



ARTIFICIAL INTELLIGENCE

Knowledge Based Agents

COMPLETE ARCHITECTURES FOR INTELLIGENCE?

Search? (Problem Solving Agents)

- Solve the problem of what to do.

Logic and inference? (Knowledge Based Agents)

- Reason about what to do.
- Encoded knowledge/“expert” systems?
 - Know what to do.

Learning? (Learning Agents)

- Learn what to do.

Modern view: It's complex & multi-faceted.

WHY DO WE NEED LOGIC?

Problem-solving agents were very inflexible: hard code every possible state.

Search is almost always exponential in the number of states.

Problem solving agents cannot infer unobserved information.

We want an algorithm that *reasons* in a way that resembles reasoning in humans

REASONING

Making an inference about something that was previously not seen/not known based on some previously stored knowledge.

For example: If I tell you

- (1) Sue is Mary's sister, and
- (2) Sue is Amy's mother, then

it necessarily follows in the world that Mary is Amy's aunt, even though *I told you nothing at all about aunts*.

This sort of reasoning pattern is what we hope to capture.

INFERENCE IN FORMAL SYMBOL SYSTEMS:

ONTOLOGY, REPRESENTATION, INFERENCE

Formal Symbol Systems

- **Symbols** correspond to things/ideas in the world
- **Pattern matching** & rewrite corresponds to inference

Ontology: What exists in the world?

- What must be represented?

Representation: Syntax vs. Semantics

- What's Said vs. What's Meant

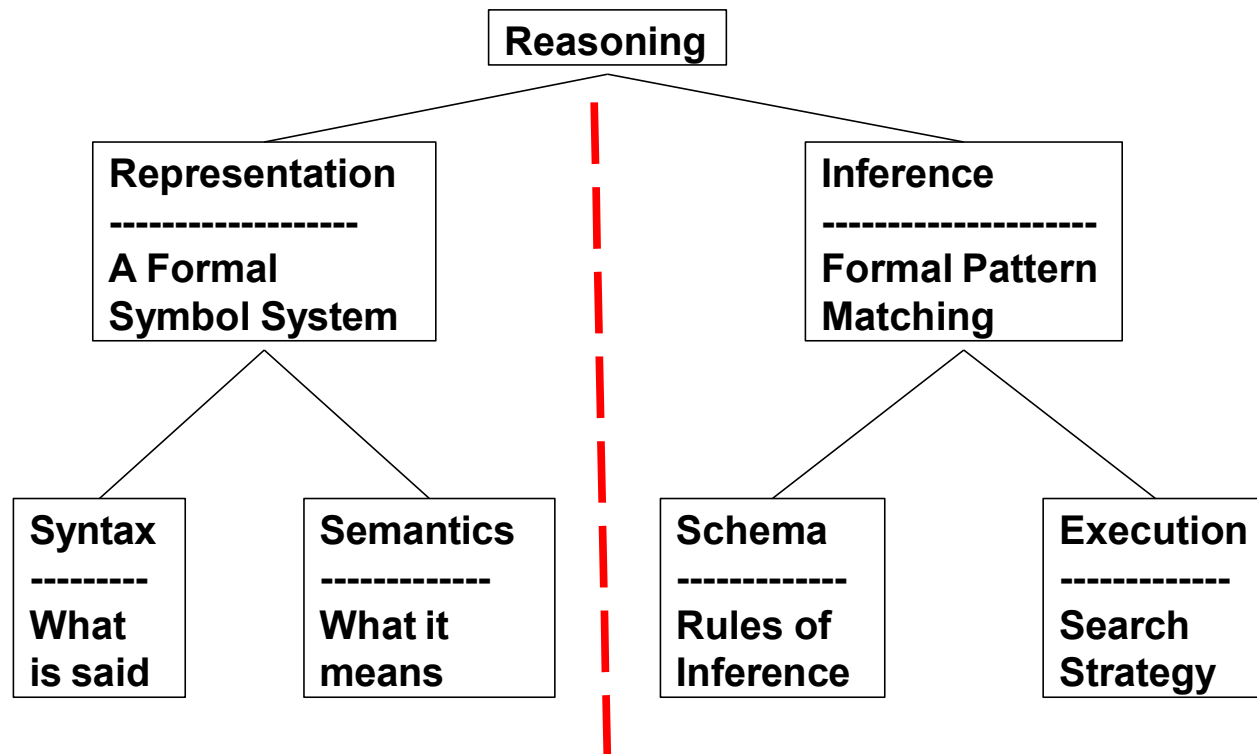
Inference: Schema vs. Mechanism

- Proof Steps vs. Search Strategy

Ontology:

What kind of things exist in the world?

What do we need to describe and reason about?



LOGICAL OR KNOWLEDGE BASED AGENTS COMPONENTS

Knowledge base / KB (facts)

Knowledge Representation Language (In what language would you tell agents the facts?)

Inference

Background Knowledge of the world

LOGICAL OR KNOWLEDGE BASED AGENTS

Basic Actions: Tell and Ask

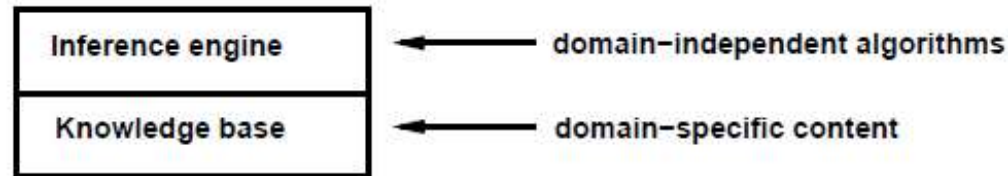
A Knowledge base keeps track of things

We can **tell** an agent facts and **ask** for inference

Example:

- **Tell:** Father of John is Bob
- **Tell:** Jane is John's Sister
- **Tell:** John's father is the same as John's sister father
- **Ask:** Who is Jane's father? (The answer requires inference on facts)

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

Agents can be viewed at the **knowledge level**

i.e., **what they know**, regardless of how implemented

Or at the **implementation level**

i.e., data structures in KB and algorithms that manipulate them

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

WUMPUS WORLD PEAS DESCRIPTION

Performance measure

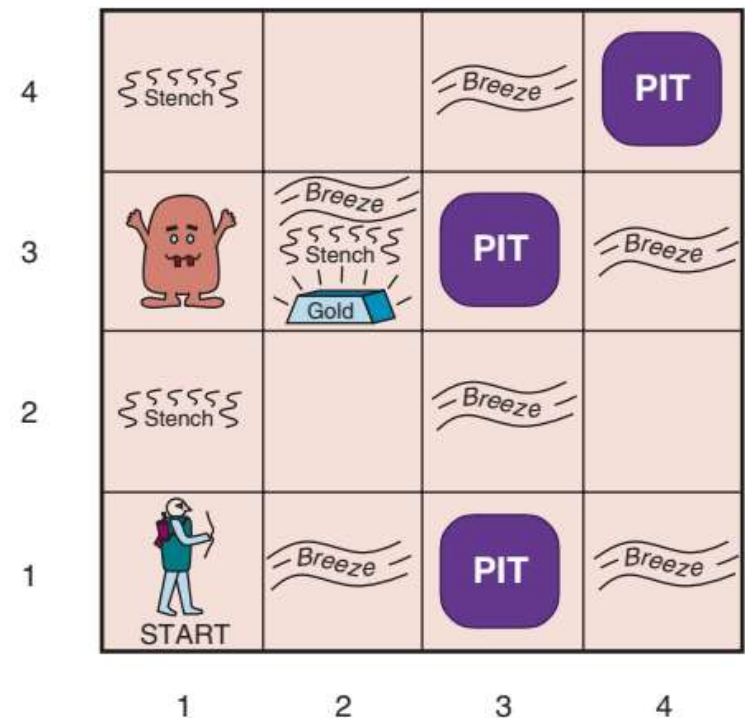
gold +1000, death -1000
-1 per step, -10 for using the arrow

