Artificial Intelligence

CS4365 --- Fall 2022

Knowledge Representation and Reasoning

Instructor: Yunhui Guo

Summary of Propositional Logic

- Syntax
    - literals
    - Connectives
    - Clauses
    - CNF
- Semantics
    - Model
- Inference
    - Semantically: Model Checking
    - Syntactically: Inference rules and axioms
        - Modus Ponens
        - Resolution

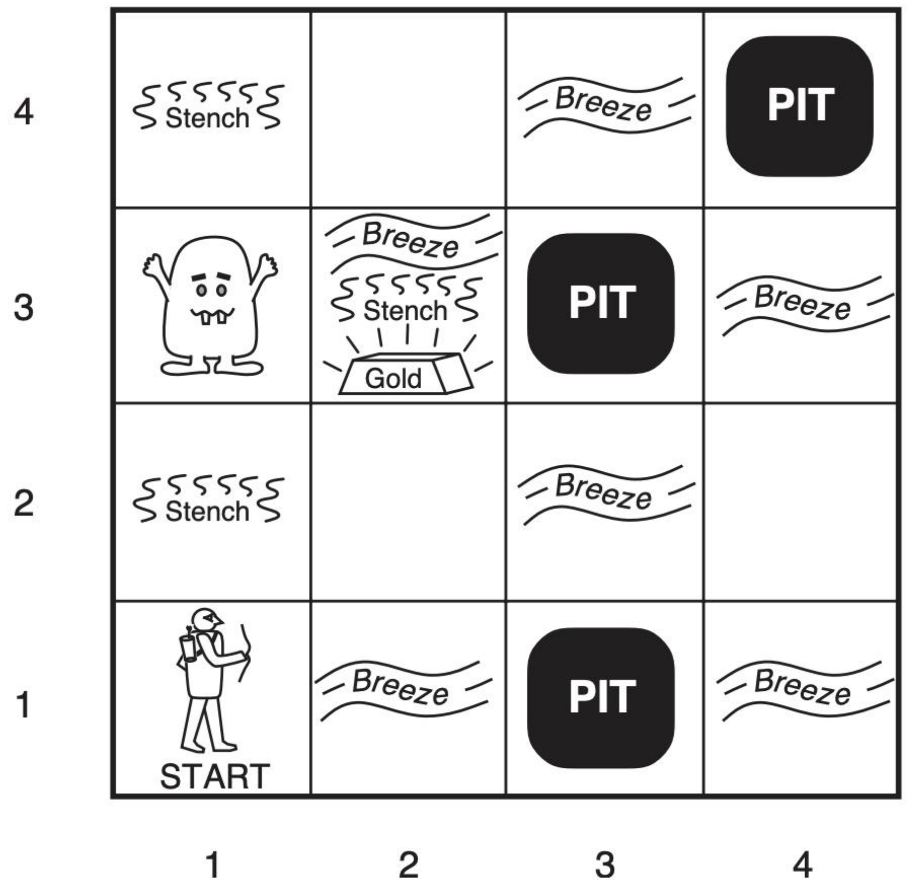
Agent Based on Propositional Logic

- Construct wumpus world agents that use propositional logic.

- Enable the agent to deduce, to the extent possible, the state of the world given its percept history

The Wumpus World

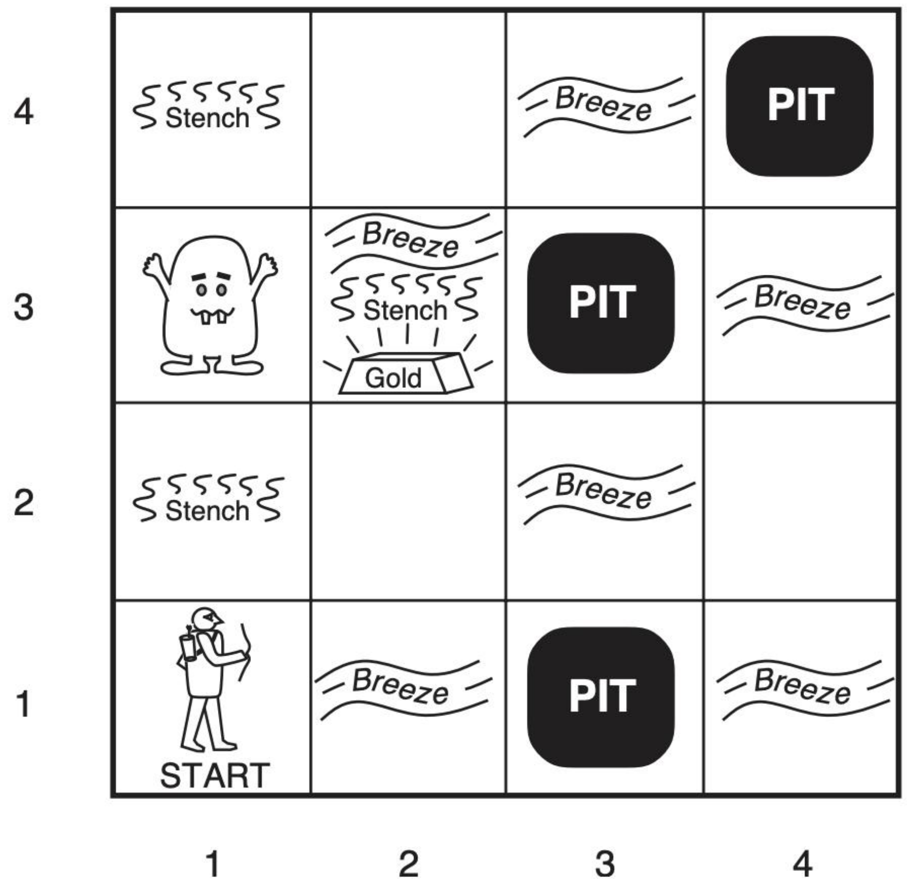
- Performance meaure:
  - gold +1000, death -1000
  - -1 per step, -10 for using the arrow



The Wumpus World

- 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: Stench, Breeze, Glitter, Bump, Scream (shot Wumpus)



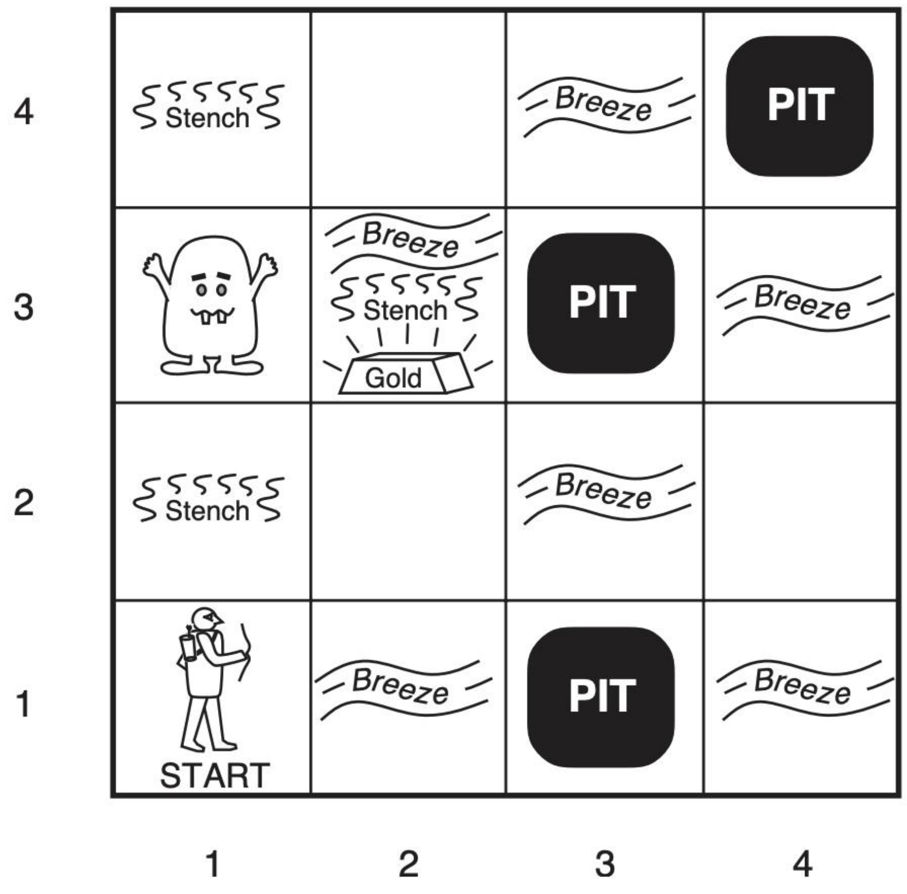
Wumpus world characterization

- 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

The Wumpus World

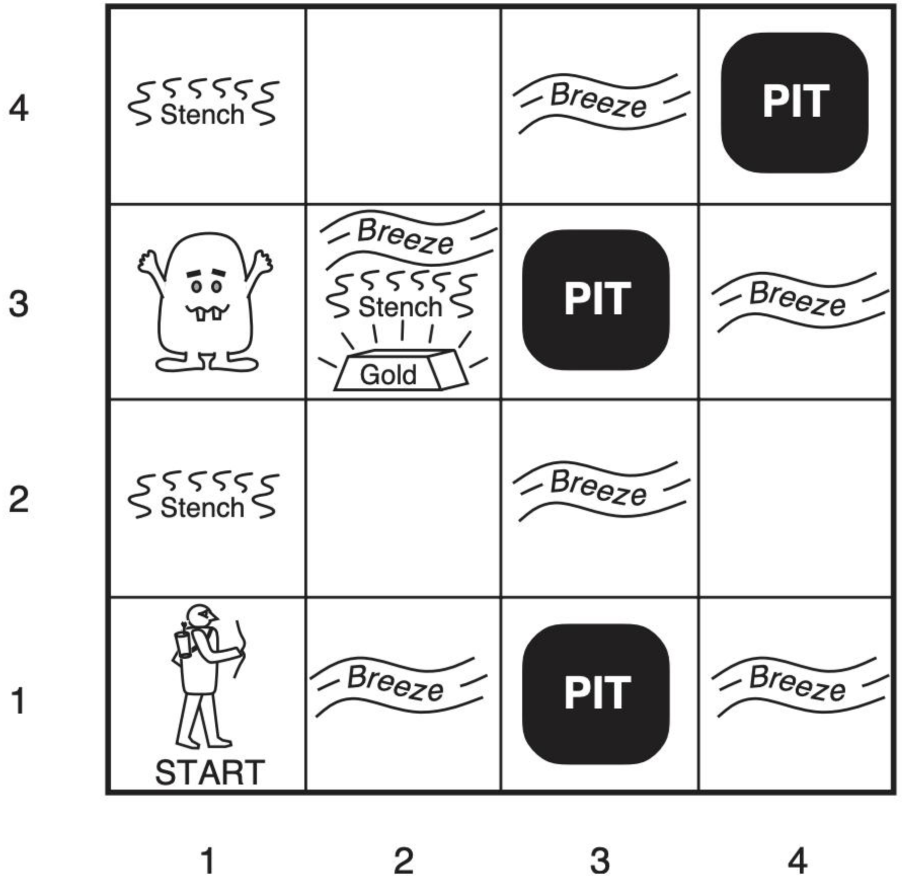
- A decision-maker needs to represent knowledge of the world and reason with it in order to safely explore this world.

- Principle Difficulty: agent is initially ignorant of the configuration of the environment – going to have to reason to figure out where the gold is without getting killed!



