ARTIFICIAL INTELLIGENCE

Lecture 13-14

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

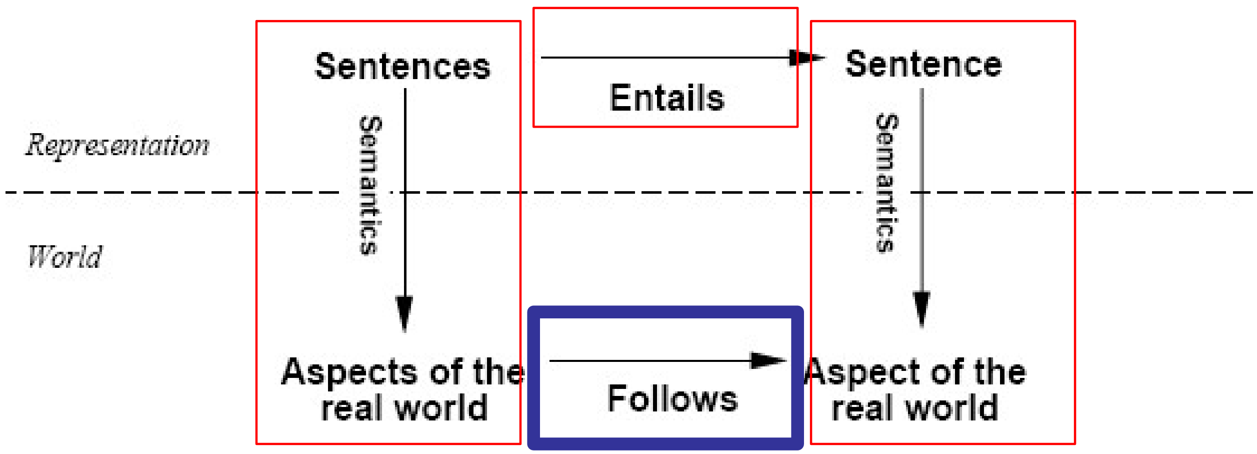
- E.g., the language of arithmetic:
  - $x+2 \geq y$ is a sentence
  - $-x2+y > \{\}$ is not a sentence

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.

$$\begin{array}{rc} & P \to Q \\ & P \\ \hline \therefore & Q \end{array}$$

"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. $(P \vee Q) \to Q$

SEMANTICS

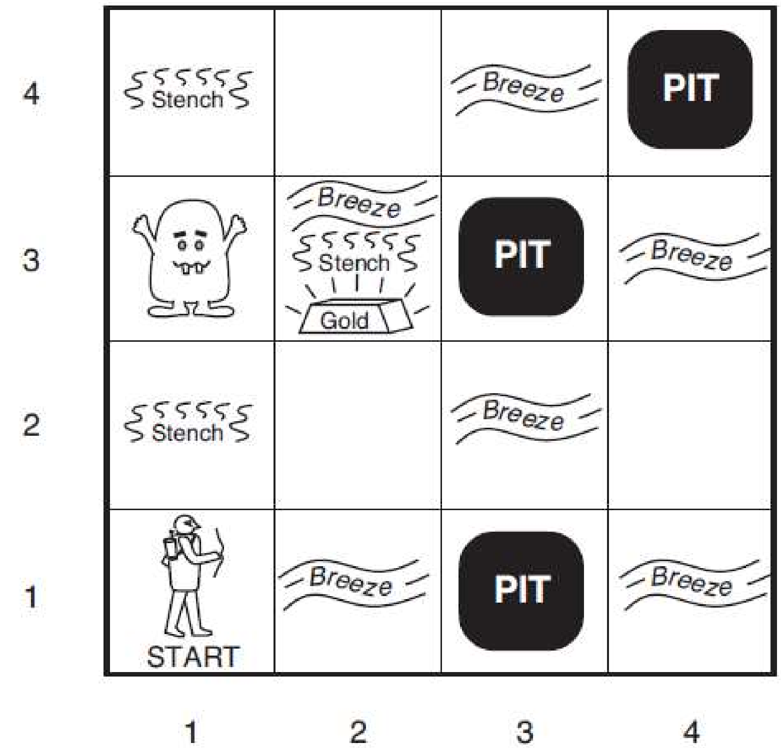
Sentences have a true value with respect to a model

E.g., model m = {$P_{1,2}$ = False, $P_{3,3}$=True, $S_{3,1}$=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(α) to mean the set of all models of α i.e., M(α) ={m1(α) ,m2(α) ,….} where all models m1,m2,…. satisfy α

