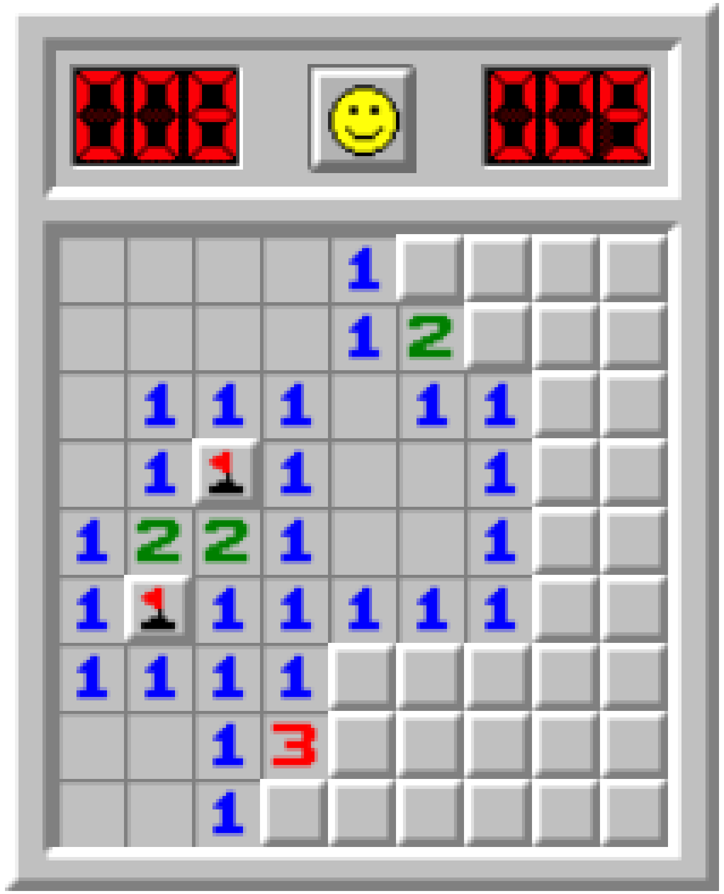
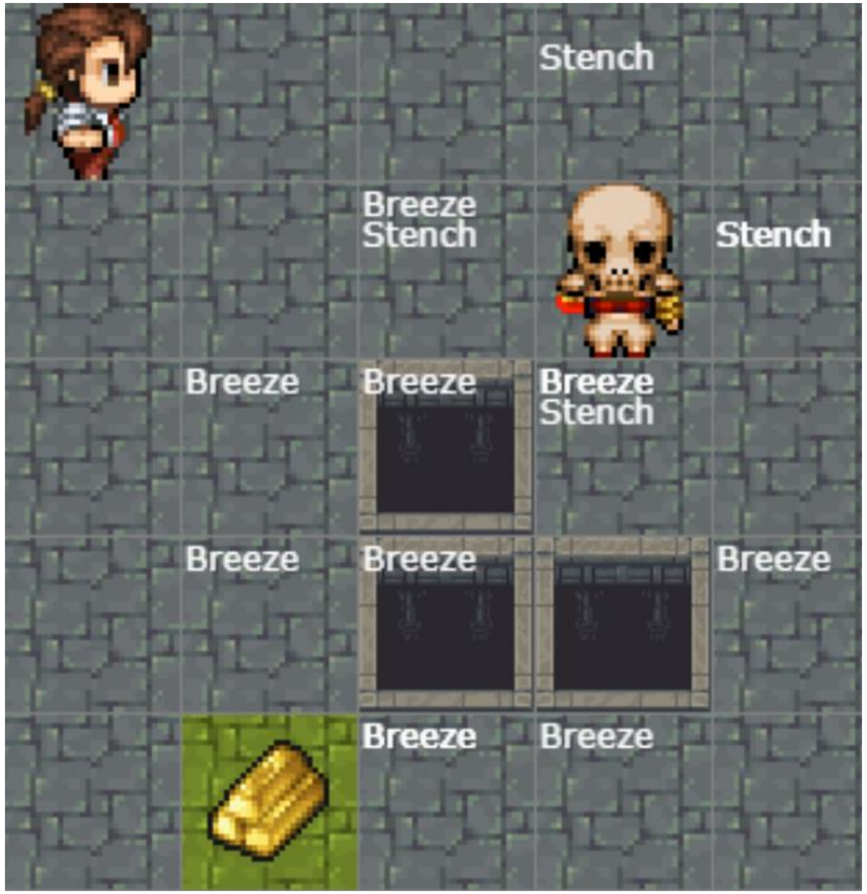
Warm-up:

Play Minesweeper or Wumpus World!





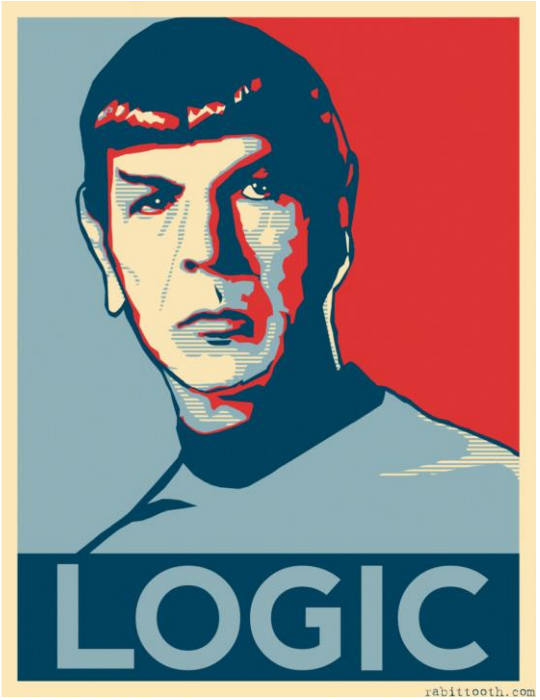
Monty Python Inference

There are ways of telling whether she is a witch

https://www.youtube.com/watch?v=rf71YotfykQ&t=52

AI: Representation and Problem Solving

Propositional Logic



Instructor: Pat Virtue

Slide credits: CMU AI, http://ai.berkeley.edu

Models and Knowledge Bases

Entailment and Satisifiability

Models and Knowledge Bases

Example: Sudoku

Model

Assignment of values to all variables

X_{ij}

$X_{ij} \neq 1$

1			
	2	1	
		3	
			4

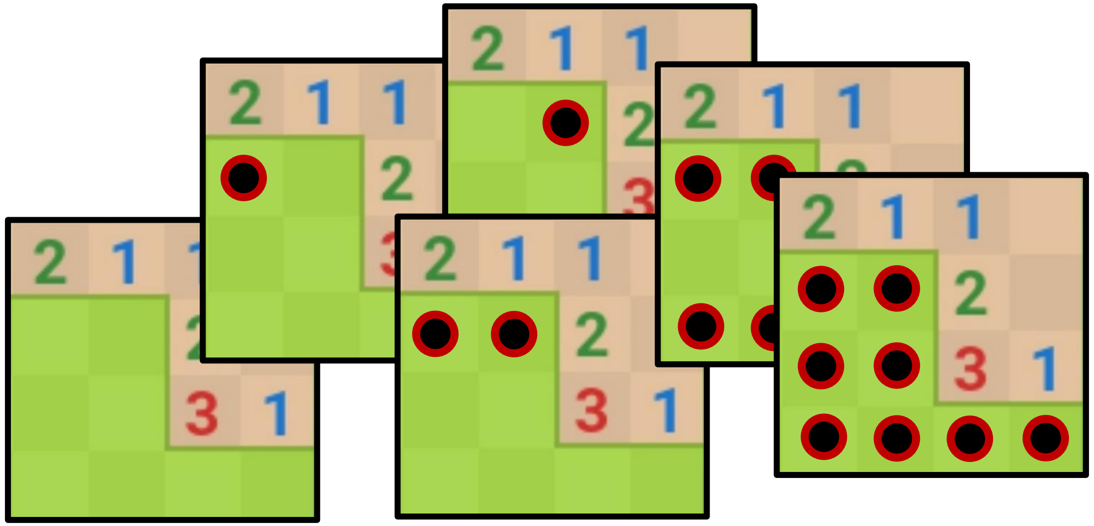
Knowledge Base

Collection of things we know to be true

- Rules of the world
- Observations
- Things we have figured out

Models and Knowledge Bases

Example: Minesweeper



Model

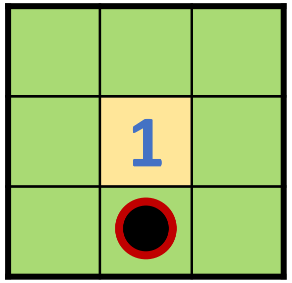
Assignment of values to all variables

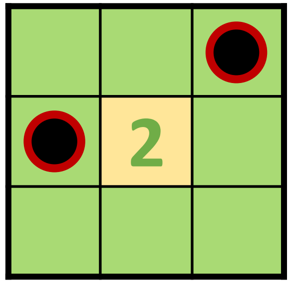
Knowledge Base

Collection of things we know to be true

- Rules of the world
- Observations
- Things we have figured out

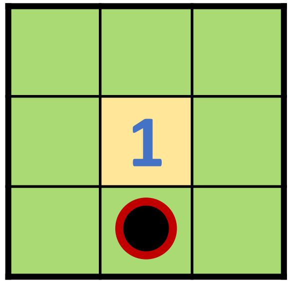
Numbers indicate how many mines

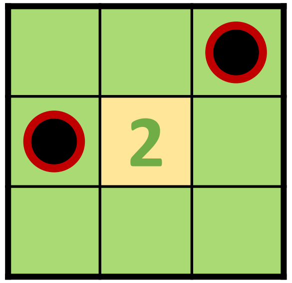




Minesweeper

Numbers indicate how many mines are in the 8 adjacent cells





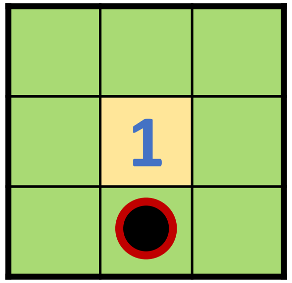
What are we trying to figure out?