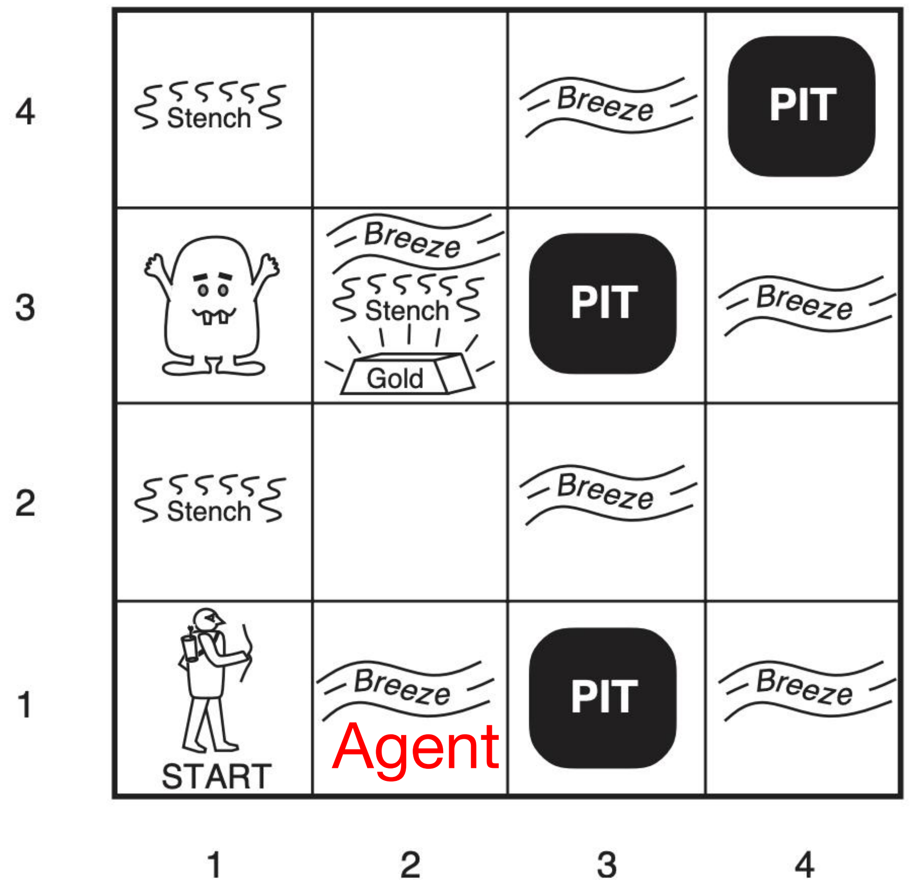
The Wumpus World

- Define the propositional symbols:
  - $P_{x,y}$: true if there is a pit in [x,y]
  - $W_{x,y}$: true if there is a wumpus in [x, y], dead or alive
  - $B_{x,y}$: true if the agent perceives a breeze in [x ,y]
  - $S_{x,y}$: true if the agent perceives a stench in [x,y]



Knowledge Base

- Build the knowledge base:
  - There is no pit in [1,1]:
    - R1: $\neg P_{1,1}$
  - For each square, it knows that the square is breezy if and only if a neighboring square has a pit;
    - R2: $B_{1,1} \Leftrightarrow (P_{1,2} \lor P_{2,1})$
    - R3: $B_{2,1} \Leftrightarrow (P_{1,1} \lor P_{2,2} \lor P_{3,1})$
  - Percepts:
    - R4: $\neg B_{1,1}$
    - R5: $B_{2,1}$

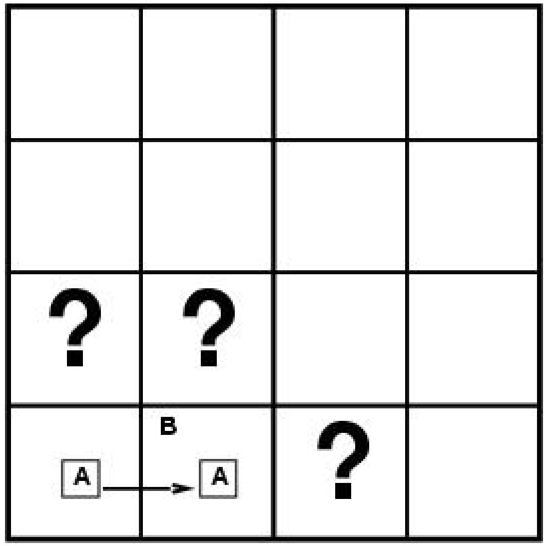


Entailment in the Wumpus World

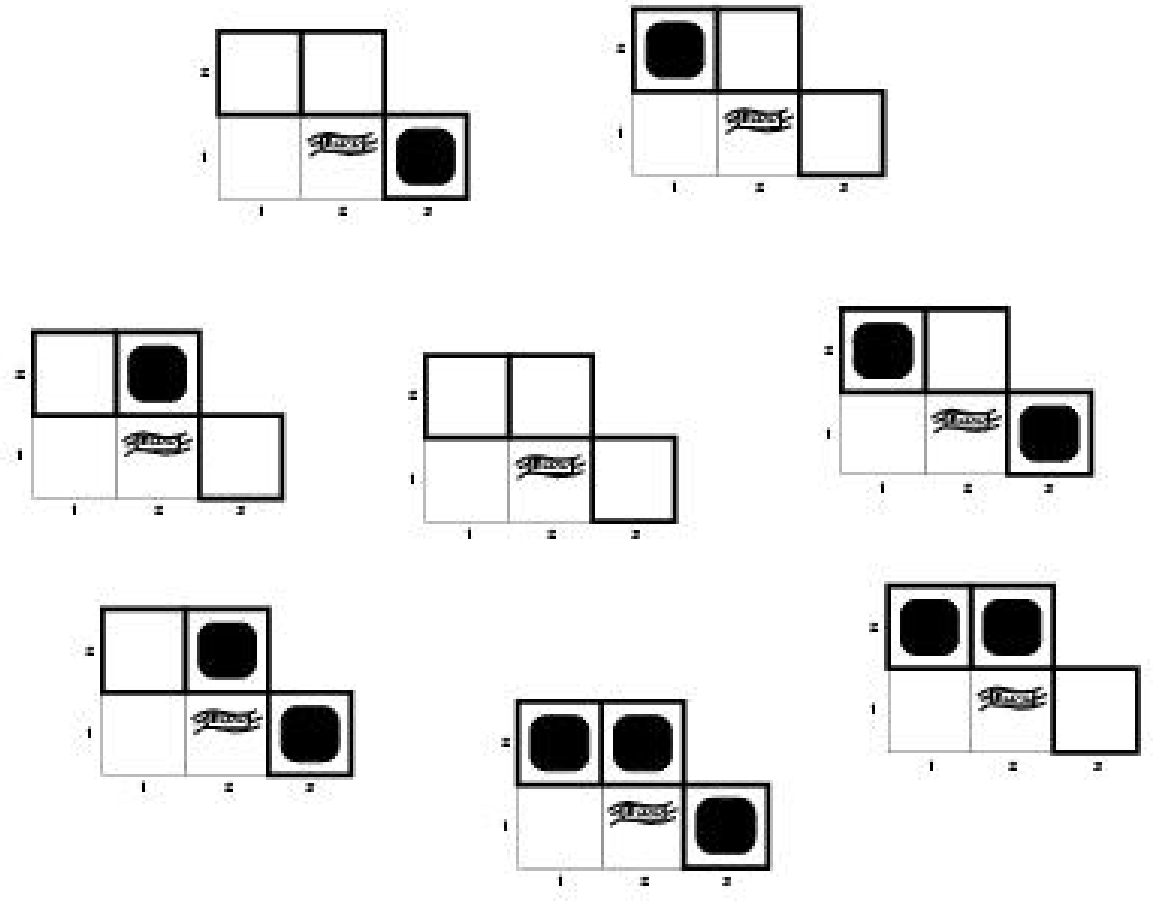
- Situation after detecting nothing in [1,1], moving right breeze in [2,1]

- Consider possible models for KB assuming only pits

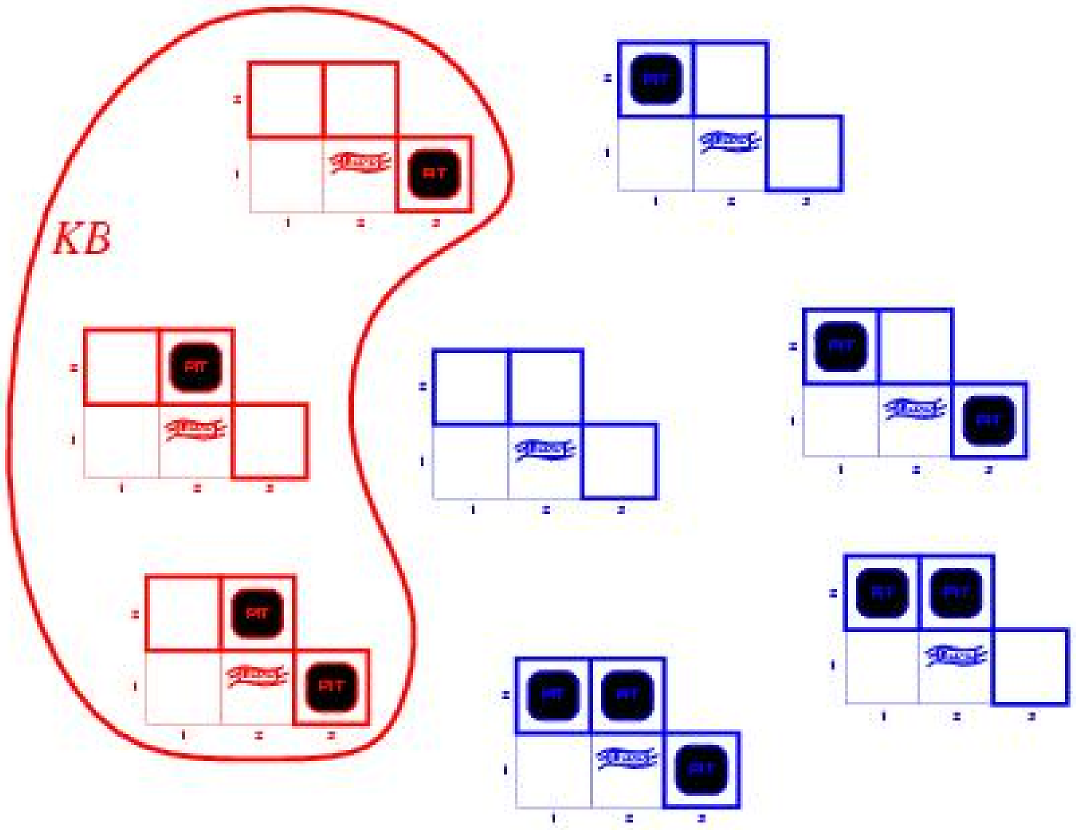
- 3 Boolean choices $\Rightarrow$ 8 possible models



