Logical agents

AIMA.3rd Chapter 7

Outline

♢ Knowledge-based agents

♢ Wumpus world

♢ Logic in general—models and entailment

♢ Propositional (Boolean) logic

♢ Equivalence, validity, satisfiability

♢ Inference rules and theorem proving
   – forward chaining
   – backward chaining
   – resolution

Knowledge bases

Knowledge base = set of sentences in a **knowledge representation language** that represents some assertion about the world.

Axiom is a sentence that are taken as given without being derived from other sentences.

Declarative approach to building an agent (or other system):

TELL sentences one by one until the agent know how to operate in its environment.

TELL: A way to add new sentences to the knowledge base

Then it can ASK itself what to do—answers should follow from the KB

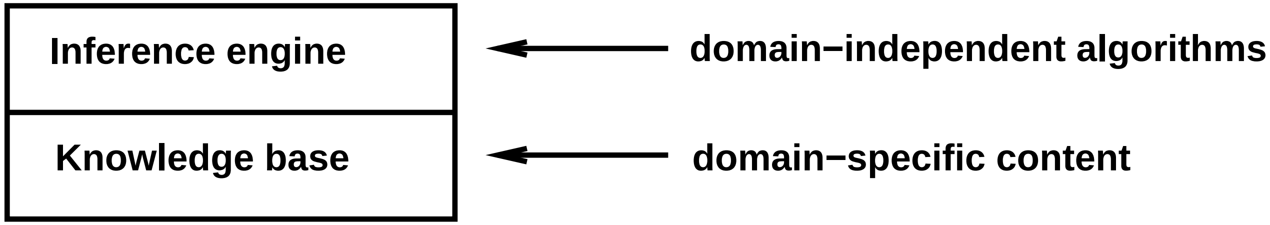
ASK: A way to query what is known.

The operations TELL and ASK may involve **inference**, i.e., deriving new sentences from old.

*NOTE: In contrast to declarative approach of system building, the procedural approach encodes desired behaviors directly as program code.

Knowledge-bases agents

In KNOWLEDGE-BASED AGENTS intelligence is embodied in the processes of reasoning that operate on internal representations of knowledge.



A simple knowledge-based agent

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

```
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
```

A simple knowledge-based agent

Each time the agent program is called, it does three things:
1. TELLs the KB what it perceives
– Make-Percept-Sentence constructs a sentence asserting that the agent perceived the given percept at the given time.
2. ASKs the KB what action it should perform.
– Make-Action-Query constructs a sentence that asks what action should be done at the current time.
3.TELLs the KB which action was chosen, and the agent executes the action
– Make-Action-Sentence constructs a sentence asserting that the chosen action was executed.

Agents can be viewed at the knowledge level

i.e., **what they know**, regardless of how implemented

Or at the implementation level

i.e., data structures in KB and algorithms that manipulate them

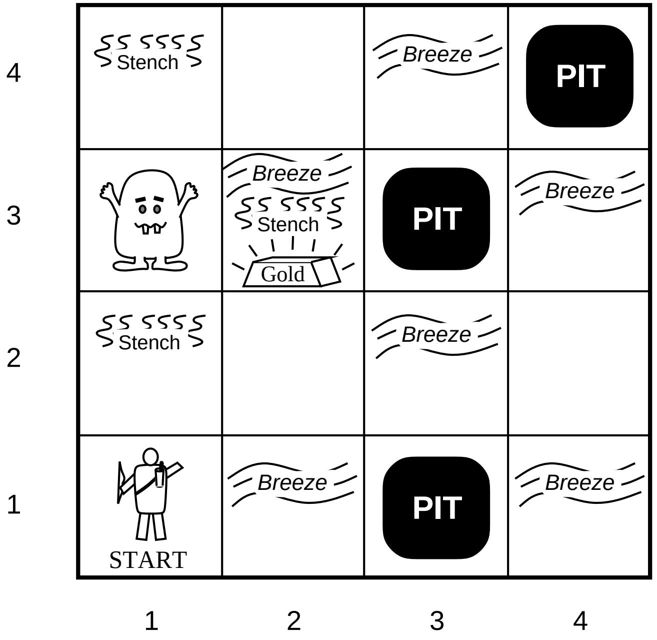
Knowledge-based agent example: Wumpus World

The wumpus world is a cave consisting of rooms connected by passageways.

Lurking somewhere in the cave is the wumpus, a beast that eats anyone who enters its room. The wumpus can be shot by an agent, but the agent has only one arrow.

