Fundamentals of Artificial Intelligence
Chapter 07: **Logical Agents**

Roberto Sebastiani

DISI, Università di Trento, Italy – roberto.sebastiani@unitn.it

https://disi.unitn.it/rseba/DIDATTICA/fai_2023/

Teaching assistants:

Mauro Dragoni, dragoni@fbk.eu, https://www.maurodragoni.com/teaching/fai/

Paolo Morettin, paolo.morettin@unitn.it, https://paolomorettin.github.io/

M.S. Course "Artificial Intelligence Systems", academic year 2023-2024

Last update: Monday 30th October, 2023, 11:59

Outline

Outline

Propositional Logic (aka Boolean Logic)

Basic Definitions and Notation

- Propositional formula (aka Boolean formula or sentence)
  - $\top, \bot$ are formulas
  - a propositional atom $A_1, A_2, A_3, \ldots$ is a formula;
  - if φ_1 and φ_2 are formulas, then
    $\neg\varphi_1, \varphi_1 \wedge \varphi_2, \varphi_1 \vee \varphi_2, \varphi_1 \rightarrow \varphi_2, \varphi_1 \leftarrow \varphi_2, \varphi_1 \leftrightarrow \varphi_2, \varphi_1 \oplus \varphi_2$
    are formulas.
- Ex: $\varphi \stackrel{\text{def}}{=} (\neg(A_1 \rightarrow A_2)) \wedge (A_3 \leftrightarrow (\neg A_1 \oplus (A_2 \vee \neg A_4)))$
- *Atoms*(φ): the set $\{A_1, \ldots, A_N\}$ of atoms occurring in φ.
- Literal: a propositional atom A_i (positive literal) or its negation $\neg A_i$ (negative literal)
  - Notation: if $l := \neg A_i$, then $\neg l := A_i$
- Clause: a disjunction of literals $\bigvee_j l_j$ (e.g., $(A_1 \vee \neg A_2 \vee A_3 \vee \ldots)$)
- Cube: a conjunction of literals $\bigwedge_j l_j$ (e.g., $(A_1 \wedge \neg A_2 \wedge A_3 \wedge \ldots)$)

Semantics of Boolean operators

Truth Table

α	β	$\neg\alpha$	$\alpha \wedge \beta$	$\alpha \vee \beta$	$\alpha \rightarrow \beta$	$\alpha \leftarrow \beta$	$\alpha \leftrightarrow \beta$	$\alpha \oplus \beta$
$\bot$	$\bot$	$\top$	$\bot$	$\bot$	$\top$	$\top$	$\top$	$\bot$
$\bot$	$\top$	$\top$	$\bot$	$\top$	$\top$	$\bot$	$\bot$	$\top$
$\top$	$\bot$	$\bot$	$\bot$	$\top$	$\bot$	$\top$	$\bot$	$\top$
$\top$	$\top$	$\bot$	$\top$	$\top$	$\top$	$\top$	$\top$	$\bot$

English Meaning of Boolean Operators

English	Logic
A and B	$A \wedge B$
A if B | A when B | A whenever B	$A \leftarrow B$
if A, then B | A implies B | A forces B | A requires B	$A \rightarrow B$
A precisely when B | A if and only if B	$A \leftrightarrow B$
A or B (or both) | A unless B	$A \vee B$ (logical or)
either A or B (but not both)	$A \oplus B$ (exclusive or)

Remark: Semantics of Implication "$\rightarrow$" (aka "$\Rightarrow$", "$\supset$")

The semantics of Implication "$\alpha \rightarrow \beta$" may be counter-intuitive

$\alpha \rightarrow \beta$: "the antecedent (aka premise) α implies the consequent (aka conclusion) β" (aka "if α holds, then β holds"), but not vice versa

- does not require causation or relevance between α and β
  - ex: "5 is odd implies Tokyo is the capital of Japan" is true in p.l. (under the standard interpretation of "5", "odd", "Tokyo", "Japan")
  - relation between antecedent & consequent: they are both true
- is true whenever its antecedent is false
  - ex: "5 is even implies Sam is smart" is true (regardless the smartness of Sam)
  - ex: "5 is even implies Tokyo is in Italy" is true (!)
  - relation between antecedent & consequent: the former is false
- does not require temporal precedence of α wrt. β
  - ex: "the grass is wet implies it must have rained" is true (the consequent precedes temporally the antecedent)

Properties Boolean Operators

- $\wedge$, $\vee$, $\leftrightarrow$ and $\oplus$ are commutative:

$$
\begin{array}{lcl}
(\alpha \wedge \beta) & \iff & (\beta \wedge \alpha) \\
(\alpha \vee \beta) & \iff & (\beta \vee \alpha) \\
(\alpha \leftrightarrow \beta) & \iff & (\beta \leftrightarrow \alpha) \\
(\alpha \oplus \beta) & \iff & (\beta \oplus \alpha)
\end{array}
$$

- $\wedge$, $\vee$, $\leftrightarrow$ and $\oplus$ are associative:

$$
\begin{array}{lclcl}
((\alpha \wedge \beta) \wedge \gamma) & \iff & (\alpha \wedge (\beta \wedge \gamma)) & \iff & (\alpha \wedge \beta \wedge \gamma) \\
((\alpha \vee \beta) \vee \gamma) & \iff & (\alpha \vee (\beta \vee \gamma)) & \iff & (\alpha \vee \beta \vee \gamma) \\
((\alpha \leftrightarrow \beta) \leftrightarrow \gamma) & \iff & (\alpha \leftrightarrow (\beta \leftrightarrow \gamma)) & \iff & (\alpha \leftrightarrow \beta \leftrightarrow \gamma) \\
((\alpha \oplus \beta) \oplus \gamma) & \iff & (\alpha \oplus (\beta \oplus \gamma)) & \iff & (\alpha \oplus \beta \oplus \gamma)
\end{array}
$$

- $\rightarrow$, $\leftarrow$ are neither commutative nor associative:

$$
\begin{array}{lcl}
(\alpha \rightarrow \beta) & \not\iff & (\beta \rightarrow \alpha) \\
((\alpha \rightarrow \beta) \rightarrow \gamma) & \not\iff & (\alpha \rightarrow (\beta \rightarrow \gamma))
\end{array}
$$

Equivalences with Boolean Operators

$$
\begin{array}{rcl}
\neg\neg\alpha & \Longleftrightarrow & \alpha \\
(\alpha \lor \beta) & \Longleftrightarrow & \neg(\neg\alpha \land \neg\beta) \\
\neg(\alpha \lor \beta) & \Longleftrightarrow & (\neg\alpha \land \neg\beta) \\
(\alpha \land \beta) & \Longleftrightarrow & \neg(\neg\alpha \lor \neg\beta) \\
\neg(\alpha \land \beta) & \Longleftrightarrow & (\neg\alpha \lor \neg\beta) \\
(\alpha \rightarrow \beta) & \Longleftrightarrow & (\neg\alpha \lor \beta) \\
\neg(\alpha \rightarrow \beta) & \Longleftrightarrow & (\alpha \land \neg\beta) \\
(\alpha \leftarrow \beta) & \Longleftrightarrow & (\alpha \lor \neg\beta) \\
\neg(\alpha \leftarrow \beta) & \Longleftrightarrow & (\neg\alpha \land \beta) \\
(\alpha \leftrightarrow \beta) & \Longleftrightarrow & ((\alpha \rightarrow \beta) \land (\alpha \leftarrow \beta)) \\
 & \Longleftrightarrow & ((\neg\alpha \lor \beta) \land (\alpha \lor \neg\beta)) \\
\neg(\alpha \leftrightarrow \beta) & \Longleftrightarrow & (\neg\alpha \leftrightarrow \beta) \\
 & \Longleftrightarrow & (\alpha \leftrightarrow \neg\beta) \\
 & \Longleftrightarrow & ((\alpha \lor \beta) \land (\neg\alpha \lor \neg\beta)) \\
(\alpha \oplus \beta) & \Longleftrightarrow & \neg(\alpha \leftrightarrow \beta)
\end{array}
$$

Boolean logic can be expressed in terms of $\{\neg, \land\}$ (or $\{\neg, \lor\}$) only!

Exercises

1. For every pair of formulas $\alpha \iff \beta$ below, show that α and β can be rewritten into each other by applying the syntactic properties of the previous slide
   - $(A_1 \wedge A_2) \vee A_3 \iff (A_1 \vee A_3) \wedge (A_2 \vee A_3)$
   - $(A_1 \vee A_2) \wedge A_3 \iff (A_1 \wedge A_3) \vee (A_2 \wedge A_3)$
   - $A_1 \to (A_2 \to (A_3 \to A_4)) \iff (A_1 \wedge A_2 \wedge A_3) \to A_4$
   - $A_1 \to (A_2 \wedge A_3) \iff (A_1 \to A_2) \wedge (A_1 \to A_3)$
   - $(A_1 \vee A_2) \to A_3 \iff (A_1 \to A_3) \wedge (A_2 \to A_3)$
   - $A_1 \oplus A_2 \iff (A_1 \vee A_2) \wedge (\neg A_1 \vee \neg A_2)$
   - $\neg A_1 \leftrightarrow \neg A_2 \iff A_1 \leftrightarrow A_2$
   - $A_1 \leftrightarrow A_2 \leftrightarrow A_3 \iff A_1 \oplus A_2 \oplus A_3$

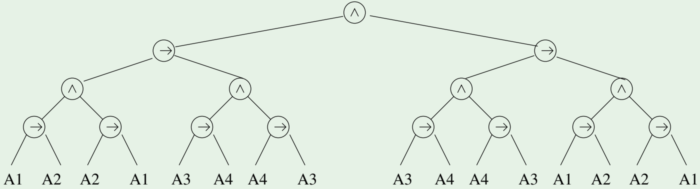
Tree & DAG Representations of Formulas

- Formulas can be represented either as trees or as DAGS (Directed Acyclic Graphs)
- DAG representation can be up to exponentially smaller
  - in particular, when $\leftrightarrow$'s are involved

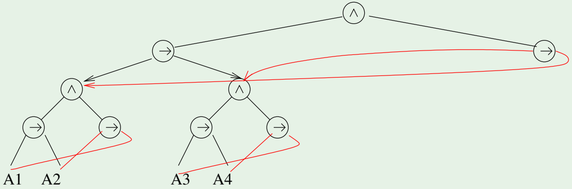
$$(A_1 \leftrightarrow A_2) \leftrightarrow (A_3 \leftrightarrow A_4)$$
$$\Downarrow$$
$$(((A_1 \leftrightarrow A_2) \rightarrow (A_3 \leftrightarrow A_4)) \wedge$$
$$((A_3 \leftrightarrow A_4) \rightarrow (A_1 \leftrightarrow A_2)))$$
$$\Downarrow$$
$$(((A_1 \rightarrow A_2) \wedge (A_2 \rightarrow A_1)) \rightarrow ((A_3 \rightarrow A_4) \wedge (A_4 \rightarrow A_3))) \wedge$$
$$(((A_3 \rightarrow A_4) \wedge (A_4 \rightarrow A_3)) \rightarrow (((A_1 \rightarrow A_2) \wedge (A_2 \rightarrow A_1))))$$

Tree & DAG Representations of Formulas: Example

$$(((A_1 \rightarrow A_2) \wedge (A_2 \rightarrow A_1)) \rightarrow ((A_3 \rightarrow A_4) \wedge (A_4 \rightarrow A_3))) \wedge$$
$$(((A_3 \rightarrow A_4) \wedge (A_4 \rightarrow A_3)) \rightarrow (((A_1 \rightarrow A_2) \wedge (A_2 \rightarrow A_1))))$$



Tree Representation



DAG Representation

Basic Definitions and Notation [cont.]