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

M(α) $\longrightarrow$ 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) -> x=0

Beta (sentence 2) -> xy=0

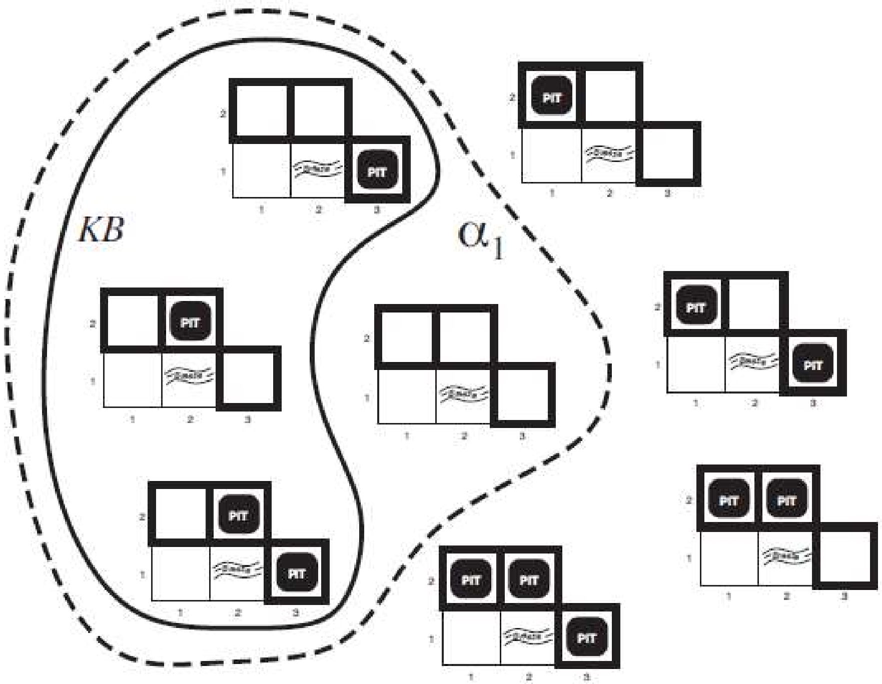
There's only one possible solution of x=0 and that is if x is 0 (x=0), therefore M(alpha) = {x=0}

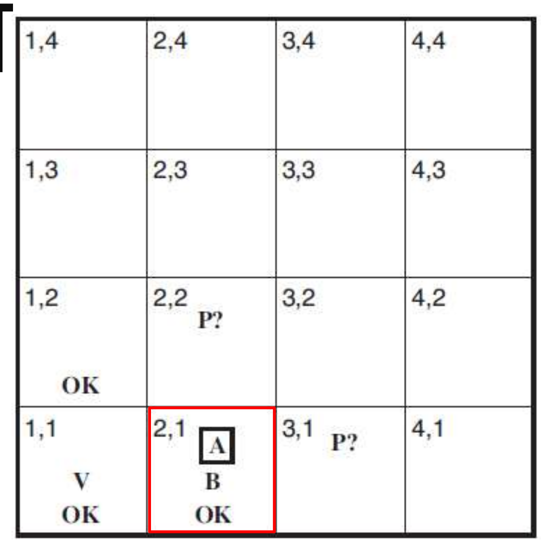
But there are many possible solutions of sentence 2, i.e., M(beta)={(x,y)=(0,0), (x,y)=(0,1), (x,y)=(1,0), …..} one of which is x=0

Since all possible solutions of alpha an 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