Some rooms contain bottomless pits that will trap anyone who wanders into these rooms (except for the wumpus, which is too big to fall in).

The only mitigating feature of this bleak environment is the possibility of nding a heap of gold.



Wumpus World PEAS description

Performance measure:

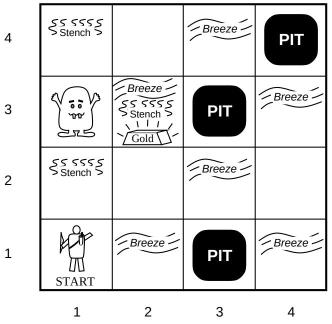
gold +1000, death -1000, -1 per step, -10 for using the arrow

Environment: 4x4 grid

– The agent always starts in [1,1], facing right.

– The gold and the wumpus is located randomly under uniform dist. excluding [1,1]

– Each square other than [1,1] can be a pit, with prob. 0.2.



Actuators: *TurnLeft, TurnRight, Forward, Grab, Shoot, Climb*

Sensors: agent has 5 sensors

- Squares adjacent to wumpus are smelly – *Stench*.
- In the squares directly adjacent to a pit, the agent will perceive a *Breeze*.
- In the square where the gold is, the agent will perceive a *Glitter*.
- When an agent walks into a wall, it will perceive a *Bump*.
- When the wumpus dies, it emits a *Scream* that can be perceived anywhere.

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

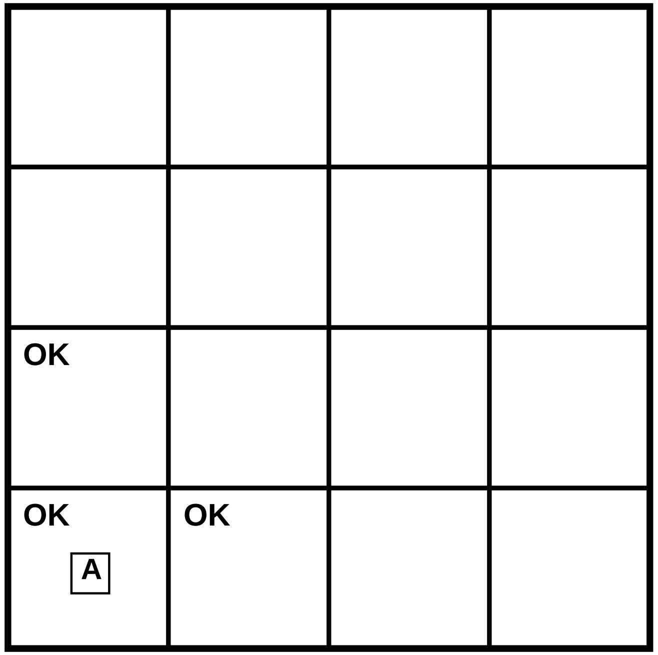
Episodic?? No—sequential at the level of actions

Static?? Yes—Wumpus and Pits do not move

Discrete?? Yes

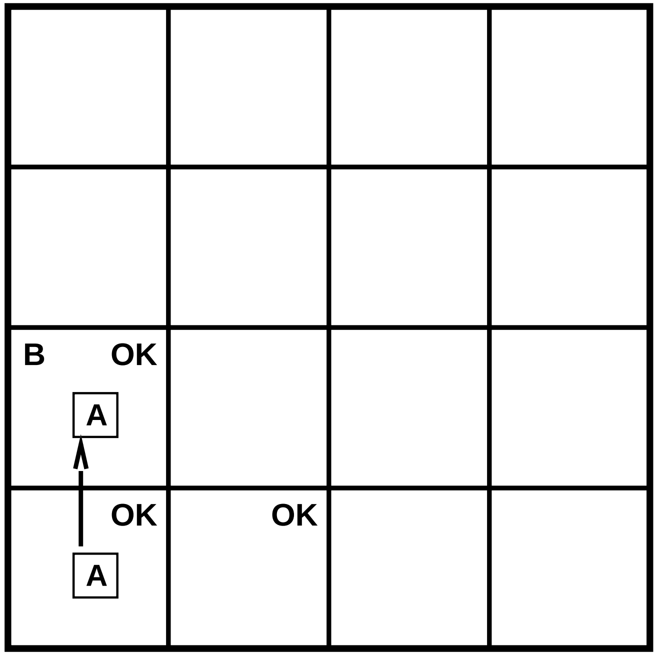
Single-agent?? Yes—Wumpus is essentially a natural feature

