CSE 473: Logic in AI

Luke Zettlemoyer

(With slides from Dan Weld, Mausam, Stuart Russell, Dieter Fox, Henry Kautz…)

There is nothing so powerful as truth, and often nothing so strange.

- Daniel Webster (1782-1852)

KR Hypothesis

Any ***intelligent process*** will have ingredients that

1) We as external observers interpret as knowledge
2) This knowledge plays a formal, causal & essential role in guiding the behavior

- Brian Smith (paraphrased)

Some KR Languages

- Propositional Logic
- Predicate Calculus
- Frame Systems
- Rules with Certainty Factors
- Bayesian Belief Networks
- Influence Diagrams
- Semantic Networks
- Concept Description Languages
- Non-monotonic Logic

Knowledge Representation

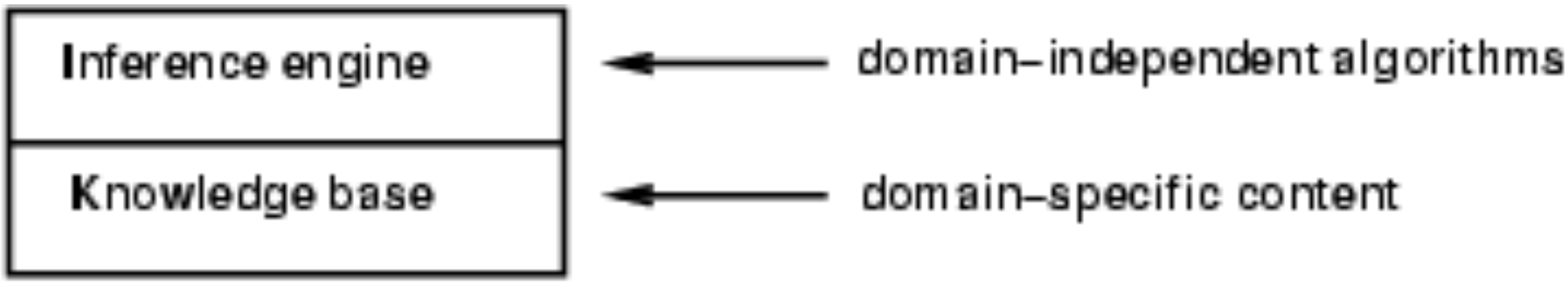
- *represent knowledge in a manner that facilitates inferencing (i.e. drawing conclusions) from knowledge.*

- Typically based on
  - Logic
  - Probability
  - Logic and Probability

Basic Idea of Logic

- By starting with true assumptions, you can deduce true conclusions.

Knowledge bases



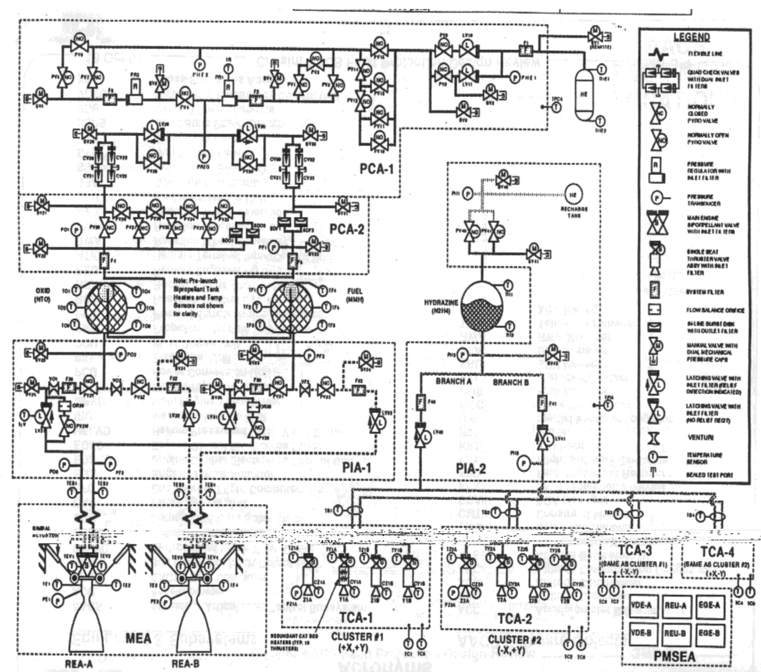
- Knowledge base = set of sentences in a formal language

- 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 should follow from the KB