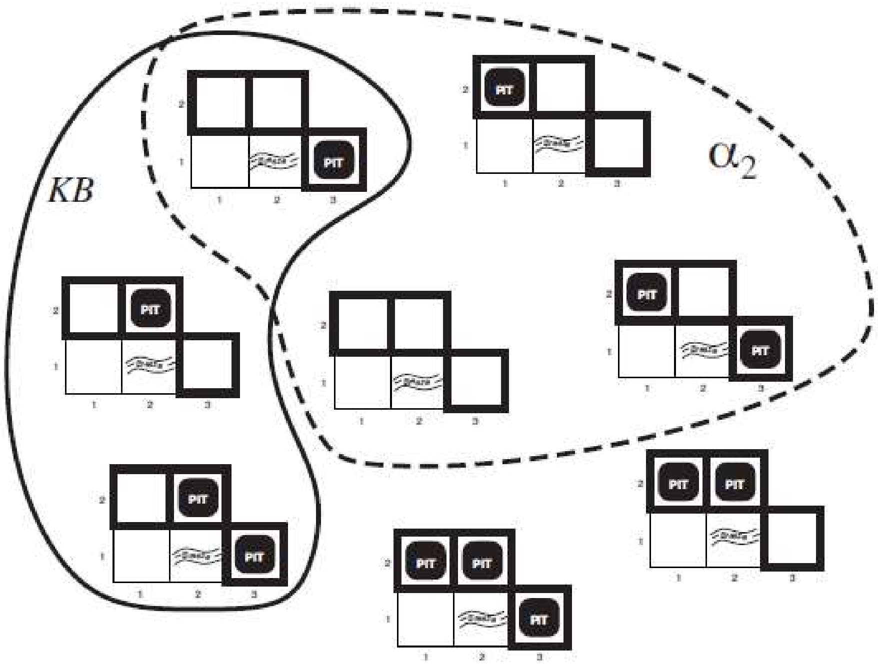


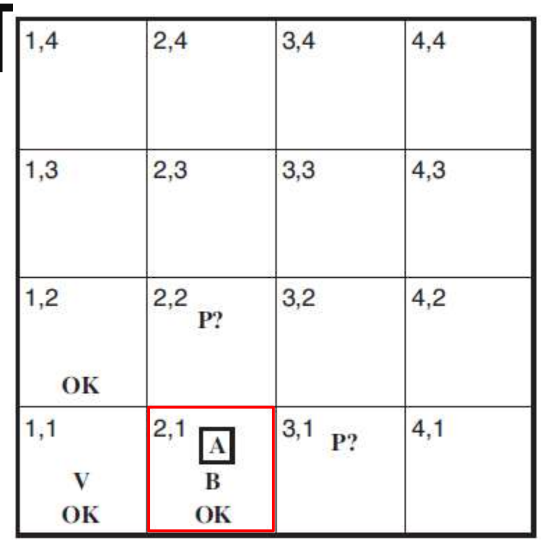
[None, Breeze, None, None, None]

KB |= α_1

Alpha 1 (sentence) -> no pits in [1, 2]

WUMPUS WORLD: ENTAILMENT





[None, Breeze, None, None, None]

Alpha 2 (sentence) -> 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.

$$
\begin{aligned}
AtomicSentence &\rightarrow True \mid False \mid P \mid Q \mid R \mid \ldots \\
ComplexSentence &\rightarrow (\ Sentence\ ) \mid [\ Sentence\ ]
\end{aligned}
$$

LOGICAL CONNECTIVITIES

¬ (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).

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

∨ (or). A sentence using ∨, 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 ∨ comes from the Latin "vel," which means "or." For most people, it is easier to remember ∨ as an upside-down ∧.)

LOGICAL CONNECTIVITIES

⇒ (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 $\rightarrow$.

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

IMPLICATION VS BICONDITIONAL

$P \Rightarrow Q$ (if P then Q)	$P \Leftrightarrow Q$ (P if and only if Q) or $Q \Leftrightarrow P$ (Q if and only if P)
If rainy then take umbrella (doesn't mean that if you have taken Umbrella, then it must rain)	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 a bullet hits my head, I'll die (doesn't mean that if I die, the bullet must have hit me)	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

$$
\begin{array}{rcl}
Sentence & \rightarrow & AtomicSentence \mid ComplexSentence \\
AtomicSentence & \rightarrow & True \mid False \mid P \mid Q \mid R \mid \ldots \\
ComplexSentence & \rightarrow & (\ Sentence\ ) \mid [\ Sentence\ ] \\
& \mid & \neg\ Sentence \\
& \mid & Sentence \land Sentence \\
& \mid & Sentence \lor Sentence \\
& \mid & Sentence \ \Rightarrow\ Sentence \\
& \mid & Sentence \ \Leftrightarrow\ Sentence \\
\\
\textsc{Operator Precedence} & : & \neg, \land, \lor, \Rightarrow, \Leftrightarrow
\end{array}
$$

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"

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

"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 = {$P_{1,2}$ =false, $P_{2,2}$ =false, $P_{3,1}$ =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

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

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

$$\neg P_{3,3} \lor S_{3,1} \Rightarrow P_{1,2} \land S_{3,1} \Leftrightarrow P_{3,3}$$