Exploring a wumpus world



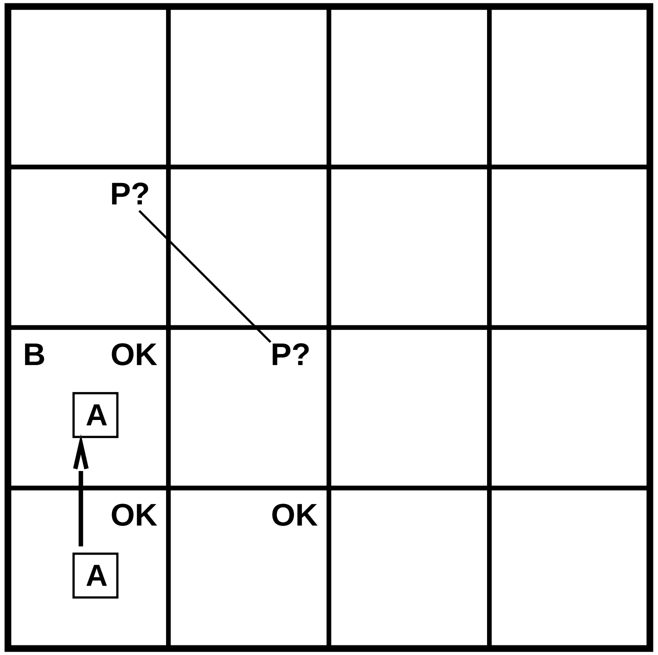
The first percept is [None,None,None,None,None], from which the agent can con clude that its neighboring squares, [1,2] and [2,1], are free of dangers.

Exploring a wumpus world



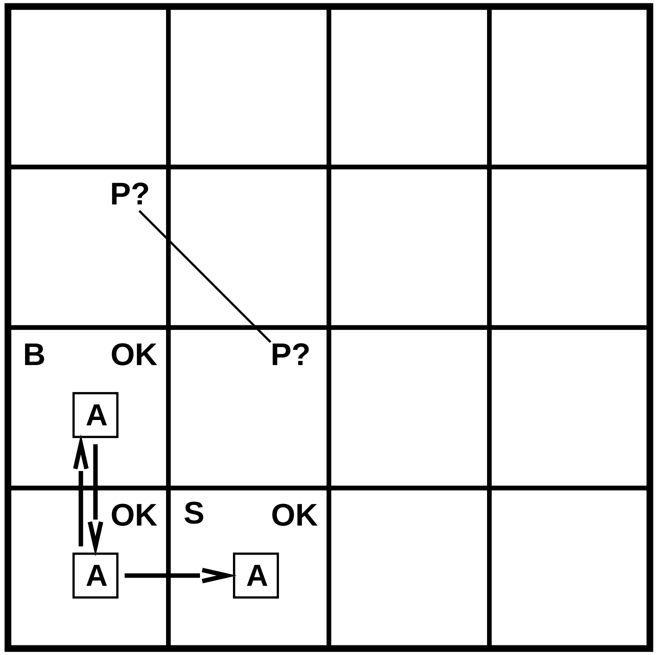
A cautious agent will move only into a square that it knows to be OK. Let us suppose the agent decides to move forward to [2,1].

Exploring a wumpus world



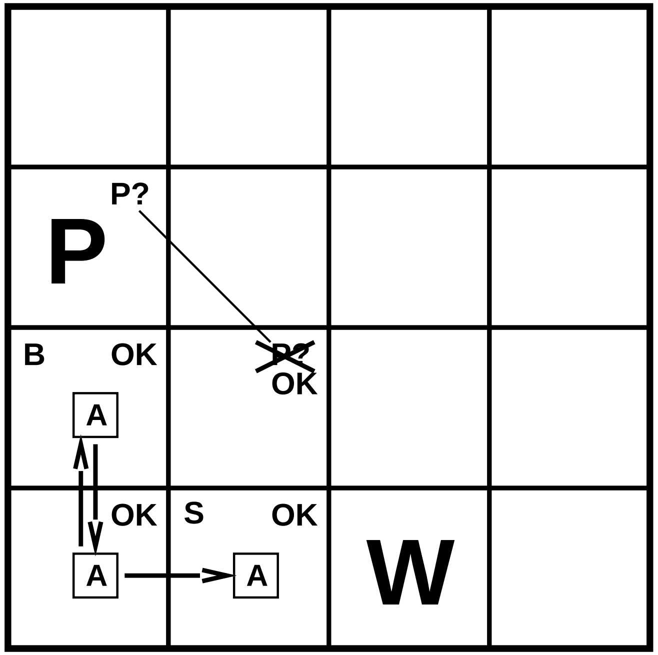
The agent perceives a $Breeze$ (denoted by B) in [2,1], so there must be a pit in a neighboring square.