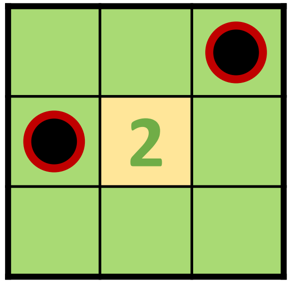
- A path (a sequence of actions)?
- A complete solution?

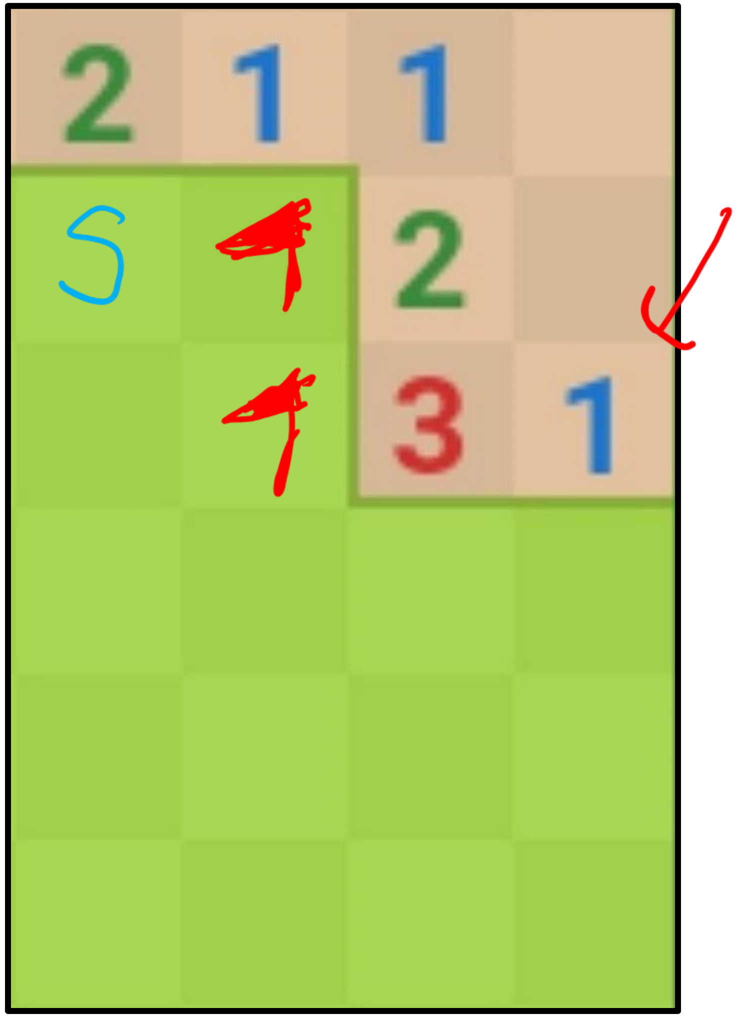
Image: Google Minesweeper game

Minesweeper

Numbers indicate how many mines are in the 8 adjacent cells





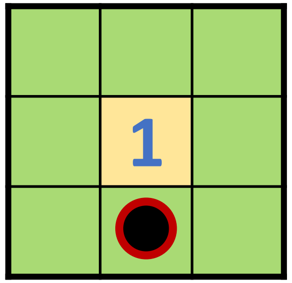


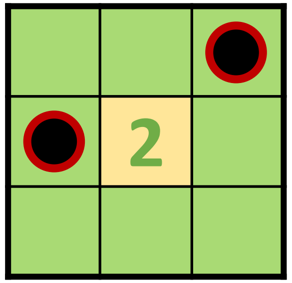
We're trying to figure out what to do next

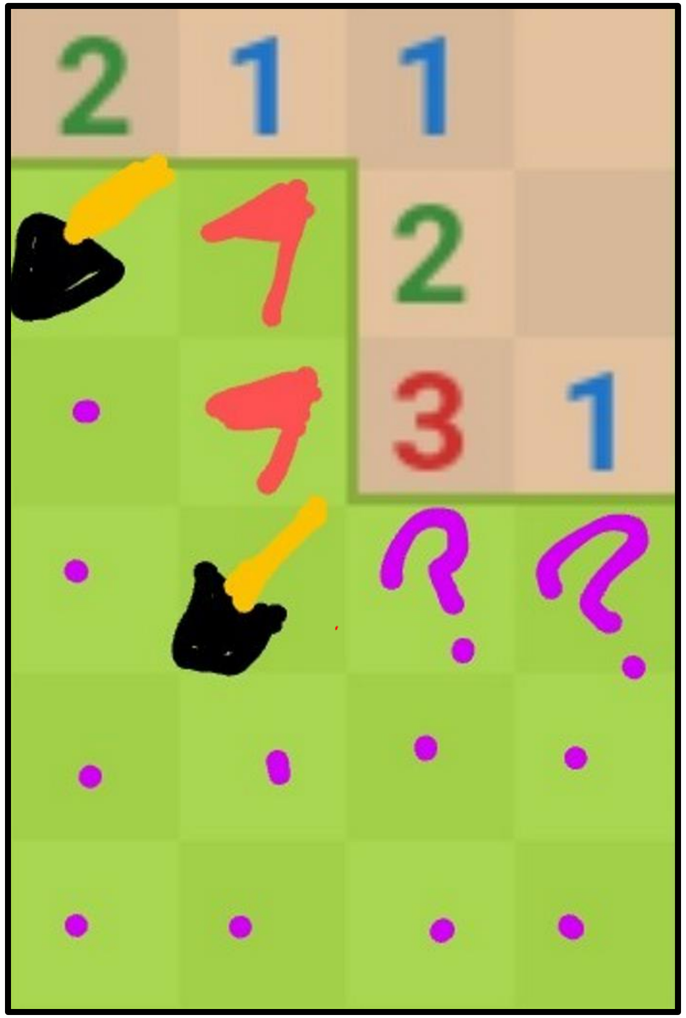
- Which unvisited spaces that are definitely safe?
- Which unvisited spaces that are definitely dangerous?
- (What about the other spaces?)

Minesweeper

Numbers indicate how many mines are in the 8 adjacent cells







We're trying to figure out what to do next

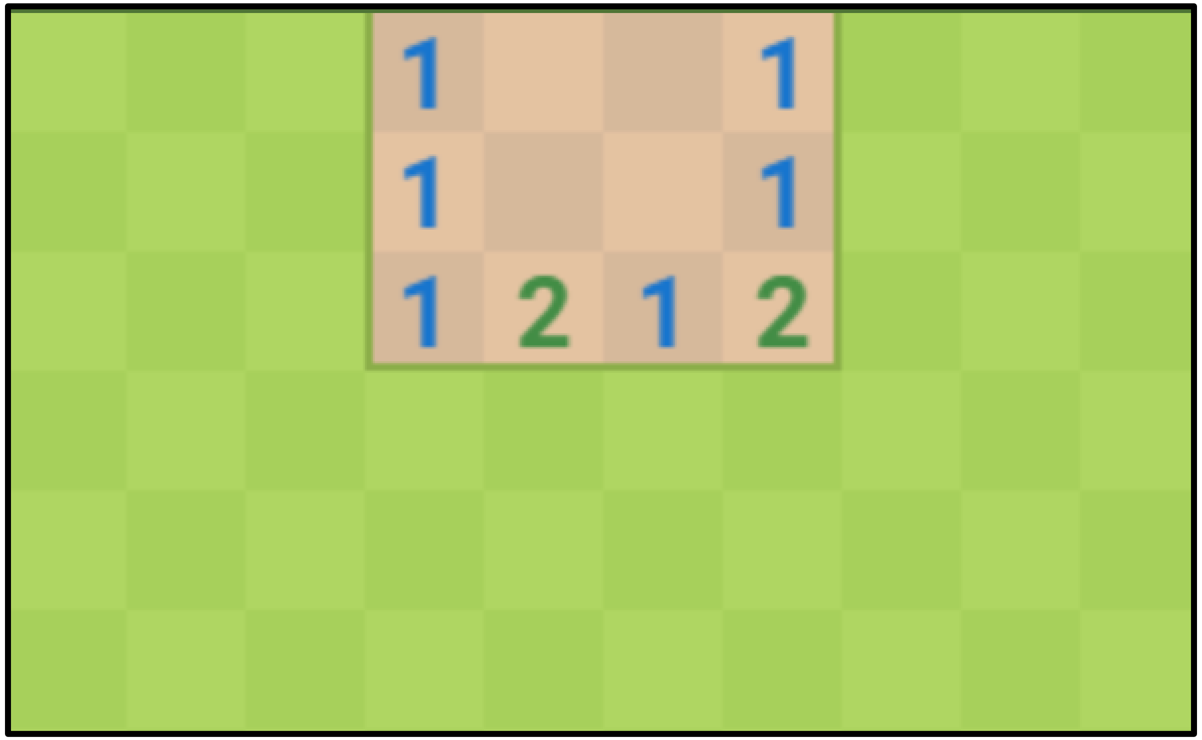
- Which unvisited spaces that are definitely safe?
- Which unvisited spaces that are definitely dangerous?
- (What about the other spaces?)