Deep Space One

- Autonomous diagnosis & repair "Remote Agent"
- Compiled schematic to 7,000 var SAT problem



Muddy Children Problem

- Mom to N children "Don't get dirty"
- While playing, K≥1 get mud on forehead
- Father: "Some of you are dirty!"
- Father: "Raise your hand if you are dirty"
  - No one raises hand
- Father: "Raise your hand if you are dirty"
  - No one raises hand

…

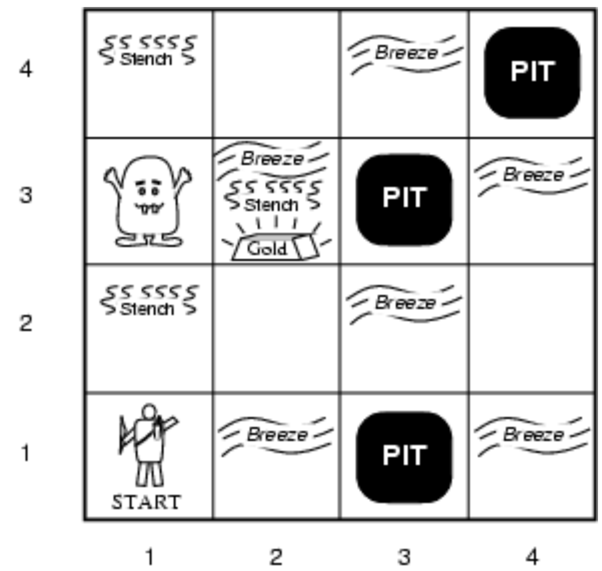
- Father: "Raise your hand if you are dirty"
  - All dirty children raise hand

} K-1 times

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



- Sensors: Stench, Breeze, Glitter, Bump, Scream
- Actuators: Left turn, Right turn, Forward, Grab, Release, Shoot

Components of KR

- Syntax: defines the sentences in the language
- Semantics: defines the "meaning" of sentences
- Inference Procedure
  - Algorithm
  - Sound?
  - Complete?
  - Complexity
- Knowledge Base

Propositional Logic

- Syntax
  - Atomic sentences: P, Q, …
  - Connectives: $\land$, $\lor$, $\neg$, $\Rightarrow$
- Semantics
  - Truth Tables
- Inference
  - Modus Ponens
  - Resolution
  - DPLL
  - GSAT

Propositional Logic: Syntax

- Atoms
  - P, Q, R, …
- Literals
  - P, $\neg$P
- Sentences
  - Any literal is a sentence
  - If S is a sentence
    - Then (S $\wedge$ S) is a sentence
    - Then (S $\vee$ S) is a sentence
- Conveniences

  P $\rightarrow$ Q same as $\neg$P $\vee$ Q

Truth tables for connectives

P	Q	$\neg P$	$P \wedge Q$	$P \vee Q$	$P \Rightarrow Q$	$P \Leftrightarrow Q$
$false$	$false$	$true$	$false$	$false$	$true$	$true$
$false$	$true$	$true$	$false$	$true$	$true$	$false$
$true$	$false$	$false$	$false$	$true$	$false$	$false$
$true$	$true$	$false$	$true$	$true$	$true$	$true$

A Knowledge Base

If the unicorn is mythical, then it is immortal, but if it is not mythical, it is a reptile. If the unicorn is either immortal or a reptile, then it is horned.

$(\neg R \vee H)$ $(\neg I \vee H)$

M = mythical
I = immortal
R = reptile
H = horned

$(M \vee R)$

$(\neg M \vee I)$

Prop. Logic: **Knowledge Engr**

1) One of the women is a biology major
2) Lisa is not next to Dave in the ranking
3) Dave is immediately ahead of Jim
4) Jim is immediately ahead of a bio major
5) Mary or Lisa is ranked first