Exploring a wumpus world



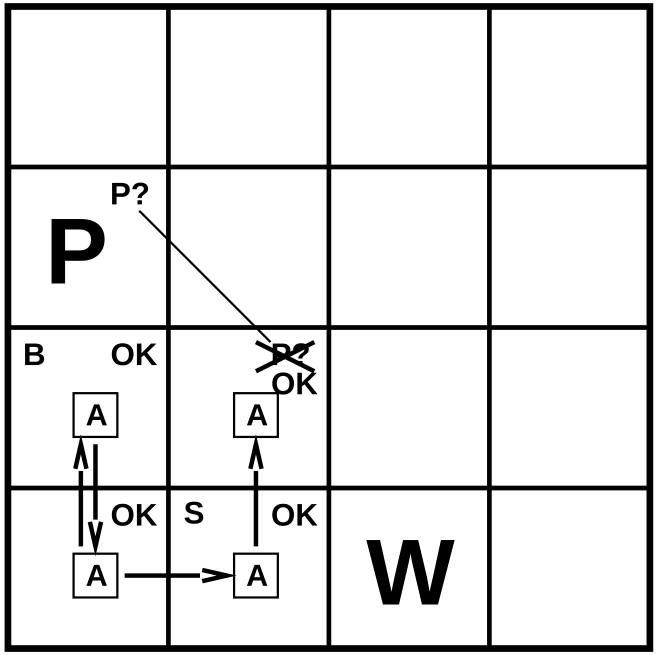
There is only one known square that is OK and that has not yet been visited. So the prudent agent will turn around, go back to [1,1], and then proceed to [1,2].

Exploring a wumpus world



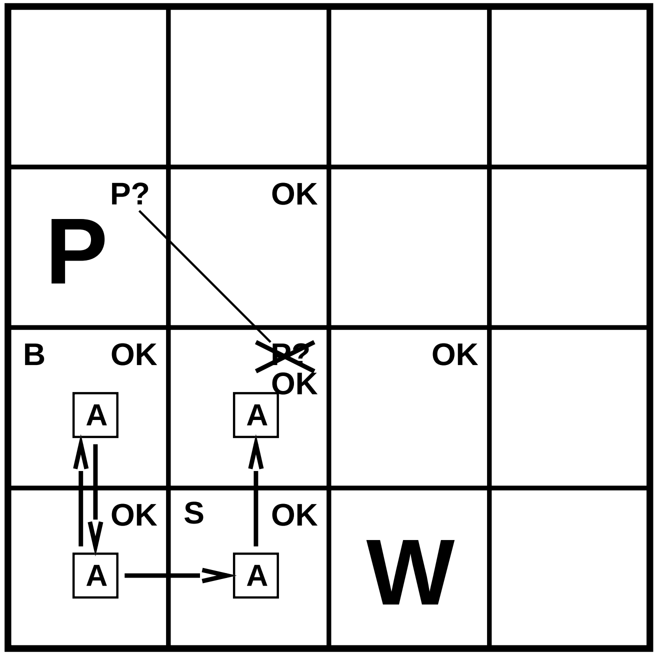
The agent perceives a *Stench* in [1,2] meaning that there must be a wumpus nearby. wumpus cannot be in [1,1], by the rules of the game, and it cannot be in [2,2] (or the agent would have detected a *Stench* when it was in [2,1]). Therefore, the agent can infer that the wumpus is in [1,3].

