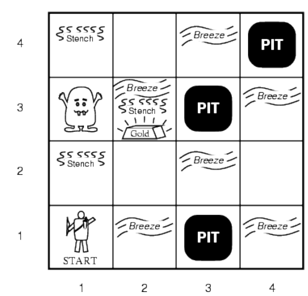
Knowledge-Based Agents and Propositional Logic

AI Class 18 (Ch. 7)



Material from Dr. Marie desJardin, Some material adopted from notes by Andreas Geyer-Schulz and Chuck Dyer

Bookkeeping

- Midterms returned Thursday
- Project designs last night
  - Should be returned Tuesday
  - Next deliverable: 11/16 @ 11:59
- HW4 due 11/7 @ 11:59

Today's Class

- Knowledge Based Agents
  - Knowledge Bases
  - Inference
- Inferential Logics
  - Propositional Logic: a Refresher
  - First-Order Logic (FOL)
- The Wumpus

Knowledge-Based Agents (Logical Agents)

Chapter 7.1-7.3

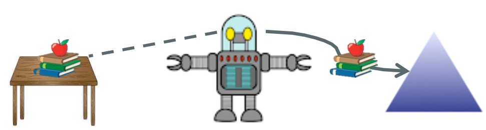
A Knowledge-Based Agent

- A knowledge-based agent needs (at least):
  - A **knowledge base**
  - An **inference system**
- A knowledge base (KB) is a set of representations of facts about the world.
  - Each individual representation is a **sentence** or **assertion**
  - Expressed in a **knowledge representation language**
  - Usually starts with some background knowledge
    - Can be general (world knowledge) or specific (domain language)
- Many existing ideas apply – is it closed-world, etc.

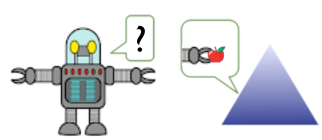
A Knowledge-Based Agent

- Operates as follows:

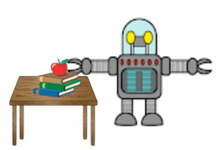
1. TELLs the knowledge base what it perceives.



2. ASKs the knowledge base what action to perform.



3. Performs the chosen action.



Architecture of a Knowledge-Based Agent

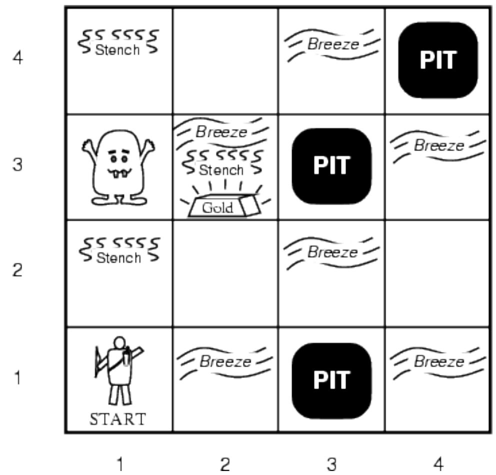
- **Knowledge Level**
  - The most abstract level
  - Describe agent by saying what it knows
    - Example: A taxi agent might know that the Golden Gate Bridge connects San Francisco with the Marin County.
- **Logical Level**
  - Level at which **knowledge** is encoded into **sentences**.
    - Example: Links(GoldenGateBridge, SanFrancisco, MarinCounty)
- **Implementation Level**
  - The physical representation of the sentences in the logical level.
    - Example: '(links goldengatebridge sanfrancisco marincounty)'

The Wumpus World Environment

- The Wumpus computer game
  - Agent explores a cave consisting of rooms connected by passageways.
  - Lurking somewhere in the cave is the Wumpus, a beast that eats any agent that enters its room.
  - Some rooms contain bottomless pits that trap any agent that wanders into the room.
  - Occasionally, there is a heap of gold in a room.
  - The goal is to collect the gold and exit the world without being eaten (or trapped).

A Typical Wumpus World

- The agent always starts in the field [1,1].
- The task of the agent is to find the gold, return to the field [1,1] and climb out of the cave.