- Total truth assignment μ for φ:
  $\mu : Atoms(\varphi) \longmapsto \{\top, \bot\}$.
  - represents a possible world or a possible state of the world
- Partial Truth assignment μ for φ:
  $\mu : \mathcal{A} \longmapsto \{\top, \bot\}$, $\mathcal{A} \subset Atoms(\varphi)$.
  - represents 2^k total assignments, k is $\#$ unassigned variables
- Notation: set and formula representations of an assignment
  - μ can be represented as a set of literals:
    EX: $\{\mu(A_1) := \top, \mu(A_2) := \bot\} \implies \{A_1, \neg A_2\}$
  - μ can be represented as a formula (cube):
    EX: $\{\mu(A_1) := \top, \mu(A_2) := \bot\} \implies (A_1 \wedge \neg A_2)$

Basic Definitions and Notation [cont.]

- A total truth assignment μ satisfies φ (μ is a model of φ, $\mu \models \varphi$):

$$\mu \models A_i \iff \mu(A_i) = \top$$
$$\mu \models \neg\varphi \iff \textit{not } \mu \models \varphi$$
$$\mu \models \alpha \wedge \beta \iff \mu \models \alpha \textit{ and } \mu \models \beta$$
$$\mu \models \alpha \vee \beta \iff \mu \models \alpha \textit{ or } \mu \models \beta$$
$$\mu \models \alpha \rightarrow \beta \iff \textit{ if } \mu \models \alpha, \textit{ then } \mu \models \beta$$
$$\mu \models \alpha \leftrightarrow \beta \iff \mu \models \alpha \textit{ iff } \mu \models \beta$$
$$\mu \models \alpha \oplus \beta \iff \mu \models \alpha \textit{ iff not } \mu \models \beta$$

- $M(\varphi) \stackrel{\text{def}}{=} \{\mu \mid \mu \models \varphi\}$ (the set of models of φ)
- A partial truth assignment μ satisfies φ iff all its total extensions satisfy φ
  - (Ex: $\{A_1\} \models (A_1 \vee A_2)$) because $\{A_1, A_2\} \models (A_1 \vee A_2)$ and $\{A_1, \neg A_2\} \models (A_1 \vee A_2)$)
- φ is satisfiable iff $\mu \models \varphi$ for some μ (i.e. $M(\varphi) \neq \emptyset$)
- α entails β ($\alpha \models \beta$) iff, for all μs, $\mu \models \alpha \Longrightarrow \mu \models \beta$
  (i.e., $M(\alpha) \subseteq M(\beta)$)
- φ is valid ($\models \varphi$) iff $\mu \models \varphi$ forall μs (i.e., $\mu \in M(\varphi)$ forall μs)

Properties & Results

Property

φ is valid iff $\neg\varphi$ is unsatisfiable

Deduction Theorem

$\alpha \models \beta$ iff $\alpha \rightarrow \beta$ is valid ($\models \alpha \rightarrow \beta$)

Corollary

$\alpha \models \beta$ iff $\alpha \wedge \neg\beta$ is unsatisfiable

Validity and entailment checking can be straightforwardly reduced to (un)satisfiability checking!

Equivalence and Equi-Satisfiability

- α and β are equivalent iff, for every μ, $\mu \models \alpha$ iff $\mu \models \beta$ (i.e., if $M(\alpha) = M(\beta)$)
- α and β are equi-satisfiable iff exists μ_1 s.t. $\mu_1 \models \alpha$ iff exists μ_2 s.t. $\mu_2 \models \beta$ (i.e., if $M(\alpha) \neq \emptyset$ iff $M(\beta) \neq \emptyset$)
- α, β equivalent

  $\Downarrow \quad \not\Uparrow$

  α, β equi-satisfiable
- EX: $A_1 \vee A_2$ and $(A_1 \vee \neg A_3) \wedge (A_3 \vee A_2)$ are equi-satisfiable, not equivalent.
  $\{\neg A_1, A_2, A_3\} \models (A_1 \vee A_2)$, but $\{\neg A_1, A_2, A_3\} \not\models (A_1 \vee \neg A_3) \wedge (A_3 \vee A_2)$
- Typically used when β is the result of applying some transformation T to α: $\beta \stackrel{\text{def}}{=} T(\alpha)$:
  - T is validity-preserving [resp. satisfiability-preserving] iff $T(\alpha)$ and α are equivalent [resp. equi-satisfiable]

Complexity

- For N variables, there are up to 2^N truth assignments to be checked.
- The problem of deciding the satisfiability of a propositional formula is NP-complete

$\Longrightarrow$ The most important logical problems (validity, inference, entailment, equivalence, ...) can be straightforwardly reduced to (un)satisfiability, and are thus (co)NP-complete.

$$\Downarrow$$

No existing worst-case-polynomial algorithm.

Conjunctive Normal Form (CNF)

- φ is in Conjunctive normal form iff it is a conjunction of disjunctions of literals:

$$\bigwedge_{i=1}^{L} \bigvee_{j_i=1}^{K_i} l_{j_i}$$

- the disjunctions of literals $\bigvee_{j_i=1}^{K_i} l_{j_i}$ are called clauses
- Easier to handle: list of lists of literals.
  $\Longrightarrow$ no reasoning on the recursive structure of the formula

Classic CNF Conversion $CNF(\varphi)$

- Every φ can be reduced into CNF by, e.g.,
  (i) expanding implications and equivalences:
  $$\alpha \rightarrow \beta \Longrightarrow \neg\alpha \vee \beta$$
  $$\alpha \leftrightarrow \beta \Longrightarrow (\neg\alpha \vee \beta) \wedge (\alpha \vee \neg\beta)$$
  (ii) pushing down negations recursively:
  $$\neg(\alpha \wedge \beta) \Longrightarrow \neg\alpha \vee \neg\beta$$
  $$\neg(\alpha \vee \beta) \Longrightarrow \neg\alpha \wedge \neg\beta$$
  $$\neg\neg\alpha \Longrightarrow \alpha$$
  (iii) applying recursively the DeMorgan's Rule: $(\alpha \wedge \beta) \vee \gamma \Longrightarrow (\alpha \vee \gamma) \wedge (\beta \vee \gamma)$
- Resulting formula worst-case exponential:
  - ex: $||\text{CNF}(\bigvee_{i=1}^{N}(l_{i1} \wedge l_{i2})|| = ||(l_{11} \vee l_{21} \vee ... \vee l_{N1}) \wedge (l_{12} \vee l_{21} \vee ... \vee l_{N1}) \wedge ... \wedge (l_{12} \vee l_{22} \vee ... \vee l_{N2})|| = 2^N$
- $Atoms(CNF(\varphi)) = Atoms(\varphi)$
- $CNF(\varphi)$ is equivalent to φ: $M(CNF(\varphi)) = M(\varphi)$
- Rarely used in practice.

Labeling CNF conversion $CNF_{label}(\varphi)$

Labeling CNF conversion $CNF_{label}(\varphi)$ (aka Tseitin's conversion)

- Every φ can be reduced into CNF by, e.g., applying recursively bottom-up the rules:

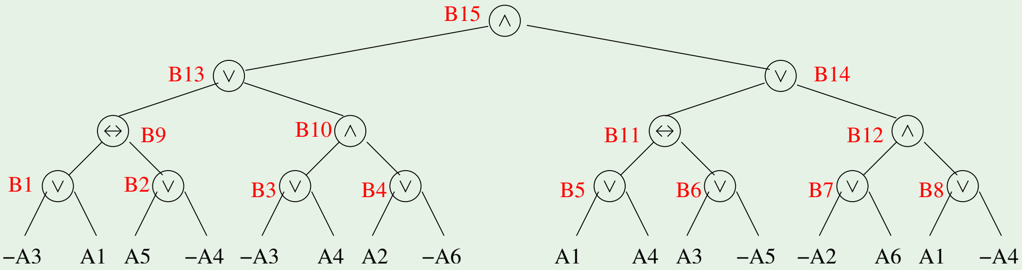
$$\varphi \implies \varphi[(l_i \vee l_j)|B] \wedge CNF(B \leftrightarrow (l_i \vee l_j))$$
$$\varphi \implies \varphi[(l_i \wedge l_j)|B] \wedge CNF(B \leftrightarrow (l_i \wedge l_j))$$
$$\varphi \implies \varphi[(l_i \leftrightarrow l_j)|B] \wedge CNF(B \leftrightarrow (l_i \leftrightarrow l_j))$$

  l_i, l_j being literals and B being a "new" variable.
- Worst-case linear!
- $Atoms(CNF_{label}(\varphi)) \supseteq Atoms(\varphi)$
- $CNF_{label}(\varphi)$ is equi-satisfiable w.r.t. φ:
  $M(CNF(\varphi)) \neq \emptyset$ iff $M(\varphi) \neq \emptyset$
- Much more used than classic conversion in practice.

Labeling CNF conversion $CNF_{label}(\varphi)$ (cont.)

$CNF(B \leftrightarrow (l_i \vee l_j))$	$\Longleftrightarrow$	$(\neg B \vee l_i \vee l_j) \wedge$ $(B \vee \neg l_i) \wedge$ $(B \vee \neg l_j)$
$CNF(B \leftrightarrow (l_i \wedge l_j))$	$\Longleftrightarrow$	$(\neg B \vee l_i) \wedge$ $(\neg B \vee l_j) \wedge$ $(B \vee \neg l_i \neg l_j)$
$CNF(B \leftrightarrow (l_i \leftrightarrow l_j))$	$\Longleftrightarrow$	$(\neg B \vee \neg l_i \vee l_j) \wedge$ $(\neg B \vee l_i \vee \neg l_j) \wedge$ $(B \vee l_i \vee l_j) \wedge$ $(B \vee \neg l_i \vee \neg l_j)$

