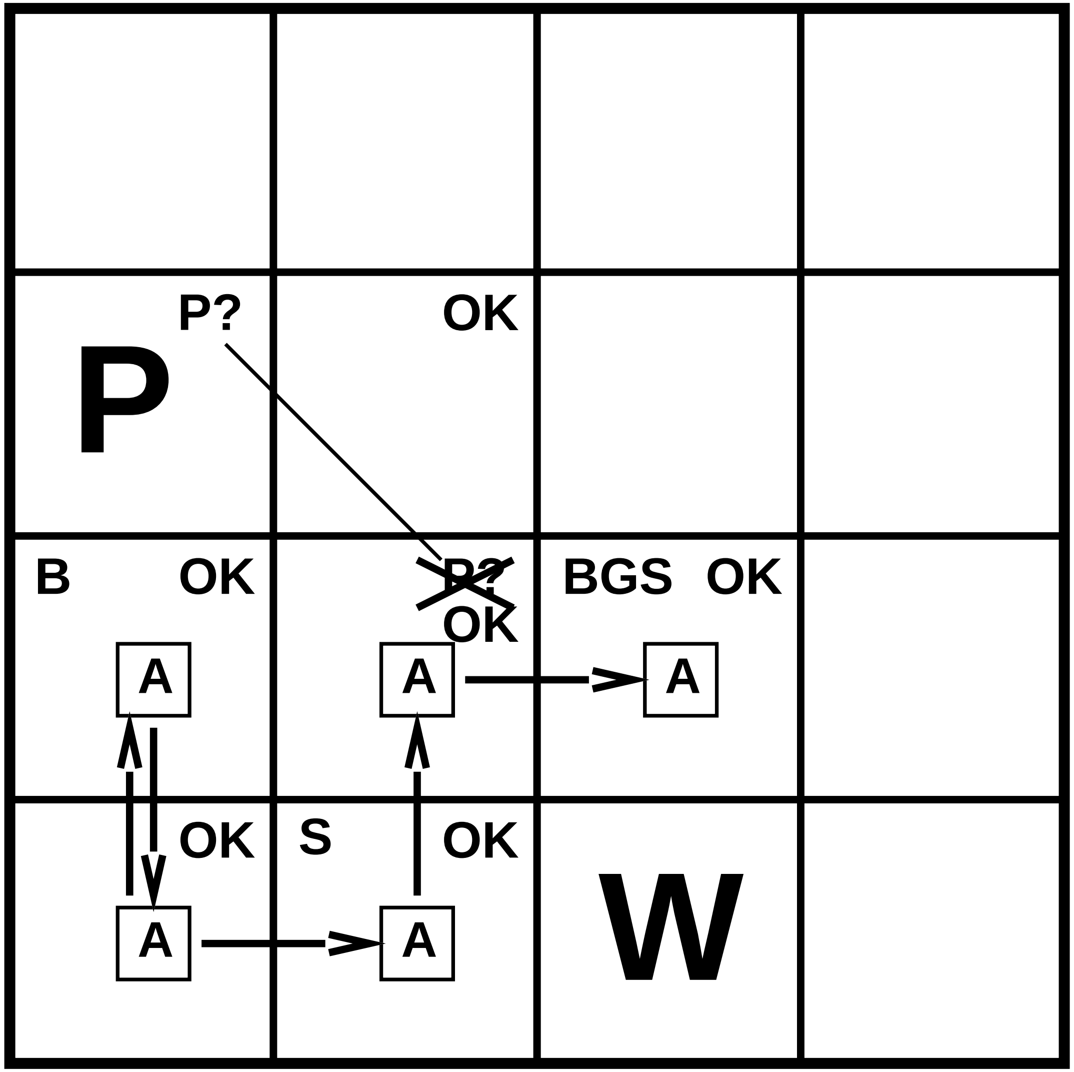
Exploring a Wumpus World



Exploring a Wumpus World



Exploring a Wumpus World



Exploring a Wumpus World

(a) (b)

- What are the safe moves from (1,1)?