Agent in a Wumpus World: Percepts

- Agent perceives
  - **Stench** in the square containing the wumpus and in adjacent squares (not diagonally)
  - **Breeze** in the squares adjacent to a pit
  - **Glitter** in the square where the gold is
  - **Bump**, if it walks into a wall
  - **Woeful** scream everywhere in the cave, if the wumpus is killed
- The percepts are given as a five-symbol list.
- If there is a stench and a breeze, but no glitter, no bump, and no scream, the percept is:

  [Stench, Breeze, None, None, None]
- The agent cannot perceive its own location

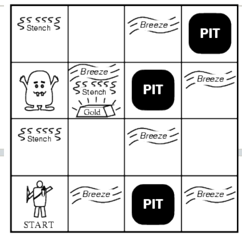
Wumpus Agent Actions

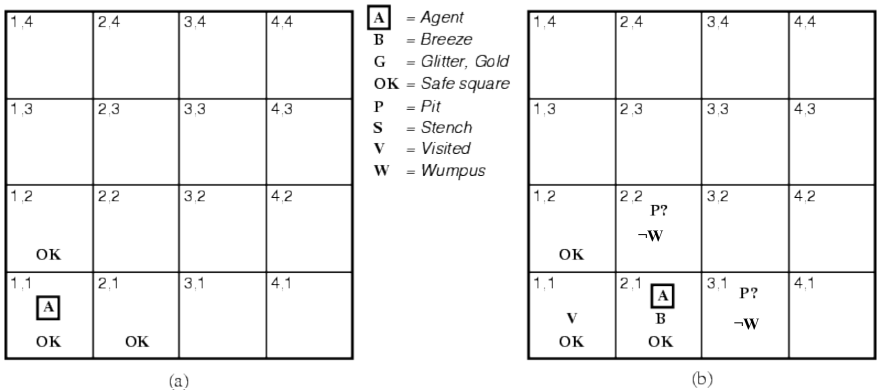
- **go forward**
- **turn right** 90 degrees
- **turn left** 90 degrees
- **grab**: Pick up an object that is in the same square as the agent
- **shoot**: Fire an arrow in a straight line in the direction the agent is facing.
  - The arrow continues until it either hits and kills the wumpus or hits the outer wall.
  - The agent has only one arrow, so only the first Shoot action has any effect
- **climb**: leave the cave. This action is only effective in the start square
- **die**: This action automatically happens if the agent enters a square with a pit or a live wumpus

Wumpus Goal

- Agent's goal is to:
  - Find the gold
  - Bring it back to the start square as quickly as possible
  - Don't get killed
- Scoring
  - 1000 points reward for climbing out with the gold
  - 1 point deducted for every action taken
  - 10000 points penalty for getting killed

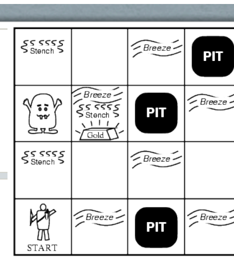
Wumpus Agent's First Step

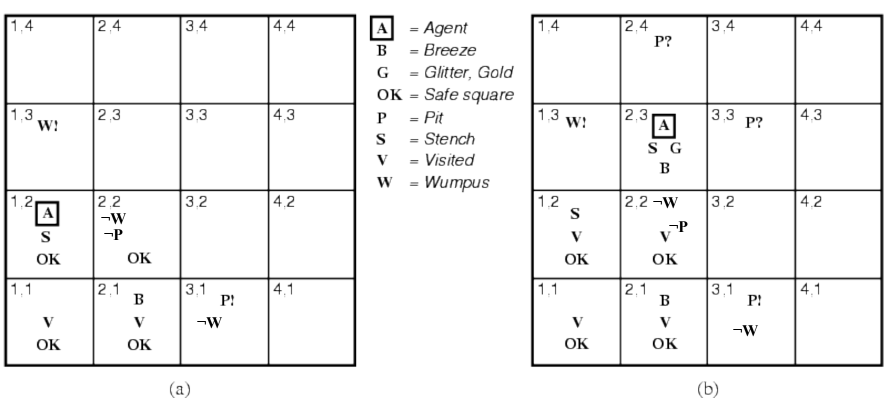




Percepts: [None, None, None, None, None]   Percepts: [None, Breeze, None, None, None]