Labeling CNF Conversion CNF_{label} – Example



$$
\begin{array}{lcl}
CNF(B_1 \leftrightarrow (\neg A_3 \vee A_1))\wedge & & (\neg B_1 \vee \neg A_3 \vee A_1) \wedge (B_1 \vee A_3) \wedge (B_1 \vee \neg A_1)\wedge \\
\ldots\wedge & & \ldots\wedge \\
CNF(B_8 \leftrightarrow (A_1 \vee \neg A_4))\wedge & & (\neg B_8 \vee A_1 \vee \neg A_4) \wedge (B_8 \vee \neg A_1) \wedge (B_8 \vee A_4)\wedge \\
CNF(B_9 \leftrightarrow (B_1 \leftrightarrow B_2))\wedge & & (\neg B_9 \vee \neg B_1 \vee B_2) \wedge (\neg B_9 \vee B_1 \vee \neg B_2)\wedge \\
 & & (B_9 \vee B_1 \vee B_2) \wedge (B_9 \vee \neg B_1 \vee \neg B_2)\wedge \\
\ldots\wedge & = & \ldots\wedge \\
CNF(B_{12} \leftrightarrow (B_7 \wedge B_8))\wedge & & (B_{12} \vee \neg B_7 \vee \neg B_8) \wedge (\neg B_{12} \vee B_7) \wedge (\neg B_{12} \vee B_8)\wedge \\
CNF(B_{13} \leftrightarrow (B_9 \vee B_{10}))\wedge & & (\neg B_{13} \vee B_9 \vee B_{10}) \wedge (B_{13} \vee \neg B_9) \wedge (B_{13} \vee \neg B_{10})\wedge \\
CNF(B_{14} \leftrightarrow (B_{11} \vee B_{12}))\wedge & & (\neg B_{14} \vee B_{11} \vee B_{12}) \wedge (B_{14} \vee \neg B_{11}) \wedge (B_{14} \vee \neg B_{12})\wedge \\
CNF(B_{15} \leftrightarrow (B_{13} \wedge B_{14}))\wedge & & (B_{15} \vee \neg B_{13} \vee \neg B_{14}) \wedge (\neg B_{15} \vee B_{13}) \wedge (\neg B_{15} \vee B_{14})\wedge \\
B_{15} & & B_{15}
\end{array}
$$

Outline

Propositional Reasoning: Generalities

- Automated Reasoning in Propositional Logic fundamental task
  - AI, formal verification, circuit synthesis, operational research,....
- Important in AI: $KB \models \alpha$: entail fact α from some knowledge base KB (aka Model Checking: $M(KB) \subseteq M(\alpha)$)
  - typically $||KB|| >> ||\alpha||$
  - sometimes KB set of variable implications $(A_1 \wedge ... \wedge A_k) \rightarrow B$
- All propositional reasoning tasks reduced to satisfiability (SAT)
  - $KB \models \alpha \Longrightarrow \text{SAT}(KB \wedge \neg\alpha) = \text{false}$
  - input formula CNF-ized and fed to a SAT solver
- Current SAT solvers dramatically efficient:
  - handle industrial problems with $10^6 - 10^7$ variables & clauses!
  - used as backend engines in a variety of systems (not only AI)

Outline

The Resolution Rule

- Resolution: deduction of a new clause from a pair of clauses with exactly one incompatible variable (resolvent):

$$\frac{(\overbrace{l_1 \vee \ldots \vee l_k}^{common} \vee \overbrace{l}^{resolvent} \vee \overbrace{l'_{k+1} \vee \ldots \vee l'_m}^{C'}) \quad (\overbrace{l_1 \vee \ldots \vee l_k}^{common} \vee \overbrace{\neg l}^{resolvent} \vee \overbrace{l''_{k+1} \vee \ldots \vee l''_n}^{C''})}{(\underbrace{l_1 \vee \ldots \vee l_k}_{common} \vee \underbrace{l'_{k+1} \vee \ldots \vee l'_m}_{C'} \vee \underbrace{l''_{k+1} \vee \ldots \vee l''_n}_{C''})}$$

- Ex: $\dfrac{(A \vee B \vee C \vee D \vee E) \quad (A \vee B \vee \neg C \vee F)}{(A \vee B \vee D \vee E \vee F)}$

- Note: many standard inference rules subcases of resolution:
(recall that $\alpha \rightarrow \beta \Longleftrightarrow \neg\alpha \vee \beta$)

$$\frac{A \rightarrow B \quad B \rightarrow C}{A \rightarrow C}\ (trans.) \qquad \frac{A \quad A \rightarrow B}{B}\ (m.\ ponens) \qquad \frac{\neg B \quad A \rightarrow B}{\neg A}\ (m.\ tollens)$$

Basic Propositional Inference: Resolution

- Assume input formula in CNF
  - if not, apply Tseitin CNF-ization first

$\Longrightarrow$ φ is represented as a set of clauses

- Search for a refutation of φ (is φ unsatisfiable?)
  - recall: $\alpha \models \beta$ iff $\alpha \wedge \neg\beta$ unsatisfiable
- Basic idea: apply iteratively the resolution rule to pairs of clauses with a conflicting literal, producing novel clauses, until either
  - a false clause is generated, or
  - the resolution rule is no more applicable
- Correct: if returns an empty clause, then φ unsat ($\alpha \models \beta$)
- Complete: if φ unsat ($\alpha \models \beta$), then it returns an empty clause
- Time-inefficient
- Very Memory-inefficient (exponential in memory)
- Many different strategies

Very-Basic PL-Resolution Procedure

function PL-RESOLUTION(KB, α) **returns** *true* or *false*
  inputs: KB, the knowledge base, a sentence in propositional logic
    α, the query, a sentence in propositional logic

  $clauses \leftarrow$ the set of clauses in the CNF representation of $KB \wedge \neg\alpha$
  $new \leftarrow \{\ \}$
  loop do
    for each pair of clauses C_i, C_j **in** $clauses$ **do**
      $resolvents \leftarrow$ PL-RESOLVE(C_i, C_j)
      if $resolvents$ contains the empty clause **then return** *true*
      $new \leftarrow new \cup resolvents$
    if $new \subseteq clauses$ **then return** *false*
    $clauses \leftarrow clauses \cup new$

Improvements: Subsumption & Unit Propagation

General "set" notation (Γ clause set):

$$\frac{\Gamma, \phi_1, ..\phi_n}{\Gamma, \phi'_1, ..\phi'_{n'}} \quad \left(e.g., \quad \frac{\Gamma, C_1 \vee p, C_2 \vee \neg p}{\Gamma, C_1 \vee p, C_2 \vee \neg p, C_1 \vee C_2,} \right)$$

- Removal of valid clauses: $\frac{\Gamma \wedge (p \vee \neg p \vee C)}{\Gamma}$
- Clause Subsumption (C clause): $\frac{\Gamma \wedge C \wedge (C \vee \bigvee_i l_i)}{\Gamma \wedge (C)}$
- Unit Resolution: $\frac{\Gamma \wedge (l) \wedge (\neg l \vee \bigvee_i l_i)}{\Gamma \wedge (l) \wedge (\bigvee_i l_i)}$
- Unit Subsumption: $\frac{\Gamma \wedge (l) \wedge (l \vee \bigvee_i l_i)}{\Gamma \wedge (l)}$
- Unit Propagation = Unit Resolution + Unit Subsumption

"Deterministic" rule: applied before other "non-deterministic" rules!

Remark

What happens with more than 1 resolvent?

- Common mistake: the following is not a correct application of the resolution rule:

$$\frac{\Gamma,\ (C_1 \vee l_1 \vee l_2),\ (C_2 \vee \neg l_1 \vee \neg l_2)}{\Gamma,\ (C_1 \vee l_1 \vee l_2),\ (C_2 \vee \neg l_1 \vee \neg l_2),\ (C_1 \vee C_2)}$$

- Rather, a correct application would be:

$$\frac{\Gamma,\ (C_1 \vee l_1 \vee l_2),\ (C_2 \vee \neg l_1 \vee \neg l_2)}{\Gamma,\ (C_1 \vee l_1 \vee l_2),\ (C_2 \vee \neg l_1 \vee \neg l_2),\ (C_1 \vee l_2 \vee C_2 \vee \neg l_2)}$$

... but $(C_1 \vee l_2 \vee C_2 \vee \vee\neg l_2)$ is valid and should be removed

$\Longrightarrow$ no clause is produced

Resolution: example

Given the following set of propositional clauses Γ:

$$
\begin{array}{l}
(A \lor D \lor \neg F) \\
(\neg C \lor E) \\
(A) \\
(B \lor E \lor \neg G) \\
(\neg G) \\
(\neg E \lor F) \\
(\neg A \lor \neg B \lor C) \\
(B) \\
(\neg B \lor \neg C \lor D) \\
(\neg B \lor \neg F \lor G)
\end{array}
$$

Produce a PL-resolution proof that Γ is unsatisfiable.

Solution:

$[(A), (\neg A \lor \neg B \lor C)] \implies (\neg B \lor C)$;
$[(B), (\neg B \lor C)] \implies (C)$;
$[(C), (\neg C \lor E)] \implies (E)$;
$[(E), (\neg E \lor F)] \implies (F)$;
$[(B), (\neg B \lor \neg F \lor G)] \implies (\neg F \lor G)$;
$[(F), (\neg F \lor G)] \implies (G)$;
$[(\neg G), (G)] \implies ()$;

Hint: resolve always unit clauses first!

Outline

The Davis-Putnam-Longemann-Loveland Procedure

- Tries to build an assignment μ satisfying φ
- At each step assigns a truth value to (all instances of) one atom
- Performs deterministic choices (mostly unit-propagation) first
- The grandfather of the most efficient SAT solvers
- Correct and complete
- Much more efficient than PL-Resolution
- Requires polynomial space

The DPLL Procedure [cont.]

```
function DPLL-SATISFIABLE?(s) returns true or false
   inputs: s, a sentence in propositional logic

   clauses ← the set of clauses in the CNF representation of s
   symbols ← a list of the proposition symbols in s
   return DPLL(clauses, symbols, { })
```

```
function DPLL(clauses, symbols, model) returns true or false

   if every clause in clauses is true in model then return true
   if some clause in clauses is false in model then return false
   P, value ← FIND-PURE-SYMBOL(symbols, clauses, model)
   if P is non-null then return DPLL(clauses, symbols – P, model ∪ {P=value})
   P, value ← FIND-UNIT-CLAUSE(clauses, model)
   if P is non-null then return DPLL(clauses, symbols – P, model ∪ {P=value})
   P ← FIRST(symbols); rest ← REST(symbols)
   return DPLL(clauses, rest, model ∪ {P=true}) or
          DPLL(clauses, rest, model ∪ {P=false}))
```


Pure-Symbol Rule out of date, no more used in modern solvers.

The DPLL Procedure [cont.]

function DPLL-SATISFIABLE?(s) **returns** *true* or *false*
  inputs: s, a sentence in propositional logic

  $clauses \leftarrow$ the set of clauses in the CNF representation of s
  $symbols \leftarrow$ a list of the proposition symbols in s
  return DPLL($clauses, symbols, \{\,\}$)

---

function DPLL($clauses, symbols, model$) **returns** *true* or *false*

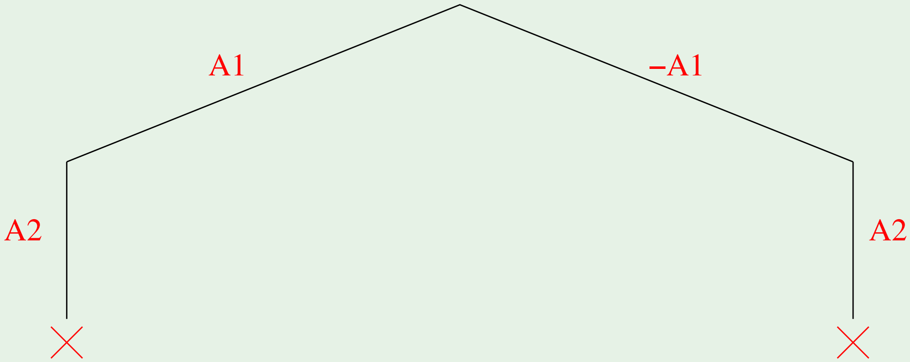
  if every clause in $clauses$ is true in $model$ **then return** *true*
  if some clause in $clauses$ is false in $model$ **then return** *false*
  ~~$P, value \leftarrow$ FIND-PURE-SYMBOL($symbols, clauses, model$)~~
  ~~**if** P is non-null **then return** DPLL($clauses, symbols - P, model \cup \{P{=}value\}$)~~
  $P, value \leftarrow$ FIND-UNIT-CLAUSE($clauses, model$)
  if P is non-null **then return** DPLL($clauses, symbols - P, model \cup \{P{=}value\}$)
  $P \leftarrow$ FIRST($symbols$); $rest \leftarrow$ REST($symbols$)
  return DPLL($clauses, rest, model \cup \{P{=}true\}$) **or**
      DPLL($clauses, rest, model \cup \{P{=}false\}$))

Pure-Symbol Rule out of date, no more used in modern solvers.

