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

Started: January 1996
Launch: October 15th, 1998
Experiment: May 17-21
Muddy Children Problem

- Mom to N children “Don’t get dirty”
- While playing, $K \geq 1$ 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: ∧, ∨, ¬, ⇒

• **Semantics**
  – Truth Tables

• **Inference**
  – Modus Ponens
  – Resolution
  – DPLL
  – GSAT
Propositional Logic: Syntax

• **Atoms**
  – P, Q, R, ...

• **Literals**
  – P, ¬P

• **Sentences**
  – Any literal is a sentence
  – If S is a sentence
    • Then (S ∧ S) is a sentence
    • Then (S ∨ S) is a sentence

• **Conveniences**
  
P → Q same as ¬P ∨ Q
Truth tables for connectives

<table>
<thead>
<tr>
<th>$P$</th>
<th>$Q$</th>
<th>$\neg P$</th>
<th>$P \land Q$</th>
<th>$P \lor Q$</th>
<th>$P \Rightarrow Q$</th>
<th>$P \Leftrightarrow Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>false</td>
<td>false</td>
<td>true</td>
<td>false</td>
<td>false</td>
<td>true</td>
<td>true</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>true</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>true</td>
</tr>
<tr>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
</tr>
</tbody>
</table>
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.

\[ \neg R \lor H \land \neg I \lor H \land \neg M \lor I \]

\[ (\neg R \lor H) \land (\neg I \lor H) \land (\neg M \lor I) \]

M = mythical
I = immortal
R = reptile
H = horned
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"
\[ B_{1,1} \iff (P_{1,2} \lor P_{2,1}) \]
\[ B_{2,1} \iff (P_{1,1} \lor P_{2,2} \lor P_{3,1}) \]
Full Encoding of Wumpus World

In propositional logic:

\[ \neg P_{1,1} \]
\[ \neg W_{1,1} \]
\[ B_{x,y} \iff (P_{x,y+1} \lor P_{x,y-1} \lor P_{x+1,y} \lor P_{x-1,y}) \]
\[ S_{x,y} \iff (W_{x,y+1} \lor W_{x,y-1} \lor W_{x+1,y} \lor W_{x-1,y}) \]
\[ W_{1,1} \lor W_{1,2} \lor \ldots \lor W_{4,4} \]
\[ \neg W_{1,1} \lor \neg W_{1,2} \]
\[ \neg W_{1,1} \lor \neg W_{1,3} \]
\[ \ldots \]

\[ \Rightarrow \quad 64 \text{ distinct proposition symbols, 155 sentences} \]
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=\{-p_{1,1}, \neg w_{1,1}, \neg b_{1,1}, \neg g_{1,1},
-p_{1,1}, \neg w_{1,1}, b_{1,1}, \neg g_{1,1},
\ldots
b_{1,1} \iff (p_{1,2} \lor p_{2,1})
\ldots \} \}

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:

\[ \text{KB}=\left\{ \neg P_{1,1}, \neg W_{1,1}, \neg B_{1,1}, \neg G_{1,1}, \right. \]
\[ \left. \neg P_{1,2}, \neg W_{1,1}, B_{1,1}, \neg G_{1,1}, \right. \]
\[ \left. \cdots \right. \]
\[ B_{1,1} \leftrightarrow (P_{1,2} \lor P_{2,1}) \]
\[ \left. \cdots \right. \]}
Wumpus Models

Models that are consistent with our KB:

\[ KB = \{ \neg P_{1,1}, \neg W_{1,1}, \neg B_{1,1}, \neg G_{1,1}, \]
\[ \neg P_{2,1}, \neg W_{2,1}, B_{1,1}, \neg G_{1,1}, \]
\[ \ldots \}
\]

\[ B_{1,1} \iff (P_{1,2} \lor P_{2,1}) \]
\[ \ldots \}

\[ KB = \text{wumpus-world rules} + \text{observations} \]
Wumpus Models

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

\[ KB = \{ \neg P_{1,1}, \neg W_{1,1}, \neg B_{1,1}, \neg G_{1,1}, \neg P_{1,2}, \neg W_{1,2}, B_{1,2}, \neg G_{1,2}, \ldots \} \]

\[ B_{1,1} \equiv (P_{1,2} \lor P_{2,1}) \]

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

\[ KB = \text{wumpus-world rules } + \text{ observations} \]

\[ \alpha_1 = \text{"[1,2] is safe"}, \quad KB \models \alpha_1, \text{ proved by model checking} \]
Wumpus Models