Later





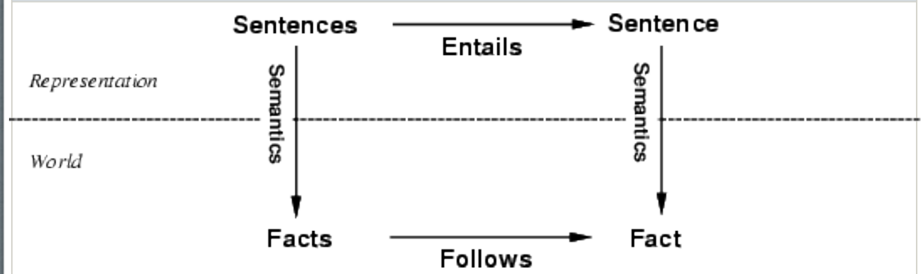
Wumpuses Online

- http://www.cs.berkeley.edu/~russell/code/doc/overview-AGENTS.html
  - Lisp version from Russell & Norvig
- http://www.dreamcodex.com/wumpus.php – Java-based version you can play online
- http://codenautics.com/wumpus/ – Downloadable Mac version

Representation, Reasoning, and Logic

- Point of knowledge representation is to express knowledge in a **computer usable** form
- Needed for agents to act on it (to do well, anyway)
- A knowledge representation language is defined by:
  - **Syntax**: all possible sequences of symbols that form sentences
    - Example: noun referents can be a single word or an adjective-then-noun
  - **Semantics:** facts in the world to which the sentences refer
    - What does it *mean*?
- Each sentence makes a claim about the world
- An agent is said to "believe" a sentence about the world

The Connection Between Sentences and Facts



Semantics maps sentences in logic to facts in the world.
The property of one fact following from another is mirrored by the property of one sentence **being entailed** by another.

"Dr M is sick with the flu" ⊨ "Dr M is sick"

Entailment and Derivation

- **Entailment: KB ⊨ Q**

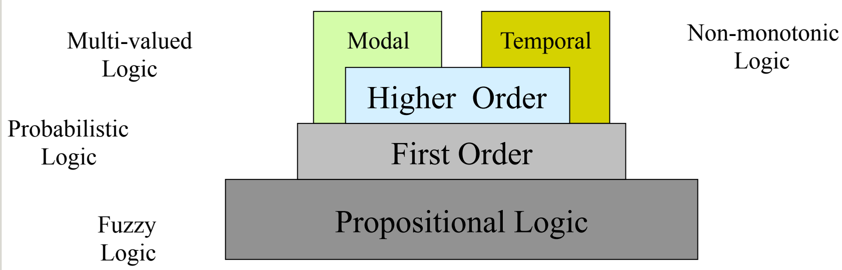
  x ⊨ y: x semantically entails y

  - Q is entailed by KB (a set of premises or assumptions) if and only if there is no logically possible world in which Q is false while all the premises in KB are true.
  - Or, stated positively, Q is entailed by KB if and only if the conclusion is true in every logically possible world in which all the premises in KB are true.
- **Derivation: KB ⊢ Q**

  x ⊢ y: y is provable from x

  - We can derive Q from KB if there is a proof consisting of a sequence of valid inference steps starting from the premises in KB and resulting in Q

Logic as a KR Language



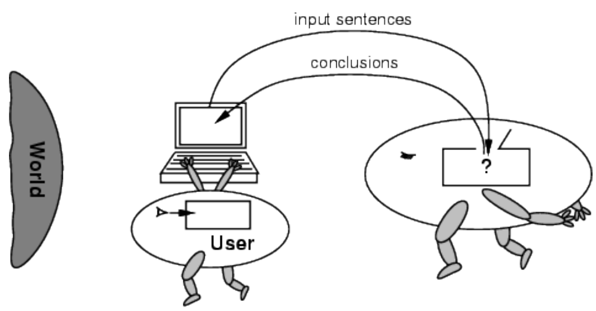
Ontology and Epistemology

• **Ontology** is the study of what there is—an inventory of what exists. An ontological commitment is a commitment to an existence claim.
• **Epistemology** is a major branch of philosophy that concerns the forms, nature, and preconditions of knowledge.