DPLL: Example

DPLL search tree

$$\varphi = (A_1 \vee A_2) \wedge (A_1 \vee \neg A_2) \wedge (\neg A_1 \vee A_2) \wedge (\neg A_1 \vee \neg A_2)$$

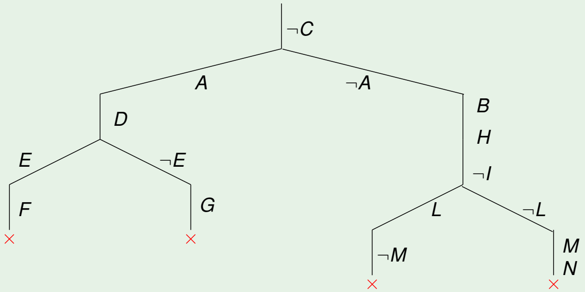


DPLL – example

DPLL (without pure-literal rule)

Here “choose-literal” selects variable in alphabetic order, selecting true first.

$$
\begin{array}{l}
(\neg C) \wedge \\
(B \vee A \vee C) \wedge \\
(\neg A \vee D) \wedge \\
(\neg E \vee \neg A \vee F) \wedge \\
(\neg E \vee \neg F \vee \neg A) \wedge \\
(G \vee \neg A \vee E) \wedge \\
(E \vee \neg G \vee \neg A) \wedge \\
(A \vee H \vee C) \wedge \\
(\neg H \vee \neg I \vee A) \wedge \\
(I \vee L \vee M) \wedge \\
(\neg L \vee C \vee \neg M) \wedge \\
(A \vee \neg L \vee M) \wedge \\
(L \vee N \vee \neg H) \wedge \\
(I \vee L \vee \neg N)
\end{array}
$$



$\Longrightarrow$ UNSAT

Remark: “choose-literal” selects only variables which still occur in the formula, after simplification. E.g., in the leftmost branch, after assigning $\neg C$, A, D, it does not select B because the clause $(B \vee A \vee C)$ has been simplified into true, and as such is no more part of the formula, so that B does not occur in the formula anymore.

Modern CDCL SAT Solvers

- Non-recursive, stack-based implementations
- Based on Conflict-Driven Clause-Learning (CDCL) schema
  - inspired to conflict-driven backjumping and learning in CSPs
  - learns implied clauses as nogoods
- Random restarts
  - abandon the current search tree and restart on top level
  - previously-learned clauses maintained
- Smart literal selection heuristics (ex: VSIDS)
  - "static": scores updated only at the end of a branch
  - "local": privileges variable in recently learned clauses
- Smart preprocessing/inprocessing technique to simplify formulas
- Smart indexing techniques (e.g. 2-watched literals)
  - efficiently do/undo assignments and reveal unit clauses
- Allow Incremental Calls (stack-based interface)
  - allow for reusing previous search on "similar" problems

Can handle industrial problems with $10^6 - 10^7$ variables and clauses!

Outline

Horn Formulas

- A Horn clause is a clause containing at most one positive literal
  - a definite clause is a clause containing exactly one positive literal
  - a goal clause is a clause containing no positive literal
- A Horn formula is a conjunction/set of Horn clauses
- Ex:
$$\begin{array}{ll} A_1 \lor \neg A_2 & \text{// definite} \\ A_2 \lor \neg A_3 \lor \neg A_4 & \text{// definite} \\ \neg A_5 \lor \neg A_3 \lor \neg A_4 & \text{// goal} \\ A_3 & \text{// definite} \end{array}$$
- Intuition: implications between positive Boolean variables:
$$\begin{array}{rl} A_2 \rightarrow & A_1 \\ (A_3 \land A_4) \rightarrow & A_2 \\ (A_5 \land A_3 \land A_4) \rightarrow & \bot \\ & A_3 \end{array}$$
- Often allow to represent knowledge-base entailment $KB \models \alpha$:
  - knowledge base KB written as sets of definite clauses
    ex: $In11;\ (\neg In11 \lor \neg MoveFrom11To12 \lor In12);$
  - goal $\neg\alpha$ as a goal clause
    ex: $\neg In12$

Tractability of Horn Formulas

Property

Checking the satisfiability of Horn formulas requires polynomial time:

- Hint:
  1. Eliminate unit clauses by propagating their value;
  2. If an empty clause is generated, return unsat
  3. Otherwise, every clause contains at least one negative literal

  $\Longrightarrow$ Assign all variables to $\bot$; return the assignment
- Alternatively: run DPLL/CDCL, selecting negative literals first

A simple polynomial procedure for Horn-SAT

function *Horn_SAT(formula* φ, *assignment &* μ*)* $\{$
    Unit_Propagate(φ, μ*)*;
    if $(\varphi == \bot)$
        then return *UNSAT*;
    else $\{$
        $\mu := \mu \cup \bigcup_{A_i \not\in \mu} \{\neg A_i\}$;
        return *SAT*;
$\}$ $\}$

function *Unit_Propagate(formula &* φ, *assignment &* μ*)*
    while $(\varphi \neq \top$ **and** $\varphi \neq \bot$ **and** {a unit clause (l) occurs in φ}) **do** $\{$
        $\varphi = assign(\varphi, l)$;
        $\mu := \mu \cup \{l\}$;
$\}$ $\}$

Example

$$
\begin{array}{lll}
\neg A_1 & \vee \ A_2 & \vee \neg A_3 \\
\ \ A_1 & \vee \neg A_3 & \vee \neg A_4 \\
\neg A_2 & \vee \neg A_4 & \\
\ \ A_3 & \vee \neg A_4 & \\
\ \ A_4 & &
\end{array}
$$

Example

$$
\begin{array}{llll}
\neg A_1 & \vee \; A_2 & \vee \neg A_3 \\
A_1 & \vee \neg A_3 & \vee \neg A_4 \\
\neg A_2 & \vee \neg A_4 & \\
A_3 & \vee \neg A_4 & \\
A_4 & &
\end{array}
$$

$\mu := \{A_4 := \top\}$

Example

$$\begin{array}{rlll} \neg A_1 & \vee \ A_2 & \vee \neg A_3 \\ A_1 & \vee \neg A_3 & \vee \neg A_4 \\ \neg A_2 & \vee \neg A_4 & \\ A_3 & \vee \neg A_4 & \\ A_4 & & \end{array}$$

$\mu := \{A_4 := \top, A_3 := \top\}$

Example

$$\begin{array}{llll} \neg A_1 & \vee A_2 & \vee \neg A_3 \\ A_1 & \vee \neg A_3 & \vee \neg A_4 \\ \neg A_2 & \vee \neg A_4 & \\ A_3 & \vee \neg A_4 & \\ A_4 & & \end{array}$$

$\mu := \{A_4 := \top, A_3 := \top, A_2 := \bot\}$

Example

$$\begin{array}{rlll} \neg A_1 & \vee \ A_2 & \vee \neg A_3 & \times \\ A_1 & \vee \neg A_3 & \vee \neg A_4 & \\ \neg A_2 & \vee \neg A_4 & & \\ A_3 & \vee \neg A_4 & & \\ A_4 & & & \end{array}$$

$\mu := \{A_4 := \top, A_3 := \top, A_2 := \bot, A_1 := \top\} \Longrightarrow$ UNSAT

Example 2

$$
\begin{array}{lll}
A_1 & \vee\neg A_2 & \\
A_2 & \vee\neg A_5 & \vee\neg A_4 \\
A_4 & \vee\neg A_3 & \\
A_3 & &
\end{array}
$$

Example 2

$$
\begin{array}{lll}
A_1 & \vee \neg A_2 & \\
A_2 & \vee \neg A_5 & \vee \neg A_4 \\
A_4 & \vee \neg A_3 & \\
A_3 & &
\end{array}
$$

$\mu := \{A_3 := \top\}$

Example 2

$$\begin{array}{lll} A_1 & \vee \neg A_2 & \\ A_2 & \vee \neg A_5 & \vee \neg A_4 \\ A_4 & \vee \neg A_3 & \\ A_3 & & \end{array}$$

$\mu := \{A_3 := \top, A_4 := \top\}$

Example 2

$$
\begin{array}{lll}
A_1 & \vee \neg A_2 & \\
A_2 & \vee \neg A_5 & \vee \neg A_4 \\
A_4 & \vee \neg A_3 & \\
A_3 & &
\end{array}
$$

$\mu := \{A_3 := \top, A_4 := \top\} \Longrightarrow$ SAT

Outline

Local Search with SAT

- Similar to Local Search for CSPs
- Input: set of clauses
- Use total truth assignments
  - allow states with unsatisfied clauses
  - "neighbour states" differ for one variable truth value
  - steps: reassign variable truth values
- Cost: # of unsatisfied clauses
- Stochastic local search [see Ch. 4] applies to SAT as well
  - random walk, simulated annealing, GAs, taboo search, ...
- The WalkSAT stochastic local search
  - Clause selection: randomly select an unsatisfied clause C
  - Variable selection:

  prob. p: flip variable from C at random
  prob. 1-p: flip variable from C causing a minimum number of unsat clauses
- Note: can detect only satisfiability, not unsatisfiability
- Many variants

The WalkSAT Procedure

function WALKSAT(*clauses*, *p*, *max_flips*) **returns** a satisfying model or *failure*
  inputs: *clauses*, a set of clauses in propositional logic
    p, the probability of choosing to do a "random walk" move, typically around 0.5
    max_flips, number of flips allowed before giving up

  model ← a random assignment of *true*/*false* to the symbols in *clauses*
  for $i = 1$ **to** *max_flips* **do**
    if *model* satisfies *clauses* **then return** *model*
    clause ← a randomly selected clause from *clauses* that is false in *model*
    with probability *p* flip the value in *model* of a randomly selected symbol from *clause*
    else flip whichever symbol in *clause* maximizes the number of satisfied clauses
  return *failure*

Outline

Outline

A Quote

You can think about deep learning as equivalent to ... our visual cortex or auditory cortex. But, of course, true intelligence is a lot more than just that, you have to recombine it into higher-level thinking and symbolic reasoning, a lot of the things classical AI tried to deal with in the 80s.

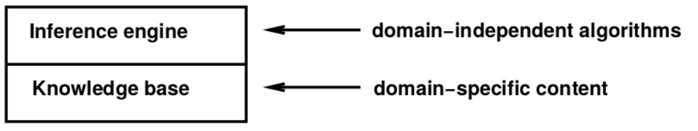
...