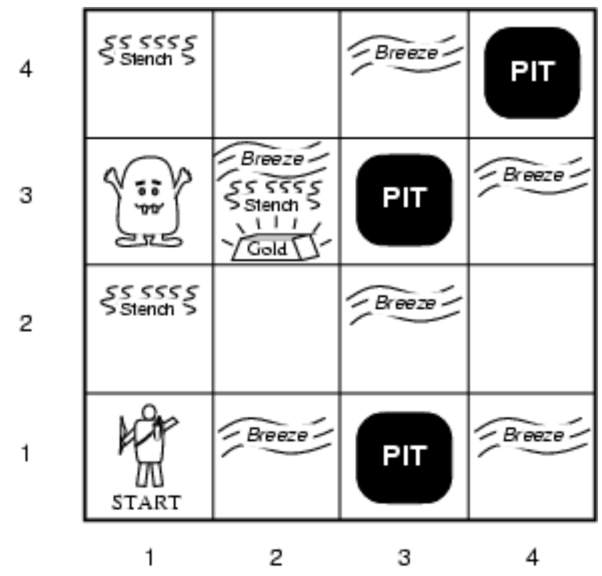
1. Choose Vocabulary

Universe: Lisa, Dave, Jim, Mary
LD = "Lisa is immediately ahead of Dave"
D = "Dave is a Bio Major"

2. Choose initial sentences

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



- Sensors: Stench, Breeze, Glitter, Bump, Scream
- Actuators: Left turn, Right turn, Forward, Grab, Release, Shoot

Wumpus world sentences: KB

Let $P_{i,j}$ be true if there is a pit in [i, j].
Let $B_{i,j}$ be true if there is a breeze in [i, j].

KB:

$\neg P_{1,1}$

$\neg B_{1,1}$

"Pits cause breezes in adjacent squares"

$B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$

$B_{2,1} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1})$

Full Encoding of Wumpus World

In propositional logic:

$\neg P_{1,1}$

$\neg W_{1,1}$

$B_{x,y} \Leftrightarrow (P_{x,y+1} \vee P_{x,y-1} \vee P_{x+1,y} \vee P_{x-1,y})$

$S_{x,y} \Leftrightarrow (W_{x,y+1} \vee W_{x,y-1} \vee W_{x+1,y} \vee W_{x-1,y})$

$W_{1,1} \vee W_{1,2} \vee \ldots \vee W_{4,4}$

$\neg W_{1,1} \vee \neg W_{1,2}$

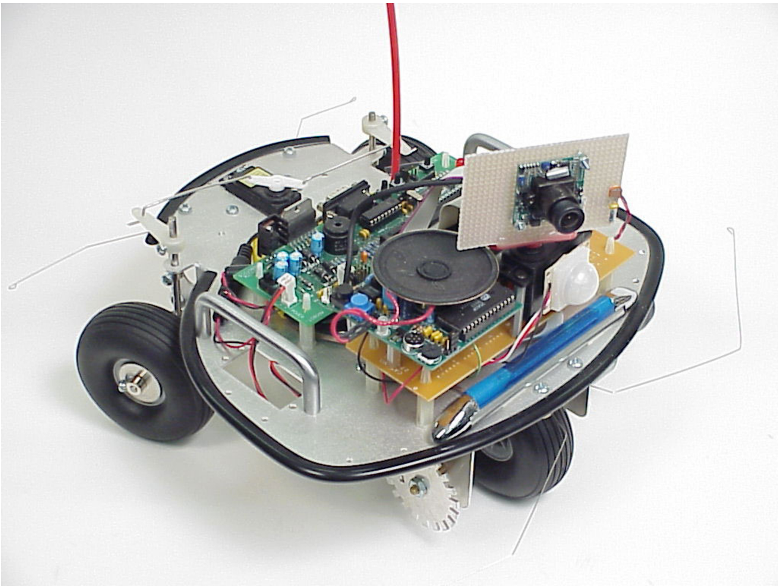
$\neg W_{1,1} \vee \neg W_{1,3}$

…

$\Rightarrow$ 64 distinct proposition symbols, 155 sentences

State Estimation

- Maintaining a KB which records what you know about the (partially observed) world state
  - Prop logic
  - First order logic
  - Probabilistic encodings