Wumpus possible models

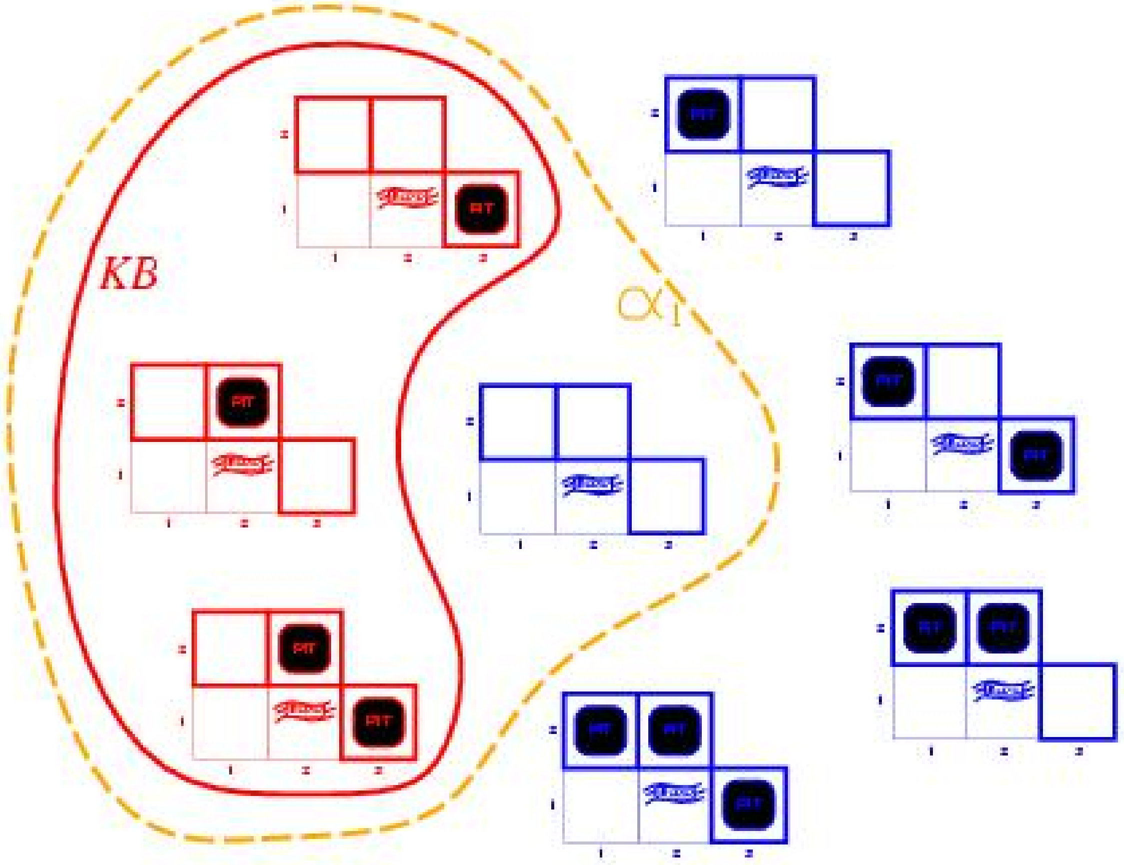


Wumpus models



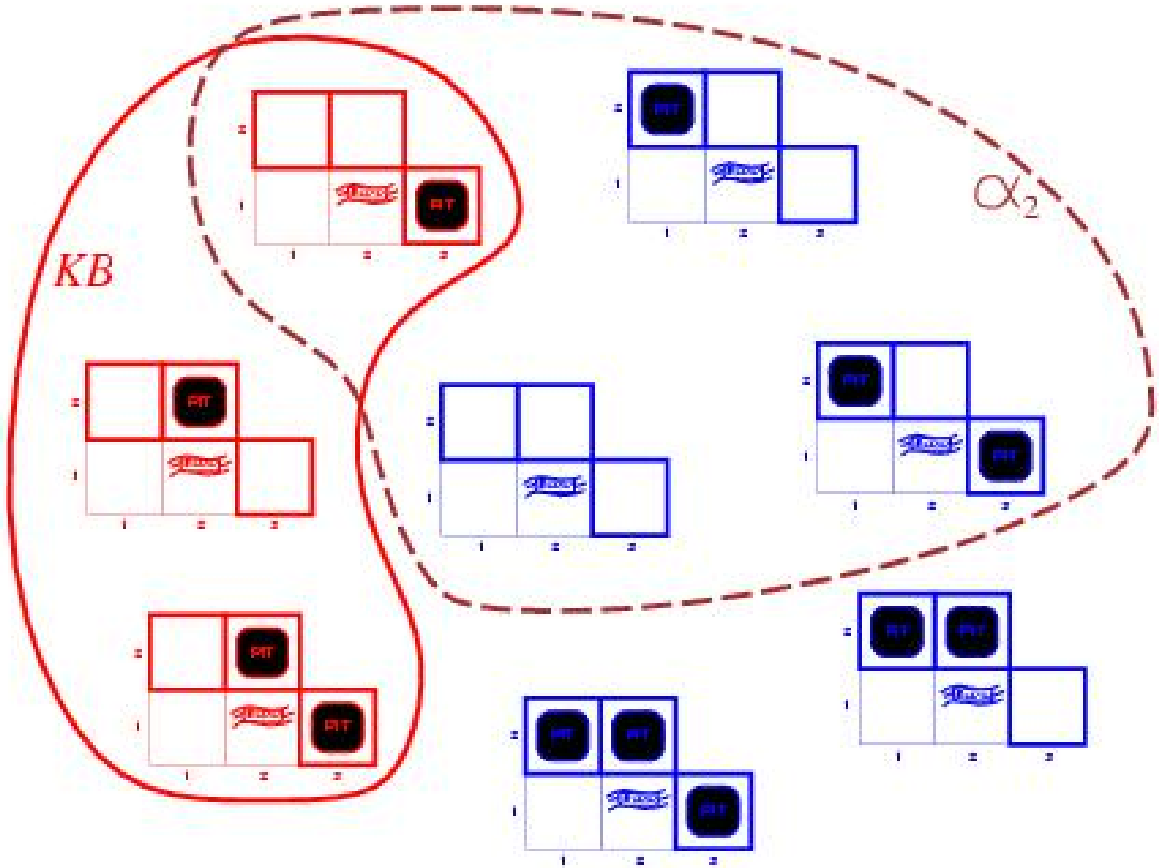
KB = wumpus-world rules + observations

Wumpus models



- KB = wumpus-world rules + observations
- α_1 = “there is no pit in [1,2]", KB $\models \alpha_1$, proved by model checking

Wumpus models



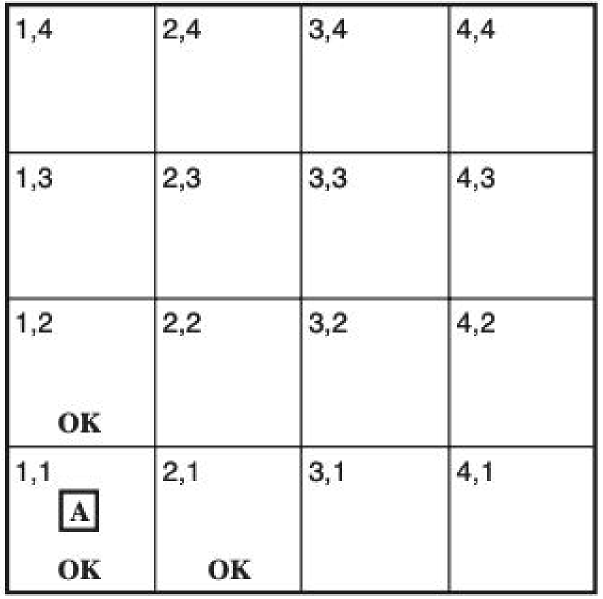
- KB = wumpus-world rules + observations
- α_2 = “there is no pit in [2,2]"

Inference

- Proof $\neg P1,2$ and $\neg P2,1$:
- R2: $B1,1 \Leftrightarrow (P1,2 \vee P2,1)$
- R6: $(B1,1 \Rightarrow (P1,2 \vee P2,1)) \wedge ((P1,2 \vee P2,1) \Rightarrow B1,1)$

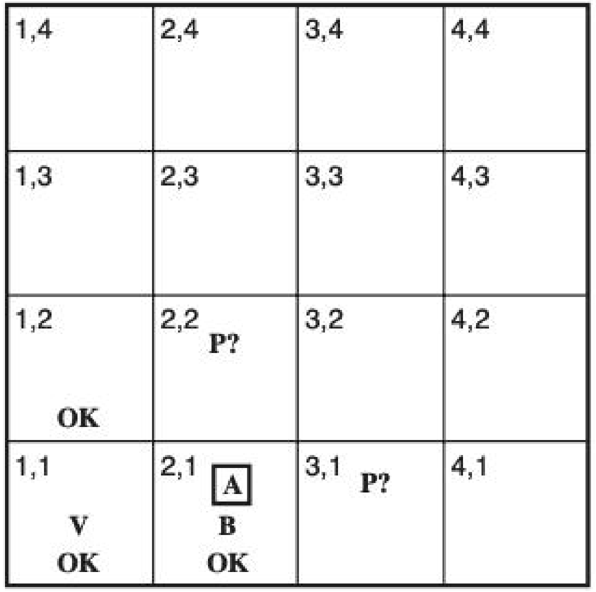
- R7: $(P1,2 \vee P2,1) \Rightarrow B1,1$ — And elimination

- R8: $\neg B1,1 \Rightarrow \neg(P1,2 \vee P2,1)$ — Contrapositive
- R9: $\neg(P1,2 \vee P2,1)$ — Modus Ponens
- R10: $\neg P1,2 \wedge \neg P2,1$ — De Morgan



Action and Percept

- Move to [2, 1]
- B2,1 is True
- There is a pit in [2,2] or [3,1]



Inference

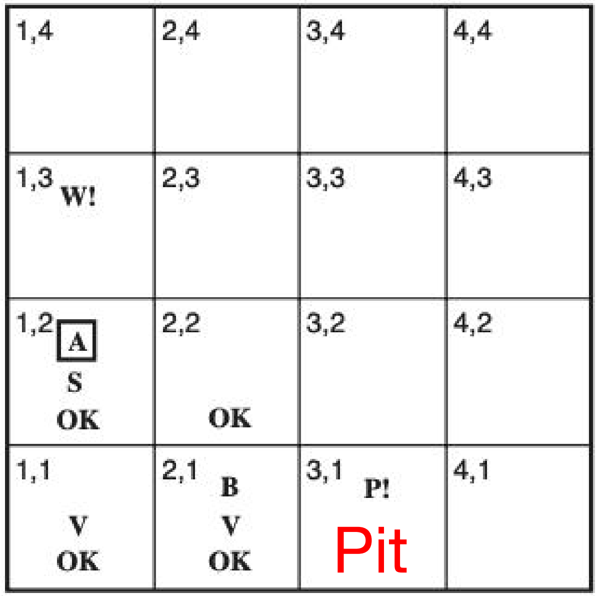
- The agent return from [2,1] to [1,1] and then moves to [1, 2]