Environment

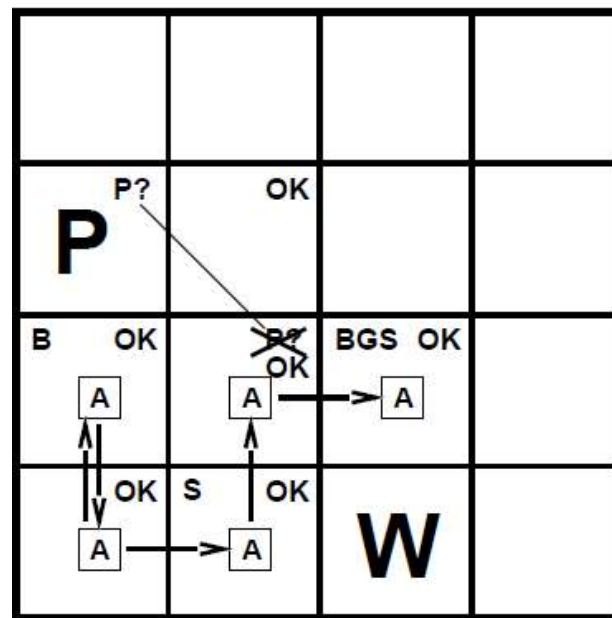
Squares adjacent to wumpus are smelly
Squares adjacent to pit are breezy
Glitter iff gold is in the same square
Shooting kills wumpus if you are facing it
Shooting uses up the only arrow
Grabbing picks up gold if in same square
Releasing drops the gold in same square

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

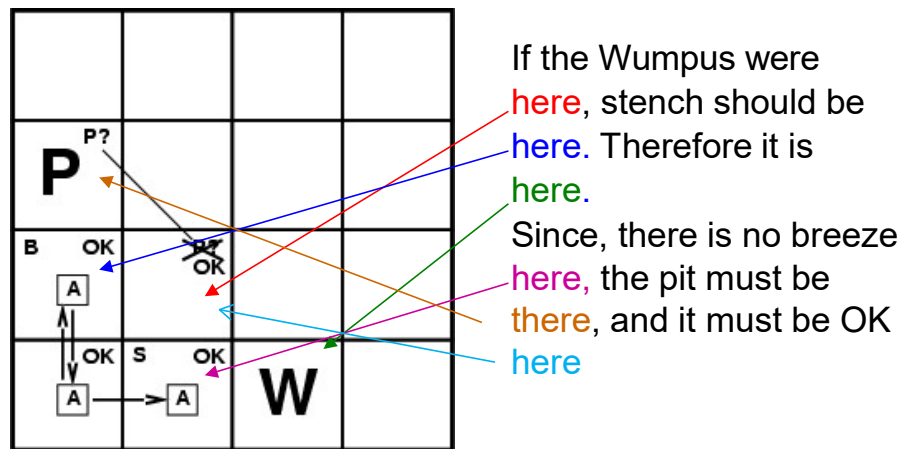
Sensors Breeze, Glitter, Smell



EXPLORING A WUMPUS WORLD



EXPLORING A WUMPUS WORLD



We need rather sophisticated reasoning here!

WUMPUS WORLD

1,4	2,4	3,4	4,4
1,3	2,3	3,3	4,3
1,2	2,2	3,2	4,2
OK			
1,1	2,1	3,1	4,1
A			
OK	OK		

[None, None, None, None, None]

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



1,4	2,4	3,4	4,4
1,3	2,3	3,3	4,3
1,2	2,2	3,2	4,2
OK	P?		
1,1	2,1	3,1	4,1
V	A	P?	
OK	B		
OK	OK		

[None, Breeze, None, None, None]

WUMPUS WORLD

1,4	2,4	3,4	4,4
1,3 W!	2,3	3,3	4,3
1,2 A S OK	2,2 OK	3,2	4,2
1,1 V OK	2,1 B V OK	3,1 P!	4,1

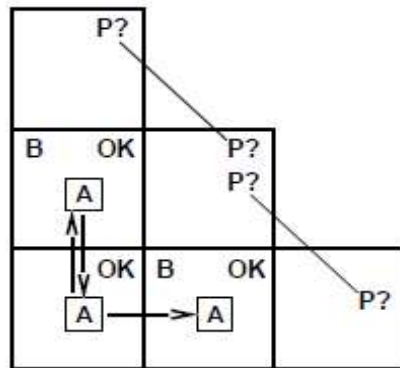
[Stench, None, None, None, None]

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

1,4	2,4 P?	3,4	4,4
1,3 W!	2,3 A S G B	3,3 P?	4,3
1,2 S V OK	2,2 V OK	3,2	4,2
1,1 V OK	2,1 B V OK	3,1 P!	4,1

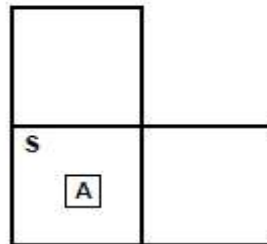
[Stench, Breeze, Glitter, None, None]

OTHER TIGHT SPOTS



Breeze in (1,2) and (2,1)
 \Rightarrow no safe actions

Assuming pits uniformly distributed,
 (2,2) has pit w/ prob 0.86, vs. 0.31



Smell in (1,1)

\Rightarrow cannot move

Can use a strategy of coercion:

shoot straight ahead

wumpus was there \Rightarrow dead \Rightarrow safe

wumpus wasn't there \Rightarrow safe

LOGIC

- We used logical reasoning to find the gold.
- **Logics**
 - ← are formal languages for representing information such that conclusions can be drawn from formal inference patterns
- **Syntax**
 - ← defines the well-formed sentences in the language
- **Semantics**
 - ← define the "meaning" or interpretation of sentences:
 - connect symbols to real events in the world
 - i.e., define **truth** of a sentence in a world
- E.g., the language of arithmetic:
 - $x+2 \geq y$ is a sentence
 - $-x^2+y > \{\}$ is not a sentence

}

→

syntax
- $x+2 \geq y$ is true in a world where $x = 7, y = 1$

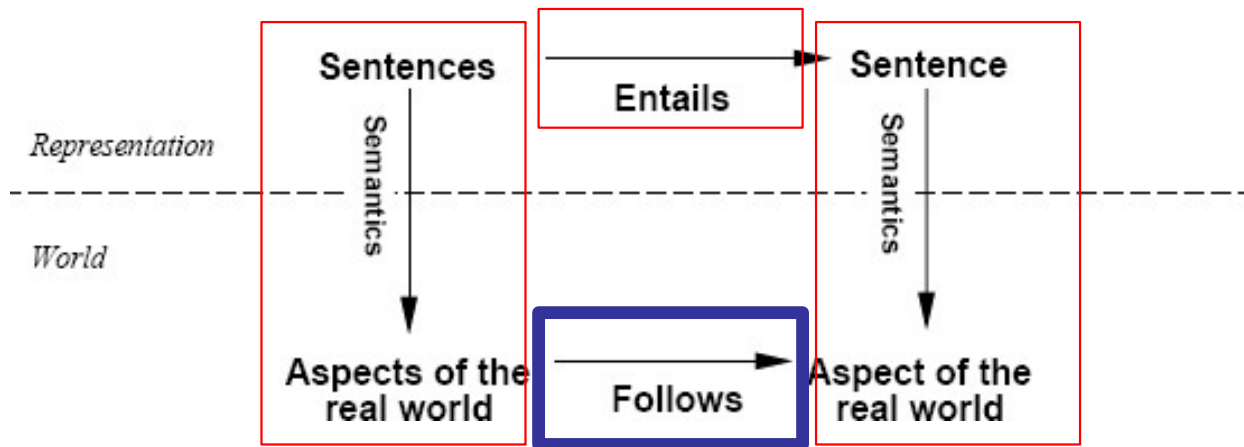
– $x+2 \geq y$ is false in a world where $x = 0, y = 6$

}

→

semantics

SCHEMATIC PERSPECTIVE



*If KB is true in the real world,
then any sentence α entailed by KB is also true
in the real world.*

For example: If I tell you (1) Sue is Mary's sister, and (2) Sue is Amy's mother, then it **necessarily follows in the world** that Mary is Amy's aunt, even though I told you nothing at all about aunts. This sort of reasoning pattern is what we hope to capture.