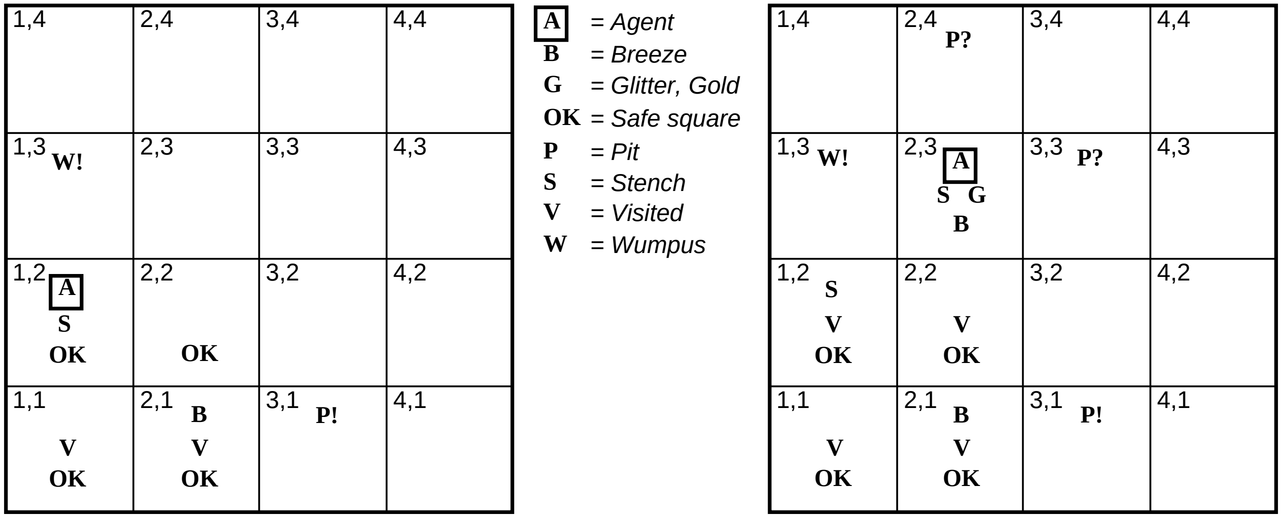
  Move to (1,2), (2,1), or stay in (1,1)

- Move to (2,1) then
- What are the safe moves from (2,1)?

  B in (2,1) $\Rightarrow$ *P in (2,2) or (3,1) or (1,1)*

- Move to (1,2) then

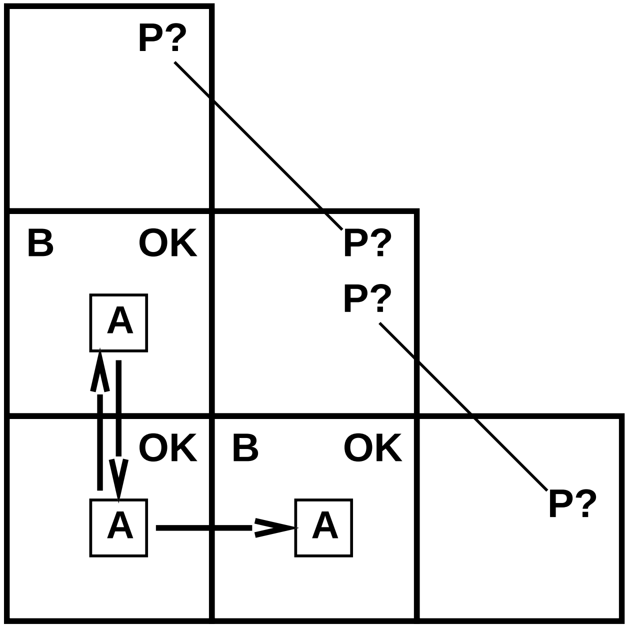
Exploring a Wumpus World



(a) (b)

- S in (1,2) $\Rightarrow$ W in (1,1) or (2,2) or (1,3)
  Survived in (1,1) and no S in (2,1) $\Rightarrow$ W in (1,3)
  No B in (1,2) $\Rightarrow$ P in (3,1)
- Move to (2,2), then to (2,3)
- G in (2,3)
- Grab G and come home

Other Considerations



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

Assuming pits uniformly distributed, (2,2) has pit w/ prob 0.86, vs. 0.31

Knowledge Representation

An (artificial) agent represents knowledge as a collection of sentences in some formal language, the knowledge representation language

A knowledge representation language is defined by its

- syntax, which describes all the possible symbol configurations that constitute a sentence,
- semantics, which maps each sentence of the language to a statement about the world

Ex: Arithmetic

- $x + y > 3$ is a sentence; $x+ > y$ is not
- $x + y > 3$ is “true” iff the number $x + y$ is greater than the number three
- the semantics of $x + y > 3$ is either the fact “true” or the fact “false”

Knowledge Representation and Reasoning

At the semantical level, reasoning is the process of deriving new facts from previous ones

At the syntactical level, this process is mirrored by that of producing new sentences from previous ones

The production of sentences from previous ones should not be arbitrary Only entailed sentences should be derivable

