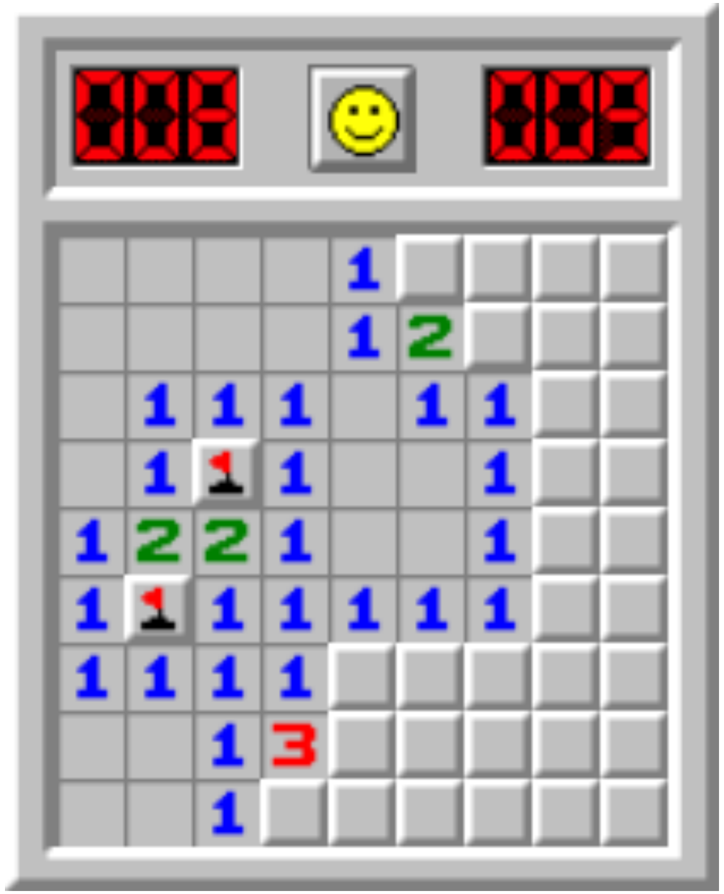
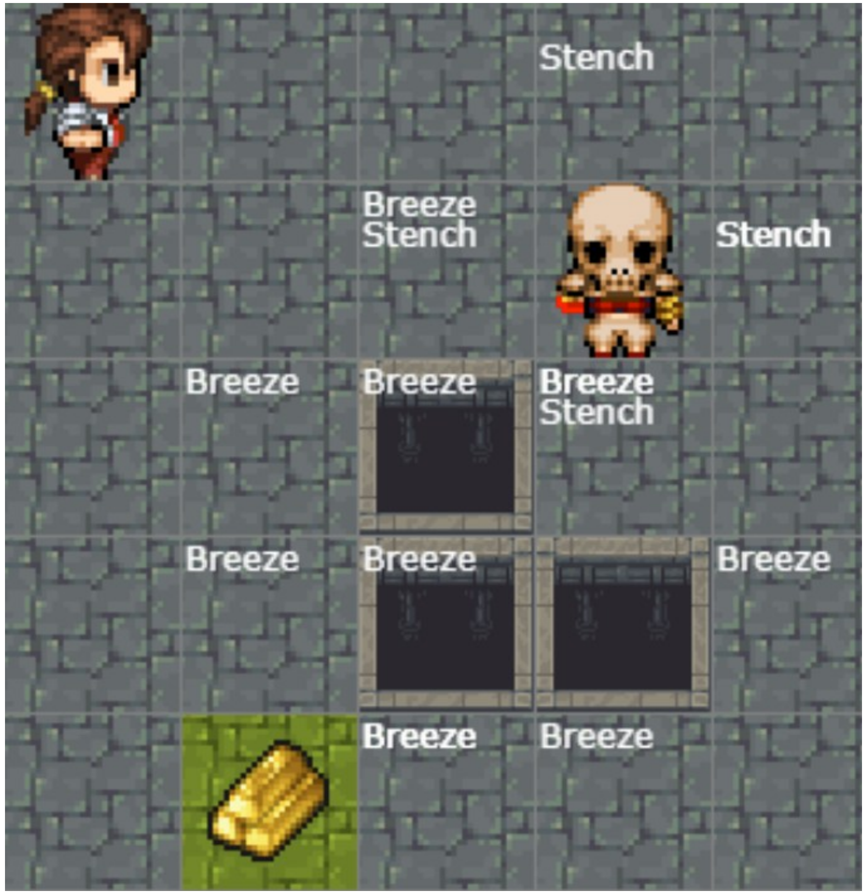
Warm-up:

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

Logical Agent Concepts

Instructor: Pat Virtue

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

AlphaGeometry

https://www.nature.com/articles/d41586-025-00406-7

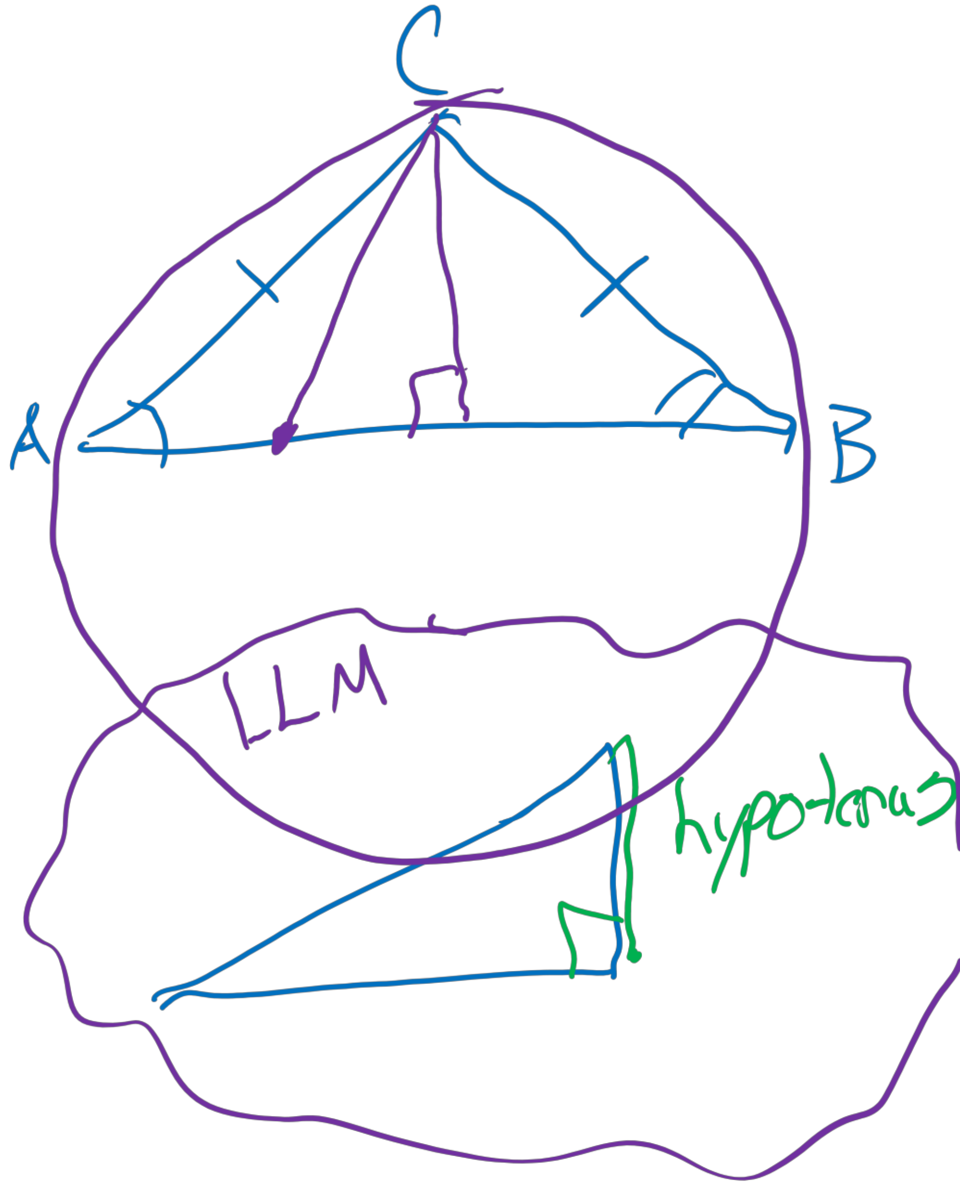
nature

NEWS | 07 February 2025

DeepMind AI crushes tough maths problems on par with top human solvers

The company's AlphaGeometry 2 reaches the level of gold-medal students in the International Mathematical Olympiad.

By Davide Castelvecchi

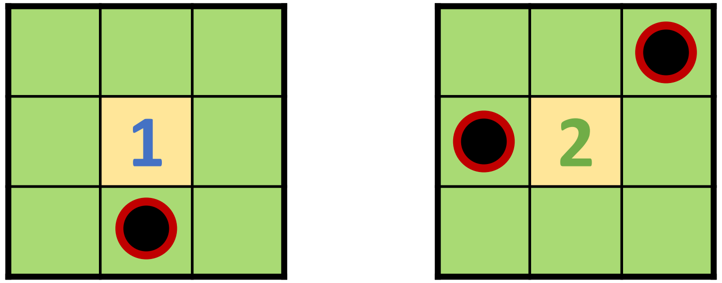


How do we solve Sudoku?

1			
	2	1	
		3	
			4

How do we play 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?