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

\[
\text{KB} = \{ \neg P_{1,1}, \neg W_{1,1}, \neg B_{1,1}, \neg G_{1,1}, \neg P_{1,2}, \neg W_{1,2}, B_{1,2}, \neg G_{1,2}, \ldots \} \\
B_{1,1} \iff (P_{1,2} \lor P_{2,1}) \\
\ldots \} \\
\alpha_2 = \neg P_{2,2} \land \neg W_{2,2}
\]

\[
\text{KB} = \text{wumpus-world rules} + \text{observations} \\
\alpha_2 = \text{“}[2,2] \text{ is safe”}, \ KB \not\models \alpha_2
\]
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 $\alpha$ if $\alpha$ is true in $m$

- $M(\alpha)$ is the set of all models of $\alpha$

- Entailment: $KB \models \alpha$ iff $M(KB) \subseteq M(\alpha)$
  - E.g. $KB = (P \lor Q) \land (\neg P \lor R)$
    $\alpha = (P \lor R)$

- How to check?
  - One way is to enumerate all elements in the truth table – slow 😞
Inference

$KB \vdash_i \alpha = \text{sentence } \alpha \text{ can be derived from } KB \text{ by procedure } i$

Consequences of $KB$ are a haystack; $\alpha$ is a needle. Entailment = needle in haystack; inference = finding it

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

**Completeness**: $i$ is complete if
evertheless $KB \models \alpha$, 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

<table>
<thead>
<tr>
<th>$B_{1,1}$</th>
<th>$B_{2,1}$</th>
<th>$P_{1,1}$</th>
<th>$P_{1,2}$</th>
<th>$P_{2,1}$</th>
<th>$P_{2,2}$</th>
<th>$P_{3,1}$</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$R_3$</th>
<th>$R_4$</th>
<th>$R_5$</th>
<th>$KB$</th>
</tr>
</thead>
<tbody>
<tr>
<td>false</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>false</td>
<td>false</td>
</tr>
<tr>
<td>false</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>true</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>false</td>
<td>false</td>
<td>false</td>
<td>true</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>true</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>true</td>
</tr>
<tr>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>false</td>
<td>true</td>
<td>true</td>
<td>true</td>
</tr>
</tbody>
</table>

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

**Problem:** exponential time and space!
Logical Equivalence

Two sentences are **logically equivalent** iff true in same models:
\[ \alpha \equiv \beta \text{ if and only if } \alpha \models \beta \text{ and } \beta \models \alpha \]

\[
\begin{align*}
(\alpha \land \beta) & \equiv (\beta \land \alpha) \quad \text{commutativity of } \land \\
(\alpha \lor \beta) & \equiv (\beta \lor \alpha) \quad \text{commutativity of } \lor \\
((\alpha \land \beta) \land \gamma) & \equiv (\alpha \land (\beta \land \gamma)) \quad \text{associativity of } \land \\
((\alpha \lor \beta) \lor \gamma) & \equiv (\alpha \lor (\beta \lor \gamma)) \quad \text{associativity of } \lor \\
\neg(\neg \alpha) & \equiv \alpha \quad \text{double-negation elimination} \\
(\alpha \Rightarrow \beta) & \equiv (\neg \beta \Rightarrow \neg \alpha) \quad \text{contraposition} \\
(\alpha \Rightarrow \beta) & \equiv (\neg \alpha \lor \beta) \quad \text{implication elimination} \\
(\alpha \leftrightarrow \beta) & \equiv ((\alpha \Rightarrow \beta) \land (\beta \Rightarrow \alpha)) \quad \text{biconditional elimination} \\
\neg(\alpha \land \beta) & \equiv (\neg \alpha \lor \neg \beta) \quad \text{De Morgan} \\
\neg(\alpha \lor \beta) & \equiv (\neg \alpha \land \neg \beta) \quad \text{De Morgan} \\
(\alpha \land (\beta \lor \gamma)) & \equiv ((\alpha \land \beta) \lor (\alpha \land \gamma)) \quad \text{distributivity of } \land \text{ over } \lor \\
(\alpha \lor (\beta \land \gamma)) & \equiv ((\alpha \lor \beta) \land (\alpha \lor \gamma)) \quad \text{distributivity of } \lor \text{ over } \land
\end{align*}
\]
Validity and Satisfiability

A sentence is valid if it is true in all models,
e.g., $\text{True}$, $A \lor \neg A$, $A \Rightarrow A$, $(A \land (A \Rightarrow B)) \Rightarrow B$

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 $\alpha$ 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