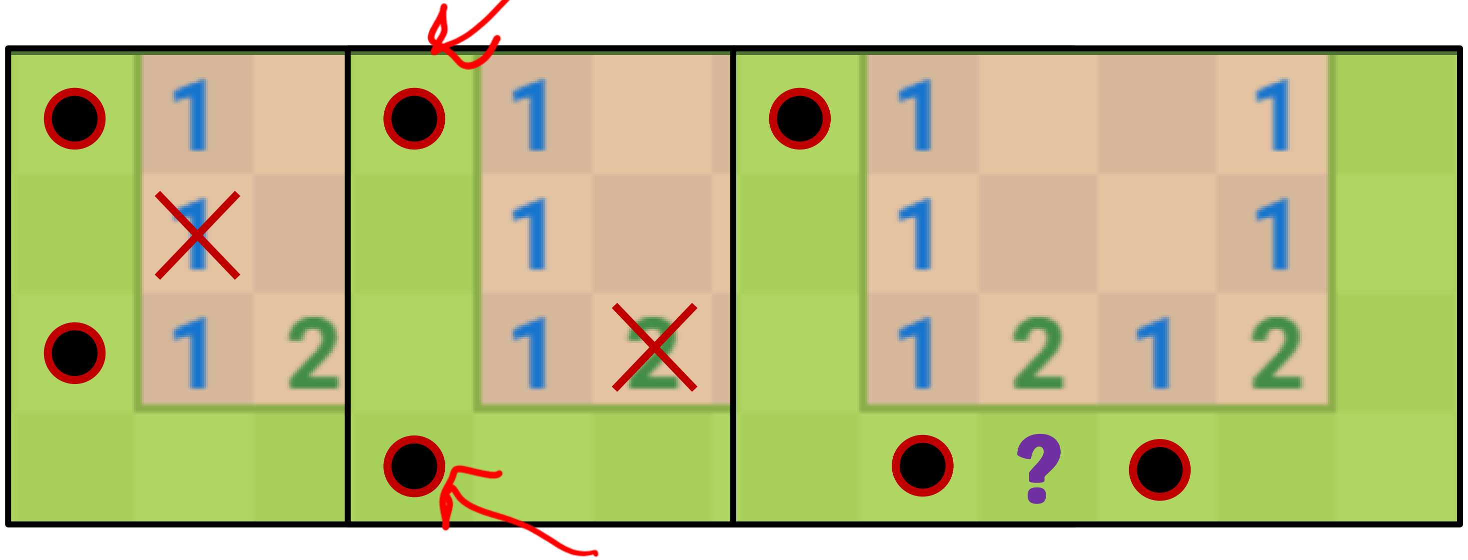
Minesweeper

It may take a few logical steps to reason about:

1) What is possible
2) What is impossible
3) What is still unknown

Example human inference steps:



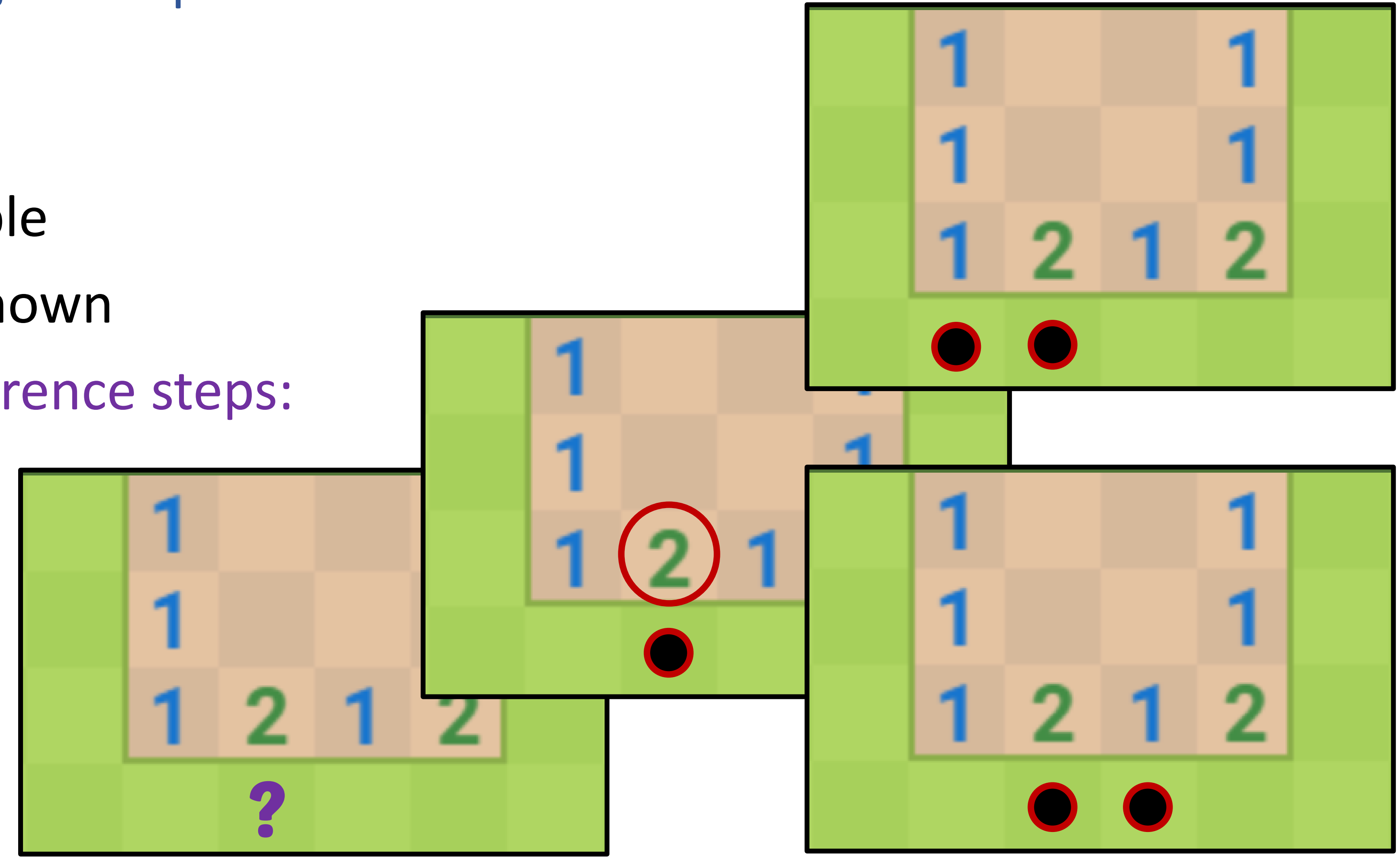


Minesweeper

It may take a few logical steps to reason about:

1) What is possible
2) What is impossible
3) What is still unknown

Example human inference steps:



Entailment and Satisfiability

What reasoning are we doing?

- Can I click here? / Is this definitely safe?
  - Yes: For all possible configurations (models), none of them have a mine in that location
  - No: There exists (at least) one possible configuration with a mine in that location

Entailment: definitely safe

Satisfiability: possibly not safe

- Is it possibly safe?
  - Yes: There exists (at least) one possible configuration with a mine in that location
  - No: For all possible configurations (models), all of them have a mine in that location → It's definitely dangerous

Satisfiability: possibly safe

Entailment: definitely not safe

Entailment and Satisfiability

More formally

- Symbol (variable)
- Models (all symbols assigned a value)
- Satisfiable: there exists (at least one) model that meets the constraints
- Entailment: statement is true for all models that meet the constraints

How do we get a computer to do this?

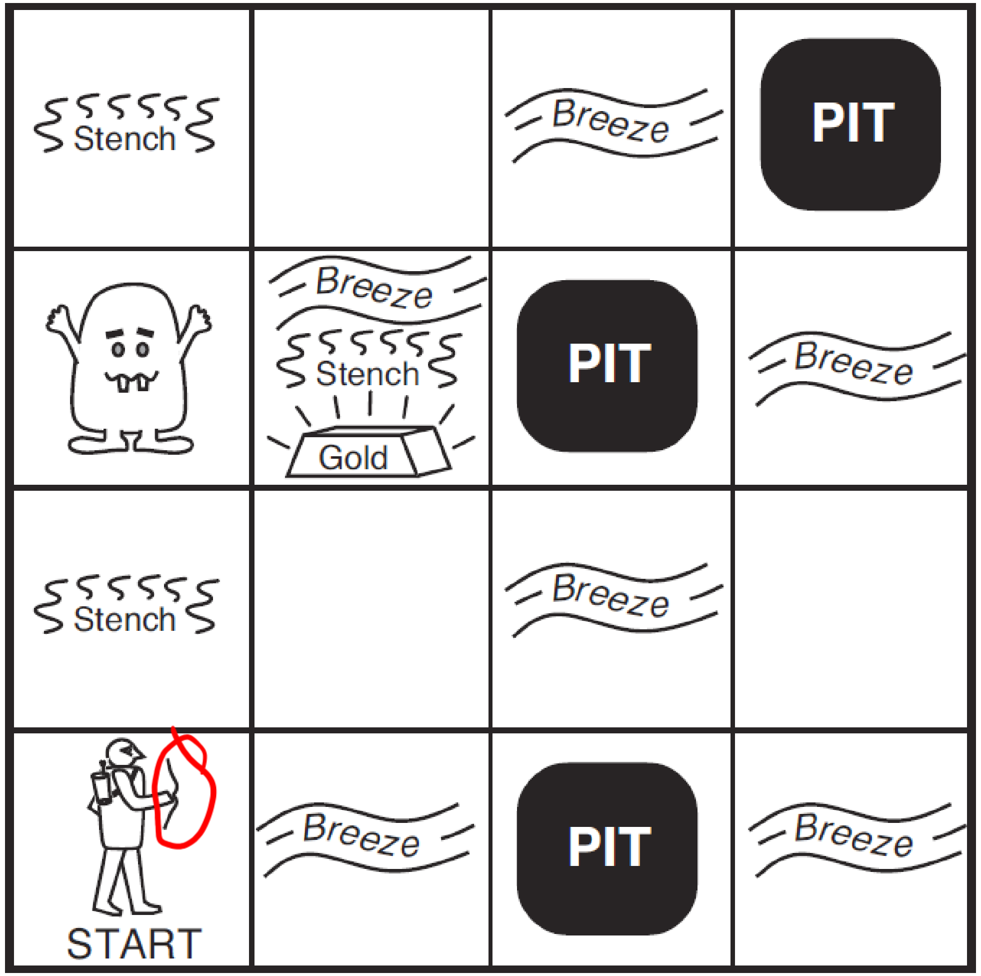
Wumpus World

We collect information as we move to a new grid in the world

- Breeze: if next to a Pit
- Stench: if next to a Wumpus
- Both
- Nothing
- Oh, and there's gold