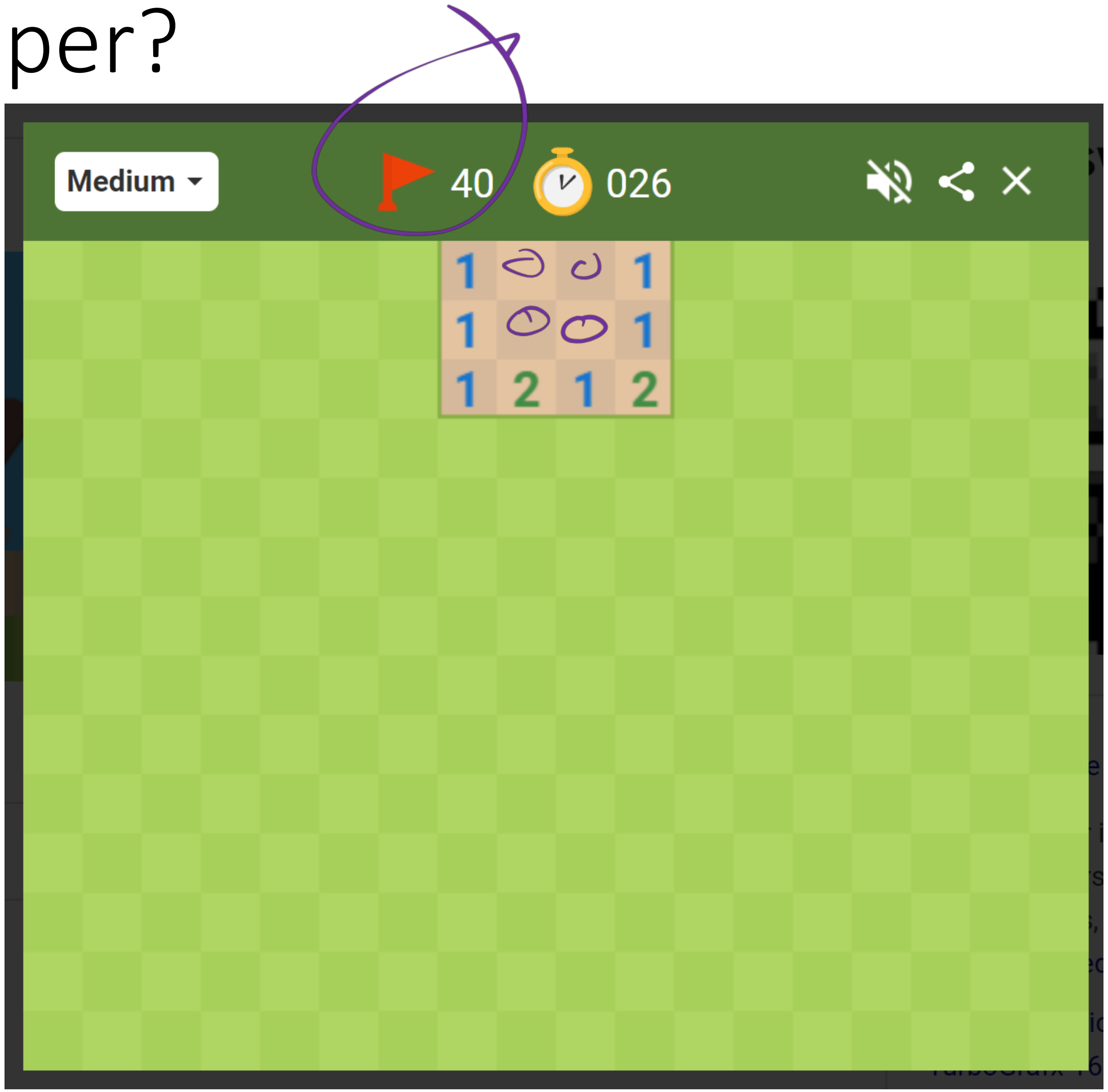
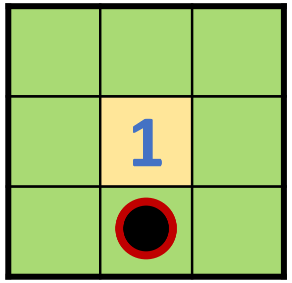
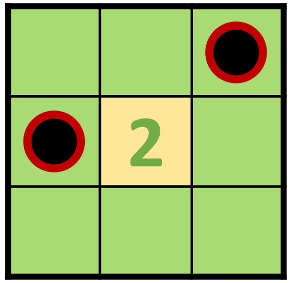


Image: Google Minesweeper game

Minesweeper

Numbers indicate how many mines are in the 8 adjacent cells



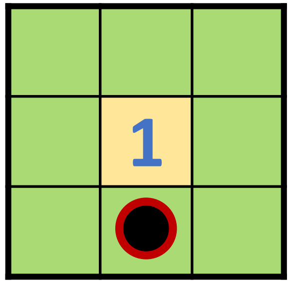


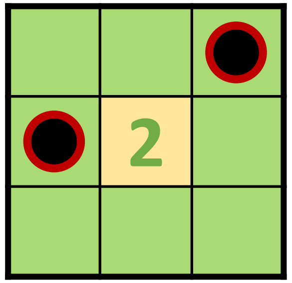
We're trying to figure out what to do next