We would like to build up to this symbolic level of reasoning - maths, language, and logic. So that's a big part of our work.

Demis Hassabis, CEO of Google Deepmind

Knowledge Representation and Reasoning

- Knowledge Representation & Reasoning (KR&R): the field of AI dedicated to representing knowledge of the world in a form a computer system can utilize to solve complex tasks
- The class of systems/agents that derive from this approach are called knowledge based (KB) systems/agents
- A KB agent maintains a knowledge base (KB) of facts
  - represent the agent's representation of the world
  - expressed in a formal language (e.g. propositional logic)
  - collection of domain-specific facts believed by the agent
  - initially contains the background knowledge
  - KB queries and updates via logical entailment, performed by an inference engine
- Inference engine allows for inferring actions and new knowledge
  - domain-independent algorithms, can answer any question



(© S. Russell & P. Norwig, AIMA)

Reasoning

- Reasoning: formal manipulation of the symbols representing a collection of beliefs to produce representations of new ones
- Logical entailment ($KB \models \alpha$) is the fundamental operation
- Ex:
  - (KB acquired fact): "Patient x is allergic to medication m"
  - (KB general rule): "Anybody allergic to m is also allergic to m'."
  - (KB general rule): "If x is allergic to m', do not prescribe m' for x."
  - (query): "Prescribe m' for x?"
  - (answer) No (because patient x is allergic to medication m')
- Other forms of reasoning (last part of this course)
  - Probablistic reasoning
- Other forms of reasoning (not addressed in this course)
  - Abductive reasoning (aka diagnosis): given KB and β, conjecture hypotheses α s.t $(KB \wedge \alpha) \models \beta$
  - Abductive reasoning: from a set of observation find a general rule

Knowledge-Based Agents (aka Logic Agents)

- Logic agents: combine domain knowledge with current percepts to infer hidden aspects of current state prior to selecting actions
  - Crucial in partially observable environments
- KB Agent must be able to:
  - represent states and actions
  - incorporate new percepts
  - update internal representation of the world
  - deduce hidden properties of the world
  - deduce appropriate actions
- Agents can be described at different levels
  - knowledge level (declarative approach):
    behaviour completely described by the sentences stored in the KB
  - implementation level (procedural approach): behaviour described as program code
- Declarative approach to building an agent (or other system):
  - Tell the KB what it needs to know (update KB)
  - Ask what to do (answers should follow logically from KB & query)

Knowledge-Based Agent: General Schema

- Given a percept, the agent
  - Tells the KB of the percept at time step t
  - ASKs the KB for the best action to do at time step t
  - Tells the KB that it has in fact taken that action
- Details hidden in three functions: MAKE-PERCEPT-SENTENCE, MAKE-ACTION-QUERY, MAKE-ACTION-SENTENCE
  - construct logic sentences
  - implement the interface between sensors/actuators and KRR core
- Tell and Ask may require complex logical inference

```
function KB-AGENT(percept) returns an action
  persistent: 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
```

Outline

Example: The Wumpus World

Task Environment: PEAS Description

Performance measure:

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

Environment:

- squares adjacent to Wumpus are stenchy
- 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

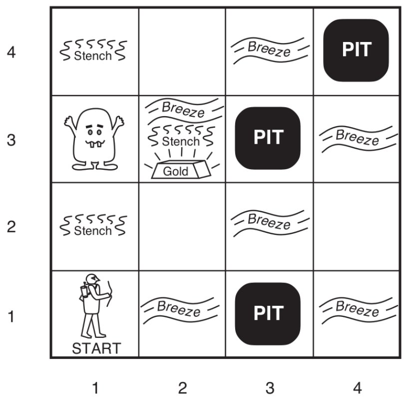
Actuators:

- Left turn, Right turn, Forward, Grab, Release, Shoot

Sensors:

- Stench, Breeze, Glitter, Bump, Scream

One possible configuration:



(© S. Russell & P. Norwig, AIMA)

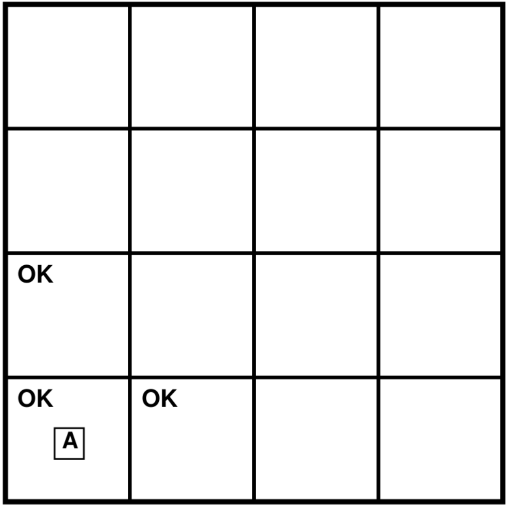
Wumpus World: Characterization

- Fully Observable? No: only local perception
- Deterministic? Yes: outcomes exactly specified
- Episodic? No: actions can have long-term consequences
- Static? Yes: Wumpus and Pits do not move
- Discrete? Yes
- Single-agent? Yes (Wumpus is essentially a natural feature)

Example: Exploring the Wumpus World

- The KB initially contains the rules of the environment.
- Agent is initially in 1,1
- Percepts:
  no stench, no breeze

$\Longrightarrow$ [1,2] and [2,1] OK



A: Agent; B: Breeze; G: Glitter; S: Stench
OK: safe square; W: Wumpus; P: Pit; BGS: bag of gold

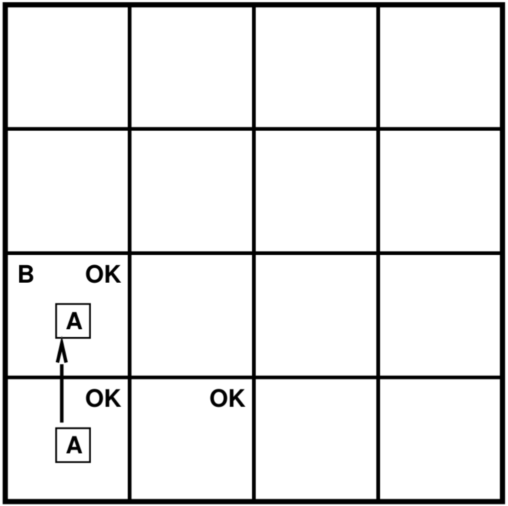
Example: Exploring the Wumpus World

- Agent moves to [2,1]
- perceives a breeze

$\Longrightarrow$ Pit in [3,1] or [2,2]

- perceives no stench

$\Longrightarrow$ no Wumpus in [3,1], [2,2]



A: Agent; B: Breeze; G: Glitter; S: Stench
OK: safe square; W: Wumpus; P: Pit; BGS: bag of gold

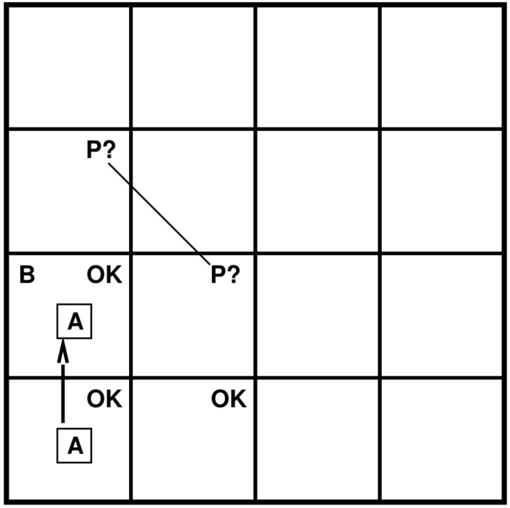
Example: Exploring the Wumpus World

- Agent moves to [2,1]
- perceives a breeze

$\Longrightarrow$ Pit in [3,1] or [2,2]

- perceives no stench

$\Longrightarrow$ no Wumpus in [3,1], [2,2]



A: Agent; B: Breeze; G: Glitter; S: Stench
OK: safe square; W: Wumpus; P: Pit; BGS: bag of gold

Example: Exploring the Wumpus World

- Agent moves to [1,1]-[1,2]
- perceives no breeze

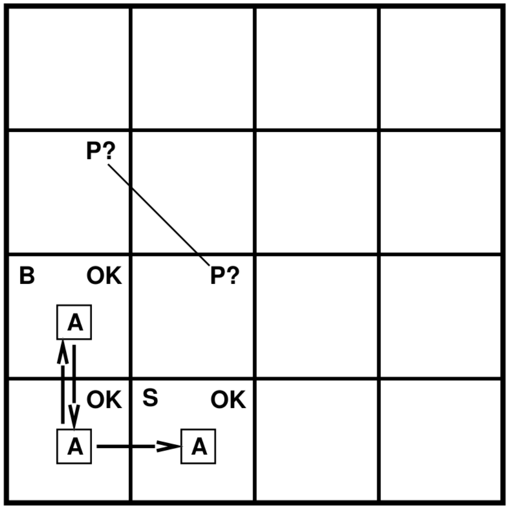
⟹ no Pit in [1,3], [2,2]

⟹ [2,2] OK

⟹ pit in [3,1]

- perceives a stench

⟹ Wumpus in ~~[2,2] or~~ [1,3]!



A: Agent; B: Breeze; G: Glitter; S: Stench
OK: safe square; W: Wumpus; P: Pit; BGS: bag of gold

Example: Exploring the Wumpus World

- Agent moves to [1,1]-[1,2]
- perceives no breeze

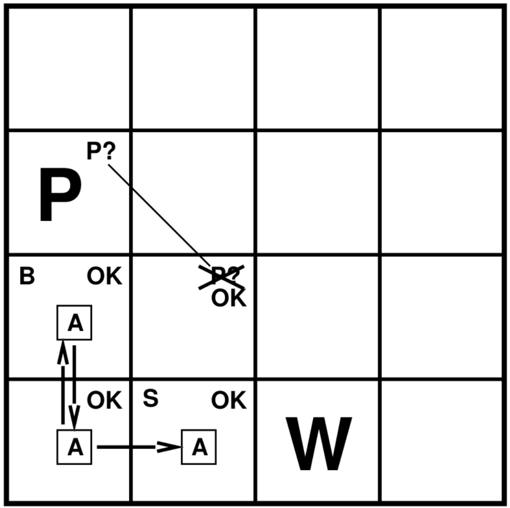
$\Longrightarrow$ no Pit in [1,3], [2,2]

$\Longrightarrow$ [2,2] OK

$\Longrightarrow$ pit in [3,1]

- perceives a stench

$\Longrightarrow$ Wumpus in ~~[2,2] or~~ [1,3]!