Operator Preference: $\neg$, $\land$, $\lor$, $\Rightarrow$, $\Leftrightarrow$

ENTAILMENT IN PROPOSITIONAL LOGIC

(x=0) |= (xy=0)

(p=True) |= (p V q)

p=True is True in 2 models both of which are subset of 3 models in which (p V q) is True

p	q	p V q
True	True	True
True	False	True
False	True	True
False	False	False

ENTAILMENT IN PROPOSITIONAL LOGIC

$(p \land q) \models (p \lor q)$

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

p	q	$p \lor q$	$p \land q$
True	True	True	True
True	False	True	False
False	True	True	False
False	False	False	False

ENTAILMENT IN PROPOSITIONAL LOGIC

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

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

p	q	r	$(p \Leftrightarrow q) \land r$	$q \Rightarrow p$
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

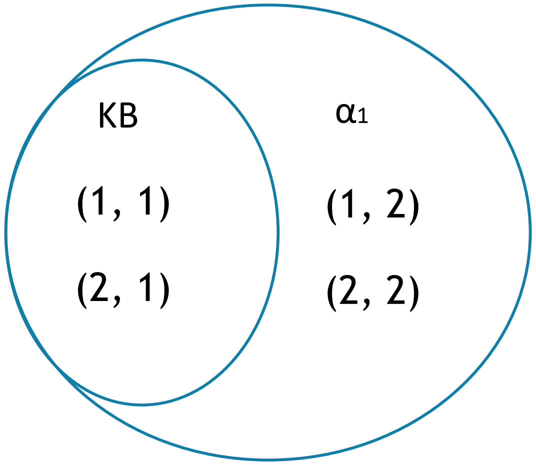
Imagine two variables, cleanliness and dependability (1 for very clean/dependabale; 3 for not at all)

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 roomates always score 1 or 2 in either cleanliness or dependability.

Question: Is John a good roommate?

KB $|= \alpha_1$?



USING ENTAILMENT FOR ANSWERS

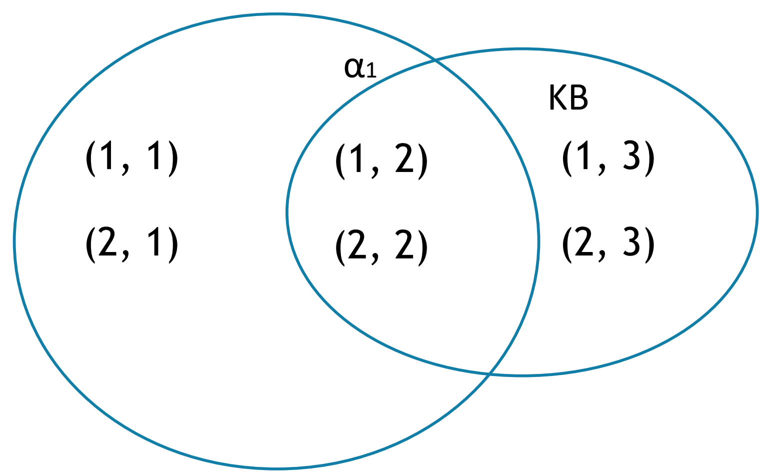
Imagine two variables, cleanliness and dependability (1 for very clean/dependabale; 3 for not at all)

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 roomates 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].

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

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

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:

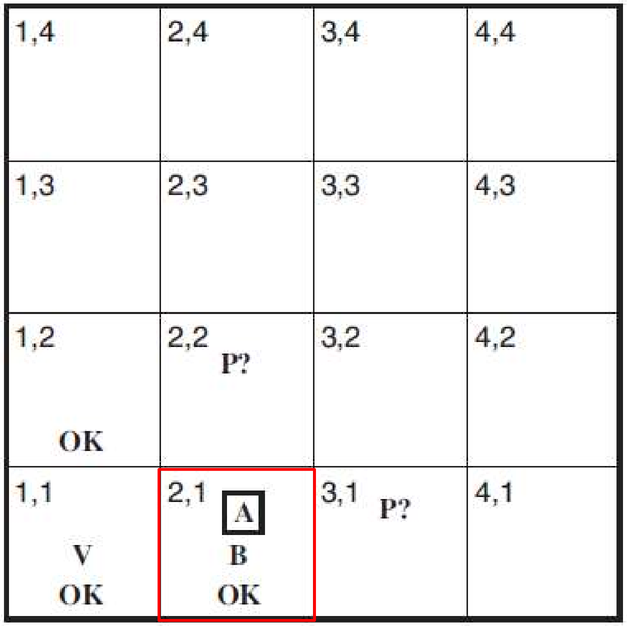
$R_2 : B_{1,1} \Leftrightarrow (P_{1,2} \lor P_{2,1})$

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

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.