EXERCISES

If Edith eats her vegetables, then she can have a cookie. Edith ate her vegetables.
Therefore Edith gets a cookie.

$$\frac{P \rightarrow Q \quad P}{\therefore Q}$$

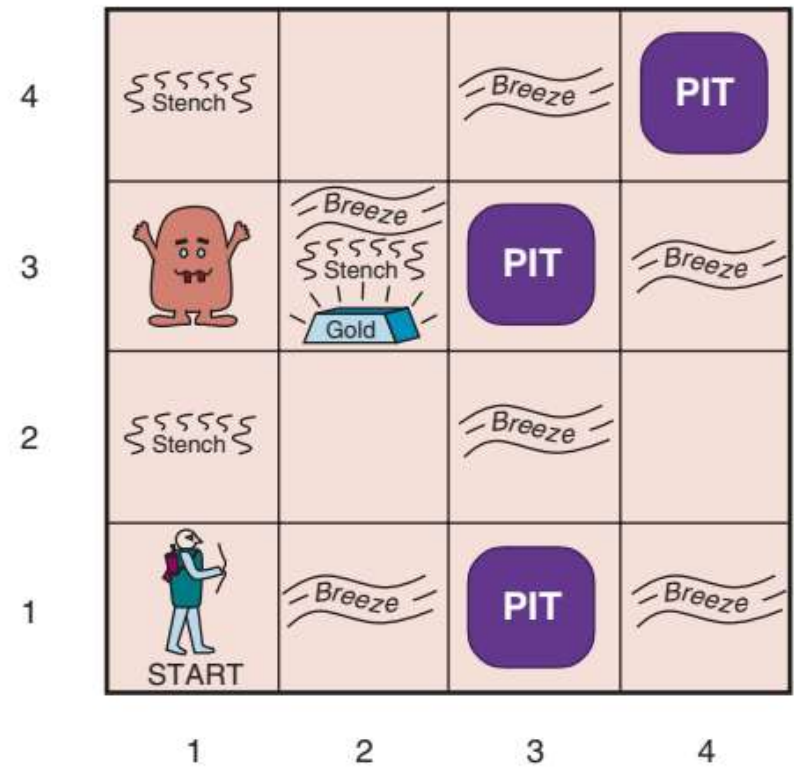
"If it's your birthday or there will be cake, then there will be cake."

Create a truth table

P: P: it's your birthday; Q: Q: there will be cake. $(P \vee Q) \rightarrow Q$

WUMPUS WORLD SYNTAX

Sentences? (using Propositional Logic)



SEMANTICS

Sentences have a true value **with respect to a model**

E.g.,

- model $m = \{P_{1,2} = \text{False}, P_{3,3} = \text{True}, S_{3,1} = \text{False}\}$

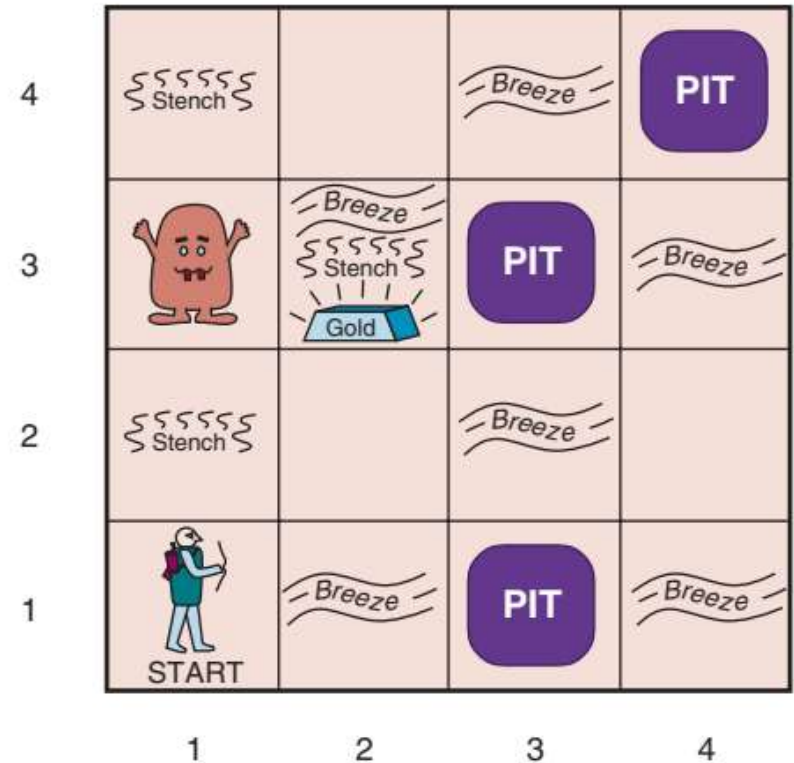
$P_{1,2}$, $P_{3,3}$, $S_{3,1}$ are symbols;

- they can mean anything.

Example Sentence: whether there isn't a pit in (3,3)

or a stench in (3,1)

$$\neg P_{3,3} \vee S_{3,1} = ?$$



MODELS

For $x^2 + y^2 = 5$, one possible model that **satisfies** the **equation/sentence** is $m1 = (x,y) = (2,1)$, another model that does so is $m2 = (1,2)$.

If a sentence α is true in model m , we say that m **satisfies** α or sometimes m is a **model** of α .

We use the notation $M(\alpha)$ to mean the set of all models of α i.e., $M(\alpha) = \{m1(\alpha), m2(\alpha), \dots\}$ where all models $m1, m2, \dots$ satisfy α

$m1(\alpha)$  refers to a single solution / model

$M(\alpha)$  refers to the set of all possible solutions / models

LOGICAL ENTAILMENT

Logical Reasoning involves the relation of logical **entailment** between sentences—the idea that a sentence *follows logically* from another sentence

$$\alpha \models \beta$$

(alpha entails beta)

$\alpha \models \beta$ if and only if, in every model in which alpha is true, beta is also true

$$\alpha \models \beta \text{ if and only if } M(\alpha) \subseteq M(\beta)$$

LOGICAL ENTAILMENT FROM MATH

Alpha (sentence 1) $\rightarrow x=0$

Beta (sentence 2) $\rightarrow xy=0$

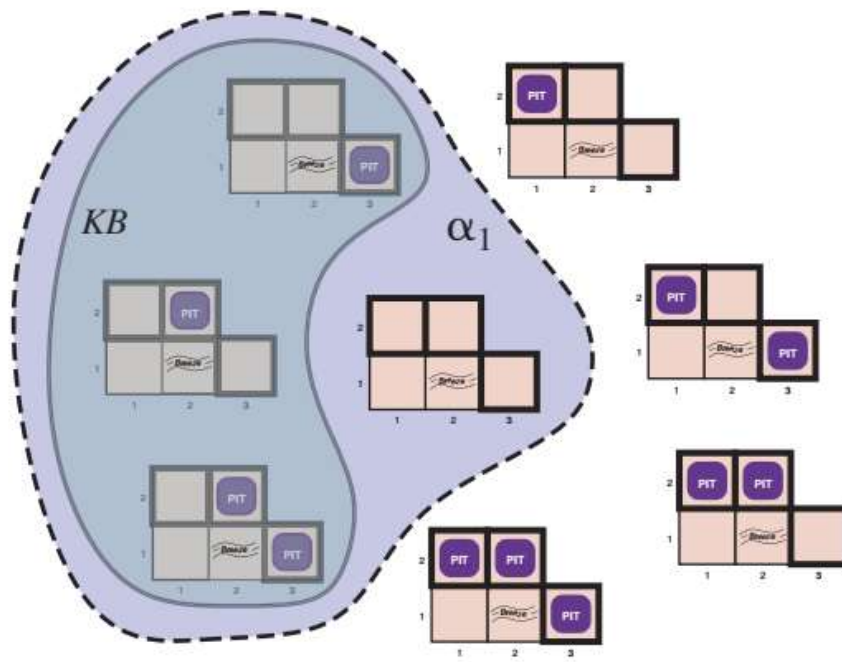
There's only one possible solution of $x=0$ and that is if x is 0 ($x=0$), therefore $M(\alpha) = \{x=0\}$

But there are many possible solutions of sentence 2, i.e., $M(\beta) = \{(x,y)=(0,0), (x,y)=(0,1), (x,y)=(1,0), \dots\}$ one of which is $x=0$

Since all possible solutions of alpha can be found in the solution set of beta, we say that alpha entails beta

$$\alpha \models \beta \text{ if and only if } M(\alpha) \subseteq M(\beta)$$

WUMPUS WORLD: ENTAILMENT



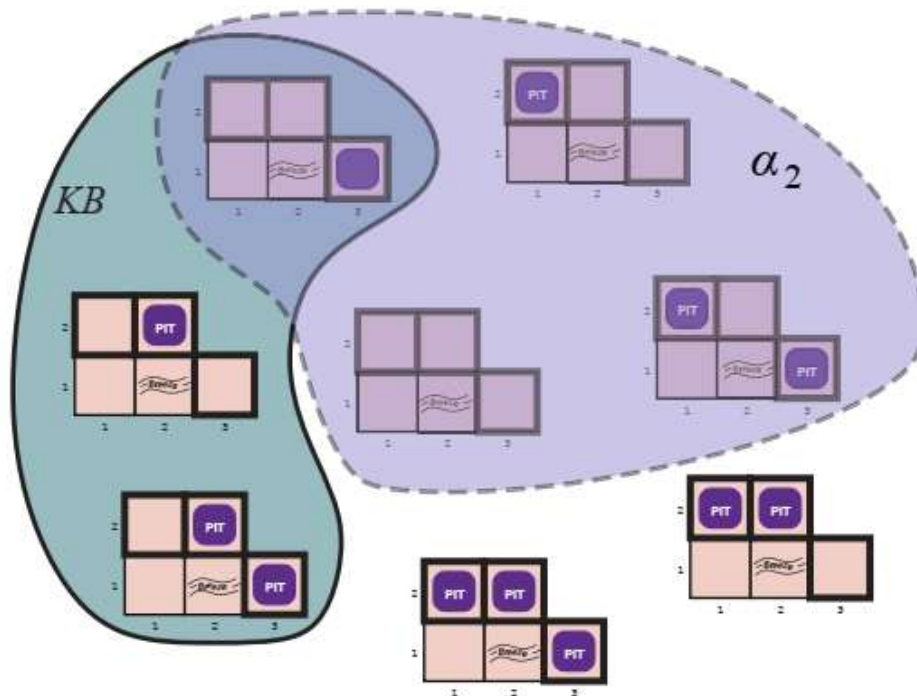
Alpha 1 (sentence) \rightarrow no pits in $[1, 2]$