We're trying to figure out what to do next

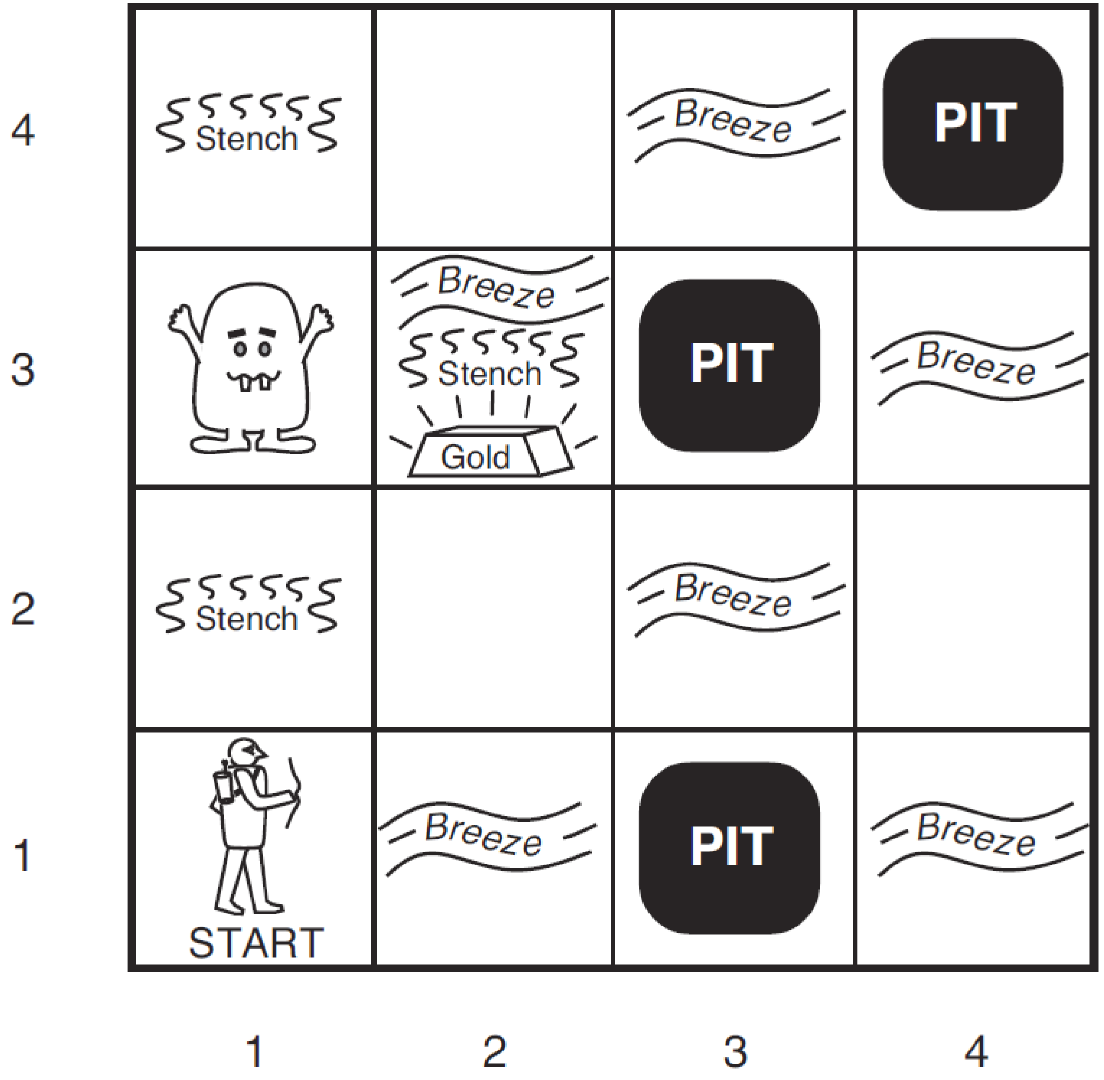
- Which unvisited spaces that are definitely safe?
- Which unvisited spaces that are definitely dangerous?
- (What about the other spaces?)



Wumpus World

Symbols for Wumpus World

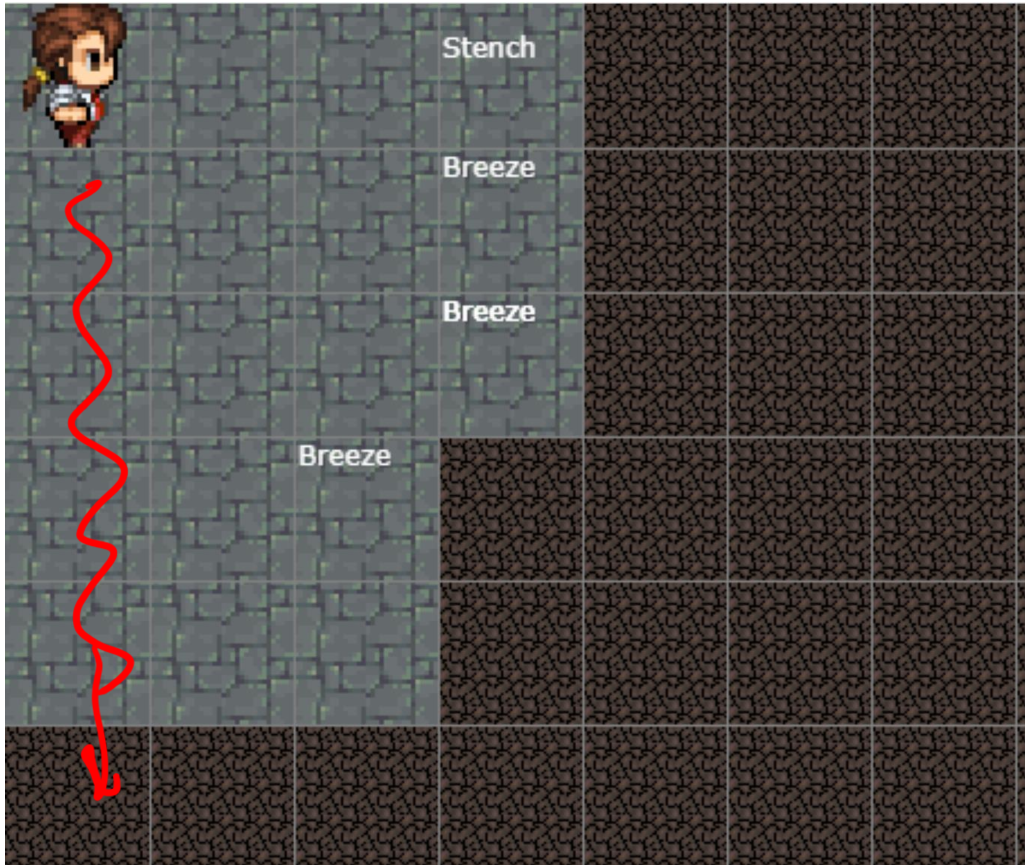
- B_{ij} = breeze felt
- S_{ij} = stench smelt
- P_{ij} = pit here
- W_{ij} = wumpus here
- G = gold



http://thiagodnf.github.io/wumpus-world-simulator/

Wumpus World

Reasoning about how to get safely get more information!