A: Agent; B: Breeze; G: Glitter; S: Stench
OK: safe square; W: Wumpus; P: Pit; BGS: bag of gold

Example: Exploring the Wumpus World

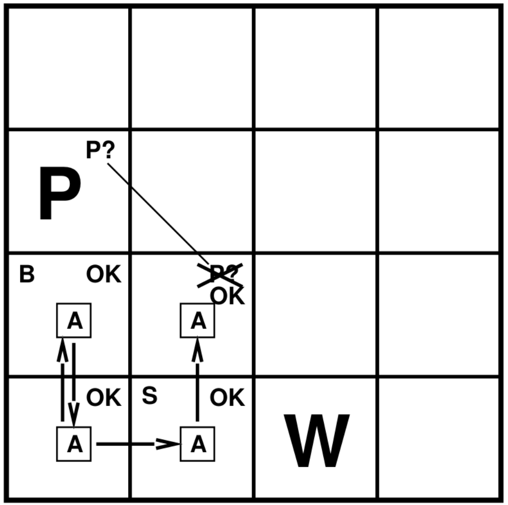
- Agent moves to [2,2]
- perceives no breeze

⟹ no pit in [3,2], [2,3]

- perceives no stench

⟹ no Wumpus in [3,2], [2,3]

⟹ [3,2] and [2,3] OK



A: Agent; B: Breeze; G: Glitter; S: Stench
OK: safe square; W: Wumpus; P: Pit; BGS: bag of gold

Example: Exploring the Wumpus World

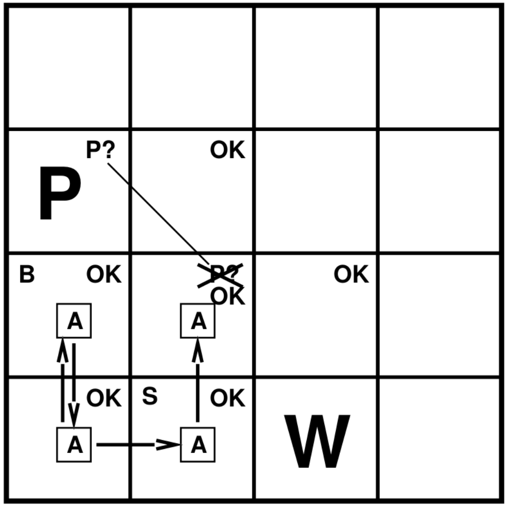
- Agent moves to [2,2]
- perceives no breeze

⟹ no pit in [3,2], [2,3]

- perceives no stench

⟹ no Wumpus in [3,2], [2,3]

⟹ [3,2] and [2,3] OK

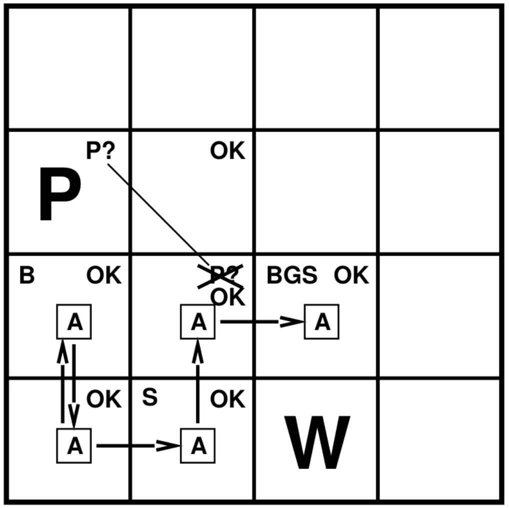


A: Agent; B: Breeze; G: Glitter; S: Stench
OK: safe square; W: Wumpus; P: Pit; BGS: bag of gold

Example: Exploring the Wumpus World

- Agent moves to [2,3]
- perceives a glitter

⟹ bag of gold!



A: Agent; B: Breeze; G: Glitter; S: Stench
OK: safe square; W: Wumpus; P: Pit; BGS: bag of gold

Example 2: Exploring the Wumpus World [see Ch 13]

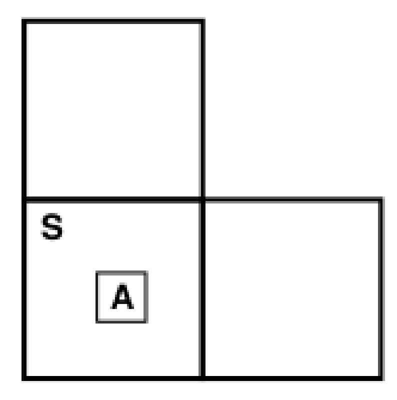
Alternative scenario: apply coercion

- Feel stench in [1,1]

$\Longrightarrow$ Wumpus [1,2] or [2,1]

$\Longrightarrow$ Cannot move

- Apply coercion: shoot ahead
  - Wumpus was there
    $\Longrightarrow$ Wumpus dead
    $\Longrightarrow$ Safe
  - Wumpus wasn't there
    $\Longrightarrow$ Safe



Example 3: Exploring the Wumpus World [see Ch. 13]

Alternative scenario: probabilistic solution (hints)

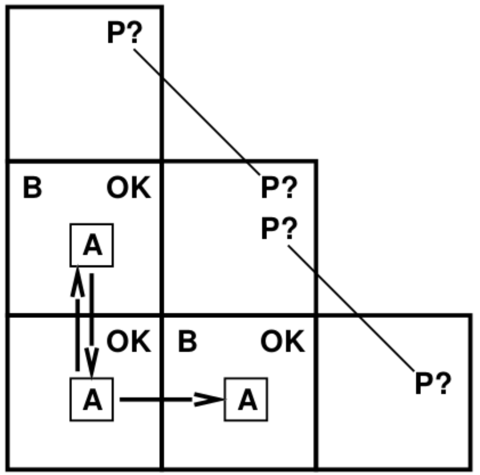
- Feel breeze in [1,2] and [2,1]

$\Longrightarrow$ pit in [1,3] or [2,2] or [3,1]

$\Longrightarrow$ no 100% safe action

- Probability analysis [see Ch 13] (assuming pits uniformly distributed):
  $P(pit \in [2,2]) = 0.86$
  $P(pit \in [1,3]) = 0.31$
  $P(pit \in [3,1]) = 0.31$

$\Longrightarrow$ better choose [1,3] or [3,1]



Outline

Outline

Propositional Logic Agents

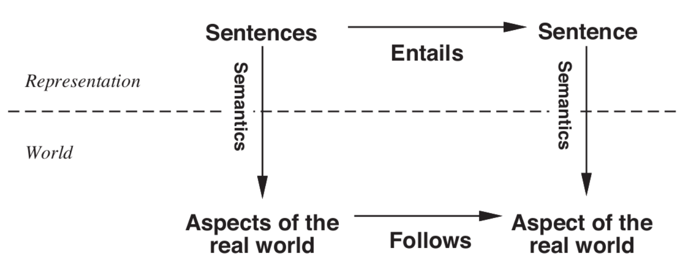
- Kind of Logic agents
- Language: propositional logic, first-order logic, ...
  - represent KB as set of propositional formulas
  - percepts and actions are (collections of) propositional atoms
  - in practice: sets of clauses
- Perform propositional logic inference
  - model checking, entailment
  - in practice: incremental calls to a SAT solver

Representation vs. World

Reasoning process (propositional entailment) sound
$\Longrightarrow$ if KB is true in the real world, then any sentence α derived from KB by a sound inference procedure is also true in the real world

- sentences are configurations of the agent
- reasoning constructs new configurations from old ones

$\Longrightarrow$ the new configurations represent aspects of the world that actually follow from the aspects that the old configurations represent



(© S. Russell & P. Norwig, AIMA)

Reasoning as Entailment

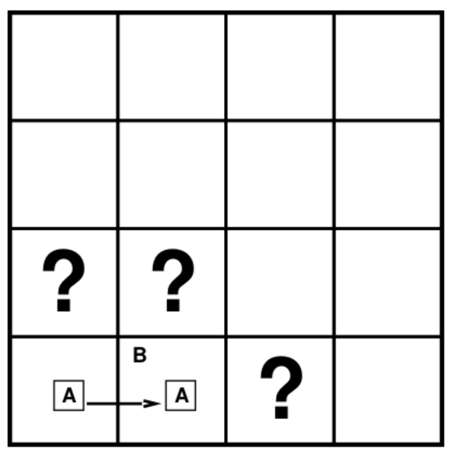
Scenario in Wumpus World

Consider pits (and breezes) only:

- initial: $\neg P_{[1,1]}$
- after detecting nothing in [1,1]: $\neg B_{[1,1]}$
- move to [2,1], detect breeze: $B_{[2,1]}$

Q: are there pits in [1,2], [2,1], [3,1]?

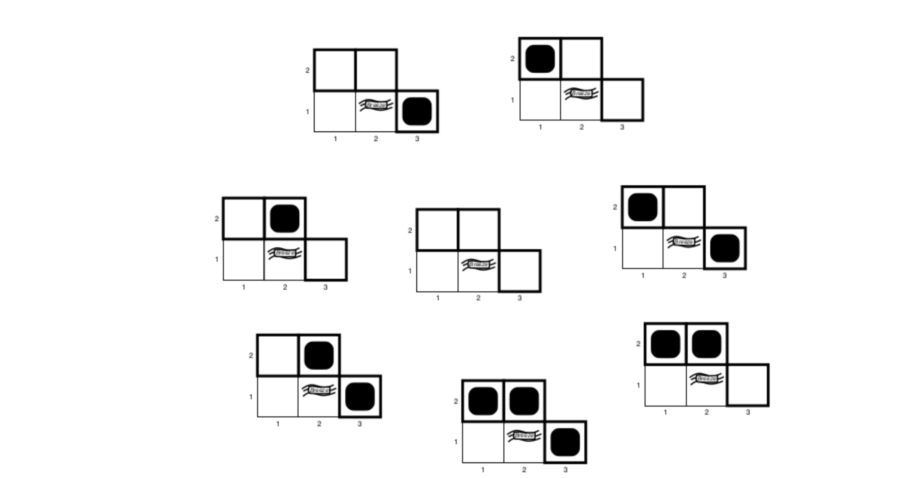
- 3 variables: $P_{[1,2]}, P_{[2,1]}, P_{[3,1]}$,
  $\Longrightarrow$ 8 possible models
  - Query α_1: $KB \models \neg P_{[1,2]}$?
  - Query α_2: $KB \models \neg P_{[2,1]}$?
  - Query α_3: $KB \models \neg P_{[3,1]}$?



A: Agent; B: Breeze; G: Glitter; S: Stench
OK: safe square; W: Wumpus; P: pit; BGS: bag of gold

Reasoning as Entailment [cont.]

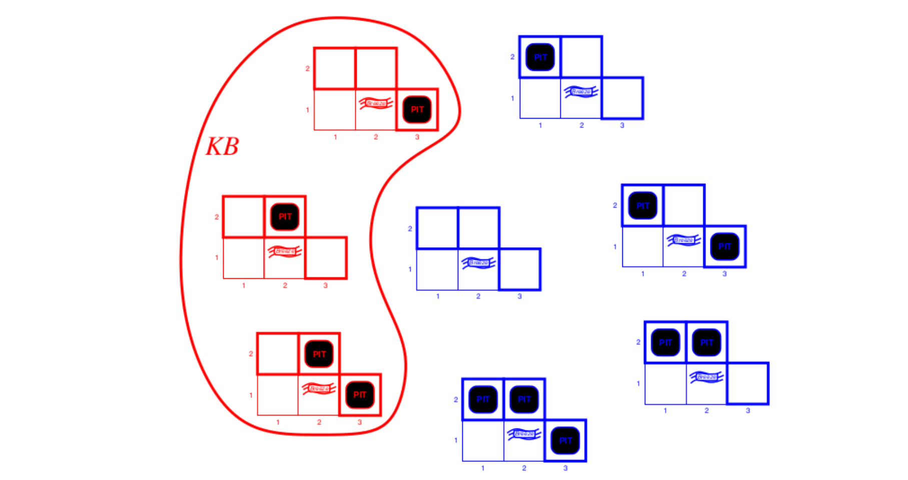
8 possible models



(© S. Russell & P. Norwig, AIMA)

Reasoning as Entailment [cont.]