$\langle B \mid = \alpha_1$

1,4	2,4	3,4	4,4
1,3	2,3	3,3	4,3
1,2	2,2 P?	3,2	4,2
OK			
1,1	2,1 A B OK	3,1 P?	4,1
V OK			

[None, Breeze, None, None, None]

WUMPUS WORLD: ENTAILMENT



1,4	2,4	3,4	4,4
1,3	2,3	3,3	4,3
1,2	2,2 P?	3,2	4,2
OK			
1,1	2,1 A B OK	3,1 P?	4,1
V OK			

[None, Breeze, None, None, None]

Alpha 2 (sentence) \rightarrow no pits in [2, 2]

$KB \not\models \alpha_2$

PROPOSITIONAL LOGIC: SYNTAX

The **atomic sentences** consists of a single **proposition symbol**.

$P, Q, R, W_{1,3}$ are **proposition symbols** that evaluate to **True** or **False**. $W_{1,3}$ is True if there's a wumpus in (1,3).

Complex sentences are constructed from simpler sentences using parenthesis and **logical connectivities**.

$$\textit{AtomicSentence} \rightarrow \textit{True} \mid \textit{False} \mid P \mid Q \mid R \mid \dots$$
$$\textit{ComplexSentence} \rightarrow (\textit{Sentence}) \mid [\textit{Sentence}]$$

LOGICAL CONNECTIVITIES

\neg (not). A sentence such as $\neg W_{1,3}$ is called the **negation** of $W_{1,3}$. A **literal** is either an atomic sentence (a **positive literal**) or a negated atomic sentence (a **negative literal**).

\wedge (and). A sentence whose main connective is \wedge , such as $W_{1,3} \wedge P_{3,1}$, is called a **conjunction**; its parts are the **conjuncts**. (The \wedge looks like an “A” for “And.”)

\vee (or). A sentence using \vee , such as $(W_{1,3} \wedge P_{3,1}) \vee W_{2,2}$, is a **disjunction** of the **disjuncts** $(W_{1,3} \wedge P_{3,1})$ and $W_{2,2}$. (Historically, the \vee comes from the Latin “vel,” which means “or.” For most people, it is easier to remember \vee as an upside-down \wedge .)

LOGICAL CONNECTIVITIES

\Rightarrow (implies). A sentence such as $(W1,3 \wedge P3,1) \Rightarrow \neg W2,2$ is called an **implication** (or conditional). Its **premise** or **antecedent** is $(W1,3 \wedge P3,1)$, and its **conclusion** or **consequent** is $\neg W2,2$. Implications are also known as **rules** or **if-then** statements. The implication symbol is sometimes written in other books as \supset or \rightarrow .

\Leftrightarrow (if and only if). The sentence $W1,3 \Leftrightarrow \neg W2,2$ is a **biconditional**. Some other books write this as Ξ .

IMPLICATION VS BICONDITIONAL

$P \Rightarrow Q$ (if P then Q)

If rainy then take umbrella
(doesn't mean that if you have taken
Umbrella, then it must rain)

If a bullet hits my head, I'll die
(doesn't mean that if I die, the bullet must
have hit me)

$P \Leftrightarrow Q$ (P if and only if Q) or
 $Q \Leftrightarrow P$ (Q if and only if P)

If pit in a square then breeze in atleast one
of the neighboring square, also if breeze in
a square then pit must be in atleast one of
the neighboring square

If you fail a mandatory course, you'll have to
retake it, if you're retaking a course then it
must be that you have failed it in the past

OPERATOR PRECEDENCE

$Sentence \rightarrow AtomicSentence \mid ComplexSentence$

$AtomicSentence \rightarrow True \mid False \mid P \mid Q \mid R \mid \dots$

$ComplexSentence \rightarrow (Sentence) \mid [Sentence]$

$\mid \neg Sentence$

$\mid Sentence \wedge Sentence$

$\mid Sentence \vee Sentence$

$\mid Sentence \Rightarrow Sentence$

$\mid Sentence \Leftrightarrow Sentence$

OPERATOR PRECEDENCE : $\neg, \wedge, \vee, \Rightarrow, \Leftrightarrow$

WUMPUS WORLD SENTENCES

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 $[i, j]$.

$$\neg P_{1,1}$$

$$\neg B_{1,1}$$

$$B_{2,1}$$

“Pits cause breezes in adjacent squares”

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

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

“A square is breezy **if and only if** there is an adjacent pit”

SEMANTICS

In propositional logic, a model simply fixes the **truth value**—true or false—for every proposition symbol e.g., $m1 = \{P_{1,2} = \text{false}, P_{2,2} = \text{false}, P_{3,1} = \text{true}\}$ says there's no pit in 2,2 and 1,2 but there's one in 3,1.

$\neg P$ is true iff P is false in m .

$P \wedge Q$ is true iff both P and Q are true in m .

$P \vee Q$ is true iff either P or Q is true in m .

$P \Rightarrow Q$ is true unless P is true and Q is false in m .

$P \Leftrightarrow Q$ is true iff P and Q are both true or both false in m .

SEMANTICS

		Negation	Conjunction	Disjunction	Implication	Biconditional
P	Q	$\neg P$	$P \wedge Q$	$P \vee Q$	$P \Rightarrow Q$	$P \Leftrightarrow Q$
<i>false</i>	<i>false</i>	<i>true</i>	<i>false</i>	<i>false</i>	<i>true</i>	<i>true</i>
<i>false</i>	<i>true</i>	<i>true</i>	<i>false</i>	<i>true</i>	<i>true</i>	<i>false</i>
<i>true</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>true</i>	<i>false</i>	<i>false</i>
<i>true</i>	<i>true</i>	<i>false</i>	<i>true</i>	<i>true</i>	<i>true</i>	<i>true</i>

Four possible models shown above

IMPLICATION

	P	Q	$P \Rightarrow Q$
m1 ←	false	false	true
m2 ←	false	true	true
m3 ←	true	false	false
m4 ←	true	true	true

\Rightarrow says if P is True; then Q must be True (m4).

There is no way that P is True and Q being False (m3), given $P \Rightarrow Q$ is True.

If P is False, then we are making no claim, Q can be either True or False (m1 and m2)

SENTENCES AND EVALUATION

Given a model $m = \{P_{1,2} = \text{False}, P_{3,3} = \text{True}, S_{3,1} = \text{False}\}$, evaluate

$$\neg P_{3,3} \vee S_{3,1} \Rightarrow P_{1,2} \wedge S_{3,1} \Leftrightarrow P_{3,3}$$

Operator Preference: $\neg, \wedge, \vee, \Rightarrow, \Leftrightarrow$

ENTAILMENT IN PROPOSITIONAL LOGIC

$(x=0) \models (xy=0)$

$(p=\text{True}) \models (p \vee q)$

$p=\text{True}$ is True in 2 models both of which are subset of 3 models in which $(p \vee q)$ is True

p	q	$p \vee q$
True	True	True
True	False	True
False	True	True
False	False	False

ENTAILMENT IN PROPOSITIONAL LOGIC

$$(p \wedge q) \models (p \vee q)$$

$p \wedge q$ is True in 1 model which is a subset of 3 models in which $(p \vee q)$ is True

p	q	$p \vee q$	$p \wedge q$
True	True	True	True
True	False	True	False
False	True	True	False
False	False	False	False