http://thiagodnf.github.io/wumpus-world-simulator/

Models and Knowledge Bases: Wumpus World

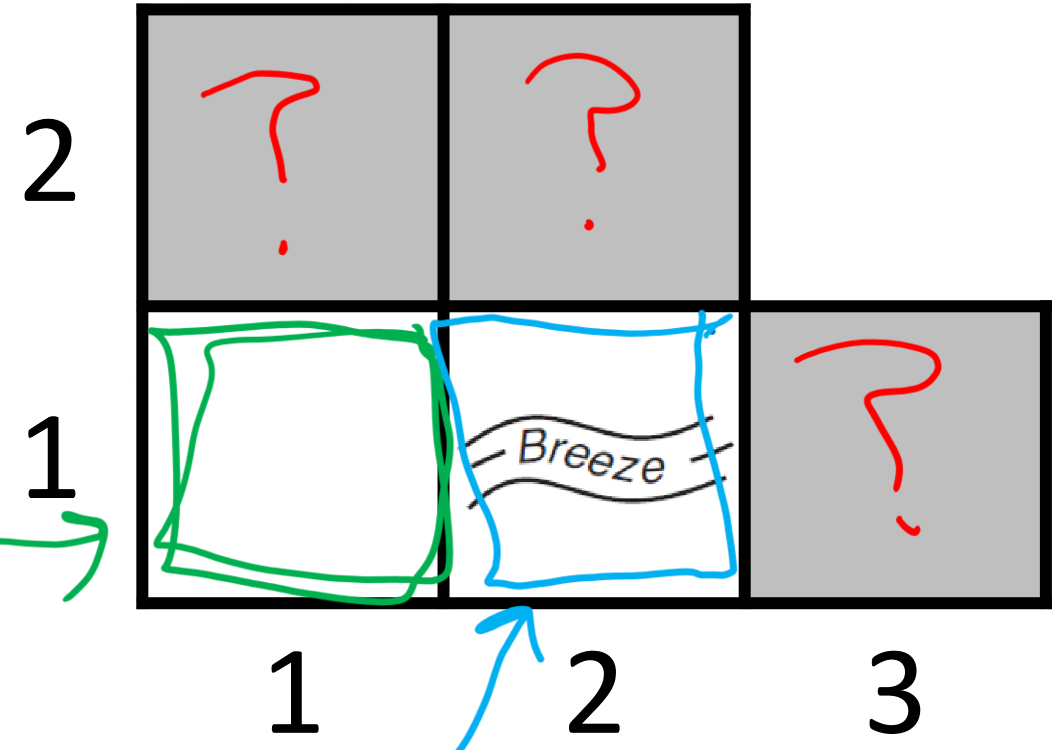
Possible Models

Symbols we are considering

- $P_{1,2}$ $P_{2,2}$ $P_{3,1}$

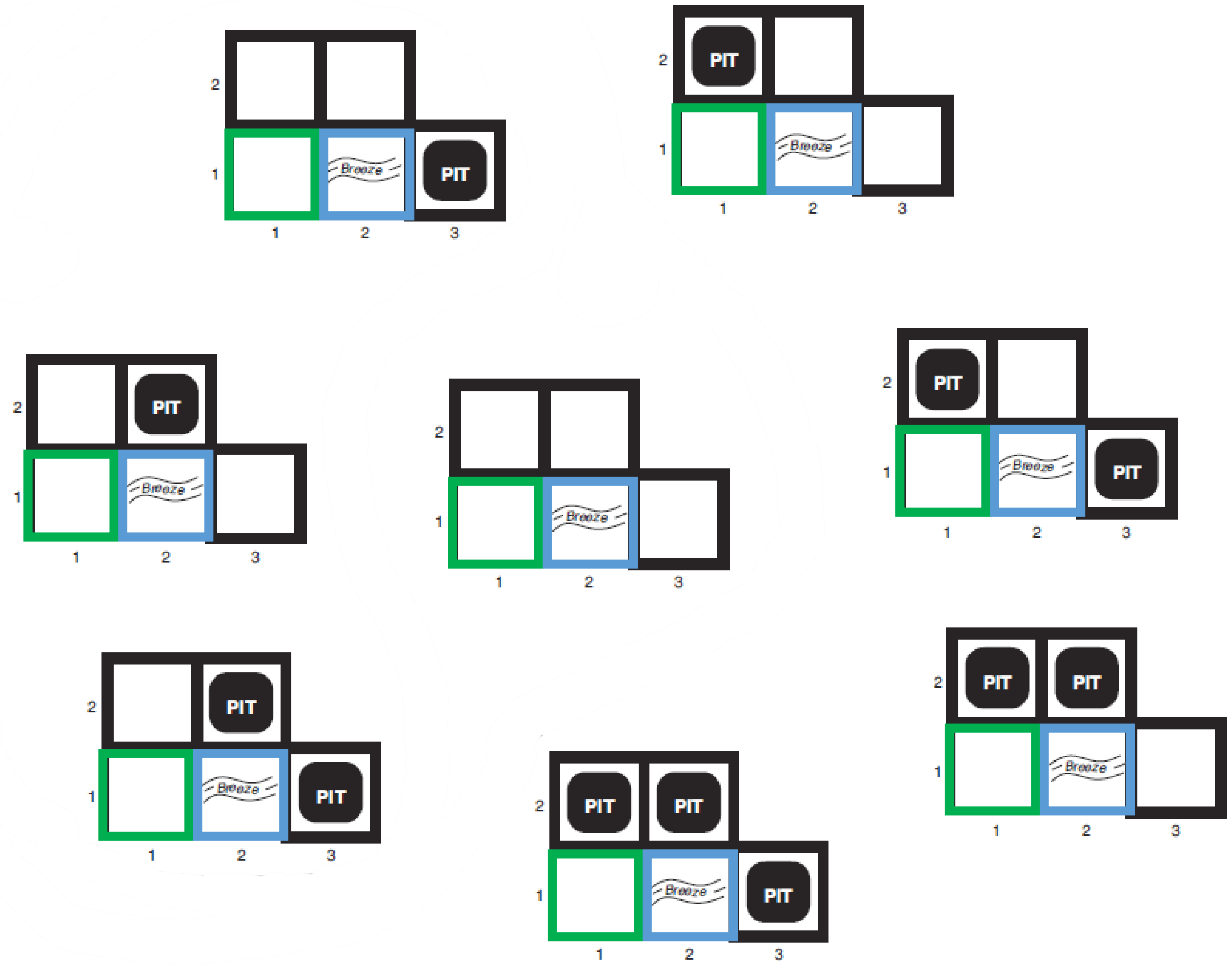
Knowledge base

- Breeze ⇒ Adjacent P
- Nothing in [1,1]
- Breeze in [2,1]



Models and Knowledge Bases: Wumpus World

Possible Models



Symbols we are considering

- $P_{1,2}$ $P_{2,2}$ $P_{3,1}$

Knowledge base

- Breeze ⇒ Adjacent Pit
- Nothing in [1,1]
- Breeze in [2,1]

Entailment: Wumpus World

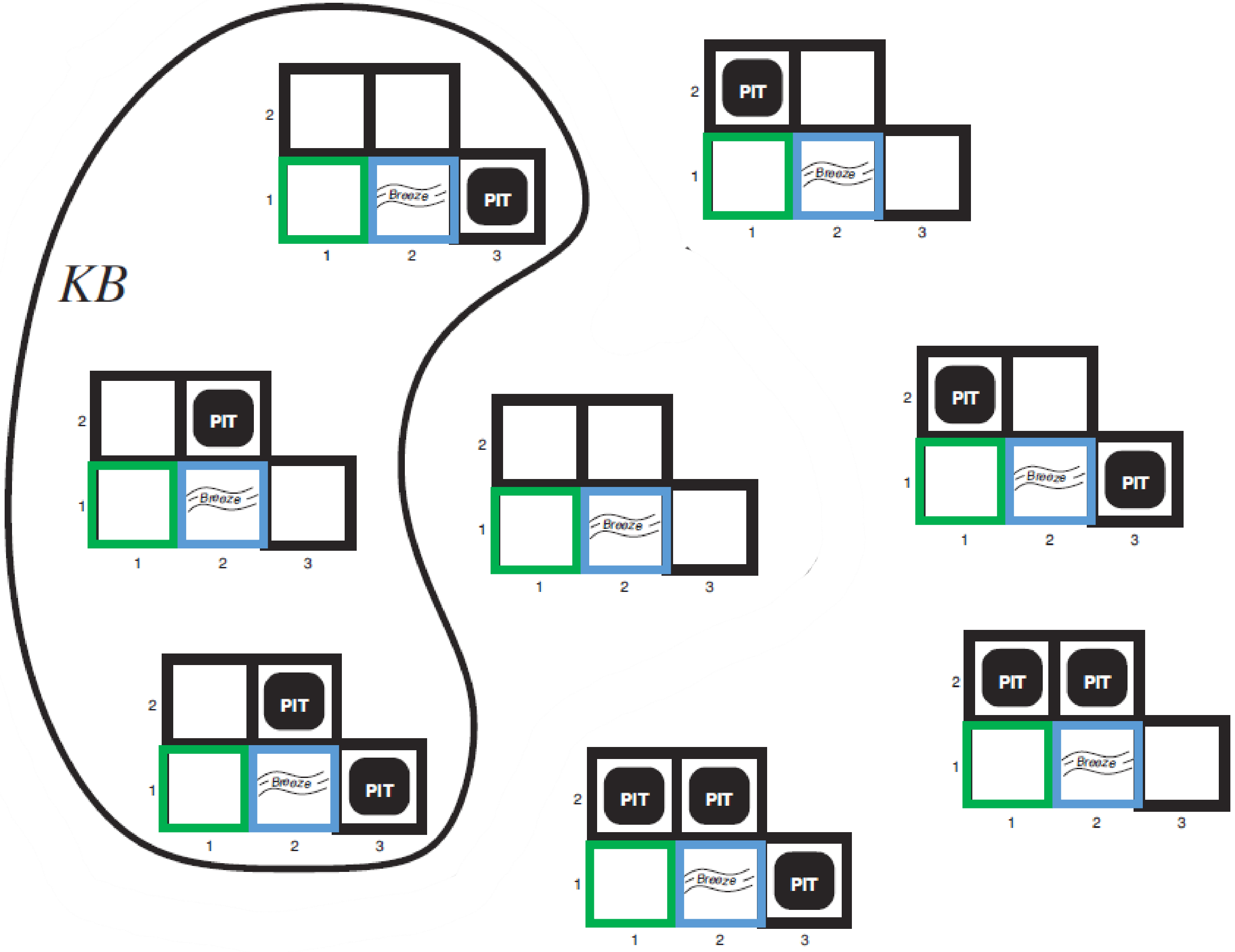
Possible Models

Symbols we are considering

- $P_{1,2}$ $P_{2,2}$ $P_{3,1}$

Knowledge base

- Breeze ⇒ Adjacent Pit
- Nothing in [1,1]
- Breeze in [2,1]