A Simple Knowledge Based Agent

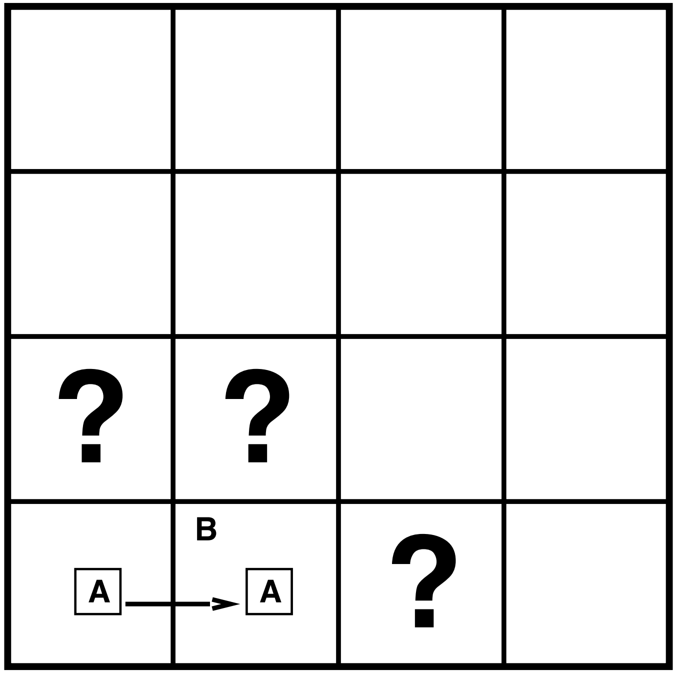
```
function KB-AGENT(percept) returns an action
    static: KB, a knowledge base
            t, a counter, initially 0, indicating time

    TELL(KB, MAKE-PERCEPT-SENTENCE(percept, t))
    action ← ASK(KB, MAKE-ACTION-QUERY(t))
    TELL(KB, MAKE-ACTION-SENTENCE(action, t))
    t ← 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

Entailment in Wumpus World

KB={$\neg P_{1,1}$, $\neg W_{1,1}$, $\neg B_{1,1}$, $\neg G_{1,1}$,
$\neg P_{1,1}$, $\neg W_{1,1}$, $B_{1,1}$, $\neg G_{1,1}$,
…
$B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$
… }



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

Consider possible models for ?s assuming only pits

3 Boolean choices $\Rightarrow$ 8 possible models

Wumpus Models

Possible assignments for the three locations which we have evidence about:

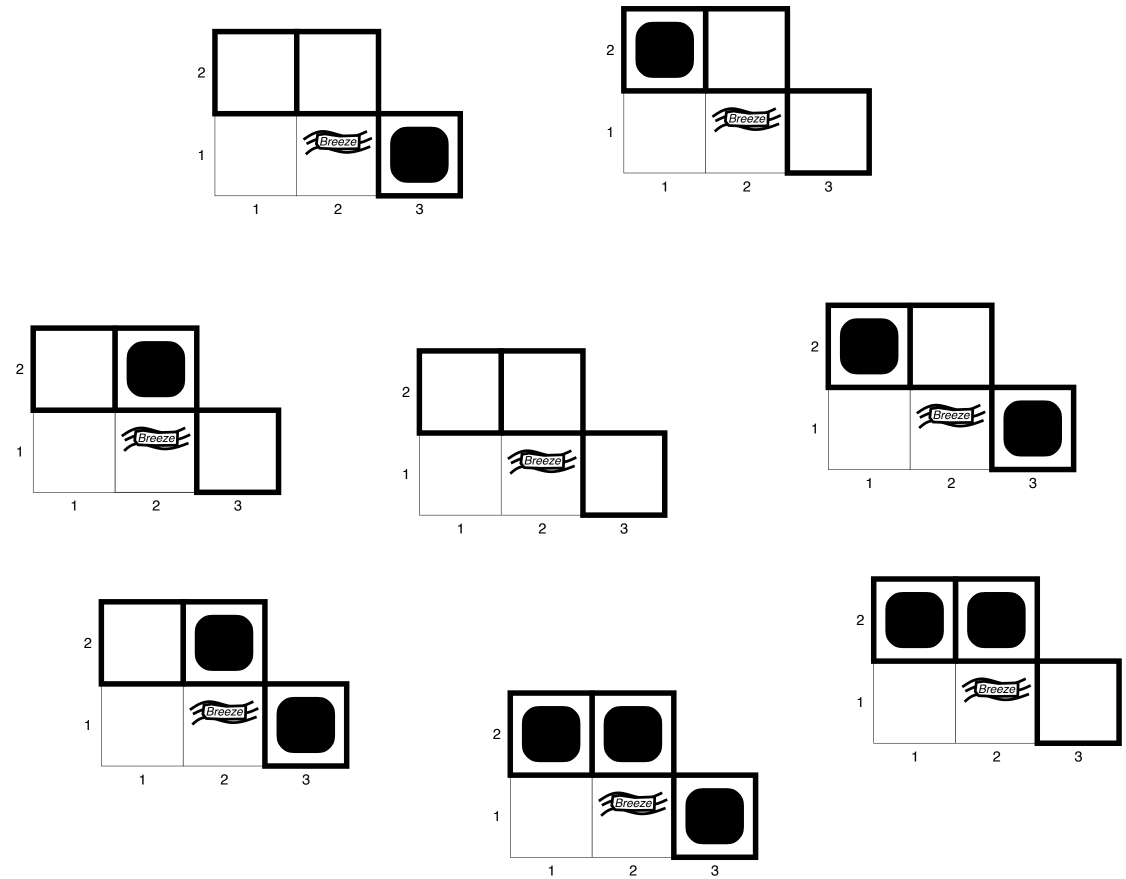
$KB = \{\neg P_{1,1}, \neg W_{1,1}, \neg B_{1,1}, \neg G_{1,1},$

$\neg P_{1,1}, \neg W_{1,1}, B_{1,1}, \neg G_{1,1},$

…

$B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$

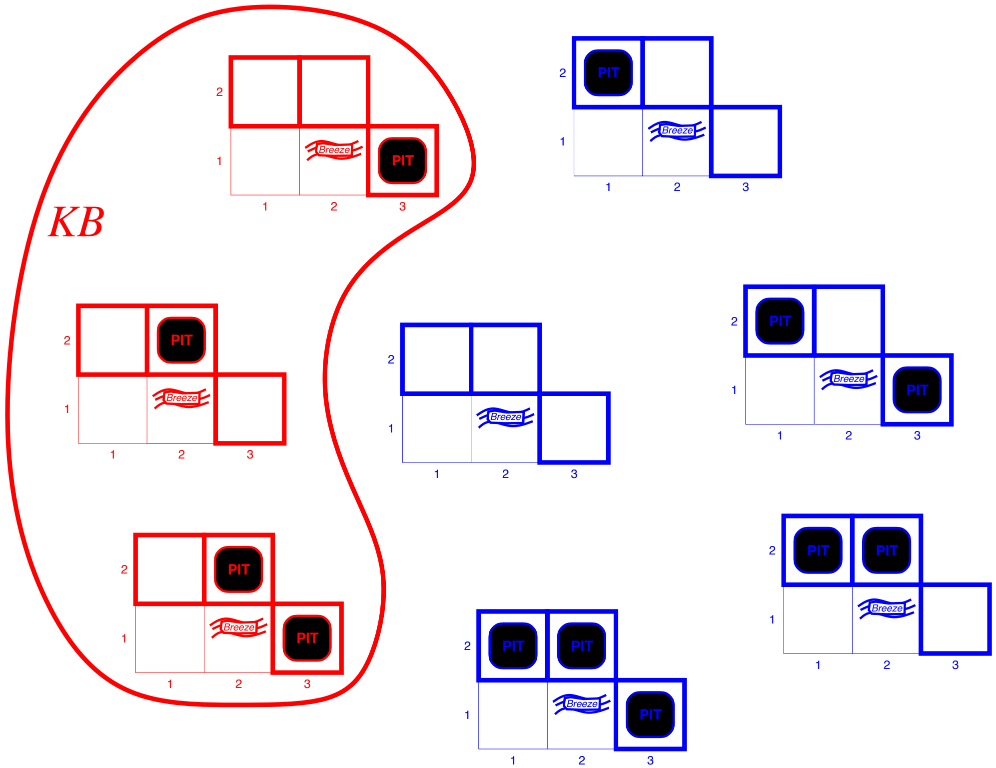
$\ldots \}$



Wumpus Models

Models that are consistent with our KB:

$$KB = \{\neg P_{1,1},\ \neg W_{1,1},\ \neg B_{1,1},\ \neg G_{1,1},$$
$$\neg P_{1,1},\ \neg W_{1,1},\ B_{1,1},\ \neg G_{1,1},$$
$$\ldots$$
$$B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$$