ENTAILMENT IN PROPOSITIONAL LOGIC

$$((p \Leftrightarrow q) \wedge r) \models (q \Rightarrow p)$$

$((p \Leftrightarrow q) \wedge r)$ is True in 2 models both of which are subset of 6 models in which $(q \Rightarrow p)$ is True

p	q	r	$(p \Leftrightarrow q) \wedge r$	$q \Rightarrow p$
False	False	False	False	True
False	False	True	True	True
False	True	False	False	False
False	True	True	False	False
True	False	False	False	True
True	False	True	False	True
True	True	False	False	True
True	True	True	True	True

USING ENTAILMENT FOR ANSWERS

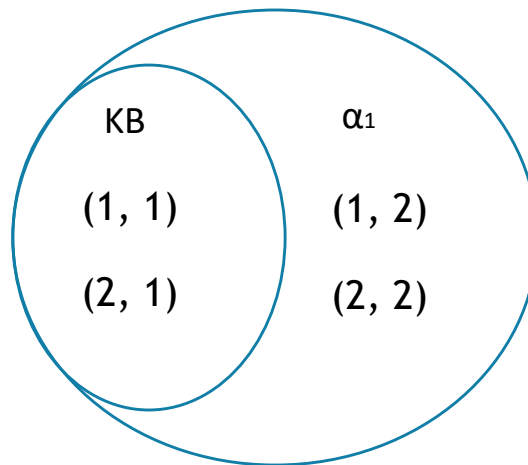
Imagine two variables, **cleanliness** and **dependability** (1 for very clean/dependable; 3 for not at all)

Knowledge base (contains info about John): I know about John through my friends that he is not messy (1 or 2) and always dependable (1)

Alpha 1: Good roommates always score 1 or 2 in either cleanliness or dependability.

Question: Is John a good roommate?

$KB \models \alpha_1$?



USING ENTAILMENT FOR ANSWERS

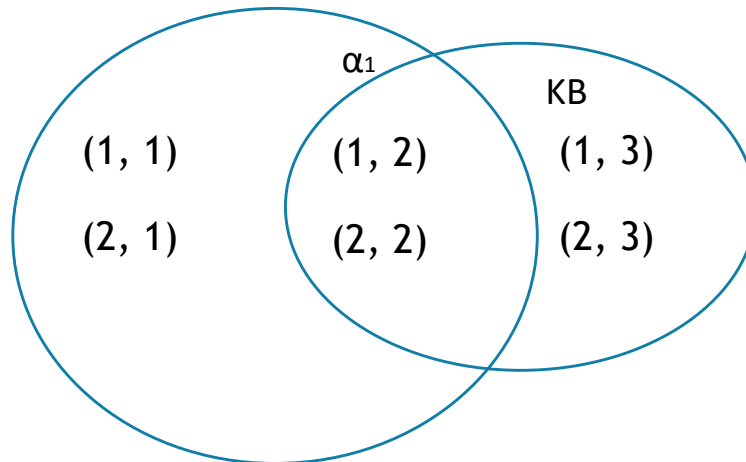
Imagine two variables, **cleanliness** and **dependability** (1 for very clean/dependable; 3 for not at all)

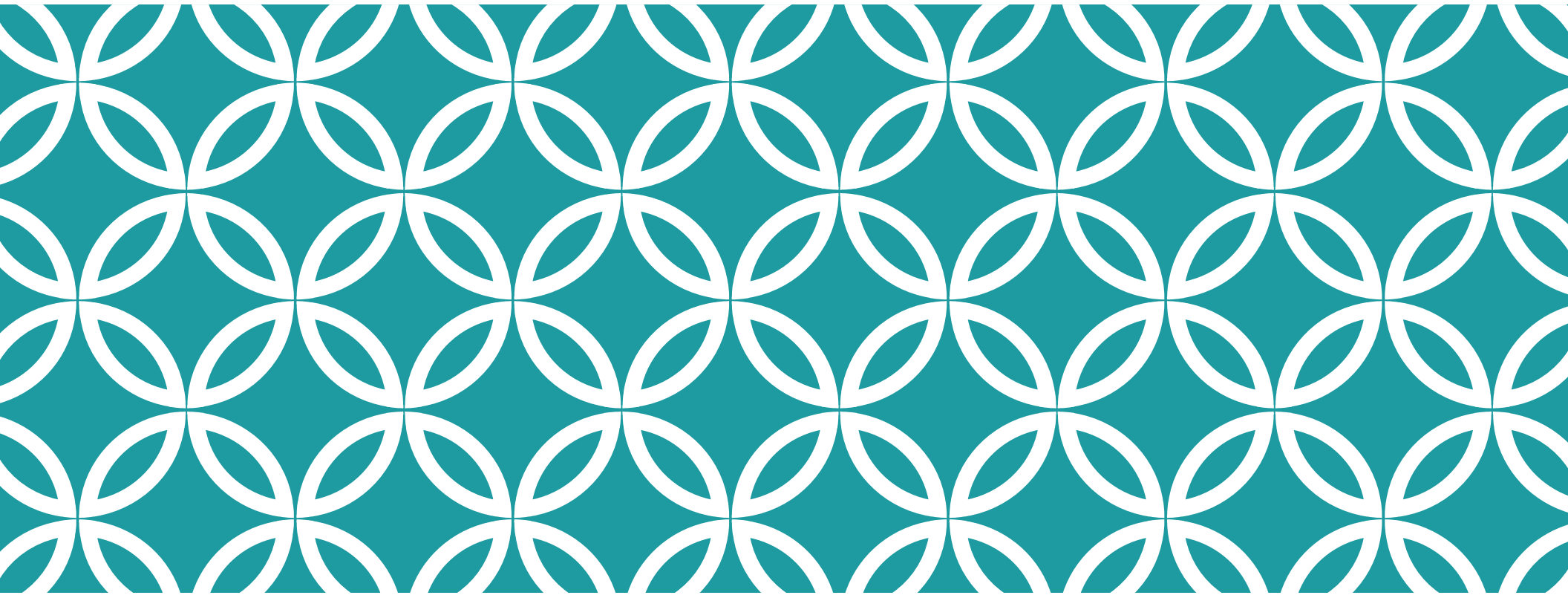
Knowledge base (contains info about John): I know about John through my friends that he is not messy (1 or 2) and he is not always dependable (2 or 3)

Alpha 1: Good roommates always score 1 or 2 in either cleanliness or dependability.

Question: Is John a good roommate?

$KB \models \alpha_1 ?$





KNOWLEDGE BASE INFERENCE



KNOWLEDGE BASE

We consider knowledge base as a set of rules for a particular problem.

We will use **TELL** and **ASK** approach. Tell agent set of rules and ask a query.

We will keep adding rules to our KB while we are searching for the solution

Once we have sufficient set of rules in our KB, we will use **inference** to find the solution (solution in this case is answer to the query)

WUMPUS WORLD: SYMBOLS

$P_{x,y}$ is true if there is a pit in $[x, y]$.

$W_{x,y}$ is true if there is a wumpus in $[x, y]$, dead or alive.

$B_{x,y}$ is true if the agent perceives a breeze in $[x, y]$.

$S_{x,y}$ is true if the agent perceives a stench in $[x, y]$.

The sentences we write will suffice to derive $\neg P_{1,2}$
(there is no pit in $[1,2]$)

1,4	2,4	3,4	4,4
1,3	2,3	3,3	4,3
1,2	2,2	3,2	4,2
1,1	2,1	3,1	4,1

TELL RULES TO KB

There is no pit in [1,1]:

$$R_1 : \neg P_{1,1}$$

A square is breezy if and only if there is a pit in a neighboring square. This has to be stated for each square; for now, we include just the relevant squares:

$$R_2 : B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$$

$$R_3 : B_{2,1} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1})$$

The preceding sentences are true in all wumpus worlds. Now we include the breeze percepts for the first two squares visited in the specific world the agent is in, leading up to the situation shown.

$$R_4 : \neg B_{1,1}$$

$$R_5 : B_{2,1}$$

1,4	2,4	3,4	4,4
1,3	2,3	3,3	4,3
1,2 OK	2,2 P?	3,2	4,2
1,1 V OK	2,1 A B OK	3,1 P?	4,1

TELL RULES TO KB

Our KB is a conjunction (\wedge) of all rules, i.e.,

KB: $R_1 \wedge R_2 \wedge R_3 \wedge R_4 \wedge R_5$

$R_1 : \neg P_{1,1}$

$R_2 : B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$

$R_3 : B_{2,1} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1})$

$R_4 : \neg B_{1,1}$

$R_5 : B_{2,1}$

Ask Query: $\neg P_{1,2}$ (there is no pit in [1,2])

KB $\models \neg P_{1,2}$?