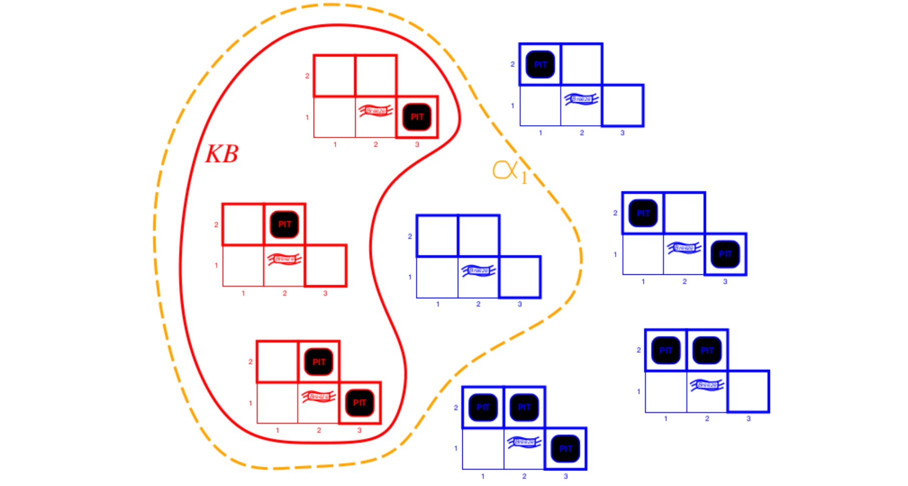
KB: Wumpus World rules + observations $\Longrightarrow$ 3 models



(© S. Russell & P. Norwig, AIMA)

Reasoning as Entailment [cont.]

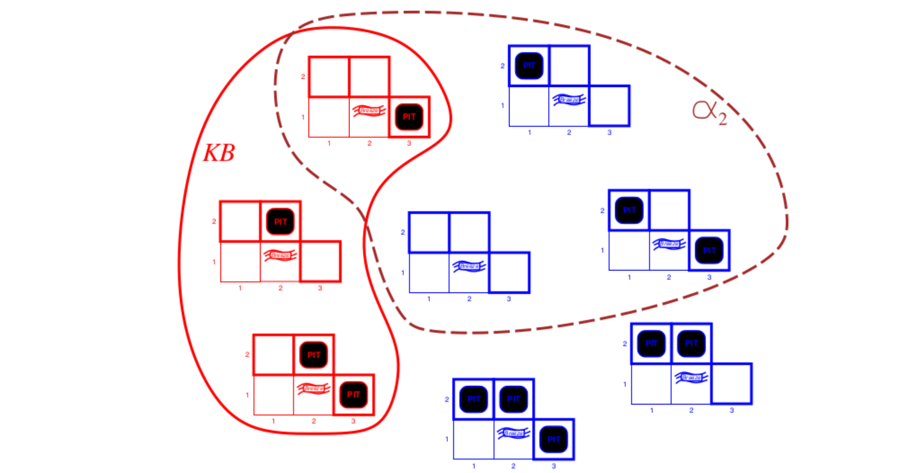
Query $\alpha_1 : \neg P_{[1,2]} \Longrightarrow KB \models \alpha_1$ (i.e $M(KB) \subseteq M(\alpha_1)$)



(© S. Russell & P. Norwig, AIMA)

Reasoning as Entailment [cont.]

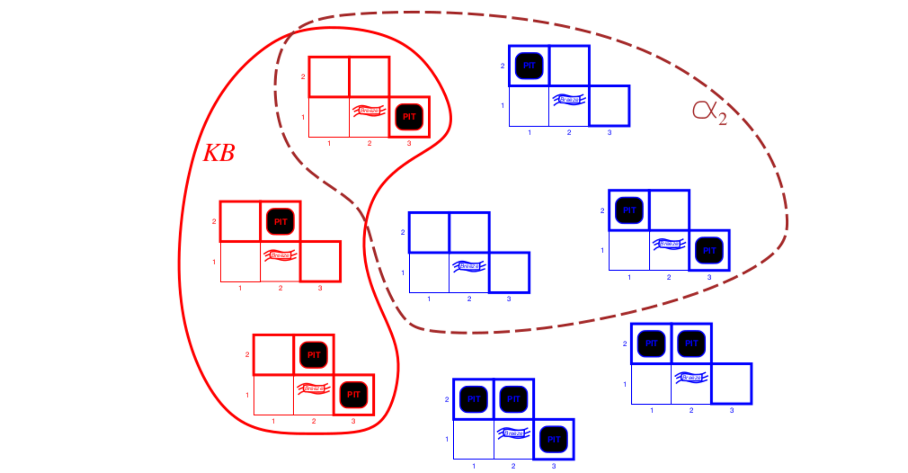
Query $\alpha_2 : \neg P_{[2,2]} \implies KB \not\models \alpha_2$ (i.e $M(KB) \not\subseteq M(\alpha_2)$)



(© S. Russell & P. Norwig, AIMA)

Reasoning as Entailment [cont.]

In practice: $DPLL(CNF(KB \wedge \neg\alpha_2)) = \text{sat}$



(© S. Russell & P. Norwig, AIMA)

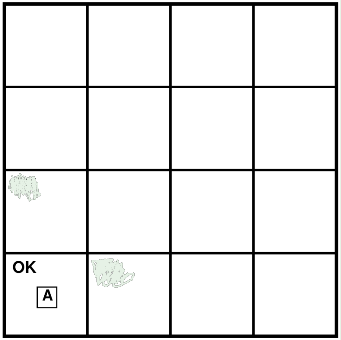
Outline

Example: Exploring the Wumpus World

KB initially contains (the CNFized versions of) the following formulas, $\forall i, j \in [1..4]$:

- breeze iff pit in neighbours
  $B_{[i,j]} \leftrightarrow (P_{[i,j-1]} \vee P_{[i+1,j]} \vee P_{[i,j+1]} \vee P_{[i-1,j]})$
- stench iff Wumpus in neighbours
  $S_{[i,j]} \leftrightarrow (W_{[i,j-1]} \vee W_{[i+1,j]} \vee W_{[i,j+1]} \vee W_{[i-1,j]})$
- safe iff no Wumpus and no pit there $OK_{[i,j]} \leftrightarrow (\neg W_{[i,j]} \wedge \neg P_{[i,j]})$
- glitter iff pile of gold there
  $G_{[i,j]} \leftrightarrow BGS_{[i,j]}$
- in $[1,1]$ no Wumpus and no pit $\Longrightarrow$ safe
  $\neg P_{[1,1]}, \neg W_{[1,1]}, OK_{[1,1]}$

(implicit: $P_{[i,j]}, W_{[i,j]}, P_{[i,j]}$ false if $i, j \notin [1..4]$)



A: Agent; B: Breeze; G: Glitter; S: Stench
OK: safe square; W: Wumpus; P: pit; BGS: bag of gold

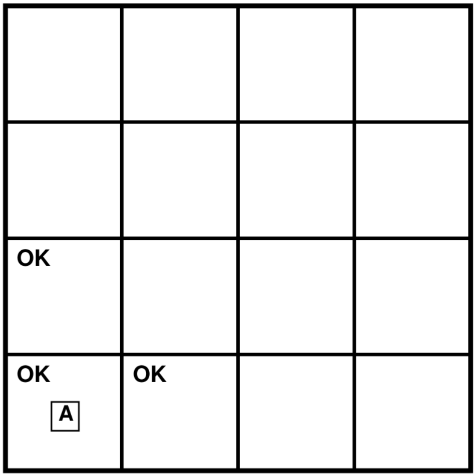
Example: Exploring the Wumpus World

- KB initially contains:
  $\neg P_{[1,1]}, \neg W_{[1,1]}, OK_{[1,1]}$
  $B_{[1,1]} \leftrightarrow (P_{[1,2]} \vee P_{[2,1]})$
  $S_{[1,1]} \leftrightarrow (W_{[1,2]} \vee W_{[2,1]})$
  $OK_{[1,2]} \leftrightarrow (\neg W_{[1,2]} \wedge \neg P_{[1,2]})$
  $OK_{[2,1]} \leftrightarrow (\neg W_{[2,1]} \wedge \neg P_{[2,1]})$
  ...
- Agent is initially in 1,1
- Percepts (no stench, no breeze): $\neg S_{[1,1]}, \neg B_{[1,1]}$

$\Longrightarrow$ $\neg W_{[1,2]}, \neg W_{[2,1]}, \neg P_{[1,2]}, \neg P_{[2,1]}$

$\Longrightarrow$ $OK_{[1,2]}, OK_{[2,1]}$ ([1,2] & [2,1] OK)

- Add all them to KB



A: Agent; B: Breeze; G: Glitter; S: Stench
OK: safe square; W: Wumpus; P: pit; BGS: glitter, bag of gold

Example: Exploring the Wumpus World

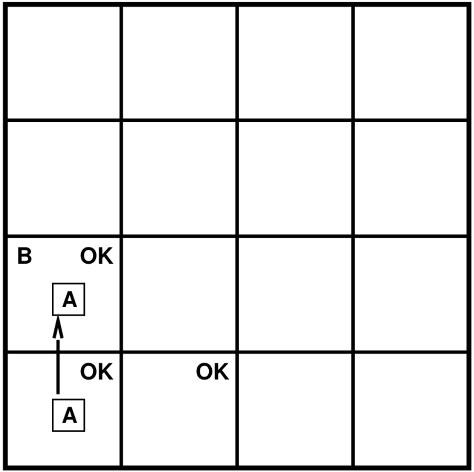
- KB initially contains:
  $\neg P_{[1,1]}, \neg W_{[1,1]}, OK_{[1,1]}$
  $B_{[2,1]} \leftrightarrow (P_{[1,1]} \vee P_{[2,2]} \vee P_{[3,1]})$
  $S_{[2,1]} \leftrightarrow (W_{[1,1]} \vee W_{[2,2]} \vee W_{[3,1]})$
  ...
- Agent moves to [2,1]
- perceives a breeze: $B_{[2,1]}$

$\Longrightarrow$ $(P_{[3,1]} \vee P_{[2,2]})$ (pit in [3,1] or [2,2])

- perceives no stench $\neg S_{[2,1]}$

$\Longrightarrow$ $\neg W_{[3,1]}, \neg W_{[2,2]}$
(no Wumpus in [3,1], [2,2])

- Add all them to KB



A: Agent; B: Breeze; G: Glitter; S: Stench
OK: safe square; W: Wumpus; P: pit; BGS: glitter, bag of gold