Entailment: Wumpus World

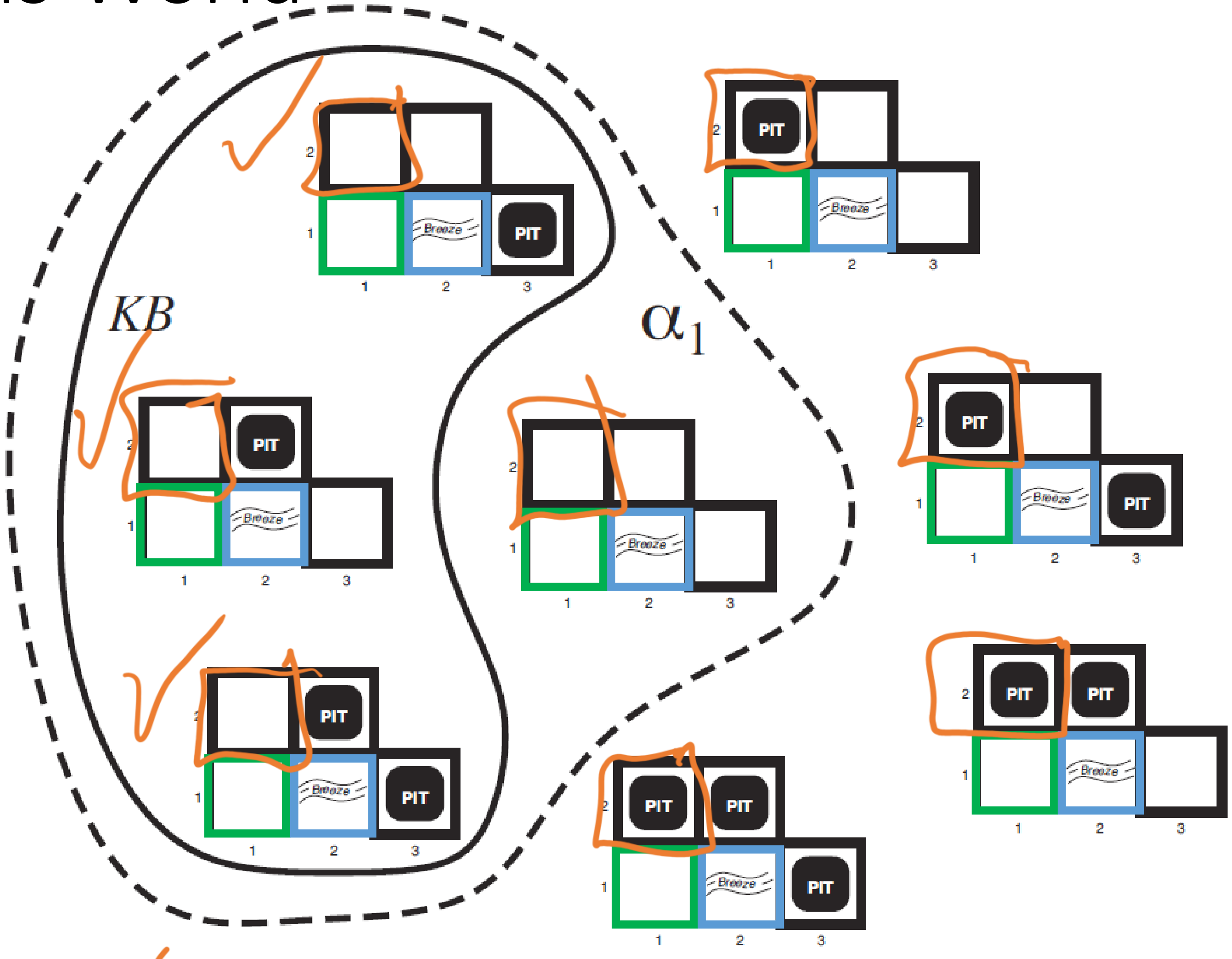
Possible Models

Symbols we are considering

- $P_{1,2}$ $P_{2,2}$ $P_{3,1}$

Knowledge base

- Breeze ⇒ Adjacent Pi
- Nothing in [1,1]
- Breeze in [2,1]



Query α_1:

- No pit in [1,2]

KB $\models \alpha_1$ Yes

Entailment: Wumpus World

Possible Models

Symbols we are considering

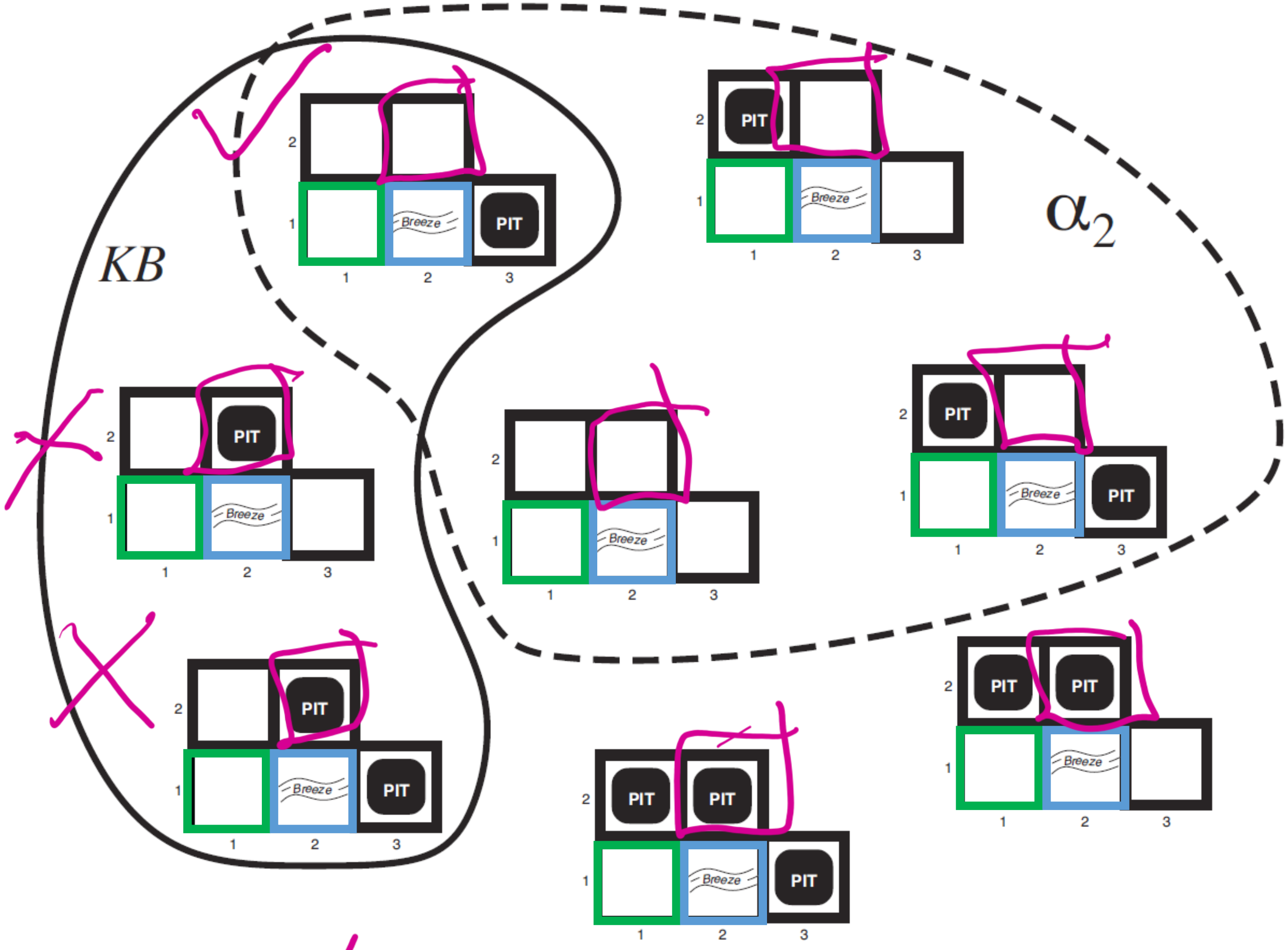
- $P_{1,2}$ $P_{2,2}$ $P_{3,1}$

Knowledge base

- Breeze ⇒ Adjacent Pit
- Nothing in [1,1]
- Breeze in [2,1]

Query α_2:

- No pit in [2,2]



$KB \models \alpha_2$ No

Entailment

Entailment: $\alpha \models \beta$ ("α entails β" or "β follows from α") iff in every world where α is true, β is also true

- I.e., the α-worlds are a subset of the β-worlds [*models*(α) $\subseteq$ *models*(β)]

Usually, we want to know if *KB* $\models$ *query*

- *models*(*KB*) $\subseteq$ *models*(*query*)
- In other words
  - *KB* removes all impossible models (any model where *KB* is false)
  - If *query* is true in all of these remaining models, we conclude that *query* must be true

Entailment and implication are very much related

- However, entailment relates two sentences, while an implication is itself a sentence (usually derived via inference to show entailment)