1,4	2,4	3,4	4,4
1,3	2,3	3,3	4,3
1,2 OK	2,2 P?	3,2	4,2
1,1 V OK	2,1 <div>⬮ B OK</div>	3,1 P?	4,1

ENTAILMENT: MODEL CHECKING

$$KB \models \neg P_{1,2}$$

Find all possible models $M(\neg P_{1,2})$ where $\neg P_{1,2}$ is True

Find all possible models $M(KB)$ where KB is True.

If $M(KB)$ is a subset of $M(\neg P_{1,2})$, then $KB \models \neg P_{1,2}$

$$KB: R_1 \wedge R_2 \wedge R_3 \wedge R_4 \wedge R_5$$

$B_{1,1}$	$B_{2,1}$	$P_{1,1}$	$P_{1,2}$	$P_{2,1}$	$P_{2,2}$	$P_{3,1}$	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
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
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	<u>true</u>
false	true	false	false	false	true	false	true	true	true	true	true	<u>true</u>
false	true	false	false	false	true	true	true	true	true	true	true	<u>true</u>
false	true	false	false	true	false	false	true	false	false	true	true	false
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
true	true	true	true	true	true	true	false	true	true	false	true	false

ENTAILMENT

Model Checking:

- discussed in previous slides where we enumerate all possible models

Theorem Proving:

- Applying rules of inference directly to the sentences in our knowledge base to construct a proof of the desired sentence without consulting model

THEOREM PROVING: LOGICAL EQUIVALENCES

$$(\alpha \wedge \beta) \equiv (\beta \wedge \alpha) \quad \text{commutativity of } \wedge$$

$$(\alpha \vee \beta) \equiv (\beta \vee \alpha) \quad \text{commutativity of } \vee$$

$$((\alpha \wedge \beta) \wedge \gamma) \equiv (\alpha \wedge (\beta \wedge \gamma)) \quad \text{associativity of } \wedge$$

$$((\alpha \vee \beta) \vee \gamma) \equiv (\alpha \vee (\beta \vee \gamma)) \quad \text{associativity of } \vee$$

$$\neg(\neg\alpha) \equiv \alpha \quad \text{double-negation elimination}$$

$$(\alpha \Rightarrow \beta) \equiv (\neg\beta \Rightarrow \neg\alpha) \quad \text{contraposition}$$

$$(\alpha \Rightarrow \beta) \equiv (\neg\alpha \vee \beta) \quad \text{implication elimination}$$

$$(\alpha \Leftrightarrow \beta) \equiv ((\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha)) \quad \text{biconditional elimination}$$

$$\neg(\alpha \wedge \beta) \equiv (\neg\alpha \vee \neg\beta) \quad \text{De Morgan}$$

$$\neg(\alpha \vee \beta) \equiv (\neg\alpha \wedge \neg\beta) \quad \text{De Morgan}$$

$$(\alpha \wedge (\beta \vee \gamma)) \equiv ((\alpha \wedge \beta) \vee (\alpha \wedge \gamma)) \quad \text{distributivity of } \wedge \text{ over } \vee$$

$$(\alpha \vee (\beta \wedge \gamma)) \equiv ((\alpha \vee \beta) \wedge (\alpha \vee \gamma)) \quad \text{distributivity of } \vee \text{ over } \wedge$$

CONCEPTS: LOGICAL EQUIVALENCE

Two sentences α and β are logically equivalent if they are true in the same set of models. We write this as $\alpha \equiv \beta$.

For example, we can easily show (using truth tables) that $P \wedge Q$ and $Q \wedge P$ are logically equivalent.

Any two sentences α and β are equivalent only if each of them entails the other: $\alpha \equiv \beta$ if and only if $\alpha \models \beta$ and $\beta \models \alpha$

CONTRAPOSITION PROOF USING TRUTH TABLE

Theorem Proving: Logical Equivalences

Contraposition

Validity

Deduction

Satisfiability

Theorems – Modus Ponens

p	q	$p \Rightarrow q$	$\neg q$	$\neg p$	$\neg q \Rightarrow \neg p$
False	False	True	True	True	True
False	True	True	False	True	True
True	False	False	True	False	False
True	True	True	False	False	True

If rainy, then take coat is logically equivalent of saying that if you haven't taken coat then it must not be raining

$$(\alpha \Rightarrow \beta) \equiv (\neg \beta \Rightarrow \neg \alpha) \quad \text{contraposition}$$

CONCEPTS: VALIDITY

Theorem Proving: Logical Equivalences

Contraposition

Validity

Deduction

Satisfiability

Theorems – Modus Ponens

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

For example, the sentence $P \vee \neg P$ is valid (irrespective of the value of P).

Valid sentences are also known as **tautologies**—they are **necessarily** true.

CONCEPTS: DEDUCTION

Theorem Proving: Logical Equivalences

Contraposition

Validity

Deduction

Satisfiability

Theorems – Modus Ponens

For any sentences α and β , $\alpha \models \beta$ if and only if the sentence $(\alpha \Rightarrow \beta)$ is valid i.e.,

$$(\alpha \models \beta) \Leftrightarrow (\alpha \Rightarrow \beta)$$

Both sides are equivalent to the assertion that there is no model in which α is true and β is false, i.e., no model in which $\alpha \Rightarrow \beta$ is false.

	P	Q	$P \Rightarrow Q$
m1 ←	false	false	true
m2 ←	false	true	true
m3 ←	true	false	false
m4 ←	true	true	true

CONCEPTS: SATISFIABILITY

Theorem Proving: Logical Equivalences

Contraposition

Validity

Deduction

Satisfiability

Theorems – Modus Ponens

A sentence is satisfiable if it is true in, or satisfied by, *some* model.

For example, the knowledge base given earlier, $(R1 \wedge R2 \wedge R3 \wedge R4 \wedge R5)$, is satisfiable because there are three models in which it is true.

Satisfiability can be checked by enumerating the possible models until one is found that satisfies the sentence.

THEOREM PROVING: PROOFS

Theorem Proving: Logical Equivalences

Contraposition

Validity

Deduction

Satisfiability

Theorems – Modus Ponens

Modus Ponens:
$$\frac{\alpha \Rightarrow \beta, \quad \alpha}{\beta} \qquad ((\alpha \Rightarrow \beta) \wedge \alpha) \Rightarrow \beta$$

If alpha implies beta, and we know alpha to be True, then B must be True

AND elimination:

- If (alpha \wedge Beta) is True then alpha must be True

$$\frac{\alpha \wedge \beta}{\alpha} \qquad \alpha \wedge \beta \Rightarrow \alpha$$

TELL RULES TO KB

Our KB is a conjunction (\wedge) of all rules, i.e.,

KB: $R_1 \wedge R_2 \wedge R_3 \wedge R_4 \wedge R_5$

$R_1 : \neg P_{1,1}$

$R_2 : B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$

$R_3 : B_{2,1} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1})$

$R_4 : \neg B_{1,1}$

$R_5 : B_{2,1}$

Ask Query: $\neg P_{1,2}$ (there is no pit in [1,2])

KB $\models \neg P_{1,2}$ or $KB \Rightarrow \neg P_{1,2}$?

1,4	2,4	3,4	4,4
1,3	2,3	3,3	4,3
1,2 OK	2,2 P?	3,2	4,2
1,1 V OK	2,1 <div><div>Λ</div><div>B</div><div>OK</div></div>	3,1 P?	4,1

KB $\Rightarrow \neg P_{1,2}$?

R₁ : $\neg P_{1,1}$

R₂ : $B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$

R₃ : $B_{2,1} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1})$

R₄ : $\neg B_{1,1}$

R₅ : $B_{2,1}$

R₆ : $(B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1})) \wedge ((P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1})$

$(\alpha \wedge \beta) \equiv (\beta \wedge \alpha)$ commutativity of \wedge

$(\alpha \vee \beta) \equiv (\beta \vee \alpha)$ commutativity of \vee

$((\alpha \wedge \beta) \wedge \gamma) \equiv (\alpha \wedge (\beta \wedge \gamma))$ associativity of \wedge

$((\alpha \vee \beta) \vee \gamma) \equiv (\alpha \vee (\beta \vee \gamma))$ associativity of \vee

$\neg(\neg\alpha) \equiv \alpha$ double-negation elimination

$(\alpha \Rightarrow \beta) \equiv (\neg\beta \Rightarrow \neg\alpha)$ contraposition

$(\alpha \Rightarrow \beta) \equiv (\neg\alpha \vee \beta)$ implication elimination

$(\alpha \Leftrightarrow \beta) \equiv ((\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha))$ biconditional elimination

$\neg(\alpha \wedge \beta) \equiv (\neg\alpha \vee \neg\beta)$ De Morgan

$\neg(\alpha \vee \beta) \equiv (\neg\alpha \wedge \neg\beta)$ De Morgan

$(\alpha \wedge (\beta \vee \gamma)) \equiv ((\alpha \wedge \beta) \vee (\alpha \wedge \gamma))$ distributivity of \wedge over \vee

$(\alpha \vee (\beta \wedge \gamma)) \equiv ((\alpha \vee \beta) \wedge (\alpha \vee \gamma))$ distributivity of \vee over \wedge

KB $\Rightarrow \neg P_{1,2}$?