Exploring a wumpus world



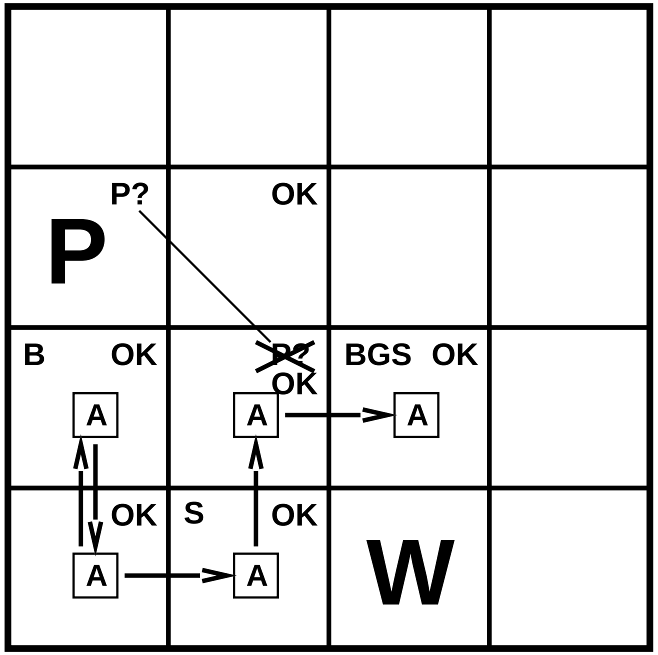
The lack of a $Breeze$ in [1,2] implies that there is no pit in [2,2].
The agent has now proved to itself that there is neither a pit nor a wumpus in [2,2], so it is OK to move there.

Exploring a wumpus world



We do not show the agents state of knowledge at [2,2]; we just assume that the agent turns and moves to [2,3],

Exploring a wumpus world



In [2,3], the agent detects a glitter, so it should grab the gold and then return home.

Logic in general

In an logical reasoning in each case for which the agent draws a conclusion from the available information, that conclusion is *guaranteed* to be correct if the available information is correct.

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 w.r.t. each possible world

Models are mathematical abstractions of the possible worlds.

Logic in general cont.

In standard logics, every sentence must be either true or false in each possible world (or model)

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 false in a world where $x = 0,\ y = 6$

If a sentence α is TRUE in model m, we say that m satifies α or sometimes m is a model of α. We use the notation $M(\alpha)$ to mean the set of all models of α.

Entailment

Entailment means that one sentence **follows logically** from another:

$\alpha \models \beta$

to mean that the setence α entails the sentence β.

E.g., $x + y = 4$ entails $4 = x + y$

Entailment is a relationship between sentences (i.e., **syntax**)
that is based on **semantics**

Note: brains process **syntax** (of some sort)

Models

Logicians typically think in terms of models, which are formally structured worlds with respect to which truth can be evaluated

We say m is a model of a sentence α if α is true in m

$M(\alpha)$ is the set of all models of α

Entailment revisited:

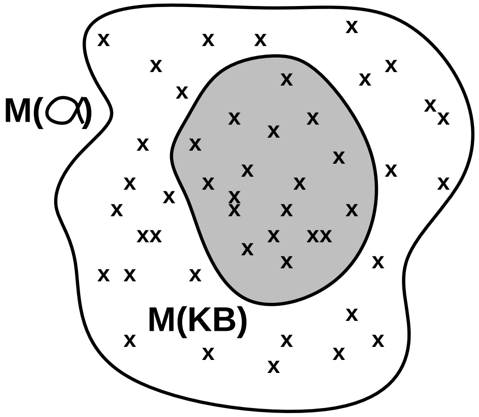
$\alpha \models \beta$ iff, in every model in which α is TRUE, β is also TRUE.

i.e. $\alpha \models \beta$ iff $M(\alpha) \subseteq M(\beta)$

E.g. $KB \models \alpha$

$KB =$ Giants won and Reds won

$\alpha =$ Giants won



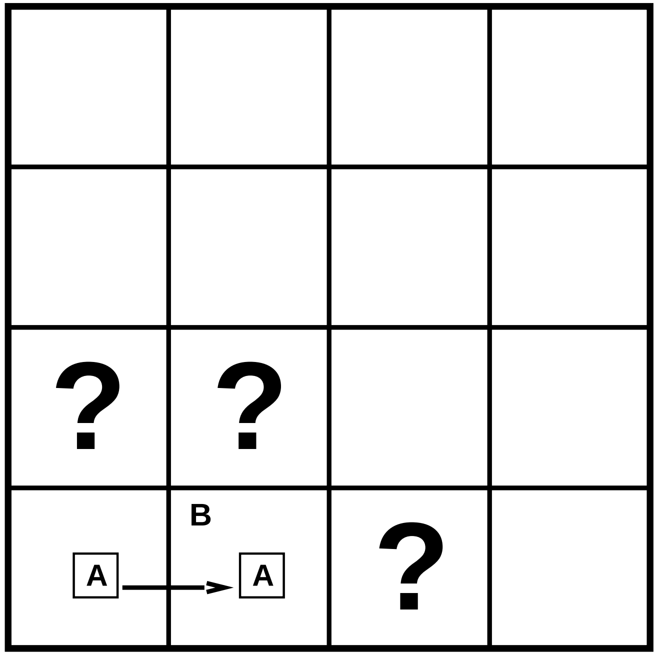
Entailment in the wumpus world

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

These percepts, combined with the agents knowledge of the rules of the wumpus world, constitute the KB.