  - R11: ¬B1,2
  - R12: B1,2 ⇔ (P1,1 ∨ P2,2 ∨ P1,3)
  - R13: ¬P2,2
  - R14: ¬P1,3

  - R3: B2,1 ⇔ (P1,1 ∨ P2,2 ∨ P3,1)
  - R5: B2,1 → R15: P1,1 ∨ P2,2 ∨ P3,1

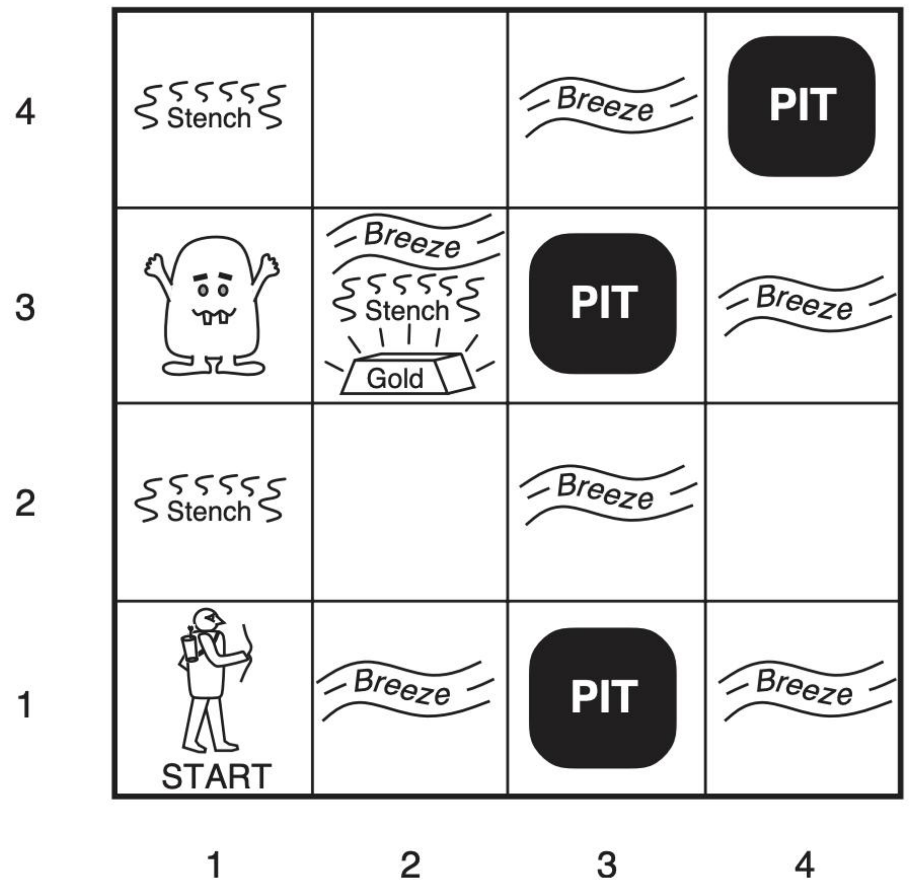
  - R16: P1,1 ∨ P3,1 (resolve R13 with R15)

  - R17: P3,1 （resolve R1 with R16)



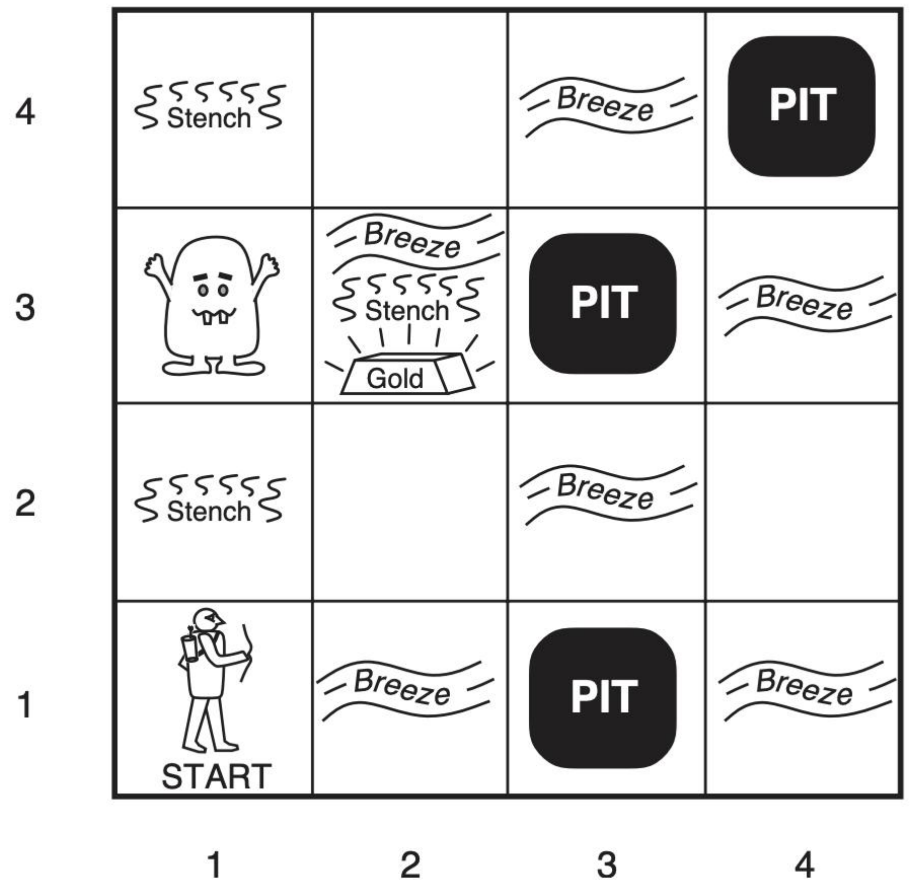
The current state of the world

- Collecting axioms:
  - ¬P1,1
  - ¬W1,1
  - For each square, it knows that the square is breezy if and only if a neighboring square has a pit:
    - B1,1 ⇔ (P1,2 ∨ P2,1)
    - ...
  - Square is smelly if and only if a neighboring square has a wumpus.
    - S1,1 ⇔ (W1,2 ∨ W2,1)
    - ...



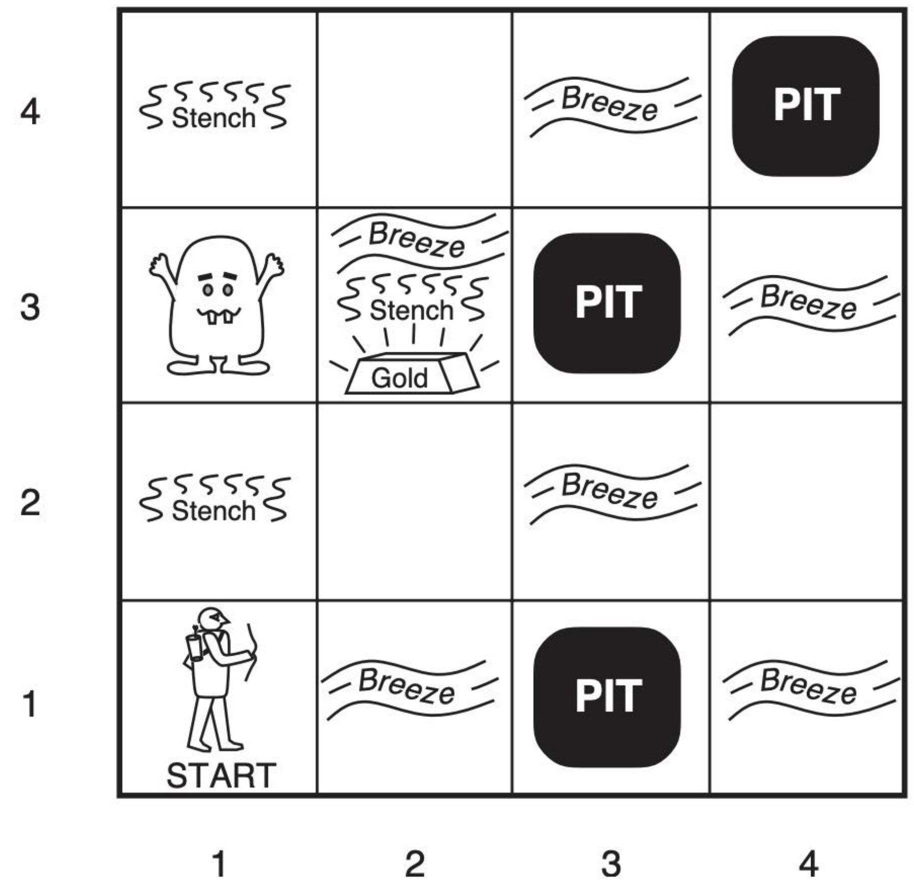
The current state of the world

- There is exactly one wumpus
  - $W1,1 \lor W1,2 \lor ..., W4,3 \lor W4,4$
- There is at most one wumpus
  - $\neg W1,1 \lor \neg W1,2$
  - $\neg W1,1 \lor \neg W1,3$
  - ...
  - $\neg W4,3 \lor \neg W4,4$



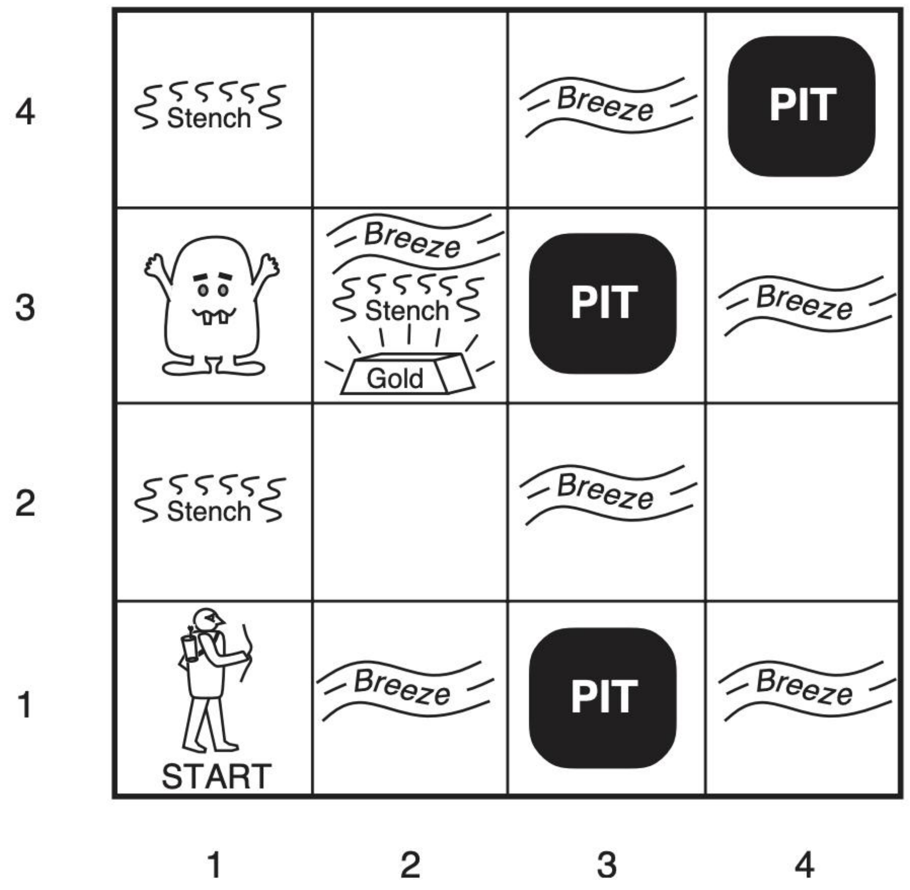
The current state of the world

- Fluent: associating propositions with time steps extends to any aspect of the world that changes over time
  - $L^0_{1,1}$: the agent is at square [1,1] at time 0
- Consider time stamp for percepts
  - $\neg Stench^3$
  - $Stench^4$