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

$R_5 : B_{2,1}$



TELL RULES TO KB

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

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

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

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

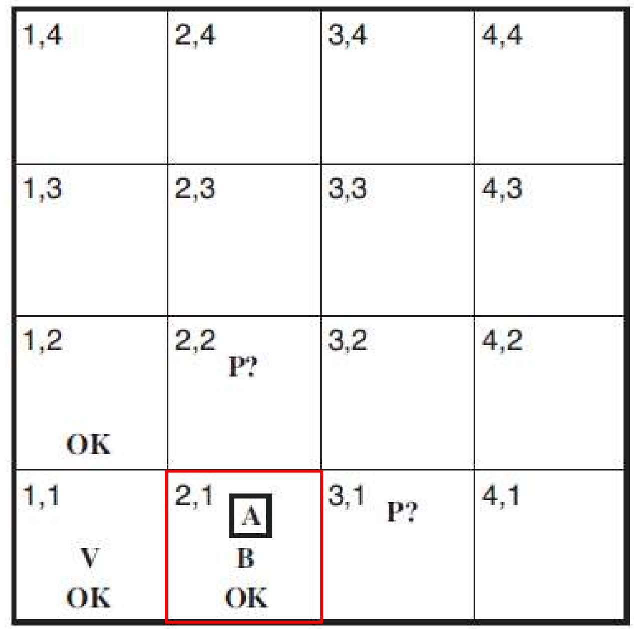
$R_3 : B_{2,1} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1})$

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

$R_5 : B_{2,1}$

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

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



ENTAILMENT: MODEL CHECKING

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

Find all possible models M(¬ P1,2) where ¬ P1,2 is True

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

If M(KB) is a subset of M(¬ P1,2), then KB $\models \neg P_{1,2}$

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

$B_{1,1}$	$B_{2,1}$	$P_{1,1}$	$P_{1,2}$	$P_{2,1}$	$P_{2,2}$	$P_{3,1}$	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*
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
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*
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
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

$$
\begin{aligned}
(\alpha \wedge \beta) &\equiv (\beta \wedge \alpha) && \text{commutativity of } \wedge \\
(\alpha \vee \beta) &\equiv (\beta \vee \alpha) && \text{commutativity of } \vee \\
((\alpha \wedge \beta) \wedge \gamma) &\equiv (\alpha \wedge (\beta \wedge \gamma)) && \text{associativity of } \wedge \\
((\alpha \vee \beta) \vee \gamma) &\equiv (\alpha \vee (\beta \vee \gamma)) && \text{associativity of } \vee \\
\neg(\neg\alpha) &\equiv \alpha && \text{double-negation elimination} \\
(\alpha \Rightarrow \beta) &\equiv (\neg\beta \Rightarrow \neg\alpha) && \text{contraposition} \\
(\alpha \Rightarrow \beta) &\equiv (\neg\alpha \vee \beta) && \text{implication elimination} \\
(\alpha \Leftrightarrow \beta) &\equiv ((\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha)) && \text{biconditional elimination} \\
\neg(\alpha \wedge \beta) &\equiv (\neg\alpha \vee \neg\beta) && \text{De Morgan} \\
\neg(\alpha \vee \beta) &\equiv (\neg\alpha \wedge \neg\beta) && \text{De Morgan} \\
(\alpha \wedge (\beta \vee \gamma)) &\equiv ((\alpha \wedge \beta) \vee (\alpha \wedge \gamma)) && \text{distributivity of } \wedge \text{ over } \vee \\
(\alpha \vee (\beta \wedge \gamma)) &\equiv ((\alpha \vee \beta) \wedge (\alpha \vee \gamma)) && \text{distributivity of } \vee \text{ over } \wedge
\end{aligned}
$$

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 ⇒ q	¬ q	¬p	¬ q ⇒ ¬ 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

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

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, ($R1 \land R2 \land R3 \land R4 \land R5$), 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

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

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

AND elimination:

- If (alpha ∧ 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: $R_1 \wedge R_2 \wedge R_3 \wedge R_4 \wedge R_5$