Language	Ontological Commitment (What exists in the world)	Epistemological Commitment (What an agent believes about facts)
Propositional logic	facts	true/false/unknown
First-order logic	facts, objects, relations	true/false/unknown
Temporal logic	facts, objects, relations, times	true/false/unknown
Probability theory	facts	degree of belief 0…1
Fuzzy logic	degree of truth	degree of belief 0…1

No Independent World Access

- The reasoning agent often gets its knowledge about the facts of the world as a *sequence of logical sentences.*
- Must draw conclusions from them without (other) access to the world.
- Thus it is very important that the agent's reasoning is sound!

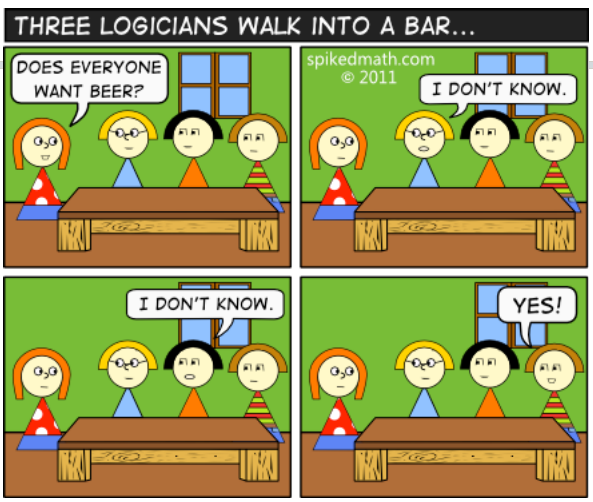


KB Agents - Summary

- Intelligent agents need **knowledge about the world** for making good decisions.
- The knowledge of an agent is stored in a knowledge base in the form of **sentences** in a **knowledge representation language**.
- A knowledge-based agent needs a **knowledge base** and an **inference mechanism**. It operates by storing sentences in its knowledge base, inferring new sentences with the inference mechanism, and using them to deduce which actions to take.
- A **representation language** is defined by its syntax and semantics, which specify structure of sentences and how they relate to world facts.
- The **interpretation** of a sentence is the fact to which it refers. If this fact is part of the actual world, then the sentence is true.

Propositional Logic

Chapter 7.4-7.8



Propositional Logic

- **Logical constants**: true, false
- **Propositional symbols**: P, Q, S, ... (**atomic sentences**)
- Wrapping **parentheses**: (…)
- Sentences are combined by **connectives**:

$\wedge$ ...and [conjunction]
$\vee$ ...or [disjunction]
$\Rightarrow$...implies [implication / conditional]
$\Leftrightarrow$..is equivalent [biconditional]
$\neg$ ...not [negation]

- **Literal**: atomic sentence or negated atomic sentence

Examples of PL Sentences

- $(P \wedge Q) \rightarrow R$

  "If it is hot and humid, then it is raining"
- $Q \rightarrow P$

  "If it is humid, then it is hot"
- Q

  "It is humid."
- A better way:
  Ho = "It is hot"
  Hu = "It is humid"
  R = "It is raining"

Propositional Logic (PL)

- A simple language useful for showing key ideas and definitions
- User defines a set of propositional symbols, like P and Q.
- User defines the **semantics** of each propositional symbol:
  - Ho means "It is hot"
  - Hu means "It is humid"
  - R means "It is raining"
- A sentence (well formed formula) is defined as follows:
  - A symbol is a sentence
  - If S is a sentence, then $\neg S$ is a sentence
  - If S is a sentence, then (S) is a sentence
  - If S and T are sentences, then $S \vee T$, $S \wedge T$, $S \rightarrow T$, and $S \leftrightarrow T$ are sentences
  - A sentence results from a finite number of applications of the above rules

Some Terms

- The meaning or **semantics** of a sentence determines its **interpretation**.
- Given the truth values of all symbols in a sentence, it can be "evaluated" to determine its **truth value** (True or False).
- A **model** for a KB is a "possible world" (assignment of truth values to propositional symbols) in which each sentence in the KB is True.
  - E.g.: it is both hot and humid.