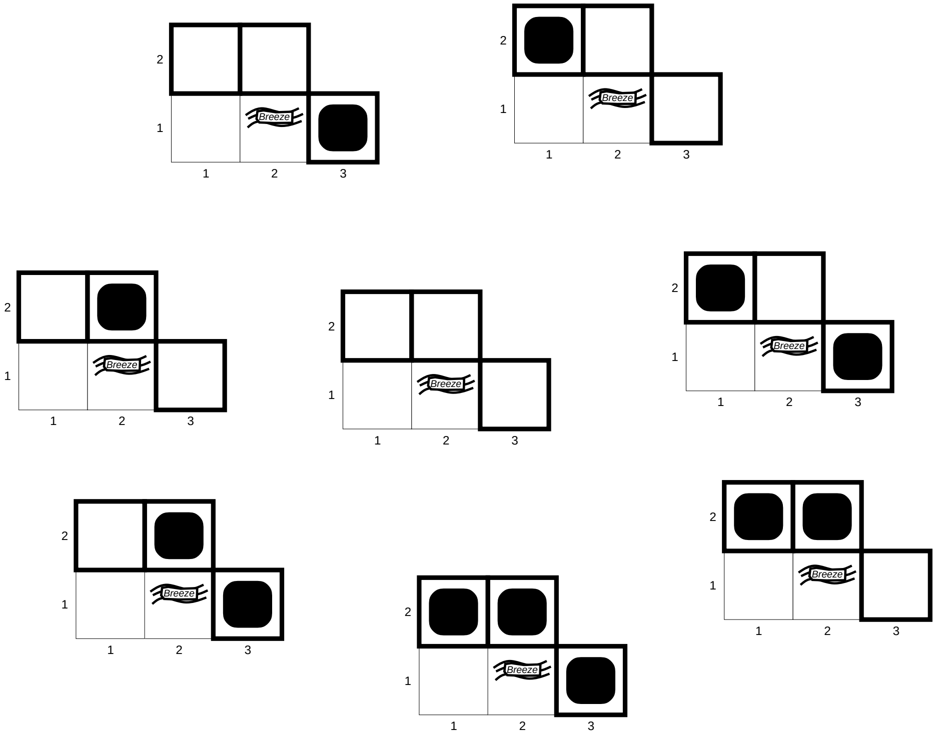
Consider possible models assuming only pits.

The agent is interested in whether the adjacent squares [1,2], [2,2], and [3,1] contain pits.