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

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

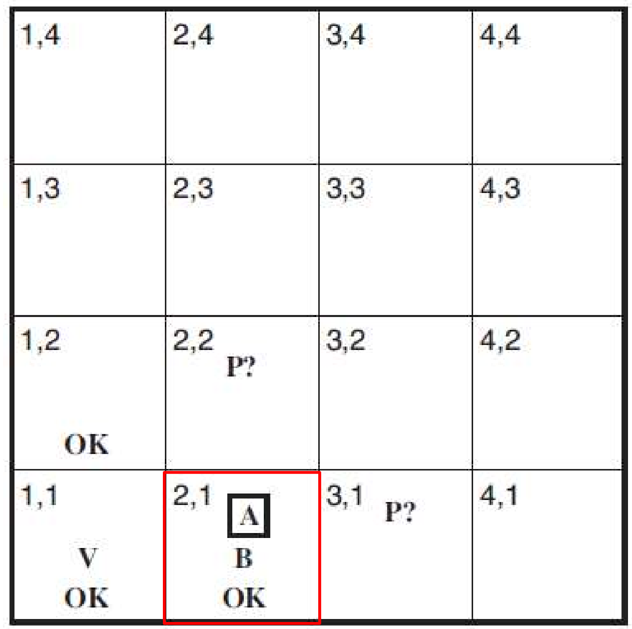
$R_3 : B_{2,1} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1})$

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

$R_5 : B_{2,1}$

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

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



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

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

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

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

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

$R_5 : B_{2,1}$

$R_6 : (B_{1,1} \Rightarrow (P_{1,2} \lor P_{2,1})) \quad \land \quad ((P_{1,2} \lor P_{2,1}) \Rightarrow B_{1,1})$

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

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

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

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

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

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

$R_5 : B_{2,1}$

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

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

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

$$
\begin{array}{rcll}
(\alpha \wedge \beta) & \equiv & (\beta \wedge \alpha) & \text{commutativity of } \wedge \\
(\alpha \vee \beta) & \equiv & (\beta \vee \alpha) & \text{commutativity of } \vee \\
((\alpha \wedge \beta) \wedge \gamma) & \equiv & (\alpha \wedge (\beta \wedge \gamma)) & \text{associativity of } \wedge \\
((\alpha \vee \beta) \vee \gamma) & \equiv & (\alpha \vee (\beta \vee \gamma)) & \text{associativity of } \vee \\
\neg(\neg \alpha) & \equiv & \alpha & \text{double-negation elimination} \\
(\alpha \Rightarrow \beta) & \equiv & (\neg \beta \Rightarrow \neg \alpha) & \text{contraposition} \\
(\alpha \Rightarrow \beta) & \equiv & (\neg \alpha \vee \beta) & \text{implication elimination} \\
(\alpha \Leftrightarrow \beta) & \equiv & ((\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha)) & \text{biconditional elimination} \\
\neg(\alpha \wedge \beta) & \equiv & (\neg \alpha \vee \neg \beta) & \text{De Morgan} \\
\neg(\alpha \vee \beta) & \equiv & (\neg \alpha \wedge \neg \beta) & \text{De Morgan} \\
(\alpha \wedge (\beta \vee \gamma)) & \equiv & ((\alpha \wedge \beta) \vee (\alpha \wedge \gamma)) & \text{distributivity of } \wedge \text{ over } \vee \\
(\alpha \vee (\beta \wedge \gamma)) & \equiv & ((\alpha \vee \beta) \wedge (\alpha \vee \gamma)) & \text{distributivity of } \vee \text{ over } \wedge
\end{array}
$$

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

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

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

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

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

$R_5 : B_{2,1}$

$R_6 : (B_{1,1} \Rightarrow (P_{1,2} \lor P_{2,1})) \land ((P_{1,2} \lor P_{2,1}) \Rightarrow B_{1,1})$