- 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

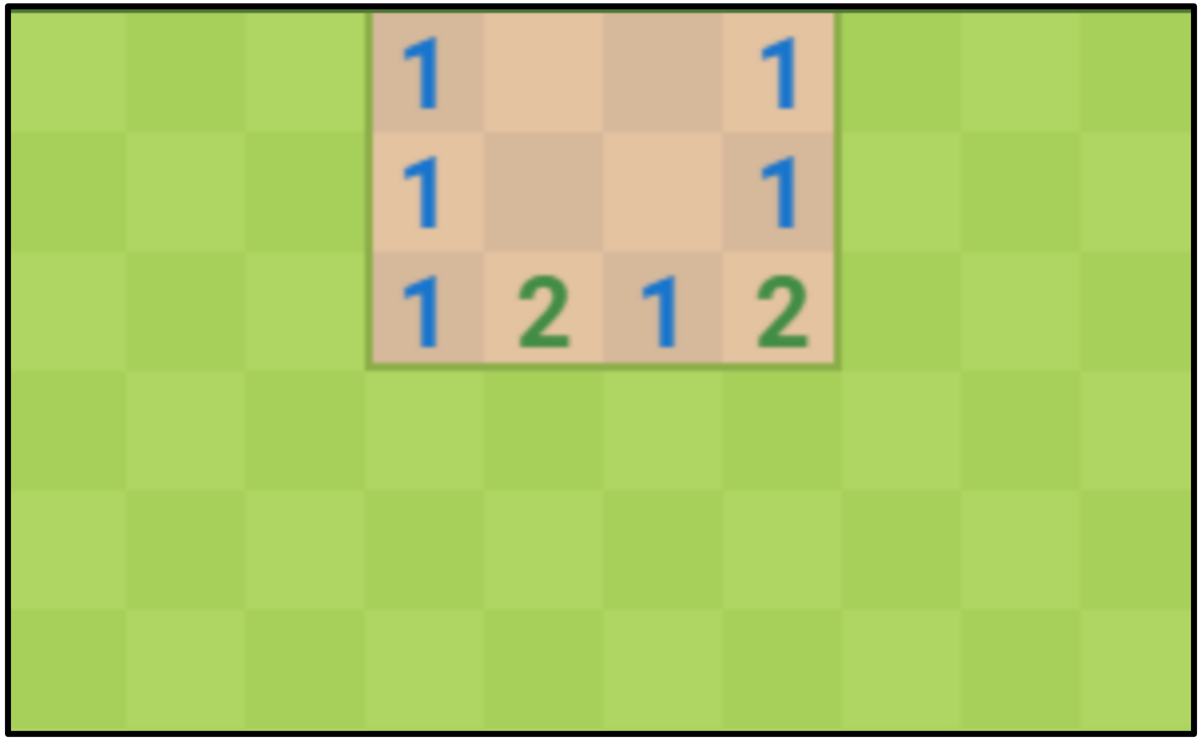
- Which unvisited spaces that are definitely safe?
- Which unvisited spaces that are definitely dangerous?
- (What about the other spaces?)

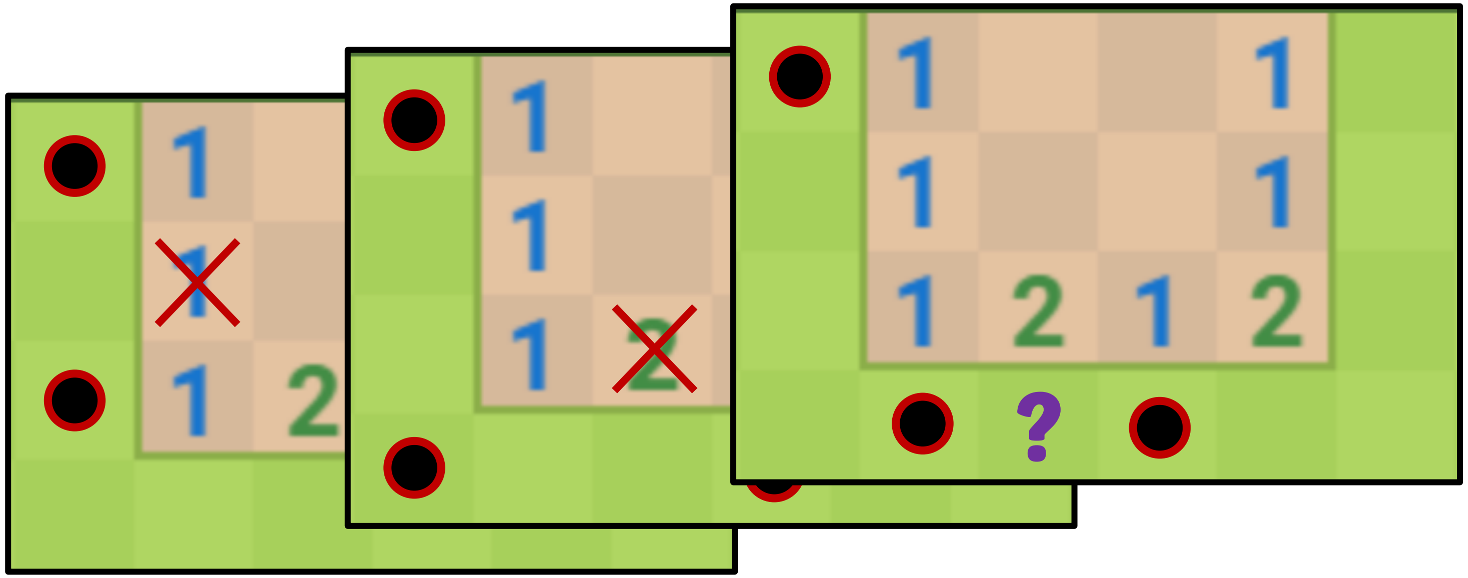
Minesweeper

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:



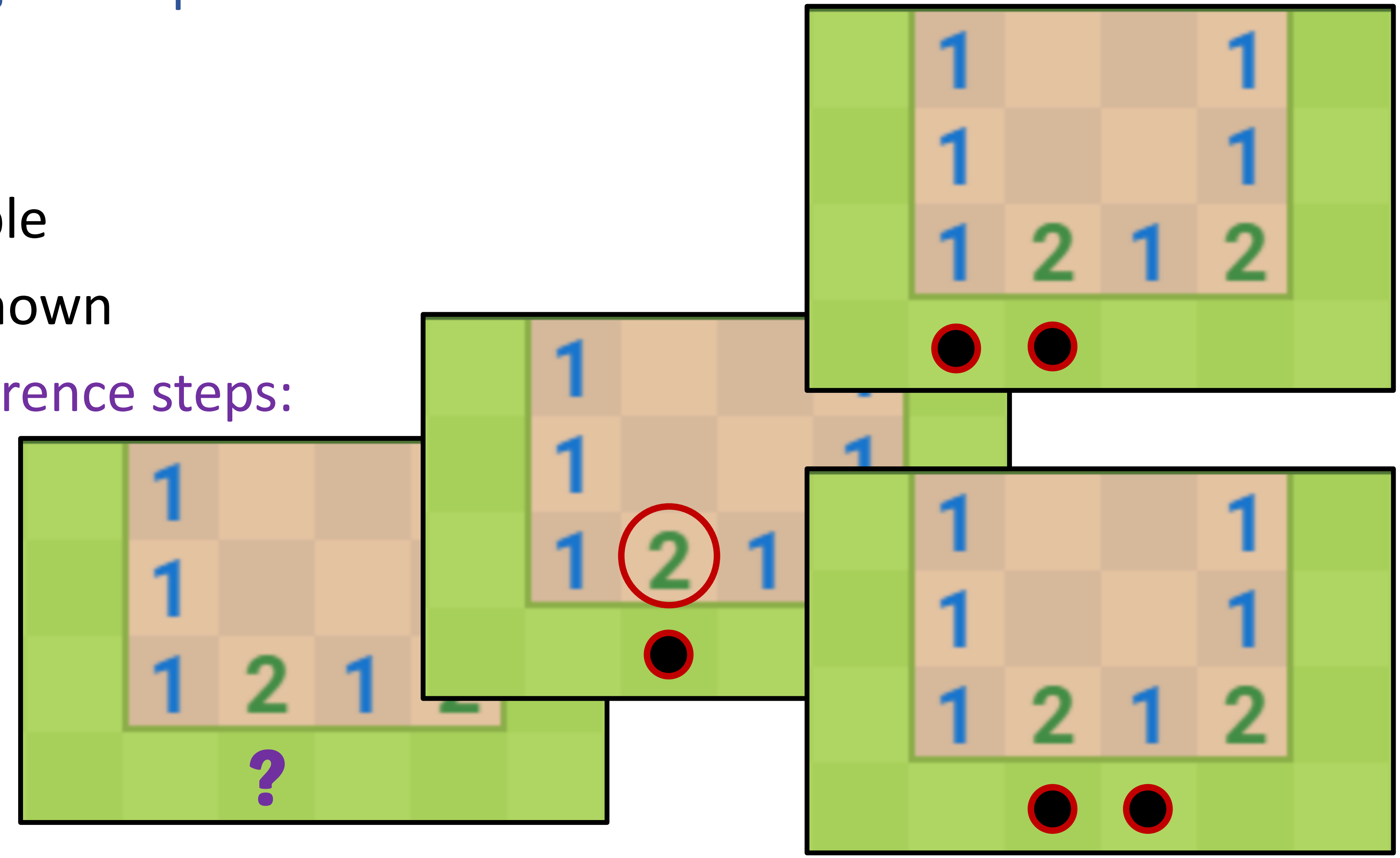


Minesweeper

It may take a few logical steps to reason about:

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

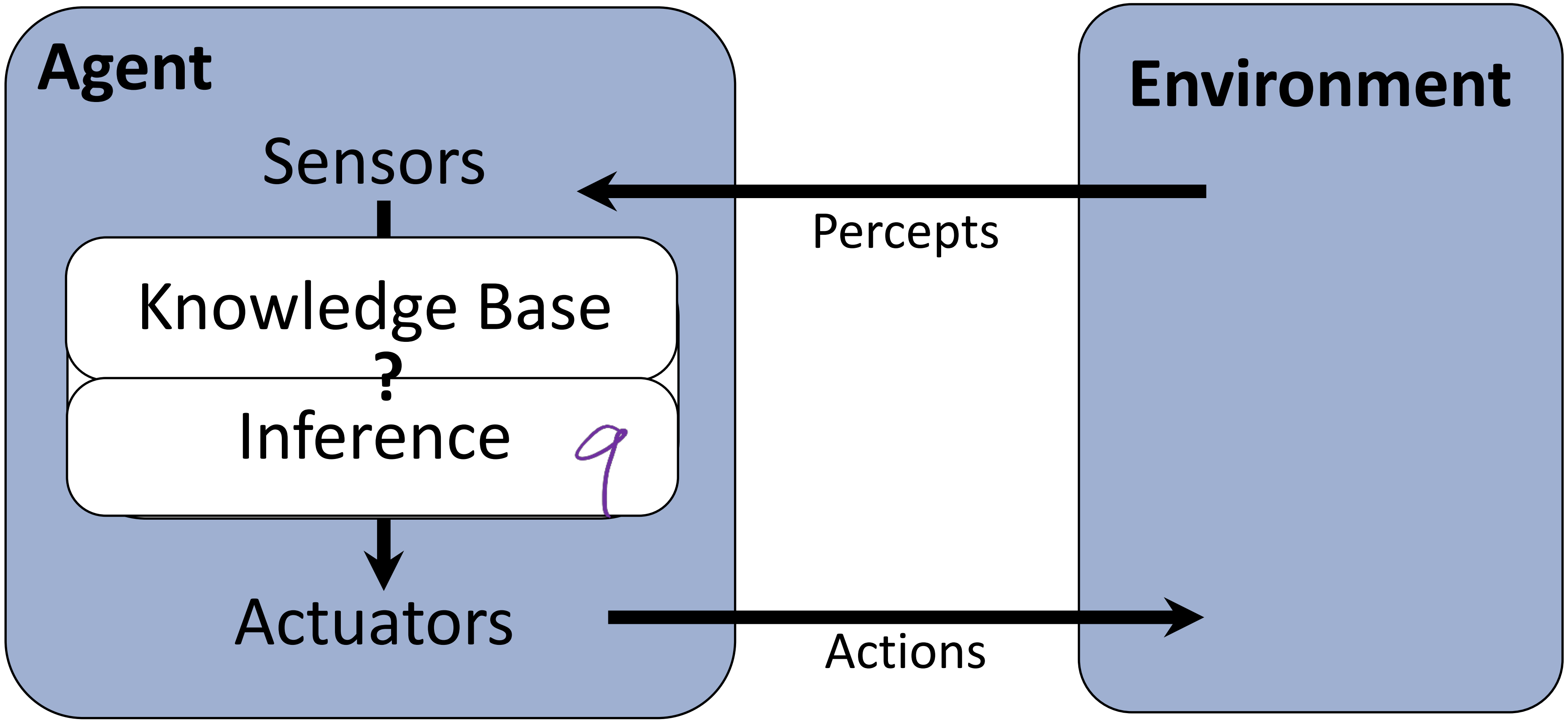
Example human inference steps:



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?