More Terms

- A **valid sentence** or **tautology** is a sentence that is True under all interpretations, no matter what the world is actually like or what the semantics is. Example: "It's raining or it's not raining."
- An **inconsistent sentence** or **contradiction** is a sentence that is False under all interpretations. The world is never like what it describes, as in "It's raining and it's not raining."
- **P entails Q**, written P |= Q, means that whenever P is True, so is Q. In other words, all models of P are also models of Q.

Truth Tables

And

p	q	$p \cdot q$
T	T	T
T	F	F
F	T	F
F	F	F

Or

p	q	$p \vee q$
T	T	T
T	F	T
F	T	T
F	F	F

If . . . then

p	q	$p \supset q$
T	T	T
T	F	F
F	T	T
F	F	T

Not

p	$\sim p$
T	F
F	T

Truth Tables II

The five logical 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 complex sentence:

P	H	$P \vee H$	$(P \vee H) \wedge \neg H$	$((P \vee H) \wedge \neg H) \Rightarrow P$
False	*False*	*False*	*False*	*True*
False	*True*	*True*	*False*	*True*
True	*False*	*True*	*True*	*True*
True	*True*	*True*	*False*	*True*

Inference Rules

- **Logical inference** is used to create new sentences that logically follow from a given set of predicate calculus sentences (KB).
- An inference **rule** is **sound** if every sentence X produced by an inference rule operating on a KB logically follows from the KB. (That is, the inference rule does not create any contradictions)
- An inference rule is **complete** if it is able to produce every expression that logically follows from (is entailed by) the KB. (Note the analogy to complete search algorithms.)

Two Important Properties for Inference

- **Soundness: If KB ⊢ Q then KB ⊨ Q**
  - If Q is derived from a set of sentences KB using a given set of rules of inference, then Q is entailed by KB.
  - Hence, inference produces only real entailments, or any sentence that follows deductively from the premises is valid.

Completeness: If KB ⊨ Q then KB ⊢ Q
  - If Q is entailed by a set of sentences KB, then Q can be derived from KB using the rules of inference.
  - Hence, inference produces all entailments, or all valid sentences can be proved from the premises.

Sound Rules of Inference

- Here are some examples of sound rules of inference
  - *A rule is sound if its conclusion is true whenever the premise is true*
- Each can be shown to be sound using a truth table

RULE	PREMISE	CONCLUSION
Modus Ponens	A, A → B	B
And Introduction	A, B	A ∧ B
And Elimination	A ∧ B	A
Double Negation	¬¬A	A
Unit Resolution	A ∨ B, ¬B	A
Resolution	**A ∨ B, ¬B ∨ C**	**A ∨ C**

Soundness of Modus Ponens

A	B	A → B	OK? (A ∧ (A→B)) → B
True	True	True	√
True	False	False	√
False	True	True	√
False	False	True	√

Proving Things

- A **proof** is a sequence of sentences, where each sentence is either a premise or a sentence derived from earlier sentences in the proof by one of the rules of inference.
- The last sentence is the **theorem** (also called goal or query) that we want to prove.
- Example for the "weather problem" given above.