Entailment: Wumpus World

Possible Models

Symbols we are considering

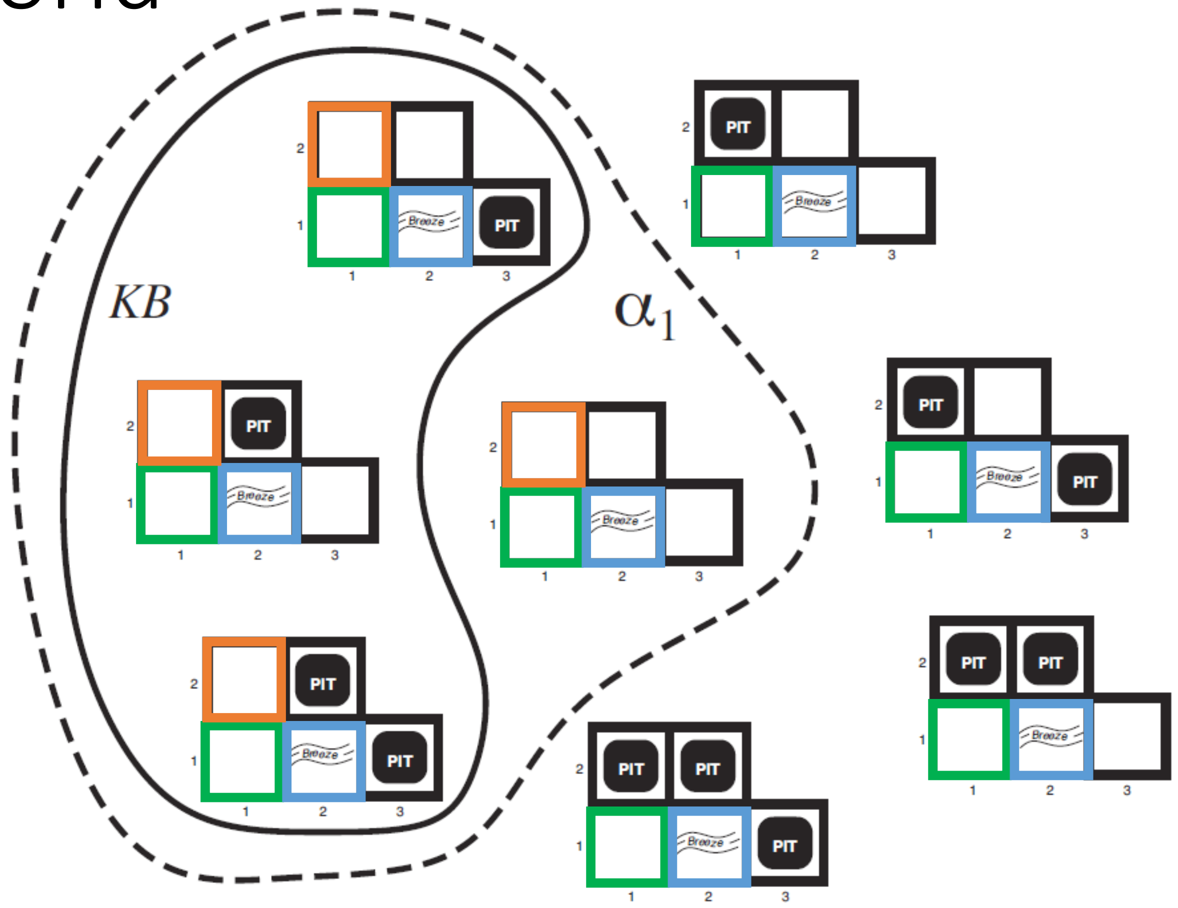
- $P_{1,2}$ $P_{2,2}$ $P_{3,1}$

Knowledge base

- Breeze ⇒ Adjacent Pit
- Nothing in [1,1]
- Breeze in [2,1]

Query α_1:

- No pit in [1,2]



Entailment: KB |= α

"KB entails α" iff in every world where KB is true, α is also true

Entailment: Wumpus World

Possible Models

Symbols we are considering

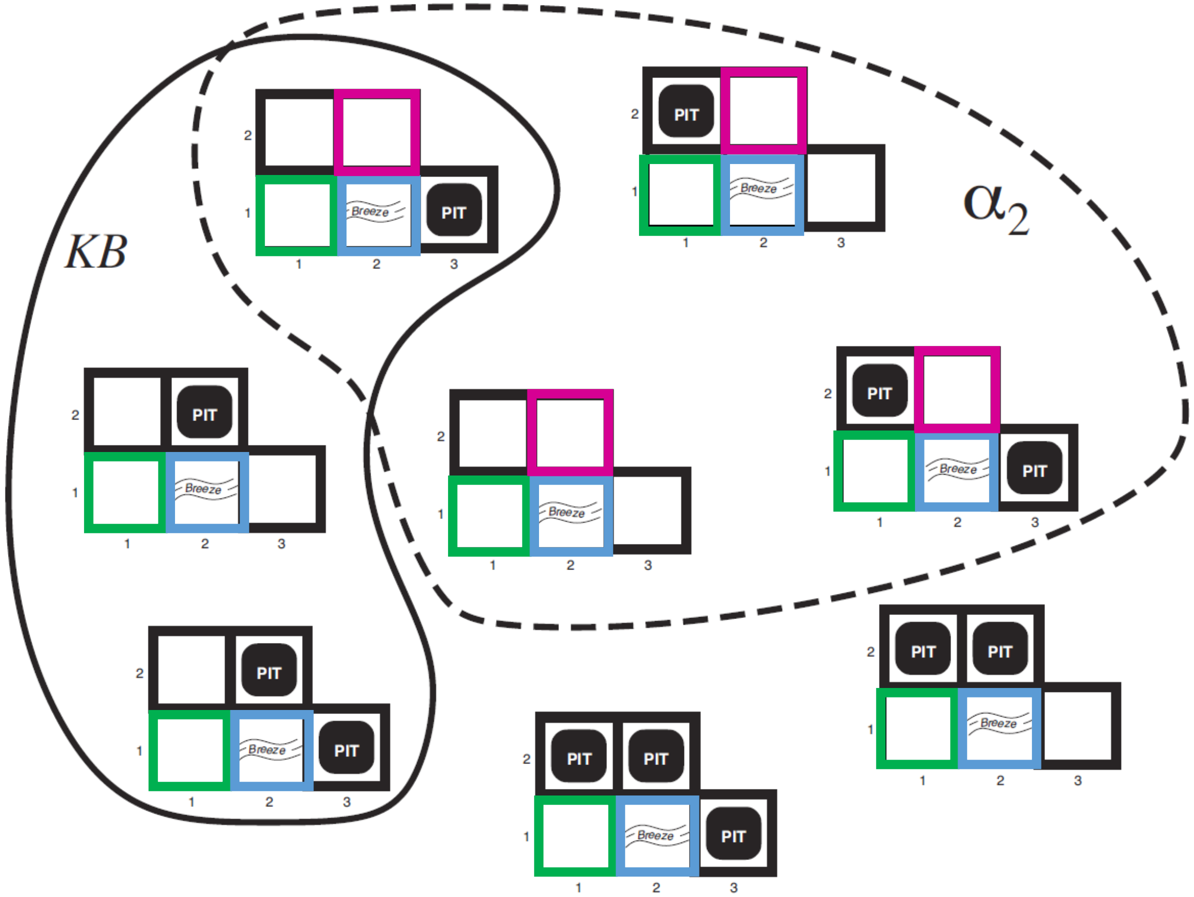
- $P_{1,2}$ $P_{2,2}$ $P_{3,1}$

Knowledge base

- Breeze ⇒ Adjacent Pit
- Nothing in [1,1]
- Breeze in [2,1]

Query α_2:

- No pit in [2,2]



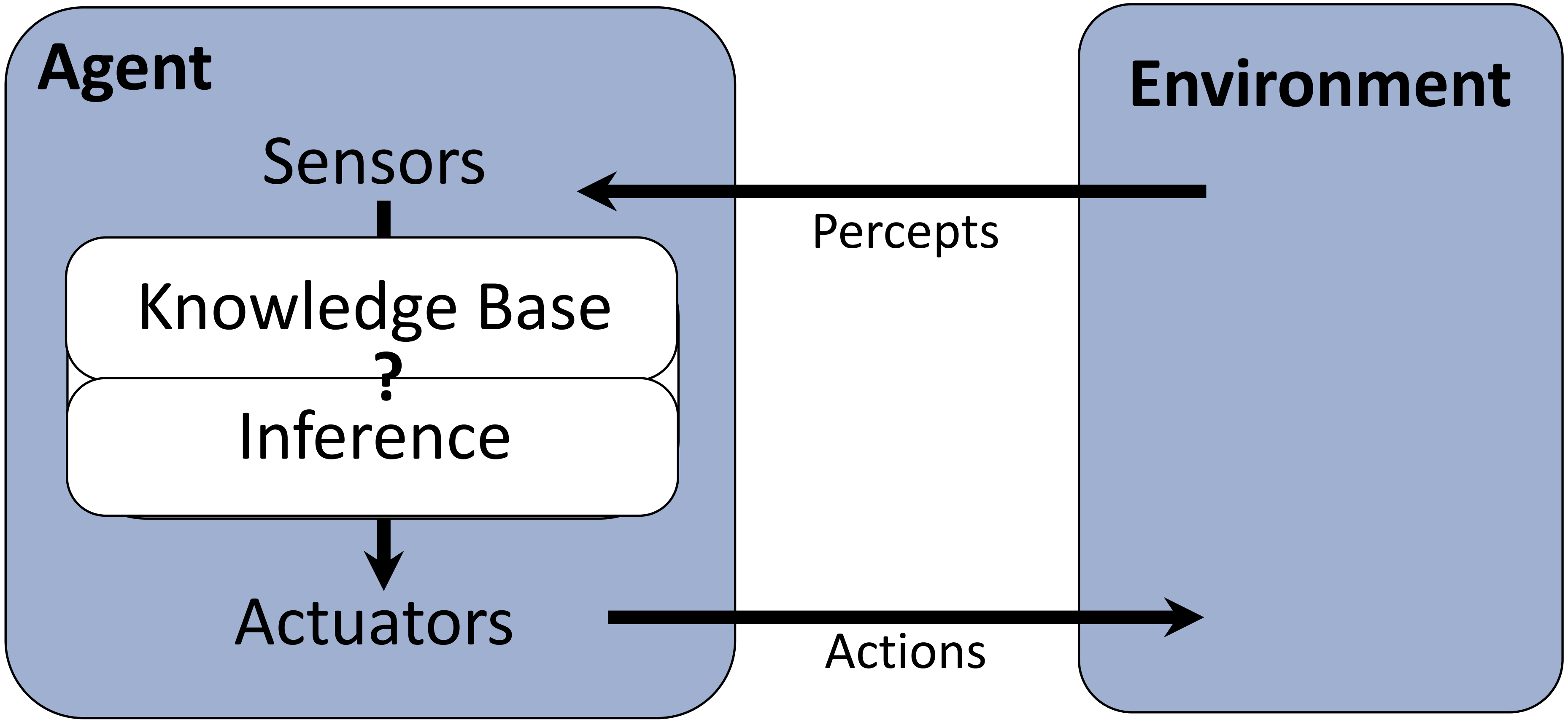
Entailment: KB |= α

"KB entails α" iff in every world where KB is true, α is also true

High-level View: Logical Agents

Logical Agents

Logical agents and environments



Logical Agents

So what do we TELL our knowledge base (KB)?