$$\ldots\}$$



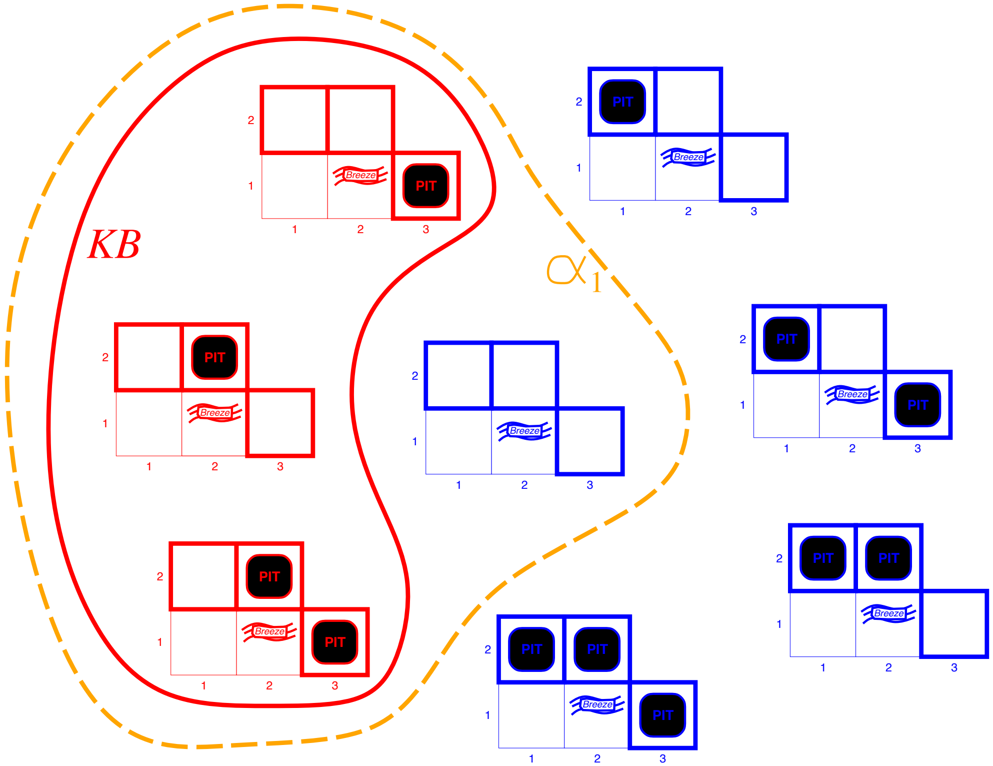
KB = wumpus-world rules + observations

Wumpus Models

This KB does entails that [1,2] is safe:

KB={$\neg P_{1,1}$, $\neg W_{1,1}$, $\neg B_{1,1}$, $\neg G_{1,1}$,
$\neg P_{1,2}$, $\neg W_{1,2}$, $B_{1,2}$, $\neg G_{1,2}$,
...
$B_{1,1} \Leftrightarrow (P_{1,2} \lor P_{2,1})$
... }

$\alpha_1 = \neg P_{1,2} \land \neg W_{1,2}$



KB = wumpus-world rules + observations

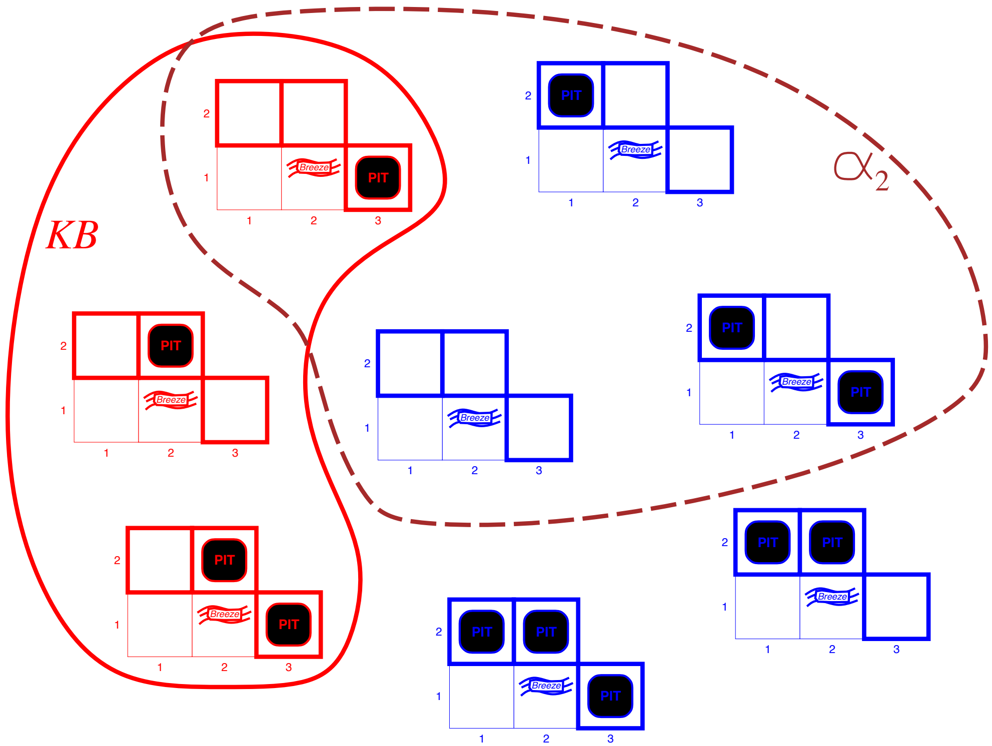
α_1 = "[1,2] is safe", $KB \models \alpha_1$, proved by model checking

Wumpus Models

This KB does not entail that [2,2] is safe:

$KB=\{\neg P_{1,1}, \neg W_{1,1}, \neg B_{1,1}, \neg G_{1,1},$
$\neg P_{1,2}, \neg W_{1,2}, B_{1,2}, \neg G_{1,2},$
$\dots$
$B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$
$\dots\}$

$\alpha_2 = \neg P_{2,2} \wedge \neg W_{2,2}$

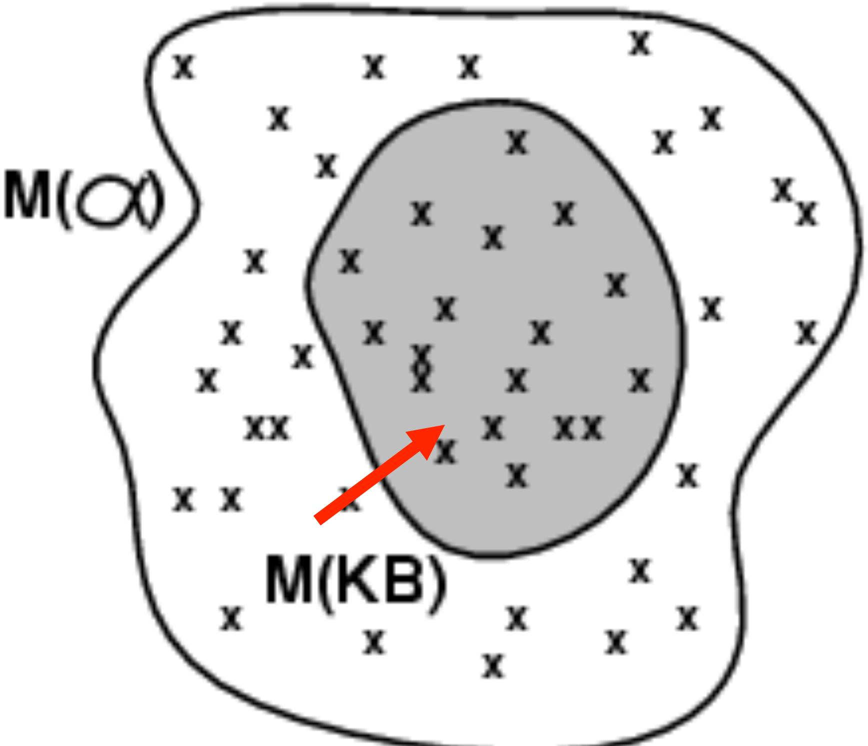


KB = wumpus-world rules + observations

α_2 = "[2,2] is safe", $KB \not\models \alpha_2$

Summary: Models

- Logicians often think in terms of *models*, which are formally structured worlds with respect to which truth can be evaluated
  - In propositional case, each model = truth assignment
  - Set of models can be enumerated in a truth table
- We say *m* is a model ***of*** a sentence α if α is true in *m*
- *M(α)* is the set of all models ***of*** α
- Entailment: KB $\models \alpha$ iff *M(KB)* $\subseteq$ *M*(α)