The current state of the world

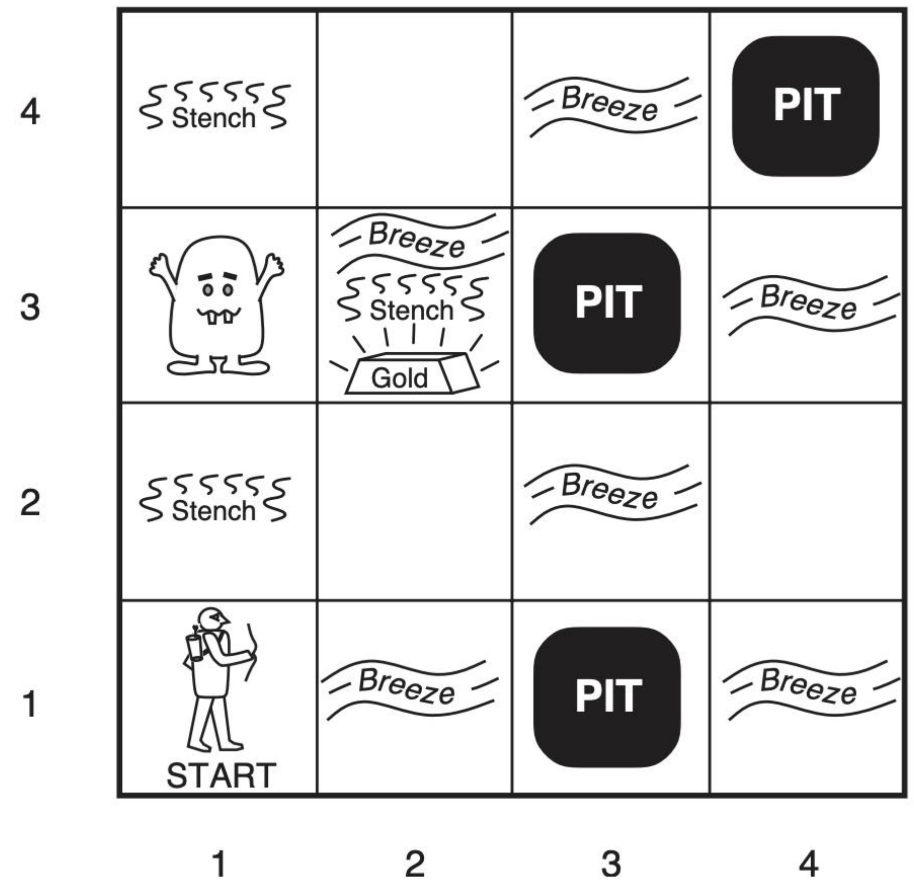
- For any time step t and any square [x, y], we have
  - $L^t_{x,y} \Rightarrow (Breeze^t \Leftrightarrow B_{x,y})$
  - $L^t_{x,y} \Rightarrow (Stench^t \Leftrightarrow S_{x,y})$



The current state of the world

- Proposition symbols for the occurrences of actions
  - $Forward_0$

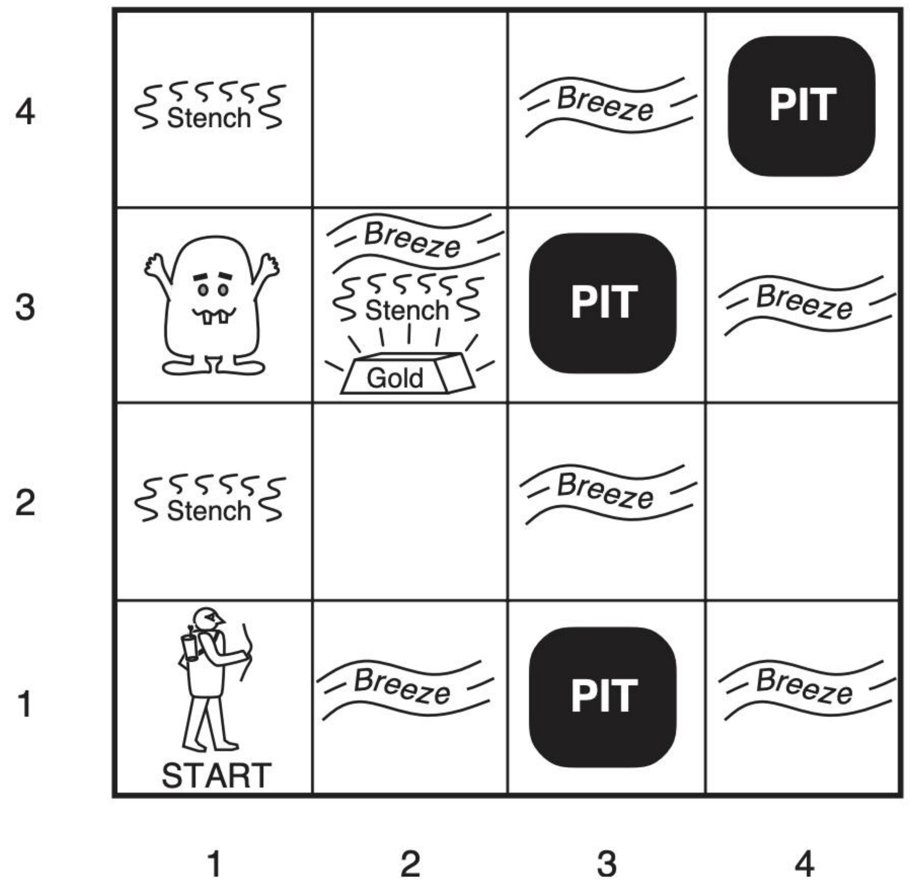
- The percept for a given time step happens first, followed by the action for that time step, followed by a transition to the next time step.



The current state of the world

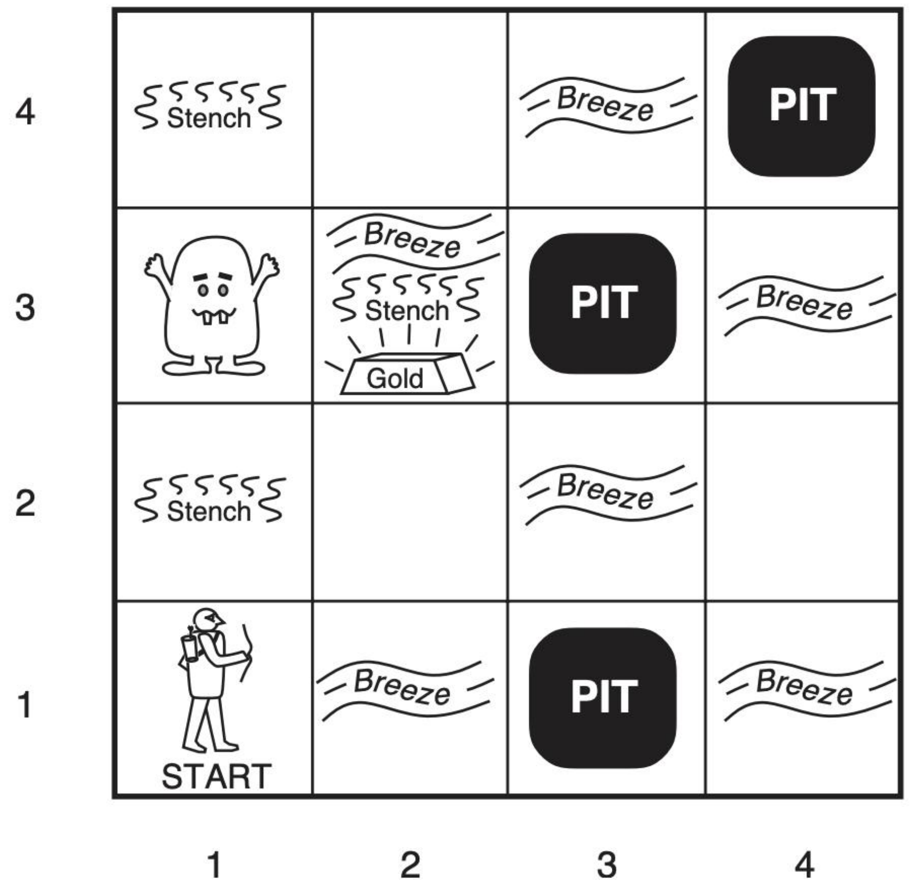
- Effect axioms: describe how the world changes,
  - $L^0_{1,1} \land FaceEast_0 \land Forward_0 \Rightarrow (L^1_{2,1} \land \neg L^1_{1,1})$

- Need to specify for each action, time step, location and orientation



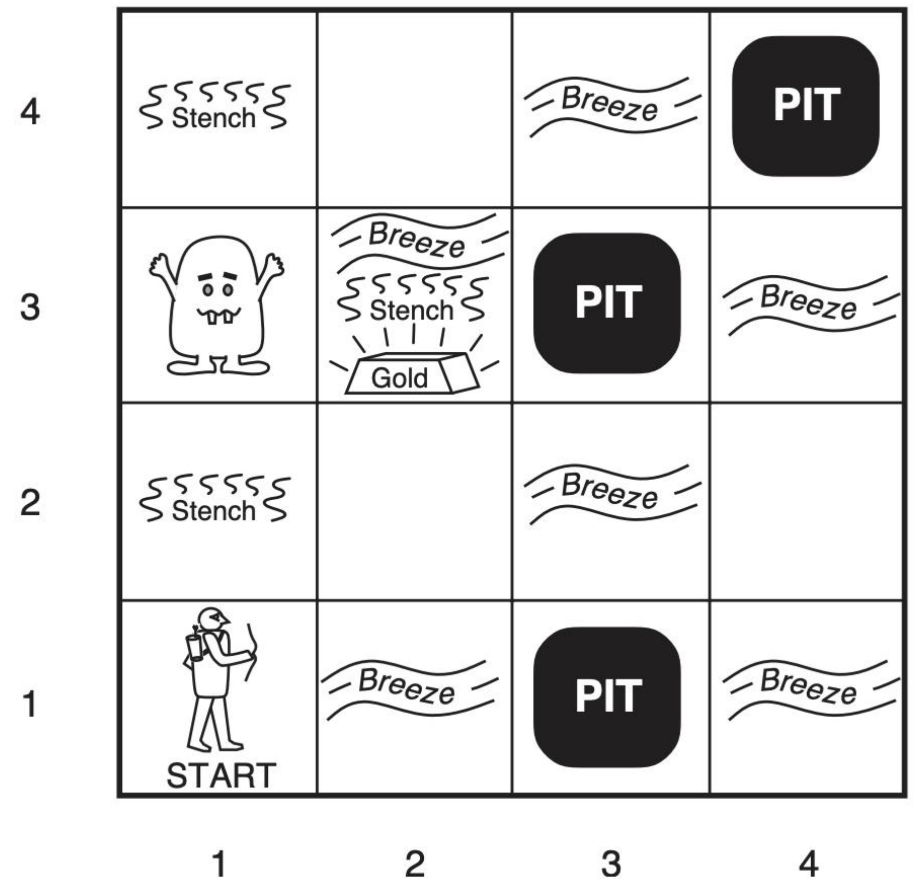
The current state of the world

- Frame problem:
  - ASK(KB, $HaveArrow^1$)
  - The information has been lost because the effect axiom fails to state what remains unchanged as the result of an action
- Add frame axioms:
  - $Forward^t \Rightarrow (HaveArrow^t \Leftrightarrow HaveArrow^{t+1})$
  - $Froward^t \Rightarrow (WumpusAlive^t \Leftrightarrow WumpusAlive^{t+1})$
- For m actions and n fluents, needs O(mn)