query

- 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
```

Models and Knowledge Bases

Entailment and Satisifiability

Logical Agent Vocab

Model

- An specific assignment of all *symbols* to True/False

Sentence

- Logical statement
- Composition of logic symbols and operators

KB

- Collection of sentences representing facts and rules we know about the world

Query

- Sentence we want to know if it is *provably* True, *provably* False, or *unsure*.

	M1	M2
A:	1	1
B:	1	1
C:	1	1
D:	1	1
S:	1	1
L:	1	0

Models and Knowledge Bases

Example: Sudoku

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

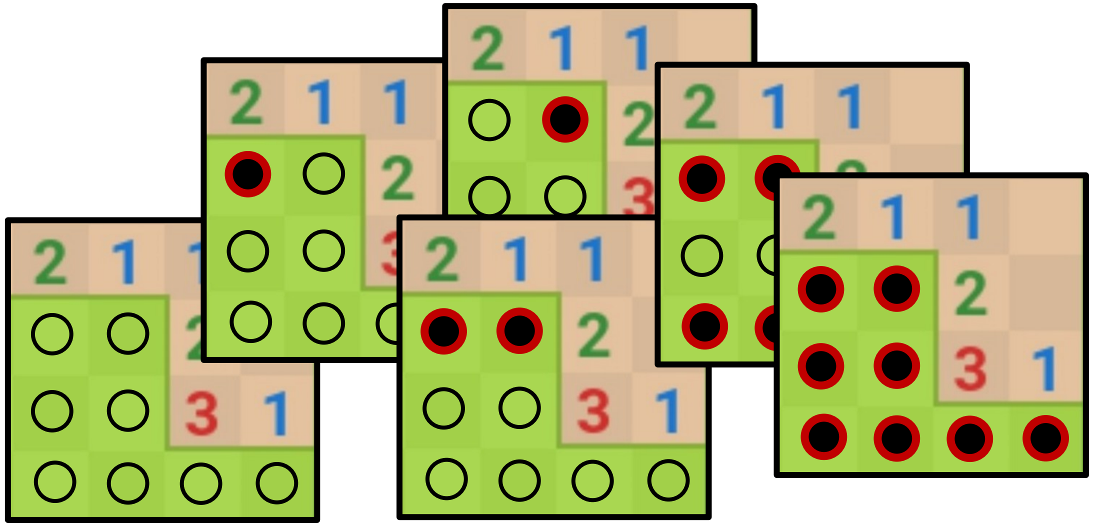
1			
	2	1	
		3	
			4

Models and Knowledge Bases

Example: Minesweeper

Model

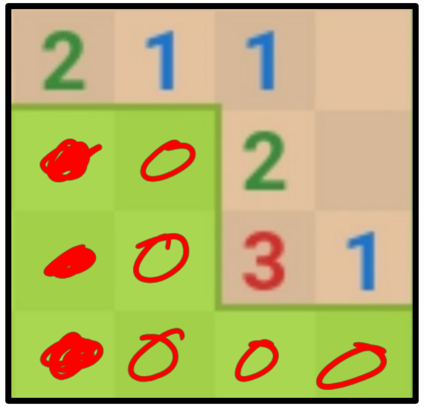
Assignment of values to all variables



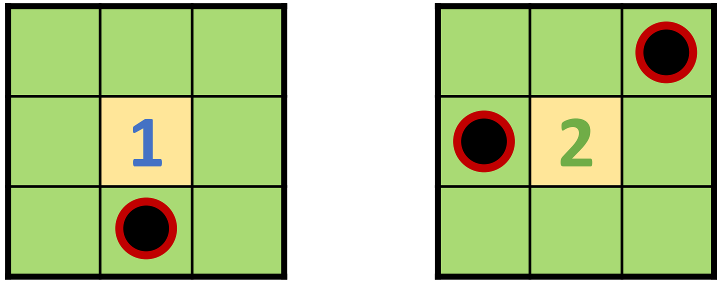
Knowledge Base

Collection of things we know to be true

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



Numbers indicate how many mines



Entailment and Satisfiability

More formally

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

How do we get a computer to do this?

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

    Entailment: definitely safe
  - No: There exists (at least) one possible configuration with a mine in that location

    Satisfiability: possibly not safe

- Is it possibly safe?
  - Yes: There exists (at least) one possible configuration with a mine in that location

    Satisfiability: possibly safe
  - No: For all possible configurations (models), all of them have a mine in that location → It's definitely dangerous

    Entailment: definitely not safe

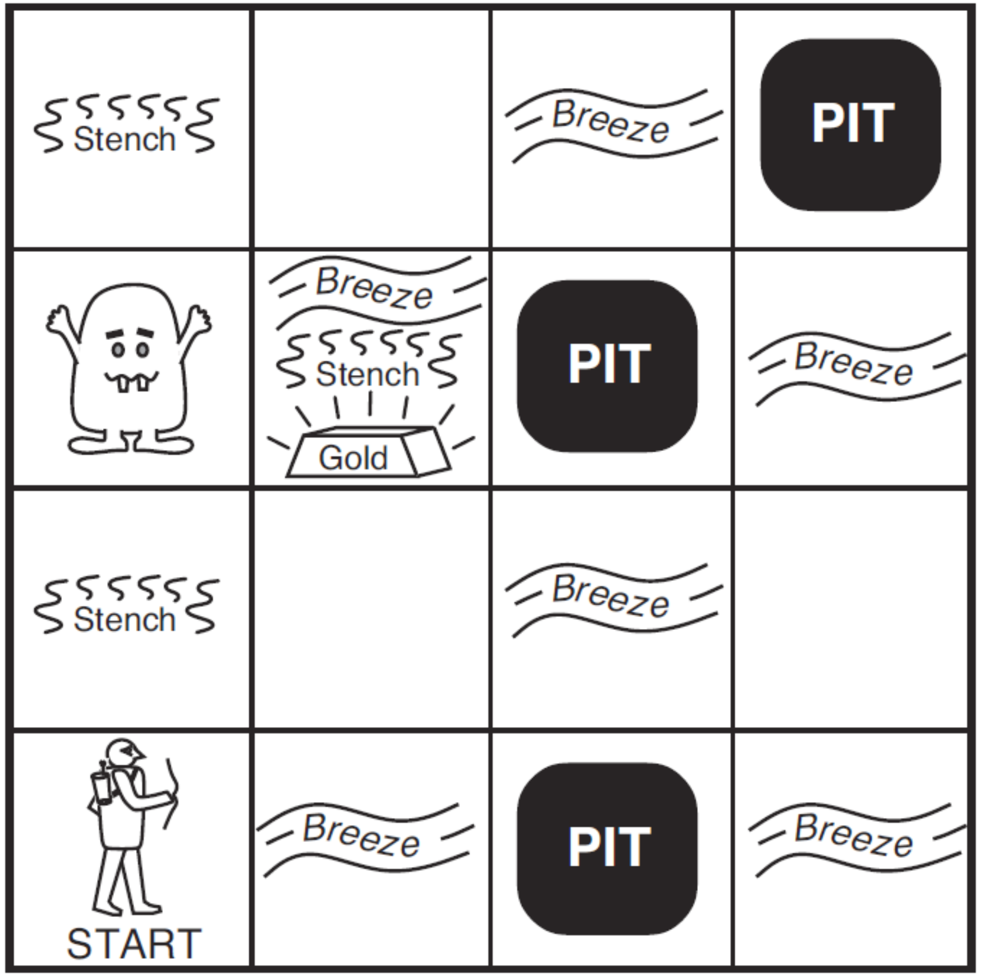
Wumpus World

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

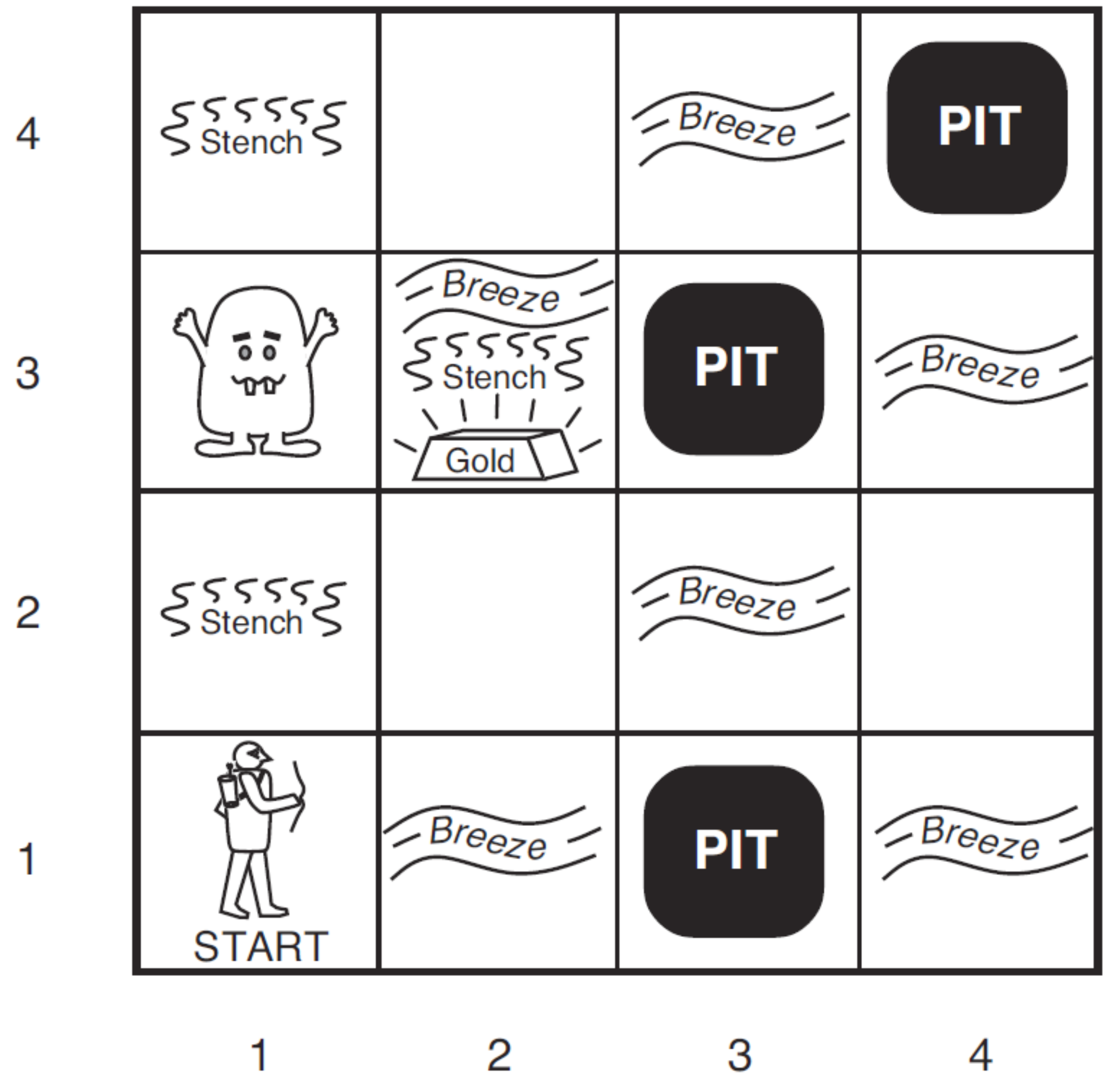
- Which unvisited spaces that are definitely safe?
- Which unvisited spaces that are definitely dangerous?
- (What about the other spaces?)



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/

Propositional Logic

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 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 $A \vee C$