  - E.g. *KB* = $(P \vee Q) \wedge (\neg P \vee R)$
    $\alpha = (P \vee R)$
- How to check?
  - One way is to enumerate all elements in the truth table – slow ☹



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: i is sound if
whenever $KB \vdash_i \alpha$, it is also true that $KB \models \alpha$

Completeness: i is complete if
whenever $KB \models \alpha$, it is also true that $KB \vdash_i \alpha$

Preview: we will define a logic (first-order 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.

Truth Tables for Inference

$B_{1,1}$	$B_{2,1}$	$P_{1,1}$	$P_{1,2}$	$P_{2,1}$	$P_{2,2}$	$P_{3,1}$	R_1	R_2	R_3	R_4	R_5	KB
false	*false*	*false*	*false*	*false*	*false*	*false*	*true*	*true*	*true*	*true*	*false*	*false*
false	*false*	*false*	*false*	*false*	*false*	*true*	*true*	*true*	*false*	*true*	*false*	*false*
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
false	*true*	*false*	*false*	*false*	*false*	*false*	*true*	*true*	*false*	*true*	*true*	*false*
false	*true*	*false*	*false*	*false*	*false*	*true*	*true*	*true*	*true*	*true*	*true*	*true*
false	*true*	*false*	*false*	*false*	*true*	*false*	*true*	*true*	*true*	*true*	*true*	*true*
false	*true*	*false*	*false*	*false*	*true*	*true*	*true*	*true*	*true*	*true*	*true*	*true*
false	*true*	*false*	*false*	*true*	*false*	*false*	*true*	*false*	*false*	*true*	*true*	*false*
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