R₁ : $\neg P_{1,1}$

R₂ : $B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$

R₃ : $B_{2,1} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1})$

R₄ : $\neg B_{1,1}$

R₅ : $B_{2,1}$

R₆ : $(B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1})) \wedge ((P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1}) \left. \vphantom{R_6} \right\} \frac{\alpha \wedge \beta}{\alpha}$

R₇ : $((P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1})$

$(\alpha \wedge \beta) \equiv (\beta \wedge \alpha)$ commutativity of \wedge

$(\alpha \vee \beta) \equiv (\beta \vee \alpha)$ commutativity of \vee

$((\alpha \wedge \beta) \wedge \gamma) \equiv (\alpha \wedge (\beta \wedge \gamma))$ associativity of \wedge

$((\alpha \vee \beta) \vee \gamma) \equiv (\alpha \vee (\beta \vee \gamma))$ associativity of \vee

$\neg(\neg\alpha) \equiv \alpha$ double-negation elimination

$(\alpha \Rightarrow \beta) \equiv (\neg\beta \Rightarrow \neg\alpha)$ contraposition

$(\alpha \Rightarrow \beta) \equiv (\neg\alpha \vee \beta)$ implication elimination

$(\alpha \Leftrightarrow \beta) \equiv ((\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha))$ biconditional elimination

$\neg(\alpha \wedge \beta) \equiv (\neg\alpha \vee \neg\beta)$ De Morgan

$\neg(\alpha \vee \beta) \equiv (\neg\alpha \wedge \neg\beta)$ De Morgan

$(\alpha \wedge (\beta \vee \gamma)) \equiv ((\alpha \wedge \beta) \vee (\alpha \wedge \gamma))$ distributivity of \wedge over \vee

$(\alpha \vee (\beta \wedge \gamma)) \equiv ((\alpha \vee \beta) \wedge (\alpha \vee \gamma))$ distributivity of \vee over \wedge

KB $\Rightarrow \neg P_{1,2}$?

R₁ : $\neg P_{1,1}$

R₂ : $B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$

R₃ : $B_{2,1} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1})$

R₄ : $\neg B_{1,1}$

R₅ : $B_{2,1}$

R₆ : $(B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1})) \wedge ((P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1})$

R₇ : $((P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1})$

R₈ : $(\neg B_{1,1} \Rightarrow \neg(P_{1,2} \vee P_{2,1}))$

$(\alpha \wedge \beta) \equiv (\beta \wedge \alpha)$ commutativity of \wedge

$(\alpha \vee \beta) \equiv (\beta \vee \alpha)$ commutativity of \vee

$((\alpha \wedge \beta) \wedge \gamma) \equiv (\alpha \wedge (\beta \wedge \gamma))$ associativity of \wedge

$((\alpha \vee \beta) \vee \gamma) \equiv (\alpha \vee (\beta \vee \gamma))$ associativity of \vee

$\neg(\neg\alpha) \equiv \alpha$ double-negation elimination

$(\alpha \Rightarrow \beta) \equiv (\neg\beta \Rightarrow \neg\alpha)$ contraposition

$(\alpha \Rightarrow \beta) \equiv (\neg\alpha \vee \beta)$ implication elimination

$(\alpha \Leftrightarrow \beta) \equiv ((\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha))$ biconditional elimination

$\neg(\alpha \wedge \beta) \equiv (\neg\alpha \vee \neg\beta)$ De Morgan

$\neg(\alpha \vee \beta) \equiv (\neg\alpha \wedge \neg\beta)$ De Morgan

$(\alpha \wedge (\beta \vee \gamma)) \equiv ((\alpha \wedge \beta) \vee (\alpha \wedge \gamma))$ distributivity of \wedge over \vee

$(\alpha \vee (\beta \wedge \gamma)) \equiv ((\alpha \vee \beta) \wedge (\alpha \vee \gamma))$ distributivity of \vee over \wedge

KB $\Rightarrow \neg P_{1,2}$?

R₁ : $\neg P_{1,1}$

R₂ : $B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$

R₃ : $B_{2,1} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1})$

R₄ : $\neg B_{1,1}$

R₅ : $B_{2,1}$

R₆ : $(B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1})) \wedge ((P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1})$

R₇ : $((P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1})$

R₈ : $(\neg B_{1,1} \Rightarrow \neg(P_{1,2} \vee P_{2,1}))$

R₉ : $\neg(P_{1,2} \vee P_{2,1})$

$(\alpha \wedge \beta) \equiv (\beta \wedge \alpha)$ commutativity of \wedge

$(\alpha \vee \beta) \equiv (\beta \vee \alpha)$ commutativity of \vee

$((\alpha \wedge \beta) \wedge \gamma) \equiv (\alpha \wedge (\beta \wedge \gamma))$ associativity of \wedge

$((\alpha \vee \beta) \vee \gamma) \equiv (\alpha \vee (\beta \vee \gamma))$ associativity of \vee

$\neg(\neg\alpha) \equiv \alpha$ double-negation elimination

$(\alpha \Rightarrow \beta) \equiv (\neg\beta \Rightarrow \neg\alpha)$ contraposition

$(\alpha \Rightarrow \beta) \equiv (\neg\alpha \vee \beta)$ implication elimination

$(\alpha \Leftrightarrow \beta) \equiv ((\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha))$ biconditional elimination

$\neg(\alpha \wedge \beta) \equiv (\neg\alpha \vee \neg\beta)$ De Morgan

$\neg(\alpha \vee \beta) \equiv (\neg\alpha \wedge \neg\beta)$ De Morgan

$(\alpha \wedge (\beta \vee \gamma)) \equiv ((\alpha \wedge \beta) \vee (\alpha \wedge \gamma))$ distributivity of \wedge over \vee

$(\alpha \vee (\beta \wedge \gamma)) \equiv ((\alpha \vee \beta) \wedge (\alpha \vee \gamma))$ distributivity of \vee over \wedge

$\left. \begin{array}{l} \alpha \Rightarrow \beta, \\ \alpha \end{array} \right\} \frac{\quad}{\beta}$

KB $\Rightarrow \neg P_{1,2}$?

$$R_1 : \neg P_{1,1}$$

$$R_2 : B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$$

$$R_3 : B_{2,1} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1})$$

$$R_4 : \neg B_{1,1}$$

$$R_5 : B_{2,1}$$

$$R_6 : (B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1})) \wedge ((P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1})$$

$$R_7 : ((P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1})$$

$$R_8 : (\neg B_{1,1} \Rightarrow \neg(P_{1,2} \vee P_{2,1}))$$

$$R_9 : \neg(P_{1,2} \vee P_{2,1}) \quad \}$$

$$R_{10} : \neg P_{1,2} \wedge \neg P_{2,1}$$

$$(\alpha \wedge \beta) \equiv (\beta \wedge \alpha) \quad \text{commutativity of } \wedge$$

$$(\alpha \vee \beta) \equiv (\beta \vee \alpha) \quad \text{commutativity of } \vee$$

$$((\alpha \wedge \beta) \wedge \gamma) \equiv (\alpha \wedge (\beta \wedge \gamma)) \quad \text{associativity of } \wedge$$

$$((\alpha \vee \beta) \vee \gamma) \equiv (\alpha \vee (\beta \vee \gamma)) \quad \text{associativity of } \vee$$

$$\neg(\neg\alpha) \equiv \alpha \quad \text{double-negation elimination}$$

$$(\alpha \Rightarrow \beta) \equiv (\neg\beta \Rightarrow \neg\alpha) \quad \text{contraposition}$$

$$(\alpha \Rightarrow \beta) \equiv (\neg\alpha \vee \beta) \quad \text{implication elimination}$$

$$(\alpha \Leftrightarrow \beta) \equiv ((\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha)) \quad \text{biconditional elimination}$$

$$\neg(\alpha \wedge \beta) \equiv (\neg\alpha \vee \neg\beta) \quad \text{De Morgan}$$

$$\neg(\alpha \vee \beta) \equiv (\neg\alpha \wedge \neg\beta) \quad \text{De Morgan}$$

$$(\alpha \wedge (\beta \vee \gamma)) \equiv ((\alpha \wedge \beta) \vee (\alpha \wedge \gamma)) \quad \text{distributivity of } \wedge \text{ over } \vee$$

$$(\alpha \vee (\beta \wedge \gamma)) \equiv ((\alpha \vee \beta) \wedge (\alpha \vee \gamma)) \quad \text{distributivity of } \vee \text{ over } \wedge$$

$$KB \Rightarrow \boxed{\neg P_{1,2} ?}$$

$$R_1 : \neg P_{1,1}$$

$$R_2 : B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$$

$$R_3 : B_{2,1} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1})$$

$$R_4 : \neg B_{1,1}$$

$$R_5 : B_{2,1}$$

$$R_6 : (B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1})) \wedge ((P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1})$$

$$R_7 : ((P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1})$$

$$R_8 : (\neg B_{1,1} \Rightarrow \neg(P_{1,2} \vee P_{2,1}))$$

$$R_9 : \neg(P_{1,2} \vee P_{2,1})$$

$$R_{10} : \boxed{\neg P_{1,2}} \wedge \neg P_{2,1}$$

In many practical cases, *finding a proof can be more efficient because the proof can ignore irrelevant propositions, no matter how many of them there are.*