- Facts (sentences)
  - The grass is green
  - The sky is blue
- Rules (sentences)
  - Eating too much candy makes you sick
  - When you’re sick you don’t go to school
- Percepts and Actions (sentences)
  - Pat ate too much candy today

What happens when we ASK the agent?

- Inference – new sentences created from old
  - Pat is not going to school today

A Knowledge-based Agent

function KB-AGENT(percept) returns an action

persistent: KB, a knowledge base

persistent: t, an integer, initially 0

TELL(KB, PROCESS-PERCEPT(percept, t))

action ← ASK(KB, PROCESS-QUERY(t))

TELL(KB, PROCESS-RESULT(action, t))

t←t+1

return action

Outline

Models and Knowledge Bases

Entailment and Satisfiability

How to get a computer to do this?

Need:

Representation: Language

- PL
- FoL $\exists$ $\forall$

Problem Solving: Algorithm

- Model checking: try them all
- Theorem proving: logical steps

Next lecture

model world
configuration

Logic Language

Natural language?

Propositional logic

- Syntax: $P \vee (\neg Q \wedge R)$; $X_1 \Leftrightarrow (\text{Raining} \Rightarrow \text{Sunny})$
- Possible model: {P=true, Q=true, R=false, S=true} or 1101
- Semantics: $\alpha \wedge \beta$ is true for a model iff is α true and β is true (etc.)

First-order logic

- Syntax: $\forall x\ \exists y\ P(x,y) \wedge \neg Q(\text{Joe},f(x)) \Rightarrow f(x)=f(y)$
- Possible model: Objects o_1, o_2, o_3; P holds for $<o_1,o_2>$; Q holds for $<o_3>$; $f(o_1)=o_1$; $\text{Joe}=o_3$; etc.
- Semantics: $\phi(\sigma)$ is true for a model if $\sigma=o_j$ and ϕ holds for o_j; etc.

Propositional Logic

Poll 1

If we know that $\overbrace{\underline{A \vee B} \text{ and } \underline{\neg B \vee C}}^{KB}$ are true,

what do we know about $\underbrace{A \vee C}_{q}$?

$(A \vee B) \wedge (\neg B \vee C)$

$\vdots$

$A \vee C$

i. $A \vee C$ is guaranteed to be true

ii. $A \vee C$ is guaranteed to be false

iii. We don't have enough information to say anything definitive about $A \vee C$

$\neg(A \vee C)$

$\neg q$

if $\neg B$

C

else if B

$\neg q \wedge KB$

Poll 1

If we know that $A \lor B$ and $\neg B \lor C$ are true, what do we know about $A \lor C$?

A	B	C	$A \lor B$	$\neg B \lor C$	$A \lor C$
false	false	false	false	true	false
false	false	true	false	true	true
false	true	false	true	false	false
false	true	true	true	true	true
true	false	false	true	true	true
true	false	true	true	true	true
true	true	false	true	false	true
true	true	true	true	true	true

Poll 1

If we know that $A \vee B$ and $\neg B \vee C$ are true, what do we know about $A \vee C$?

KB ⊨ q

A	B	C	$A \vee B$	$\neg B \vee C$	$A \vee C$
false	false	false	false	true	false
false	false	true	false	true	true
false	true	false	true	false	false
false	true	true	true	true	true
true	false	false	true	true	true
true	false	true	true	true	true
true	true	false	true	false	true
true	true	true	true	true	true

Poll 1

If we know that $A \vee B$ and $\neg B \vee C$ are true,

what do we know about $A \vee C$?

i. $A \vee C$ is guaranteed to be true

ii. $A \vee C$ is guaranteed to be false

iii. We don't have enough information to say anything definitive about $A \vee C$

Poll 2

If we know that $A \vee B$ and $\neg B \vee C$ are true,

what do we know about A?

i. A is guaranteed to be true

ii. A is guaranteed to be false

iii. We don't have enough information to say anything definitive about A

Poll 2

A ?

If we know that $A \lor B$ and $\neg B \lor C$ are true, what do we know about A?

KB

A	B	C	$A \lor B$	$\neg B \lor C$	$A \lor C$
false	false	false	false	true	false
false	false	true	false	true	true
false	true	false	true	false	false
false	true	true	true	true	true
true	false	false	true	true	true
true	false	true	true	true	true
true	true	false	true	false	true
true	true	true	true	true	true

Poll 2

If we know that $A \vee B$ and $\neg B \vee C$ are true,

what do we know about A?

i. A is guaranteed to be true
ii. A is guaranteed to be false
iii. We don't have enough information to say anything definitive about A

Propositional Logic

Symbol:

- Variable that can be true or false
- We'll try to use capital letters, e.g. A, B, $P_{1,2}$
- Often include True and False

Operators:

- $\neg$ A: not A
- A $\wedge$ B: A and B (conjunction)
- A $\vee$ B: A or B (disjunction) Note: this is not an "exclusive or"
- A $\Rightarrow$ B: A implies B (implication). If A then B
- A $\Leftrightarrow$ B: A if and only if B (biconditional)

Sentences

Propositional Logic Syntax

Given: a set of proposition symbols $\{X_1, X_2, \ldots, X_n\}$

- (we often add True and False for convenience)

X_i is a sentence

If α is a sentence then $\neg\alpha$ is a sentence

If α and β are sentences then $\alpha \wedge \beta$ is a sentence

If α and β are sentences then $\alpha \vee \beta$ is a sentence

If α and β are sentences then $\alpha \Rightarrow \beta$ is a sentence

If α and β are sentences then $\alpha \Leftrightarrow \beta$ is a sentence

And p.s. there are no other sentences!

Notes on Operators

$\boldsymbol{\alpha} \lor \boldsymbol{\beta}$ is inclusive or, not exclusive

Truth Tables

$\alpha \vee \beta$ is inclusive or, not exclusive

α	β	$\alpha \wedge \beta$
F	F	F
F	T	F
T	F	F
T	T	T

α	β	$\alpha \vee \beta$
F	F	F
F	T	T
T	F	T
T	T	T

Notes on Operators

$\boldsymbol{\alpha} \vee \boldsymbol{\beta}$ is inclusive or, not exclusive

$\boldsymbol{\alpha} \Rightarrow \boldsymbol{\beta}$ is equivalent to $\neg\boldsymbol{\alpha} \vee \boldsymbol{\beta}$

- Says who?

Truth Tables

$\alpha \Rightarrow \beta$ is equivalent to $\neg\alpha \vee \beta$

α	β	$\alpha \Rightarrow \beta$	$\neg\alpha$	$\neg\alpha \vee \beta$
F	F	T	T	T
F	T	T	T	T
T	F	F	F	F
T	T	T	F	T

Notes on Operators

$\alpha \lor \beta$ is inclusive or, not exclusive

$\alpha \Rightarrow \beta$ is equivalent to $\neg\alpha \lor \beta$

- Says who?

$\alpha \Leftrightarrow \beta$ is equivalent to $(\alpha \Rightarrow \beta) \land (\beta \Rightarrow \alpha)$

- Prove it!

Truth Tables

$\alpha \Leftrightarrow \beta$ is equivalent to $(\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha)$

α	β	$\alpha \Leftrightarrow \beta$	$\alpha \Rightarrow \beta$	$\beta \Rightarrow \alpha$	$(\alpha\Rightarrow\beta) \wedge (\beta\Rightarrow\alpha)$
F	F	T	T	T	T
F	T	F	T	F	F
T	F	F	F	T	F
T	T	T	T	T	T

Equivalence: it's true in all models. Expressed as a logical sentence:

$$(\alpha \Leftrightarrow \beta) \Leftrightarrow [(\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha)]$$

Propositional Logical Vocab

Vocab Alert!

Literal

- Atomic sentence: True, False, Symbol, ¬Symbol

Clause

- Disjunction of literals: $A \vee B \ \vee \ \neg C$

Definite clause

- Disjunction of literals, *exactly one* is positive
- $\neg A \vee B \ \vee \ \neg C$

Horn clause

- Disjunction of literals, *at most one* is positive
- All definite clauses are Horn clauses

Propositional Logic

Check if sentence is true in given model

In other words, does the model *satisfy* the sentence?

function PL-TRUE?(α,model) returns true or false

  if α is a symbol then return Lookup(α, model)

  if Op(α) = $\neg$ then return **not**(PL-TRUE?(Arg1(α),model))

  if Op(α) = $\wedge$ then return **and**(PL-TRUE?(Arg1(α),model),
                                      PL-TRUE?(Arg2(α),model))

  etc.

(Sometimes called “recursion over syntax”)

Outline

Models and Knowledge Bases

Entailment and Satisfiability

How to get a computer to do this?

Need:

Representation: Language

  - PL
  - FoL

Problem Solving: Algorithm

- Model checking: try them all
- Theorem proving: logical steps