true	*true*	*true*	*true*	*true*	*true*	*true*	*false*	*true*	*true*	*false*	*true*	*false*

Enumerate rows (different assignments to symbols), if KB is true in row, check that α is too

Problem: exponential time and space!

Logical Equivalence

Two sentences are logically equivalent iff true in same models:

$\alpha \equiv \beta$ if and only if $\alpha \models \beta$ and $\beta \models \alpha$

$(\alpha \land \beta)$	$\equiv$	$(\beta \land \alpha)$	commutativity of $\land$
$(\alpha \lor \beta)$	$\equiv$	$(\beta \lor \alpha)$	commutativity of $\lor$
$((\alpha \land \beta) \land \gamma)$	$\equiv$	$(\alpha \land (\beta \land \gamma))$	associativity of $\land$
$((\alpha \lor \beta) \lor \gamma)$	$\equiv$	$(\alpha \lor (\beta \lor \gamma))$	associativity of $\lor$
$\neg(\neg\alpha)$	$\equiv$	α	double-negation elimination
$(\alpha \Rightarrow \beta)$	$\equiv$	$(\neg\beta \Rightarrow \neg\alpha)$	contraposition
$(\alpha \Rightarrow \beta)$	$\equiv$	$(\neg\alpha \lor \beta)$	implication elimination
$(\alpha \Leftrightarrow \beta)$	$\equiv$	$((\alpha \Rightarrow \beta) \land (\beta \Rightarrow \alpha))$	biconditional elimination
$\neg(\alpha \land \beta)$	$\equiv$	$(\neg\alpha \lor \neg\beta)$	De Morgan
$\neg(\alpha \lor \beta)$	$\equiv$	$(\neg\alpha \land \neg\beta)$	De Morgan
$(\alpha \land (\beta \lor \gamma))$	$\equiv$	$((\alpha \land \beta) \lor (\alpha \land \gamma))$	distributivity of $\land$ over $\lor$