1. Hu	Premise	"It is humid"
2. Hu→Ho	Premise	"If it is humid, it is hot"
3. Ho	Modus Ponens(1,2)	"It is hot"
4. (Ho∧Hu)→R	Premise	"If it's hot & humid, it's raining"
5. Ho∧Hu	And Introduction(1,3)	"It is hot and humid"
6. R	Modus Ponens(4,5)	"It is raining"

Horn Sentences

- A **Horn sentence** or **Horn clause** has the form:

$P1 \wedge P2 \wedge P3 \ldots \wedge Pn \rightarrow Q$

(P → Q) = (¬P ∨ Q)

or alternatively

$\neg P1 \vee \neg P2 \vee \neg P3 \ldots \vee \neg Pn \vee Q$

where Ps and Q are non-negated atoms

- To get a proof for Horn sentences, apply Modus Ponens repeatedly until nothing can be done
- We will use the Horn clause form later

Propositional Logic is a Weak Language

- Hard to identify "individuals" (e.g., Mary, 3)
- Can't directly talk about properties of individuals or relations between individuals (e.g., "Bill is tall")
- Generalizations, patterns, regularities can't easily be represented (e.g., "all triangles have 3 sides")
- First-Order Logic (abbreviated FOL or FOPC) is expressive enough to concisely represent this kind of information
- FOL adds relations, variables, and quantifiers, e.g.,
  - *"Every elephant is gray":* ∀ x (elephant(x) → gray(x))
  - *"There is a white alligator":* ∃ x (alligator(X) ^ white(X))

Example

- Consider the problem of representing the following information:
  - Every person is mortal.
  - Confucius is a person.
  - Confucius is mortal.
- How can these sentences be represented so that we can infer the third sentence from the first two?

Example II

- In PL we have to create propositional symbols to stand for all or part of each sentence. For example, we might have:
  P = "person"; Q = "mortal"; R = "Confucius"
- so the above 3 sentences are represented as:
  $P \rightarrow Q$; $R \rightarrow P$; $R \rightarrow Q$
- Although the third sentence is entailed by the first two, we needed an explicit symbol, R, to represent an individual, Confucius, who is a member of the classes "person" and "mortal"
- To represent other individuals we must introduce separate symbols for each one, with some way to represent the fact that all individuals who are "people" are also "mortal"

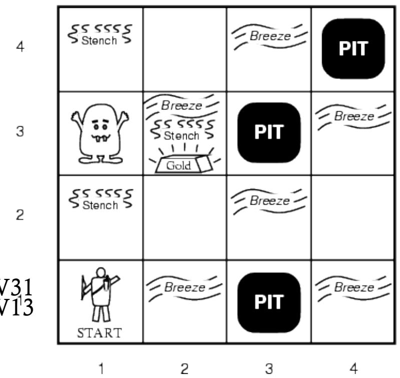
The "Hunt the Wumpus" Agent

- Some atomic propositions:

S12 = There is a stench in cell (1,2)
B34 = There is a breeze in cell (3,4)
W13 = The Wumpus is in cell (1,3)
V11 = We have visited cell (1,1)
OK11 = Cell (1,1) is safe.
etc

- Some rules:

(R1) $\neg S11 \rightarrow \neg W11 \wedge \neg W12 \wedge \neg W21$
(R2) $\neg S21 \rightarrow \neg W11 \wedge \neg W21 \wedge \neg W22 \wedge \neg W31$
(R3) $\neg S12 \rightarrow \neg W11 \wedge \neg W12 \wedge \neg W22 \wedge \neg W13$
(R4) $S12 \rightarrow W13 \vee W12 \vee W22 \vee W11$
etc.



- Note that the lack of variables requires us to give similar rules for each cell

Prove it!

- **YOUR MISSION**
- Prove that the Wumpus is in (1,3) and there is a pit in (3,1), given the observations shown and these rules:
- If there is no stench in a cell, then there is no wumpus in any adjacent cell
- If there is a stench in a cell, then there is a wumpus in some adjacent cell
- If there is no breeze in a cell, then there is no pit in any adjacent cell
- If there is a breeze in a cell, then there is a pit in some adjacent cell
- If a cell has been visited, it has neither a wumpus nor a pit
- **FIRST** write the propositional rules for the relevant cells
- **NEXT** write the proof steps and indicate what inference rules you used in each step