Example: Exploring the Wumpus World

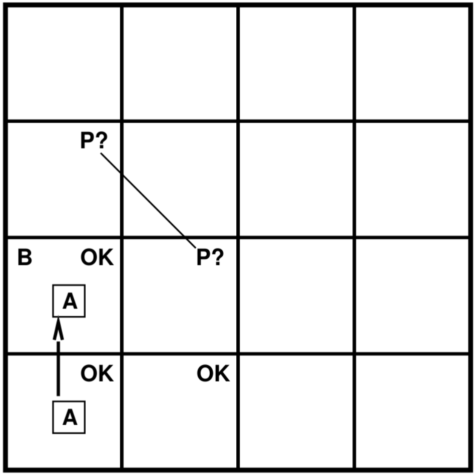
- KB initially contains:
  $\neg P_{[1,1]}, \neg W_{[1,1]}, OK_{[1,1]}$
  $B_{[2,1]} \leftrightarrow (P_{[1,1]} \vee P_{[2,2]} \vee P_{[3,1]})$
  $S_{[2,1]} \leftrightarrow (W_{[1,1]} \vee W_{[2,2]} \vee W_{[3,1]})$
  ...
- Agent moves to [2,1]
- perceives a breeze: $B_{[2,1]}$

$\Longrightarrow$ $(P_{[3,1]} \vee P_{[2,2]})$ (pit in [3,1] or [2,2])

- perceives no stench $\neg S_{[2,1]}$

$\Longrightarrow$ $\neg W_{[3,1]}, \neg W_{[2,2]}$
(no Wumpus in [3,1], [2,2])

- Add all them to KB



A: Agent; B: Breeze; G: Glitter; S: Stench
OK: safe square; W: Wumpus; P: pit; BGS: glitter, bag of gold

Example: Exploring the Wumpus World

- KB initially contains:
  $\neg P_{[1,1]}, \neg W_{[1,1]}, OK_{[1,1]}$
  $(P_{[3,1]} \lor P_{[2,2]}), \neg W_{[3,1]}, \neg W_{[2,2]}$
  $B_{[1,2]} \leftrightarrow (P_{[1,1]} \lor P_{[2,2]} \lor P_{[1,3]})$
  $S_{[1,2]} \leftrightarrow (W_{[1,1]} \lor W_{[2,2]} \lor W_{[1,3]})$
  $OK_{[2,2]} \leftrightarrow (\neg W_{[2,2]} \land \neg P_{[2,2]})$
- Agent moves to [1,1]-[1,2]
- perceives no breeze: $\neg B_{[1,2]}$

$\Longrightarrow$ $\neg P_{[2,2]}, \neg P_{[1,3]}$ (no pit in [2,2], [1,3])

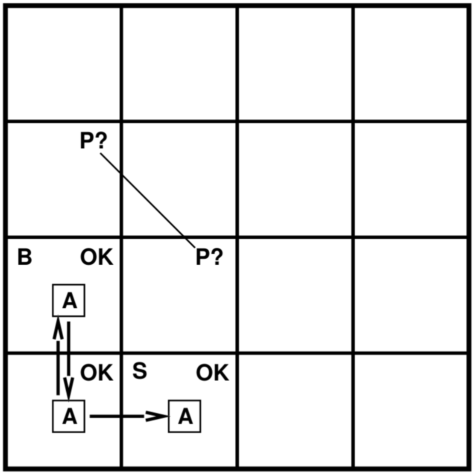
$\Longrightarrow$ $P_{[3,1]}$ (pit in [3,1])

- perceives a stench: $S_{[1,2]}$

$\Longrightarrow$ $W_{[1,3]}$ (Wumpus in [1,3]!)

$\Longrightarrow$ $OK_{[2,2]}$ ([2,2] OK)

- Add all them to KB



A: Agent; B: Breeze; G: Glitter; S: Stench
OK: safe square; W: Wumpus; P: pit; BGS: glitter, bag of gold

Example: Exploring the Wumpus World

- KB initially contains:
  $\neg P_{[1,1]}, \neg W_{[1,1]}, OK_{[1,1]}$
  $(P_{[3,1]} \vee P_{[2,2]}), \neg W_{[3,1]}, \neg W_{[2,2]}$
  $B_{[1,2]} \leftrightarrow (P_{[1,1]} \vee P_{[2,2]} \vee P_{[1,3]})$
  $S_{[1,2]} \leftrightarrow (W_{[1,1]} \vee W_{[2,2]} \vee W_{[1,3]})$
  $OK_{[2,2]} \leftrightarrow (\neg W_{[2,2]} \wedge \neg P_{[2,2]})$
- Agent moves to [1,1]-[1,2]
- perceives no breeze: $\neg B_{[1,2]}$

$\Longrightarrow$ $\neg P_{[2,2]}, \neg P_{[1,3]}$ (no pit in [2,2], [1,3])

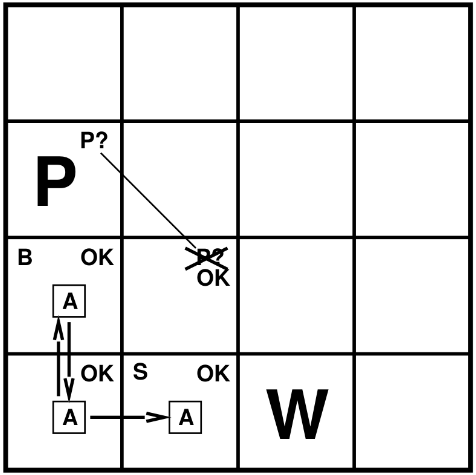
$\Longrightarrow$ $P_{[3,1]}$ (pit in [3,1])

- perceives a stench: $S_{[1,2]}$

$\Longrightarrow$ $W_{[1,3]}$ (Wumpus in [1,3]!)

$\Longrightarrow$ $OK_{[2,2]}$ ([2,2] OK)

- Add all them to KB



A: Agent; B: Breeze; G: Glitter; S: Stench
OK: safe square; W: Wumpus; P: pit; BGS: glitter, bag of gold

Example: Exploring the Wumpus World

- KB initially contains:
  $B_{[2,2]} \leftrightarrow (P_{[2,1]} \vee P_{[3,2]} \vee P_{[2,3]} \vee P_{[1,2]})$
  $S_{[2,2]} \leftrightarrow (W_{[2,1]} \vee W_{[3,2]} \vee W_{[2,3]} \vee W_{[1,2]})$
  $OK_{[3,2]} \leftrightarrow (\neg W_{[3,2]} \wedge \neg P_{[3,2]})$
  $OK_{[2,3]} \leftrightarrow (\neg W_{[2,3]} \wedge \neg P_{[2,3]})$
- Agent moves to [2,2]
- perceives no breeze: $\neg B_{[2,2]}$

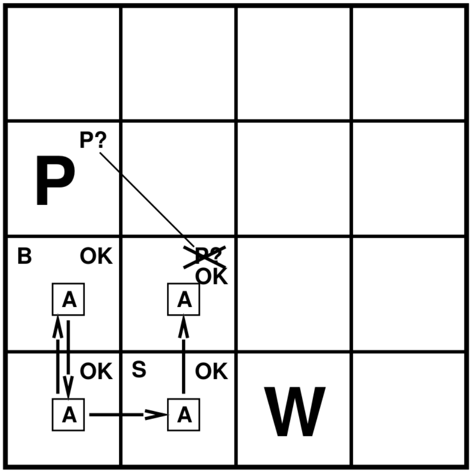
$\Longrightarrow$ $\neg P_{[3,2]}, \neg P_{[2,3]}$ (no pit in [3,2], [2,3])

- perceives no stench: $\neg S_{[2,2]}$

$\Longrightarrow$ $\neg W_{[3,2]}, \neg W_{[3,2]}$ (no Wumpus in [3,2], [2,3])

$\Longrightarrow$ $OK_{[3,2]}, OK_{[2,3]}$, ([3,2] and [2,3] OK)

- Add all them to KB



A: Agent; B: Breeze; G: Glitter; S: Stench
OK: safe square; W: Wumpus; P: pit; BGS: glitter, bag of gold

Example: Exploring the Wumpus World

- KB initially contains:
  $B_{[2,2]} \leftrightarrow (P_{[2,1]} \vee P_{[3,2]} \vee P_{[2,3]} \vee P_{[1,2]})$
  $S_{[2,2]} \leftrightarrow (W_{[2,1]} \vee W_{[3,2]} \vee W_{[2,3]} \vee W_{[1,2]})$
  $OK_{[3,2]} \leftrightarrow (\neg W_{[3,2]} \wedge \neg P_{[3,2]})$
  $OK_{[2,3]} \leftrightarrow (\neg W_{[2,3]} \wedge \neg P_{[2,3]})$
- Agent moves to [2,2]
- perceives no breeze: $\neg B_{[2,2]}$

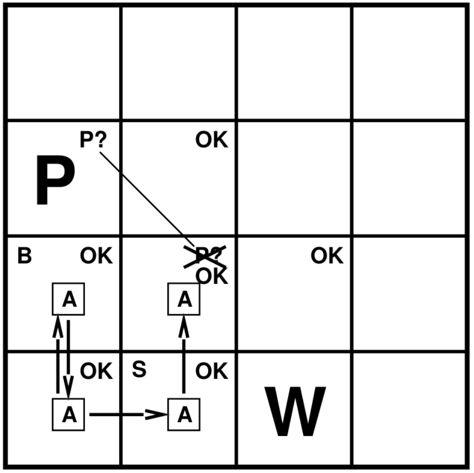
$\Longrightarrow$ $\neg P_{[3,2]}, \neg P_{[2,3]}$ (no pit in [3,2], [2,3])

- perceives no stench: $\neg S_{[2,2]}$

$\Longrightarrow$ $\neg W_{[3,2]}, \neg W_{[3,2]}$ (no Wumpus in [3,2], [2,3])

$\Longrightarrow$ $OK_{[3,2]}, OK_{[2,3]}$, ([3,2] and [2,3] OK)

- Add all them to KB



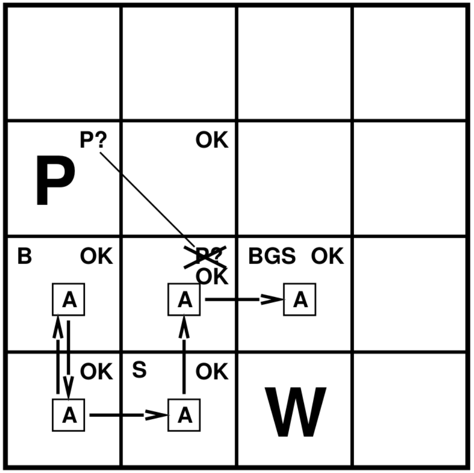
A: Agent; B: Breeze; G: Glitter; S: Stench
OK: safe square; W: Wumpus; P: pit; BGS: glitter, bag of gold

Example: Exploring the Wumpus World

- KB initially contains:
  $G_{[2,3]} \leftrightarrow BGS_{[2,3]}$
- Agent moves to [2,3]
- perceives a glitter: $G_{[2,3]}$

$\Longrightarrow$ $BGS_{[2,3]}$ (bag of gold!)

- Add it them to KB



A: Agent; B: Breeze; G: Glitter; S: Stench
OK: safe square; W: Wumpus; P: pit; BGS: glitter, bag of gold

Exercise

Consider the previous example.

1. Convert all formulas from KB into CNF
2. Execute all steps in the example as resolution calls
3. Execute all steps in the example as DPLL calls