$(\alpha \lor (\beta \land \gamma))$	$\equiv$	$((\alpha \lor \beta) \land (\alpha \lor \gamma))$	distributivity of $\lor$ over $\land$

Validity and Satisfiability

A sentence is valid if it is true in **all** models,

e.g., $True$, $A \lor \neg A$, $A \Rightarrow A$, $(A \land (A \Rightarrow B)) \Rightarrow B$

Validity is connected to inference via the Deduction Theorem:

$KB \models \alpha$ if and only if $(KB \Rightarrow \alpha)$ is valid

A sentence is satisfiable if it is true in **some** model

e.g., $A \lor B$, C

A sentence is unsatisfiable if it is true in **no** models

e.g., $A \land \neg A$

Satisfiability is connected to inference via the following:

$KB \models \alpha$ if and only if $(KB \land \neg\alpha)$ is unsatisfiable

i.e., prove α by *reductio ad absurdum*

Proof Methods

Proof methods divide into (roughly) two kinds:

Application of inference rules

- Legitimate (sound) generation of new sentences from old
- Proof $=$ a sequence of inference rule applications
  Can use inference rules as operators in a standard search alg.
- Typically require translation of sentences into a normal form

Model checking

truth table enumeration (always exponential in n)
improved backtracking, e.g., Davis–Putnam–Logemann–Loveland
heuristic search in model space (sound but incomplete)
    e.g., min-conflicts-like hill-climbing algorithms