INFERENCE RULES

Modus Ponens
$A, A \rightarrow B$
ergo B

And Introduction
A, B
ergo $A \wedge B$

And Elimination
$A \wedge B$
ergo A

Double Negation
$\neg\neg A$
ergo A

Unit Resolution
$A \vee B, \neg B$
ergo A

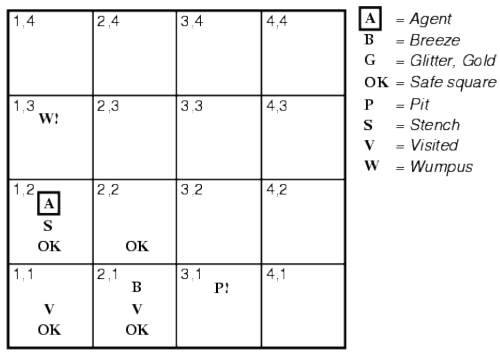
Resolution
$A \vee B, \neg B \vee C$
ergo $A \vee C$

A = *Agent*
B = *Breeze*
G = *Glitter, Gold*
OK = *Safe square*
P = *Pit*
S = *Stench*
V = *Visited*
W = *Wumpus*

V12 S12 -B12	V22 -S22 -B22		
V11 -S11 -B11	V21 B21 -S21		

After the Third Move

- We can prove that the Wumpus is in (1,3) using the four rules given.
- See R&N section 7.5

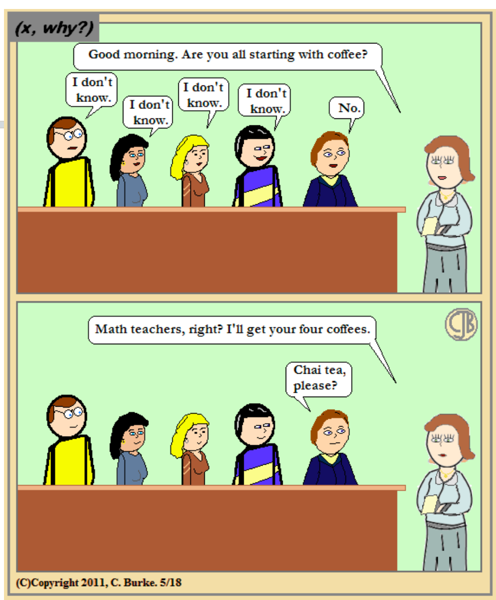


Proving W13

- Apply MP with ¬S11 and R1:
  ¬ W11 ∧ ¬ W12 ∧ ¬ W21
- Apply And-Elimination to this, yielding three sentences:
  ¬ W11, ¬ W12, ¬ W21
- Apply MP to ~S21 and R2, then apply And-Elimination:
  ¬ W22, ¬ W21, ¬ W31
- Apply MP to S12 and R4 to obtain:
  W13 ∨ W12 ∨ W22 ∨ W11
- Apply Unit Resolution on (W13 ∨ W12 ∨ W22 ∨ W11) and ¬W11:
  W13 ∨ W12 ∨ W22
- Apply Unit Resolution with (W13 ∨ W12 ∨ W22) and ¬W22:
  W13 ∨ W12
- Apply UR with (W13 ∨ W12) and ¬W12:
  W13
- QED

Problems with the Propositional Wumpus Hunter

- Lack of variables prevents stating more general rules
  - We need a set of similar rules for each cell
- Change of the KB over time is difficult to represent
  - Standard technique is to index facts with the time when they're true
  - This means we have a separate KB for every time point



Summary

- The process of deriving new sentences from old one is called **inference**.
  - **Sound** inference processes derives true conclusions given true premises
  - **Complete** inference processes derive all true conclusions from a set of premises
- A **valid sentence** is true in all worlds under all interpretations
- If an implication sentence can be shown to be valid, then—given its premise—its consequent can be derived
- Different logics make different **commitments** about what the world is made of and what kind of beliefs we can have regarding the facts
  - Logics are useful for the commitments they do not make because lack of commitment gives the knowledge base engineer more freedom
- **Propositional logic** commits only to the existence of facts that may or may not be the case in the world being represented
  - It has a simple syntax and simple semantics. It suffices to illustrate the process of inference
  - Propositional logic quickly becomes impractical, even for very small worlds