if B

$\neg B \vee C \rightarrow F \vee C$

$F \vee C \rightarrow C$

$C \rightarrow A \vee C$

else

$\neg B$

$A \vee B \rightarrow A \vee F$

$A \vee F \rightarrow A$

$A \rightarrow A \vee C$

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 $A \vee C$

Assume $\neg(A \vee C)$

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

$\neg A \vee B \rightarrow F \vee B$

$\rightarrow B$

$\neg B \vee C \rightarrow \neg B \vee F$

$\rightarrow \neg B$

$B \wedge \neg B \rightarrow$ contradict

$\therefore$ assumption incorrect; thus $A \vee C$

Poll 1

KB

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

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 2

KB

If we know that $\boxed{A \vee B \text{ and } \neg B \vee C}$ are true,

KB A

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

KB

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

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
X 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

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 T F 1 0

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

literal

symbol

$\neg$symbol

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

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

Truth Tables

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

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

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

Notes on Operators

$\boldsymbol{\alpha} \vee \boldsymbol{\beta}$ is <u>inclusive or</u>, 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

$\boldsymbol{\alpha} \lor \boldsymbol{\beta}$ is <u>inclusive or</u>, not exclusive

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

- Says who?

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

- Prove it!

Truth Tables

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

α	β	$\alpha \Leftrightarrow \beta$	$\alpha \Rightarrow \beta$	$\beta \Rightarrow \alpha$	$(\alpha\Rightarrow\beta) \land (\beta\Rightarrow\alpha)$
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:

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