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

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

$$
\begin{array}{rcll}
(\alpha \land \beta) & \equiv & (\beta \land \alpha) & \text{commutativity of } \land \\
(\alpha \lor \beta) & \equiv & (\beta \lor \alpha) & \text{commutativity of } \lor \\
((\alpha \land \beta) \land \gamma) & \equiv & (\alpha \land (\beta \land \gamma)) & \text{associativity of } \land \\
((\alpha \lor \beta) \lor \gamma) & \equiv & (\alpha \lor (\beta \lor \gamma)) & \text{associativity of } \lor \\
\neg(\neg\alpha) & \equiv & \alpha & \text{double-negation elimination} \\
\boxed{(\alpha \Rightarrow \beta) \equiv (\neg\beta \Rightarrow \neg\alpha) \quad \text{contraposition}} & & & \\
(\alpha \Rightarrow \beta) & \equiv & (\neg\alpha \lor \beta) & \text{implication elimination} \\
(\alpha \Leftrightarrow \beta) & \equiv & ((\alpha \Rightarrow \beta) \land (\beta \Rightarrow \alpha)) & \text{biconditional elimination} \\
\neg(\alpha \land \beta) & \equiv & (\neg\alpha \lor \neg\beta) & \text{De Morgan} \\
\neg(\alpha \lor \beta) & \equiv & (\neg\alpha \land \neg\beta) & \text{De Morgan} \\
(\alpha \land (\beta \lor \gamma)) & \equiv & ((\alpha \land \beta) \lor (\alpha \land \gamma)) & \text{distributivity of } \land \text{ over } \lor \\
(\alpha \lor (\beta \land \gamma)) & \equiv & ((\alpha \lor \beta) \land (\alpha \lor \gamma)) & \text{distributivity of } \lor \text{ over } \land
\end{array}
$$

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

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

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

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

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

$R_5 : B_{2,1}$

$R_6 : (B_{1,1} \Rightarrow (P_{1,2} \lor P_{2,1})) \land ((P_{1,2} \lor P_{2,1}) \Rightarrow B_{1,1})$

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

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

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

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

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

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

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

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

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

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

$R_5 : B_{2,1}$

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

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

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

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

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

$(\alpha \wedge \beta)$	$\equiv$	$(\beta \wedge \alpha)$	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 $\vee$
$\neg(\neg\alpha)$	$\equiv$	α	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 $\wedge$

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

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

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

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

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

$R_5 : B_{2,1}$

$R_6 : (B_{1,1} \Rightarrow (P_{1,2} \lor P_{2,1})) \quad \land \quad ((P_{1,2} \lor P_{2,1}) \Rightarrow B_{1,1})$

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

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

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

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

R_{10}: neither [1,2] nor [2,1] contains a pit.

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 $\neg P_{1,2} \land \neg P_{2,1}$ does not mention the propositions $B_{2,1}$, $P_{1,1}$, $P_{2,2}$, or $P_{3,1}$.

EXERCISE

Prove $\neg P2,1$

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

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

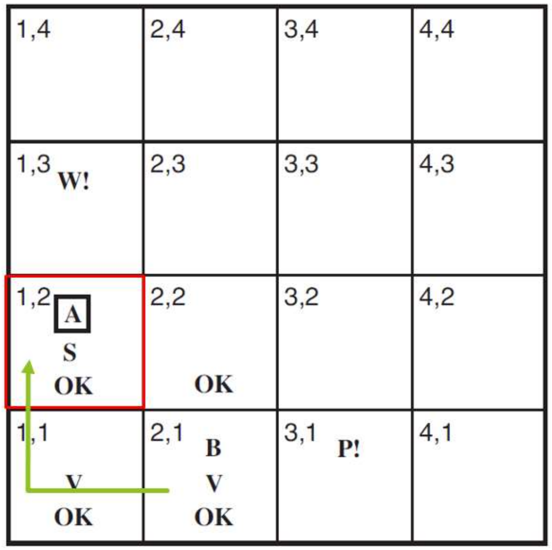
$R_3 : B_{2,1} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1})$