The current state of the world

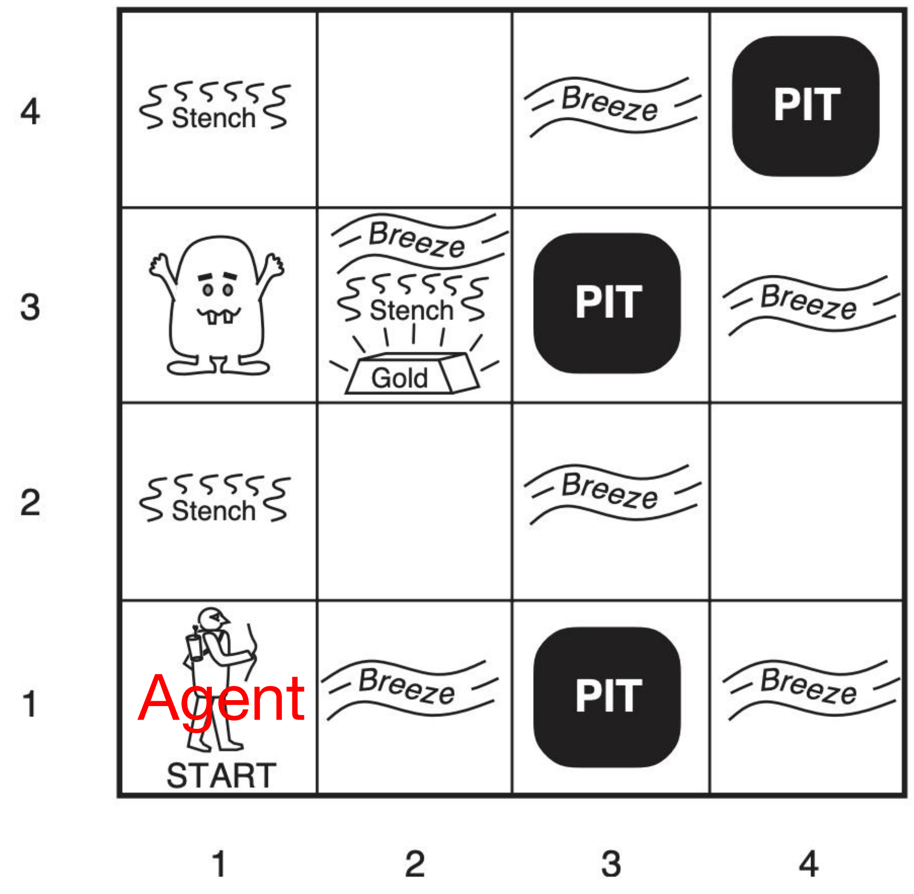
- Writing axioms about fluents
- Successor-state axiom:
  - $F^{t+1} \Leftrightarrow \text{ActionCausesF}^{t} \lor (F^{t} \land \text{ActionCausesNotF}^{t})$
  - ...
  - $\text{HaveArrow}^{t+1} \Leftrightarrow (\text{HaveArrow}^{t} \Leftrightarrow \neg\text{Shoot}^{t})$
  - $L^{t+1}_{1,1} \Leftrightarrow (L^{t}_{1,1} \land (\neg\text{Forward}^{t} \lor \text{Bump}^{t+1})) \lor (L^{t}_{1,2} \land (\text{South}^{t} \land \text{Forward}^{t})) \lor (L^{t}_{2,1} \land (\text{West}^{t} \land \text{Forward}^{t}))$



The current state of the world

- $\neg Stench^0 \wedge \neg Breeze^0 \wedge \neg Glitter^0 \wedge \neg Bump^0 \wedge \neg Screem^0$

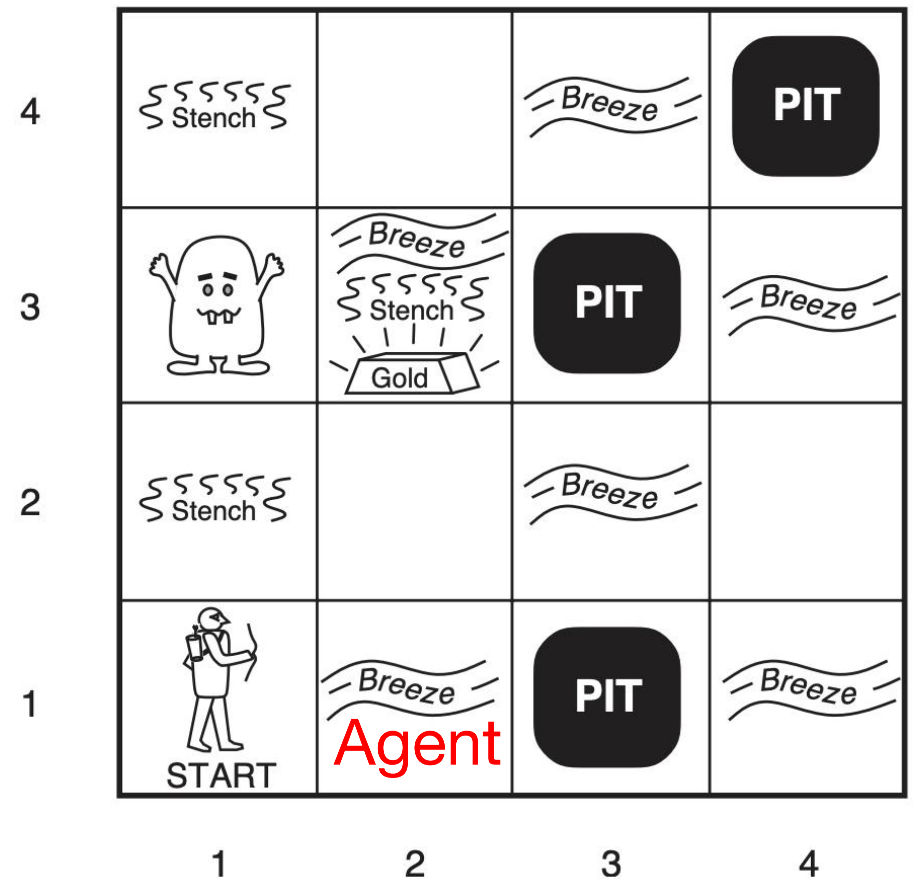
- $Forward^0$



The current state of the world

- $\neg Stench^1 \wedge Breeze^1 \wedge \neg Glitter^1 \wedge \neg Bump^1 \wedge \neg Screem^1$