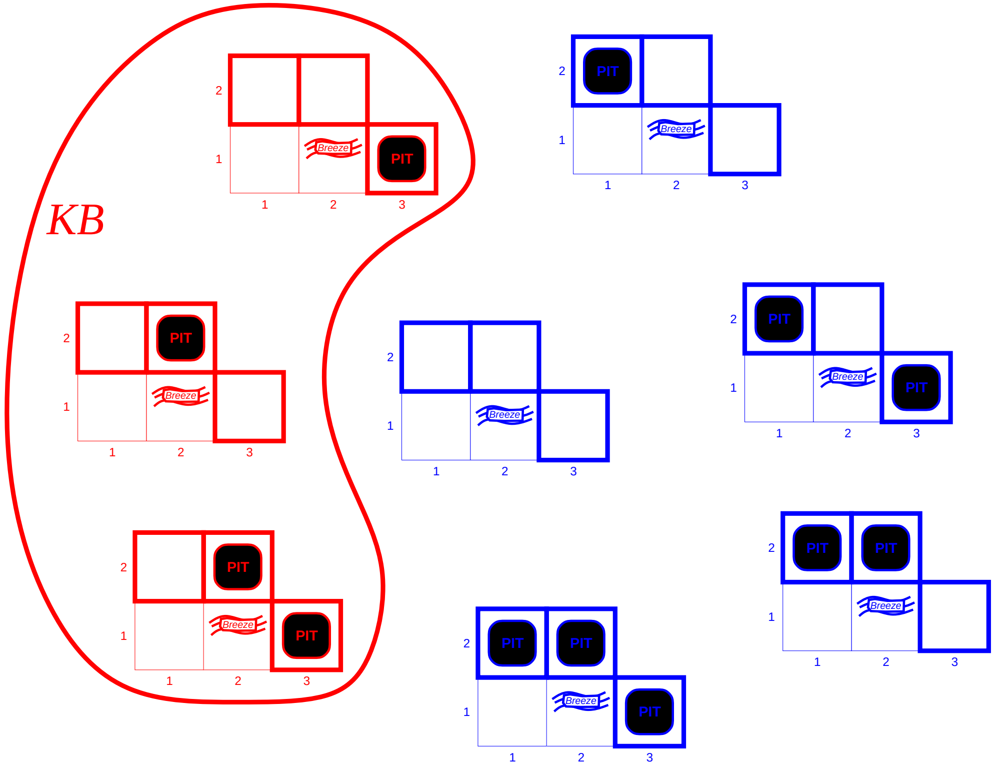


Wumpus models

3 squares each may or may not contain pits $\Rightarrow$ $2^3 = 8$ possible models



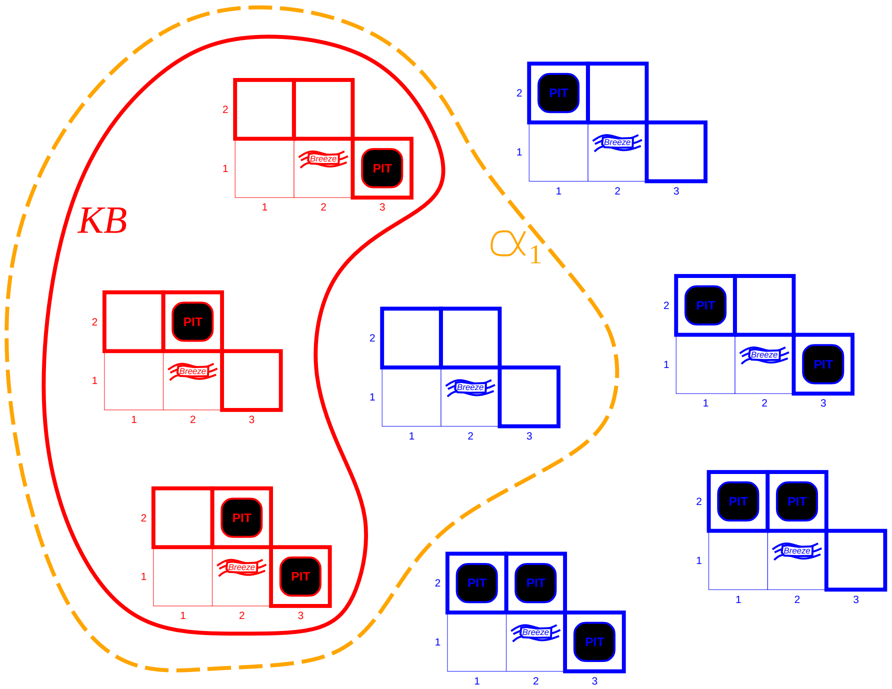
Wumpus models



KB = wumpus-world rules + observations

KB corresponding to the observations of nothing in [1,1] and a breeze in [2,1] in solid line.