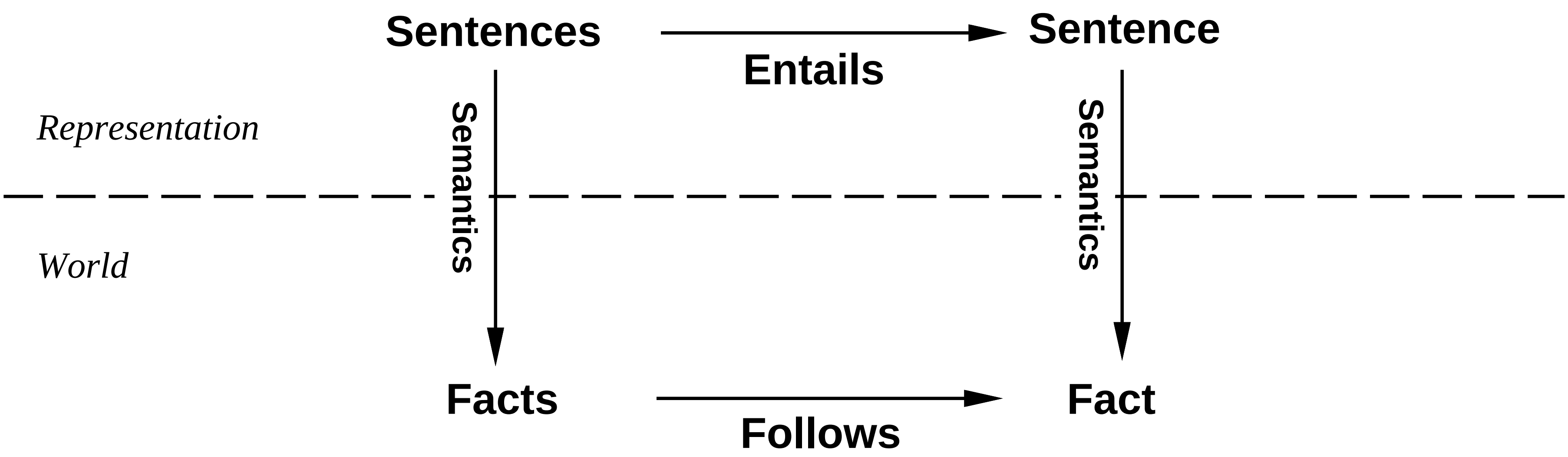
Entailment

Informally,

a sentence φ is entailed by a set of sentences Γ

iff

the fact denoted by φ *follows logically* from the facts denoted by Γ



Entailment

Notation: $\Gamma \models \varphi$ if the set of sentences Γ entail the sentence φ

Intuitive reading of $\Gamma \models \varphi$:

Whenever Γ is true in the world, φ is also true

Examples: Let Γ consist of the axioms of arithmetic

$$
\begin{array}{lcl}
\{x = y,\ y = z\} & \models & x = z \\
\Gamma \cup \{x + y \geq 0\} & \models & x \geq -y \\
\Gamma \cup \{x + y = 3,\ x - y = 1\} & \models & x = 2 \\
\Gamma \cup \{x + y = 3\} & \not\models & x = 2
\end{array}
$$

Inference Systems

At the knowledge representation level, reasoning is achieved by an inference system I, a *computational device* able to derive new sentences from previous ones

Notation: $\Gamma \vdash_I \varphi$ if I can derive the sentence φ from the set Γ

To be useful at all, an inference system must be sound:

if $\Gamma \vdash_I \varphi$ then $\Gamma \models \varphi$ holds as well

Ideally, an inference system is also complete:

if $\Gamma \models \varphi$ then $\Gamma \vdash_I \varphi$ holds as well

Inference Rules

An inference system is typically described as a set of inference (or derivation) rules

Each derivation rule has the form:

$$\frac{P_1, \ldots, P_n}{C} \qquad \begin{array}{l} \longleftarrow \textit{premises} \\ \longleftarrow \textit{conclusion} \end{array}$$

Derivation Rules and Soundness

A derivation rule is sound if it derives true conclusions from true premises

All men are mortal Aristotle is a man ――――――――― Aristotle is mortal	Sound Inference
All men are mortal Aristotle is mortal ――――――――― All men are Aristotle	Unsound Inference!

Knowledge Representation Languages

Why don't we use natural language (e.g., English) to represent knowledge?

- Natural language is certainly expressive enough!
- But it is also too ambiguous for automated reasoning

  Ex: I saw the boy on the hill with the telescope

Why don't we use programming languages?

- They are well-defined and unambiguous
- But they are not expressive enough

Knowledge Representation and Logic

The field of Mathematical Logic provides powerful, formal knowledge representation languages and inference systems to build reasoning agents

We will consider two languages, and associated inference systems, from mathematical logic:

- Propositional Logic
- First-order Logic