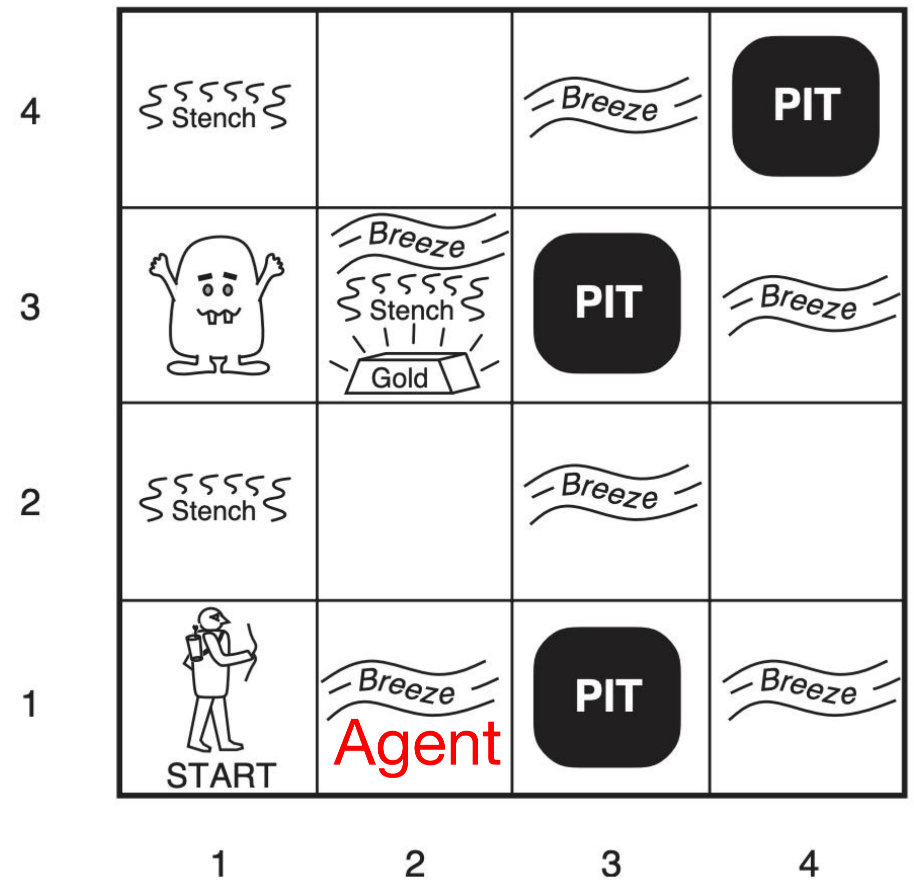
- $TurnRight^1$



The current state of the world

- ¬Stench^2 ∧ Breeze^2 ∧ ¬Glitter^2 ∧ ¬Bump^2 ∧ ¬Screem^2

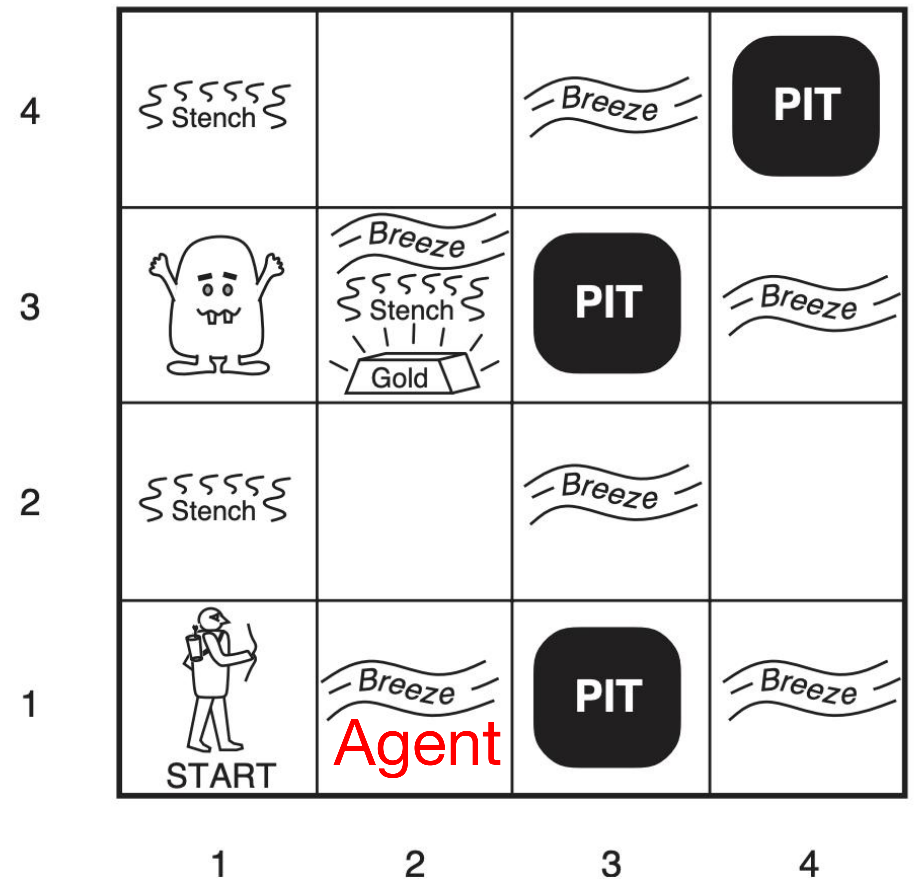
- TurnRight^2



The current state of the world

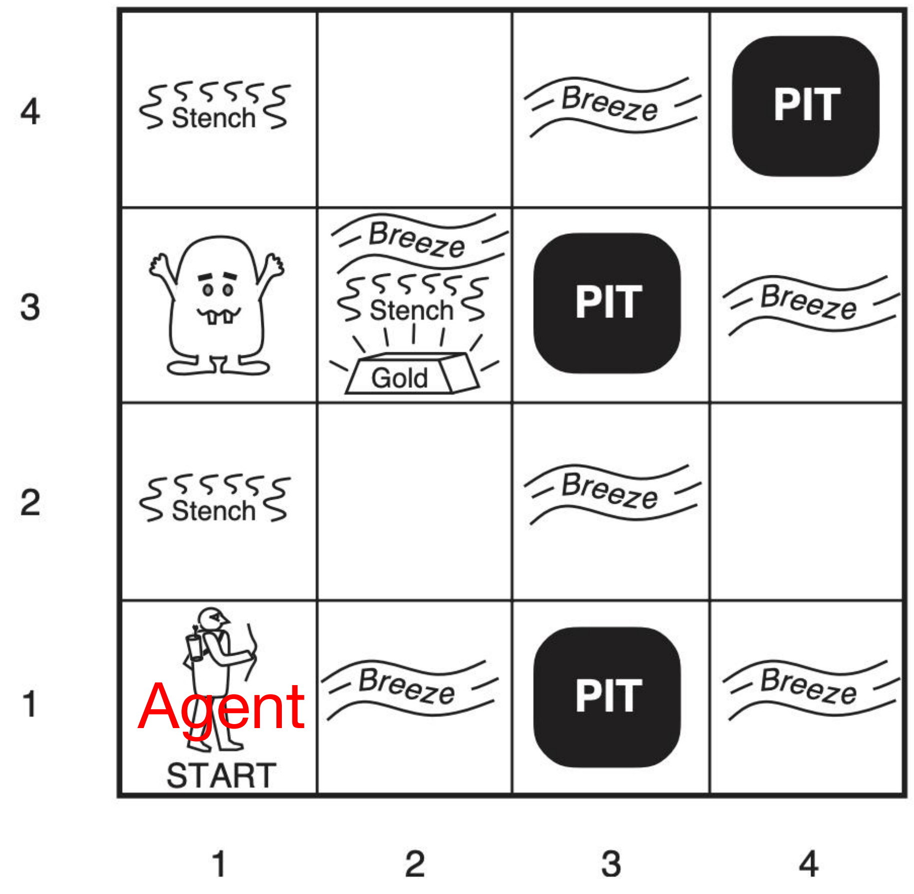
- $\neg Stench^3 \wedge Breeze^3 \wedge \neg Glitter^3 \wedge \neg Bump^3 \wedge \neg Screem^3$

- $Forward^3$



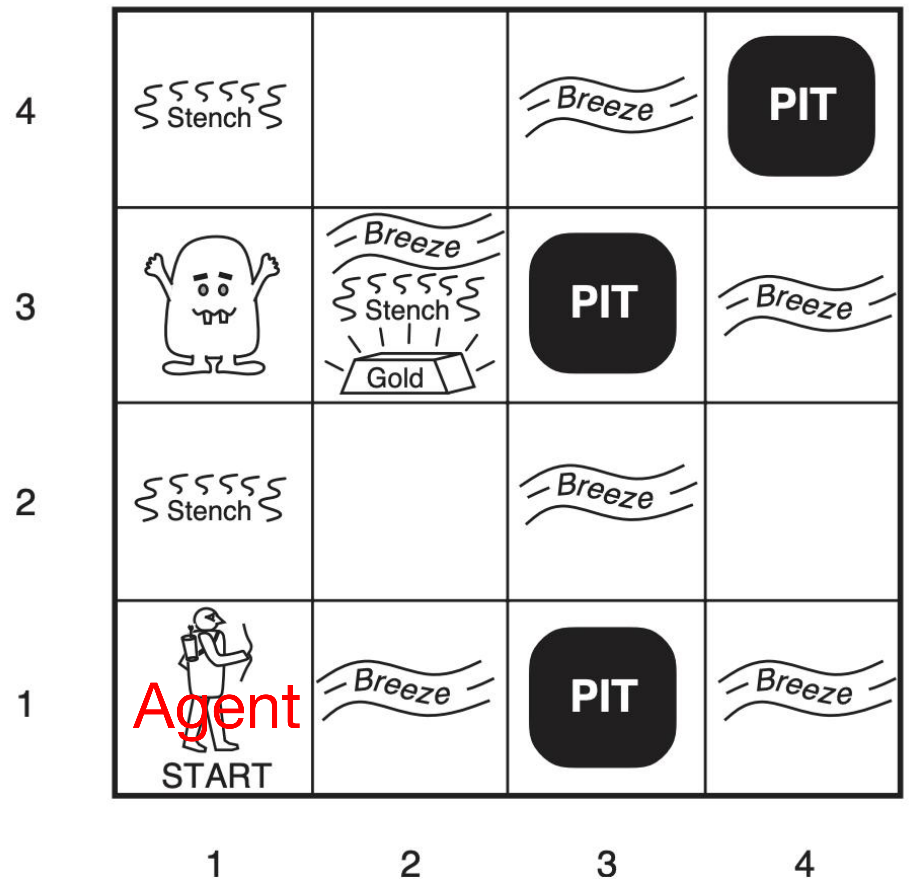
The current state of the world

- $\neg Stench^4 \land \neg Breeze^4 \land \neg Glitter^4 \land \neg Bump^4 \land \neg Screem^4$
- $TurnRight^4$



The current state of the world

- $\neg Stench^5 \wedge \neg Breeze^5 \wedge \neg Glitter^5 \wedge \neg Bump^5 \wedge \neg Screem^5$
- $Forward^5$



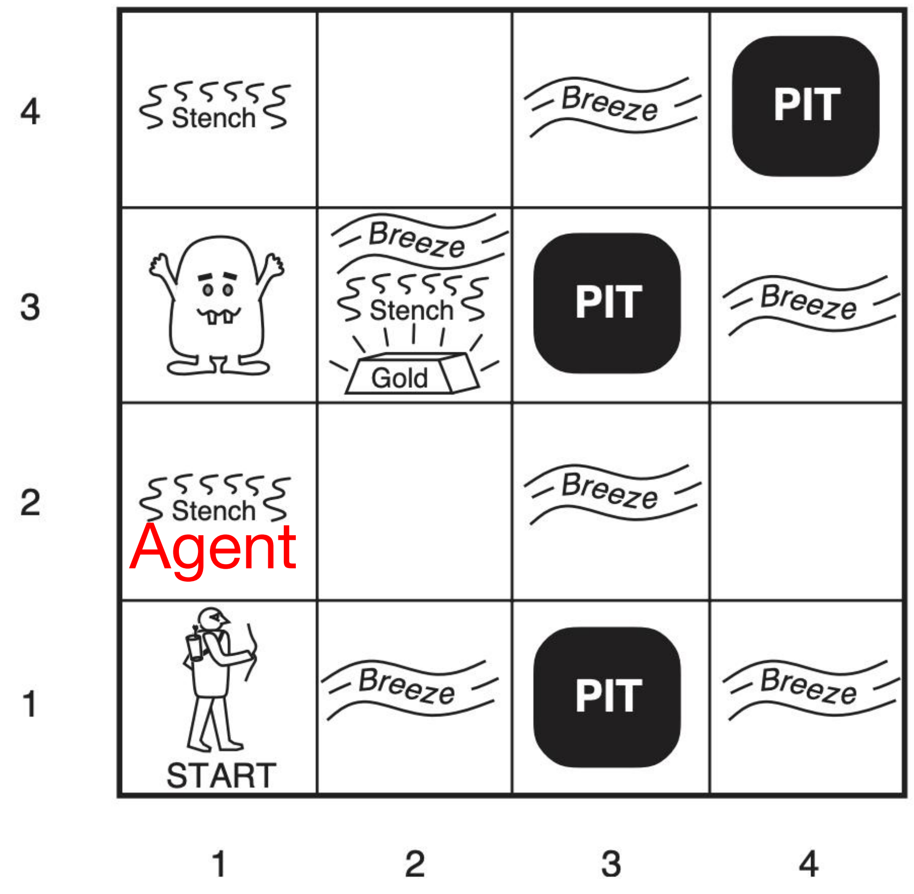
The current state of the world

- $Stench^6 \land \neg Breeze^6 \land \neg Glitter^6 \land \neg Bump^6 \land \neg Screem^6$