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

$R_5 : B_{2,1}$

$$
\begin{aligned}
(\alpha \wedge \beta) &\equiv (\beta \wedge \alpha) && \text{commutativity of } \wedge \\
(\alpha \vee \beta) &\equiv (\beta \vee \alpha) && \text{commutativity of } \vee \\
((\alpha \wedge \beta) \wedge \gamma) &\equiv (\alpha \wedge (\beta \wedge \gamma)) && \text{associativity of } \wedge \\
((\alpha \vee \beta) \vee \gamma) &\equiv (\alpha \vee (\beta \vee \gamma)) && \text{associativity of } \vee \\
\neg(\neg \alpha) &\equiv \alpha && \text{double-negation elimination} \\
(\alpha \Rightarrow \beta) &\equiv (\neg \beta \Rightarrow \neg \alpha) && \text{contraposition} \\
(\alpha \Rightarrow \beta) &\equiv (\neg \alpha \vee \beta) && \text{implication elimination} \\
(\alpha \Leftrightarrow \beta) &\equiv ((\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha)) && \text{biconditional elimination} \\
\neg(\alpha \wedge \beta) &\equiv (\neg \alpha \vee \neg \beta) && \text{De Morgan} \\
\neg(\alpha \vee \beta) &\equiv (\neg \alpha \wedge \neg \beta) && \text{De Morgan} \\
(\alpha \wedge (\beta \vee \gamma)) &\equiv ((\alpha \wedge \beta) \vee (\alpha \wedge \gamma)) && \text{distributivity of } \wedge \text{ over } \vee \\
(\alpha \vee (\beta \wedge \gamma)) &\equiv ((\alpha \vee \beta) \wedge (\alpha \vee \gamma)) && \text{distributivity of } \vee \text{ over } \wedge
\end{aligned}
$$

INFERENCE BY RESOLUTION



[Stench, None, None, None, None]

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] :

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

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

INFERENCE BY RESOLUTION

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

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

$R_{15} : (B_{2,1} \Rightarrow (P_{1,1} \lor P_{2,2} \lor P_{3,1})) \land ((P_{1,1} \lor P_{2,2} \lor P_{3,1}) \Rightarrow B_{2,1})$

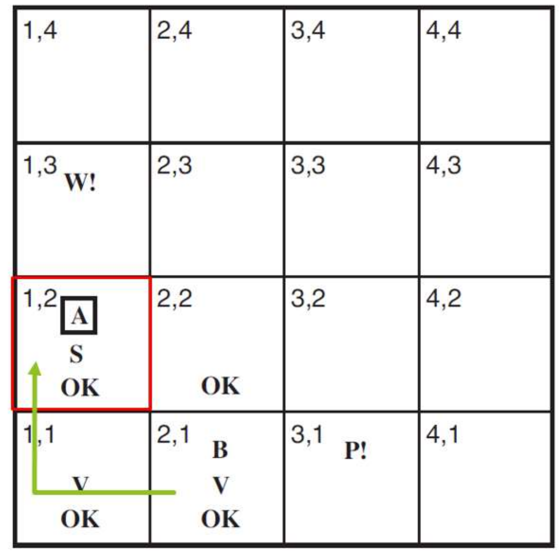
$R_{16} : B_{2,1} \Rightarrow (P_{1,1} \lor P_{2,2} \lor P_{3,1})$ (and elimination of R15)

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

$R_5 : B_{2,1}$

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

$R_{17} : (P_{1,1} \lor P_{2,2} \lor P_{3,1})$



[Stench, None, None, None, None]

UNIT RESOLUTION

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

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

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

- Combining R1 and R17 ,

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

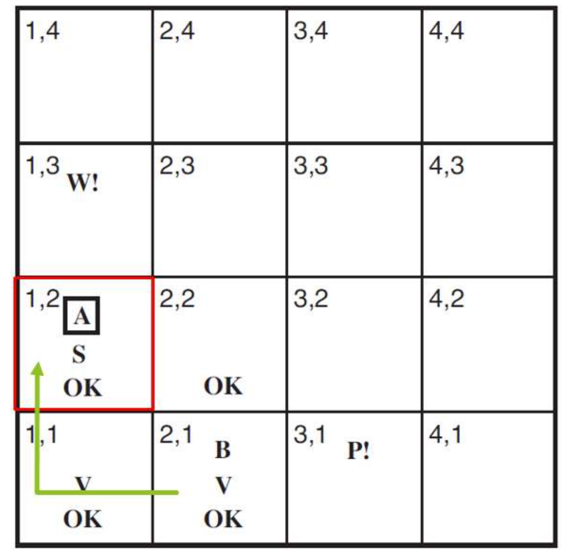
R18 : $(P_{2,2} \lor P_{3,1})$

- Combining R13 and R18

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

R19 : $P_{3,1}$

- Hence, pit in (3,1)



[Stench, None, None, None, None]

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

$$\boxed{\frac{\ell_1 \lor \cdots \lor \ell_k, \qquad m}{\ell_1 \lor \cdots \lor \ell_{i-1} \lor \ell_{i+1} \lor \cdots \lor \ell_k}}$$

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

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

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

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

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

1. Eliminate $\Leftrightarrow$, replacing $\alpha \Leftrightarrow \beta$ 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.

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