For example, the proof given earlier leading to $\neg P_{1,2} \wedge \neg P_{2,1}$ does not mention the propositions $B_{2,1}$, $P_{1,1}$, $P_{2,2}$, or $P_{3,1}$.

R10: neither [1,2] nor [2,1] contains a pit.

EXERCISE

Prove $\neg P_{2,1}$

R₁ : $\neg P_{1,1}$

R₂ : $B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$

R₃ : $B_{2,1} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1})$

R₄ : $\neg B_{1,1}$

R₅ : $B_{2,1}$

$(\alpha \wedge \beta) \equiv (\beta \wedge \alpha)$ commutativity of \wedge

$(\alpha \vee \beta) \equiv (\beta \vee \alpha)$ commutativity of \vee

$((\alpha \wedge \beta) \wedge \gamma) \equiv (\alpha \wedge (\beta \wedge \gamma))$ associativity of \wedge

$((\alpha \vee \beta) \vee \gamma) \equiv (\alpha \vee (\beta \vee \gamma))$ associativity of \vee

$\neg(\neg\alpha) \equiv \alpha$ double-negation elimination

$(\alpha \Rightarrow \beta) \equiv (\neg\beta \Rightarrow \neg\alpha)$ contraposition

$(\alpha \Rightarrow \beta) \equiv (\neg\alpha \vee \beta)$ implication elimination

$(\alpha \Leftrightarrow \beta) \equiv ((\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha))$ biconditional elimination

$\neg(\alpha \wedge \beta) \equiv (\neg\alpha \vee \neg\beta)$ De Morgan

$\neg(\alpha \vee \beta) \equiv (\neg\alpha \wedge \neg\beta)$ De Morgan

$(\alpha \wedge (\beta \vee \gamma)) \equiv ((\alpha \wedge \beta) \vee (\alpha \wedge \gamma))$ distributivity of \wedge over \vee

$(\alpha \vee (\beta \wedge \gamma)) \equiv ((\alpha \vee \beta) \wedge (\alpha \vee \gamma))$ distributivity of \vee over \wedge

INFERENCE BY RESOLUTION

Suppose the agent returns from $[2,1]$ to $[1,1]$ and then goes to $[1,2]$, where it perceives a stench, but no breeze.

We add the following facts to the knowledge base:



$$R_{11} : \neg B_{1,2} .$$

$$R_{12} : B_{1,2} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{1,3})$$

By the same process that led to R_{10} earlier, we can now derive the absence of pits in $[2,2]$ and $[1,3]$:

$$R_{13} : \neg P_{2,2} \text{ (Do this as an exercise)}$$

$$R_{14} : \neg P_{1,3} \text{ (Do this as an exercise)}$$

1,4	2,4	3,4	4,4
1,3 W!	2,3	3,3	4,3
1,2  S OK	2,2 OK	3,2	4,2
1,1  v OK	2,1 B V OK	3,1 P!	4,1

[Stench, None, None, None, None]

INFERENCE BY RESOLUTION

R3 : $B_{2,1} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1})$

$(\alpha \Leftrightarrow \beta) \equiv ((\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha))$ biconditional elimination

R15 : $(B_{2,1} \Rightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1})) \wedge ((P_{1,1} \vee P_{2,2} \vee P_{3,1}) \Rightarrow B_{2,1})$

R16 : $B_{2,1} \Rightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1})$ (and elimination of R15)

But we already know that there's breeze in (2,1) via

R5 : $B_{2,1}$

Therefore $(P_{1,1} \vee P_{2,2} \vee P_{3,1})$ must be True (Modus Ponens), i.e.,

R17 : $(P_{1,1} \vee P_{2,2} \vee P_{3,1})$

1,4	2,4	3,4	4,4
1,3 w!	2,3	3,3	4,3
1,2 A S OK	2,2 OK	3,2	4,2
1,1 v OK	2,1 B V OK	3,1 P!	4,1

[Stench, None, None, None, None]

UNIT RESOLUTION

R1 : $\neg P_{1,1}$

R13 : $\neg P_{2,2}$ (Do this as an exercise)

R17 : $(P_{1,1} \vee P_{2,2} \vee P_{3,1})$

► Combining R1 and R17 ,

R18 : $(P_{2,2} \vee P_{3,1})$

► Combining R13 and R18

R19 : $P_{3,1}$

► Hence, pit in (3,1)

$$\frac{(P_{1,1} \vee P_{2,2} \vee P_{3,1}) , \neg P_{1,1}}{(P_{2,2} \vee P_{3,1})}$$

$$\frac{(P_{2,2} \vee P_{3,1}) , \neg P_{2,2}}{(P_{3,1})}$$

where each ℓ is a literal and ℓ_i and m are **complementary literals**

1,4	2,4	3,4	4,4
1,3 W!	2,3	3,3	4,3
1,2 A S OK	2,2	3,2	4,2
1,1 v OK	2,1 B v OK	3,1 P!	4,1

[Stench, None, None, None, None]

$$\frac{\ell_1 \vee \dots \vee \ell_k, \quad m}{\ell_1 \vee \dots \vee \ell_{i-1} \vee \ell_{i+1} \vee \dots \vee \ell_k}$$

Unit resolution

FULL RESOLUTION

This says that resolution takes two clauses and produces a new clause containing all the literals of the two original clauses *except* the two complementary literals

$$\frac{P_{1,1} \vee \neg P_{3,1}, \quad \neg P_{1,1} \vee \neg P_{2,2}}{P_{3,1} \vee \neg P_{2,2}}$$

$$\frac{\ell_1 \vee \dots \vee \ell_k, \quad m_1 \vee \dots \vee m_n}{\ell_1 \vee \dots \vee \ell_{i-1} \vee \ell_{i+1} \vee \dots \vee \ell_k \vee m_1 \vee \dots \vee m_{j-1} \vee m_{j+1} \vee \dots \vee m_n}$$

where ℓ_i and m_j are complementary literals.

CONJUNCTIVE NORMAL FORM

The resolution rules that we studied in last couple of slides apply only on disjunction (\vee) of literals (or clauses as they are called).

Every sentence of propositional logic is logically equivalent to a conjunction of clauses.

A sentence expressed as a conjunction of clauses is said to be in **conjunctive normal form** or **CNF**.

Example: $(P_{3,3} \vee S_{3,1}) \wedge (\neg S_{3,1} \vee P_{1,2} \vee P_{3,3}) \wedge \dots$

$B_{1,1} \iff (P_{1,2} \vee P_{2,1})$ INTO CNF?

1. Eliminate \iff , replacing $\alpha \iff \beta$ with $(\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha)$.

$$(B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1})) \wedge ((P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1}) .$$

2. Eliminate \Rightarrow , replacing $\alpha \Rightarrow \beta$ with $\neg\alpha \vee \beta$:

$$(\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge (\neg(P_{1,2} \vee P_{2,1}) \vee B_{1,1}) .$$

3. CNF requires \neg to appear only in literals, so we “move \neg inwards”

$$(\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge ((\neg P_{1,2} \wedge \neg P_{2,1}) \vee B_{1,1}) .$$

4. Now we have a sentence containing nested \wedge and \vee operators applied to literals. We apply the distributivity law from Figure 7.11, distributing \vee over \wedge wherever possible.

$$(\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge (\neg P_{1,2} \vee B_{1,1}) \wedge (\neg P_{2,1} \vee B_{1,1}) .$$