Wumpus models



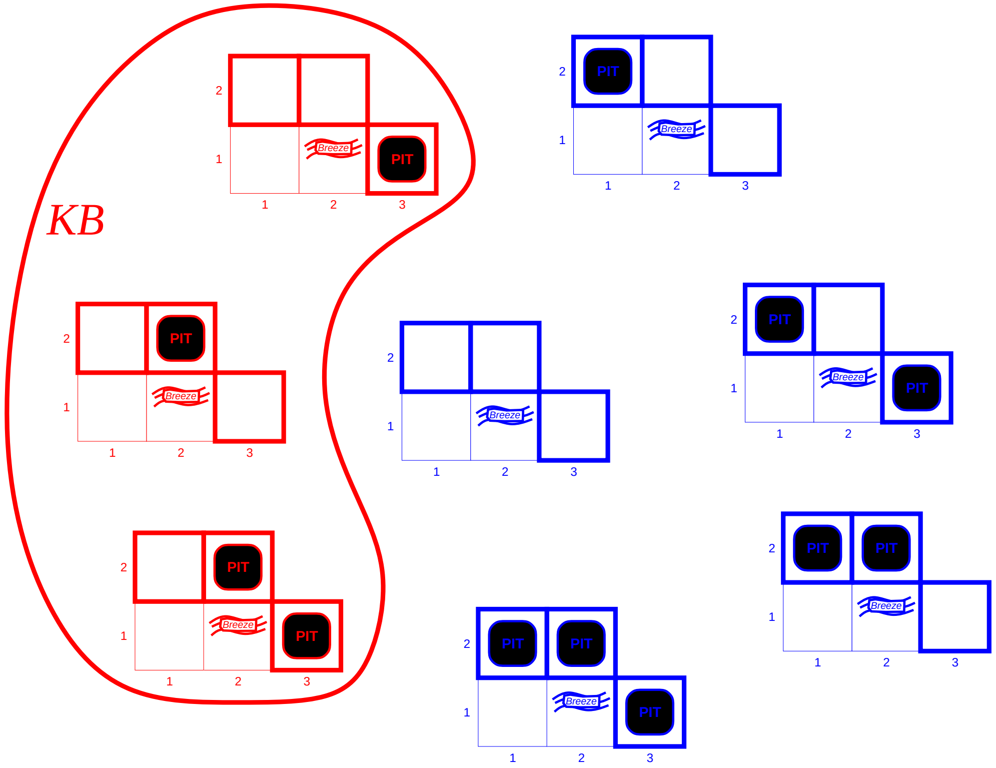
KB = wumpus-world rules + observations

possible conclusion 1: α_1 = “no pit in [1,2]”

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

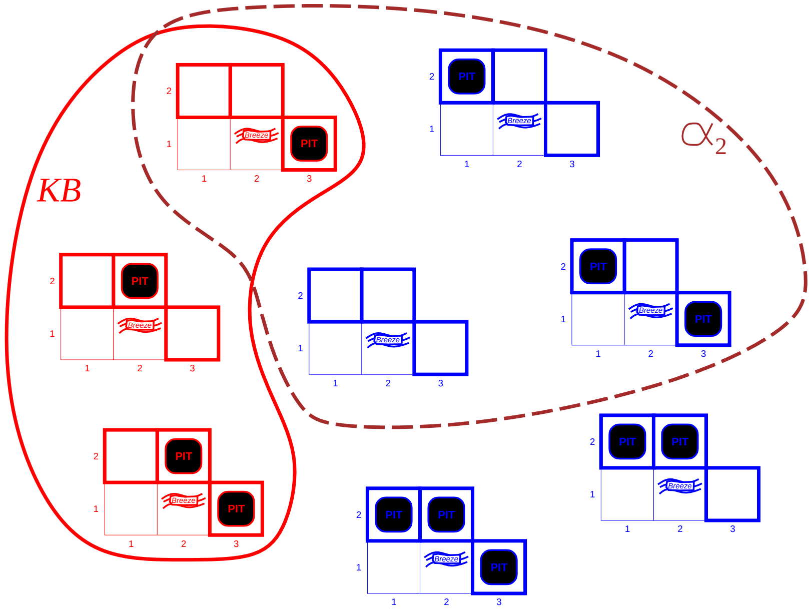
$KB \models \alpha_1$, proved by **model checking**

Wumpus models



KB = wumpus-world rules + observations

Wumpus models



KB = wumpus-world rules + observations

possible conclusion 2: α_2 = “no pit in [2,2]”

in some models in which KB is true, α_2 is false: $KB \not\models \alpha_2$
thus, the agent cannot conclude the existance of pit in [2,2]

Inference

$KB \vdash_i \alpha$: sentence α can be derived from KB by procedure i

Consequences of KB are a haystack; α is a needle.
Entailment = needle in haystack; inference = finding it

Soundness (truth preserving): inference algo. i is sound if
whenever $KB \vdash_i \alpha$, it is also true that $KB \models \alpha$

Completeness: inference algo. i is complete if it can derive any sentence that is entailed.
whenever $KB \models \alpha$, it is also true that $KB \vdash_i \alpha$

Preview: we will define a logic which is expressive enough to say almost anything of interest, and for which there exists a sound and complete inference procedure.

That is, the procedure will answer any question whose answer follows from what is known by the KB.