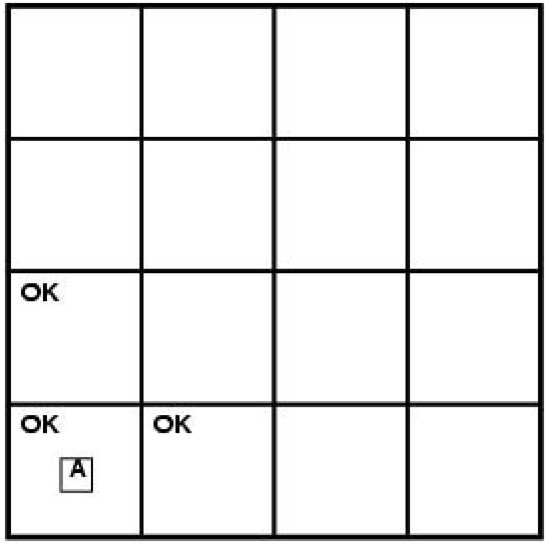
- $Ask(KB, L^6_{1,2}) = True$

- $Ask(KB, W_{1,3}) = True$

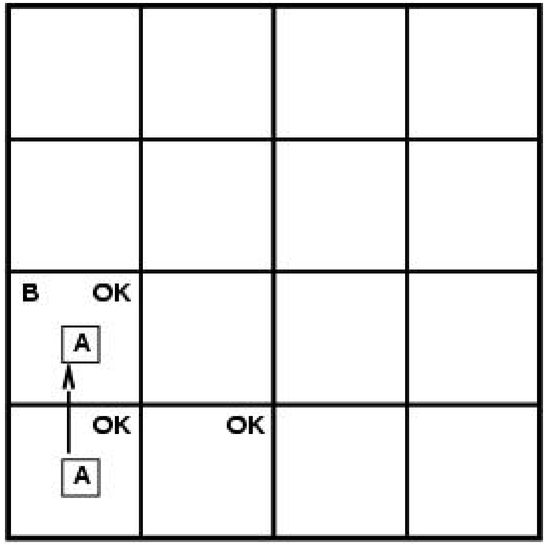
- $Ask(KB, P_{3,1}) = True$



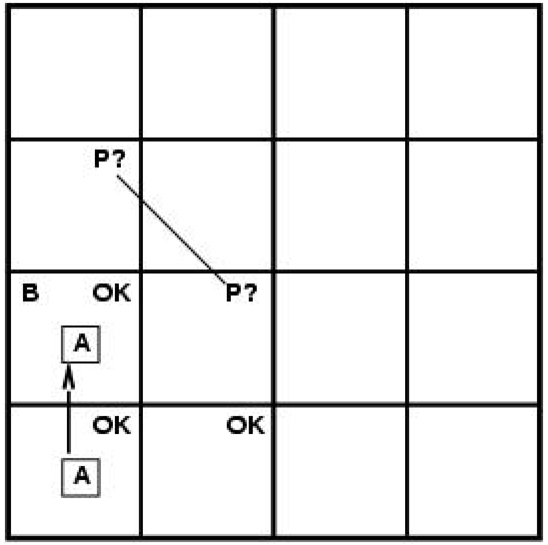
The current state of the world



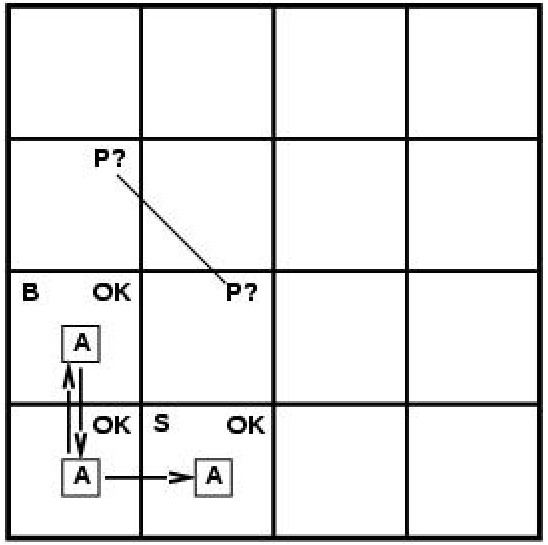
The current state of the world



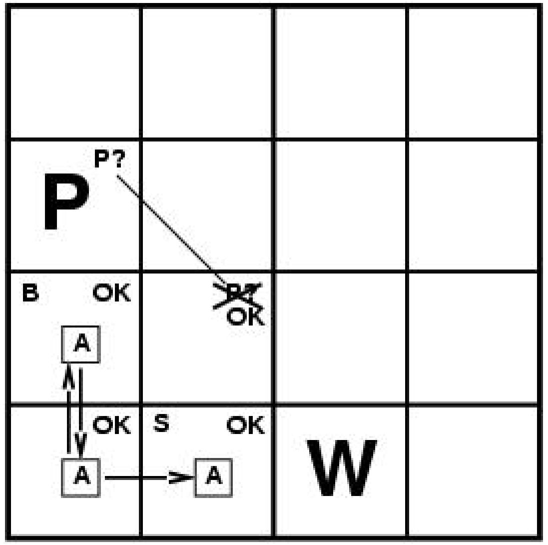
The current state of the world



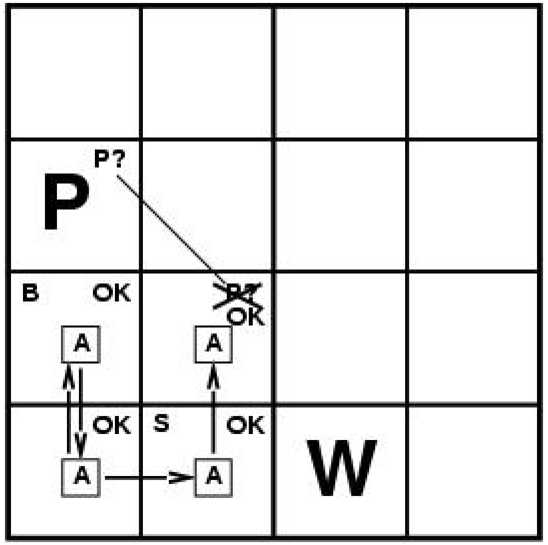
The current state of the world



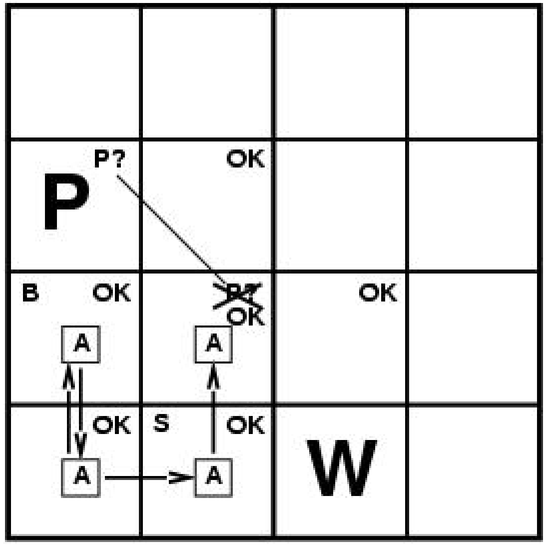
The current state of the world



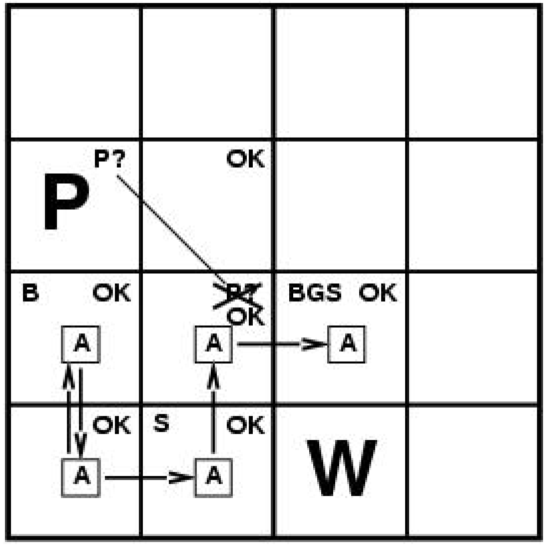
The current state of the world



The current state of the world

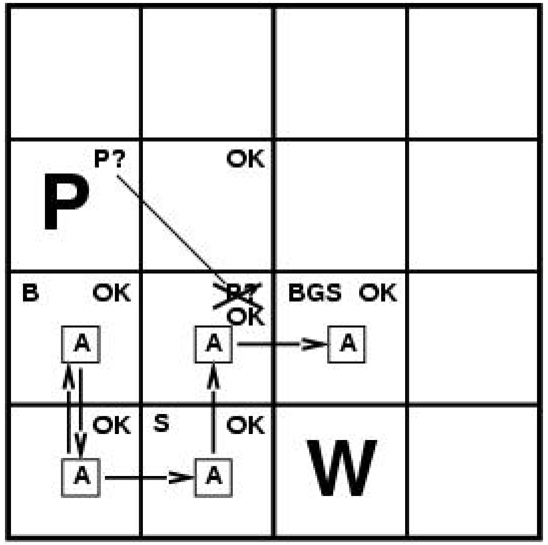


The current state of the world



The current state of the world

- In each case where the agent draws a conclusion from the available Information, that conclusion is guaranteed to be correct if the available Information is correct.
- This is a fundamental property of logical reasoning



A Hybrid Agent

- The agent program maintains and updates a knowledge base as well as a current plan.

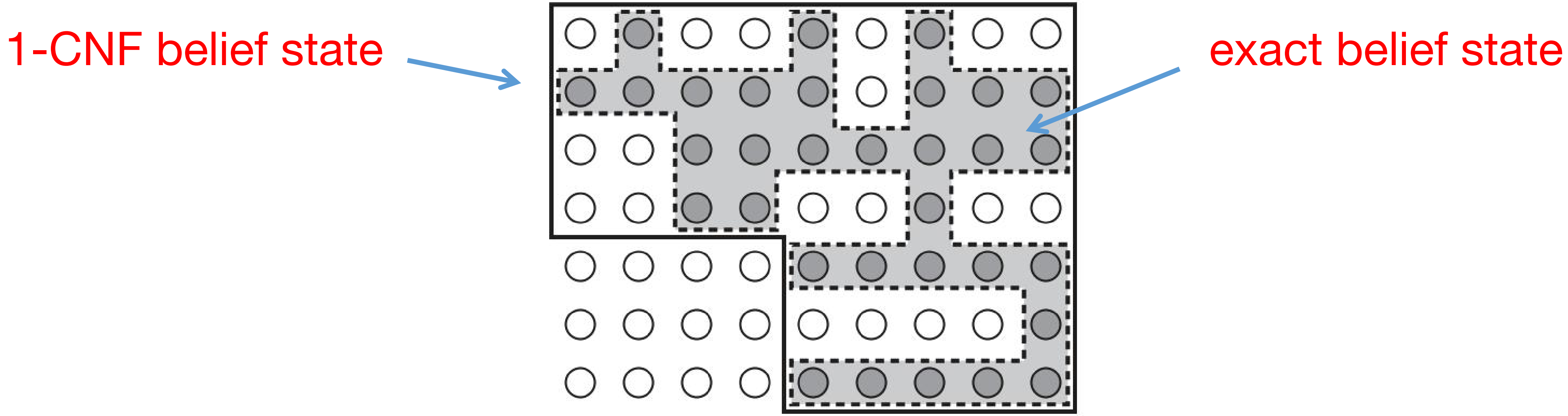
- A* search
  - First, if there is a glitter, the program constructs a plan to grab the gold, follow a route back to the initial location.
  - Otherwise, if there is no current plan, the program plans a route to the closest safe square that it has not visited yet, making sure the route goes through only safe squares.

Logical State Estimation

- Constant update time of the knowledge base
  - Using a cache to store the results of inference.
- Belief state:
  - the short-term memory of the agent.
  - all of the information the agent has remembered from the previous time at time t
  - every possible set of physical states.
- The set of states of at time 1:
  - $WumpusAlive \wedge L^1_{2,1} \wedge B_{2,1} \wedge (P_{2,1} \vee P_{2,2})$
- State estimation
  - The process of updating the belief state as new percepts arrive

Logical State Estimation

- Approximate state estimation
    - represent belief states as conjunctions of literals, that is, 1-CNF formulas
- Prove X_t and $\neg X_t$ for each symbol X_t given the belief state at t - 1



Example

- Which of the following is correct?
  - False $\models$ True True
  - True $\models$ False False
  - $(A \wedge B) \models (A \Leftrightarrow B)$ True
  - $A \Leftrightarrow B \models A \vee B$ False
  - $A \Leftrightarrow B \models \neg A \vee B$ True
  - $(A \vee B) \wedge \neg(A \Rightarrow B)$ is satisfiable True
  - $(A \Leftrightarrow B) \wedge (\neg A \vee B)$ is satisfiable True

Example

- Consider a vocabulary with only four propositions, A, B, C, D. How many models are there for the following sentences?
  - $B \lor C$ 12
  - $\neg A \lor \neg B \lor \neg C \lor \neg D$ 15

Example

- Prove, or find a counterexample to, each of the following assertions

1. If $\alpha \vDash \gamma$ or $\beta \vDash \gamma$ (or both) then $(\alpha \land \beta) \vDash \gamma$ True

2. If $\alpha \vDash (\beta \land \gamma)$ then $\alpha \vDash \gamma$ and $\alpha \vDash \beta$ True

3. If $\alpha \vDash (\beta \lor \gamma)$ then $\alpha \vDash \gamma$ or $\alpha \vDash \beta$ (or both) False

Example

- A propositional 2-CNF expression is a conjunction of clauses, each containing exactly 2 literals, e.g.,

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

1. Prove using resolution that the above sentence entails G

- Add $\neg G$. Resolve with the last two clauses to produce $\neg C$ and $\neg D$.
- Resolve with the second and third produce $\neg A$ and $\neg B$
- Resolve with the first clause to produce the empty clause

Example

- A propositional 2-CNF expression is a conjunction of clauses, each containing exactly 2 literals, e.g.,

$$(A \lor B) \land (\neg A \lor C) \land (\neg B \lor D) \land (\neg C \lor G) \land (\neg D \lor G)$$

2. Two clauses are semantically distinct if they are not logically equivalent. How many semantically distinct 2-CNF clauses can be constructed from n proposition symbols?

- There are $C(2n, 2) = (2n)(2n-1)/2 = 2n^2-n$ clauses with two different literals
- $\neg A \lor A$ ... are equivalent. So we have $2n^2-2n+1$
- Add $A \lor A$ ..., so we have $2n^2+1$