

# Knowledge-Based Systems

# Announcements

- Review sessions
- CS 4701 – focus on AI

# Schedule

- Search
- Machine learning
- Knowledge based systems
- Discovery



# History of AI

## 1943 – 1969 The Beginnings

1943 McCulloch and Pitts show networks of neurons can compute and learn any function

1950 Shannon and Turing wrote chess programs

1951 Minsky and Edmonds build the first neural network computer (SNARC)

1956 Dartmouth Conference – Newell and Simon brought a reasoning program “The Logic Theorist” which proved theorems.

1952 Samuel’s checkers player

1958 McCarthy designed LISP, helped invent time-sharing and created Advice Taker (a domain independent reasoning system)

1960’s Microworlds – solving limited problems: SAINT (1963), ANALOGY (1968), STUDENT (1967), blocksworld invented.

1962 Perceptron Convergence Theorem is proved.

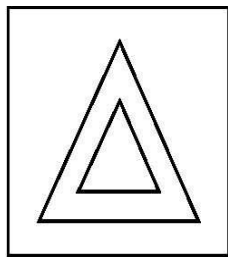


1952 Samuel's checkers player o TV

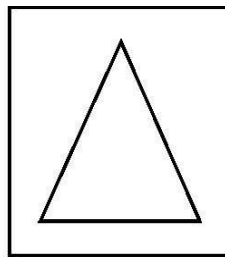


Arthur Samuel (1901-1990)

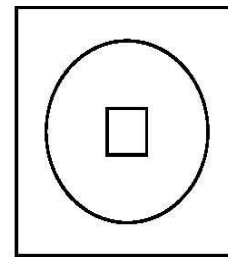
# Example ANALOGY Problem



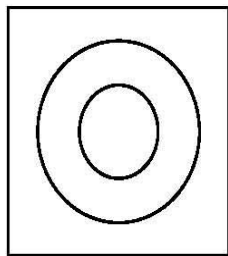
is to



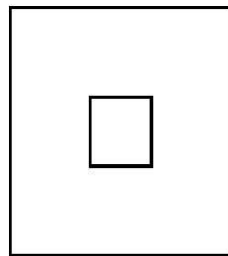
as



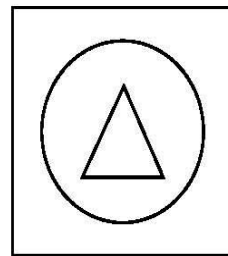
is to:



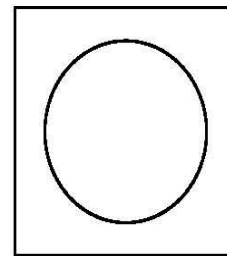
1



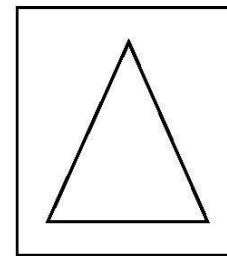
2



3

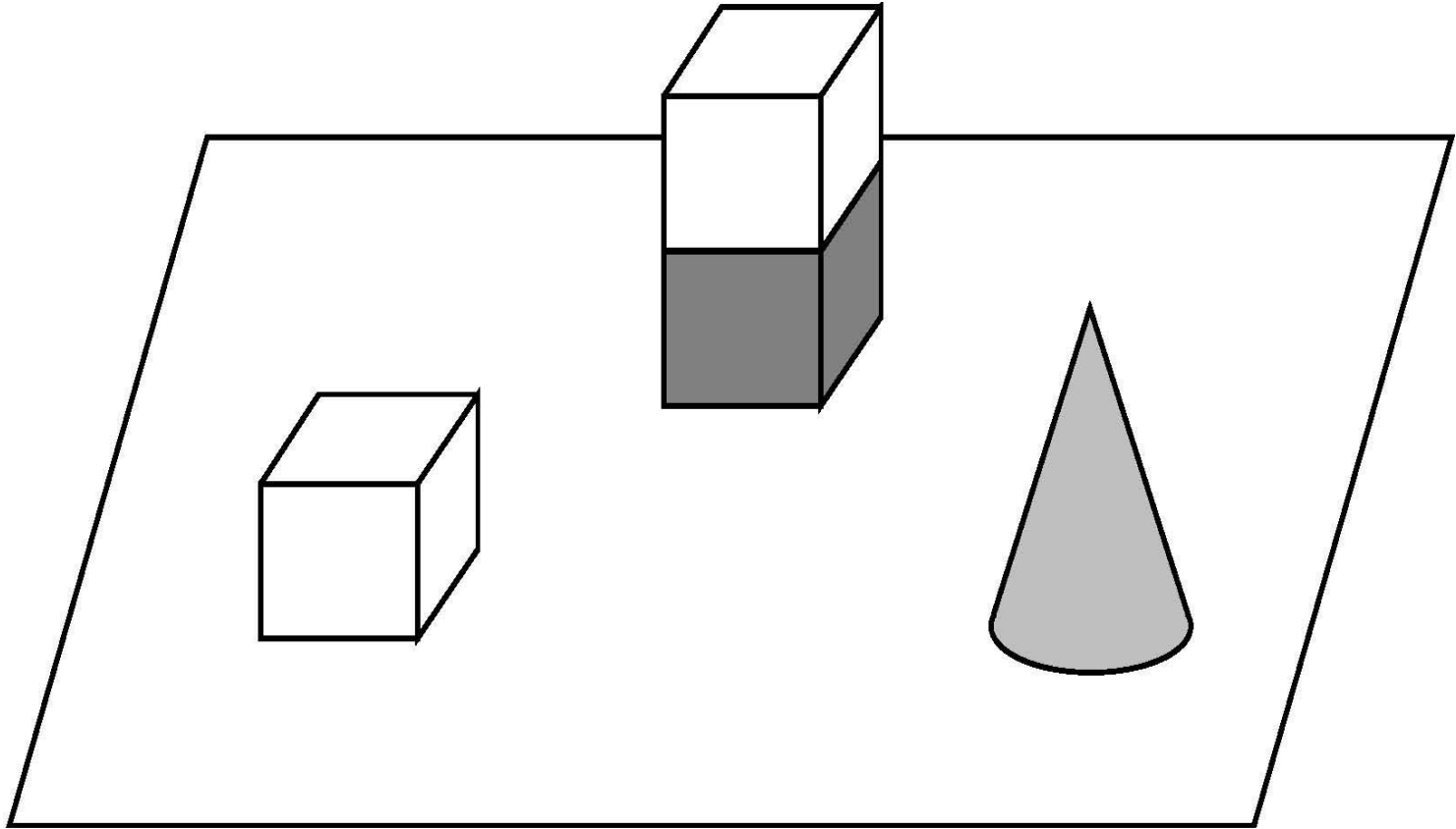


4



5

# Blocksworld





# History of AI

## 1966 – 1974 Recognizing Lack of Knowledge

- Herb Simon (1957): Computer chess program will be world chess champion within 10 years.
  - Intractable problems, lack of computing power (Lighthill Report, 1973)
  - Machine translation
  - Limitations in knowledge representation (Minsky and Papert, 1969)
- ➔ Knowledge-poor programs

# Knowledge Representation

- Human intelligence relies on a lot of background knowledge
  - the more you know, the easier many tasks become
  - “knowledge is power”
  - E.g. SEND + MORE = MONEY puzzle.
- Natural language understanding
  - Time flies like an arrow.
  - Fruit flies like a banana.
  - John saw the diamond through the window and coveted **it**
  - John threw the brick through the window and broke **it**
  - The spirit is willing but the flesh is weak. (English)
  - The vodka is good but the meat is rotten. (Russian)
- Or: Plan a trip to L.A.

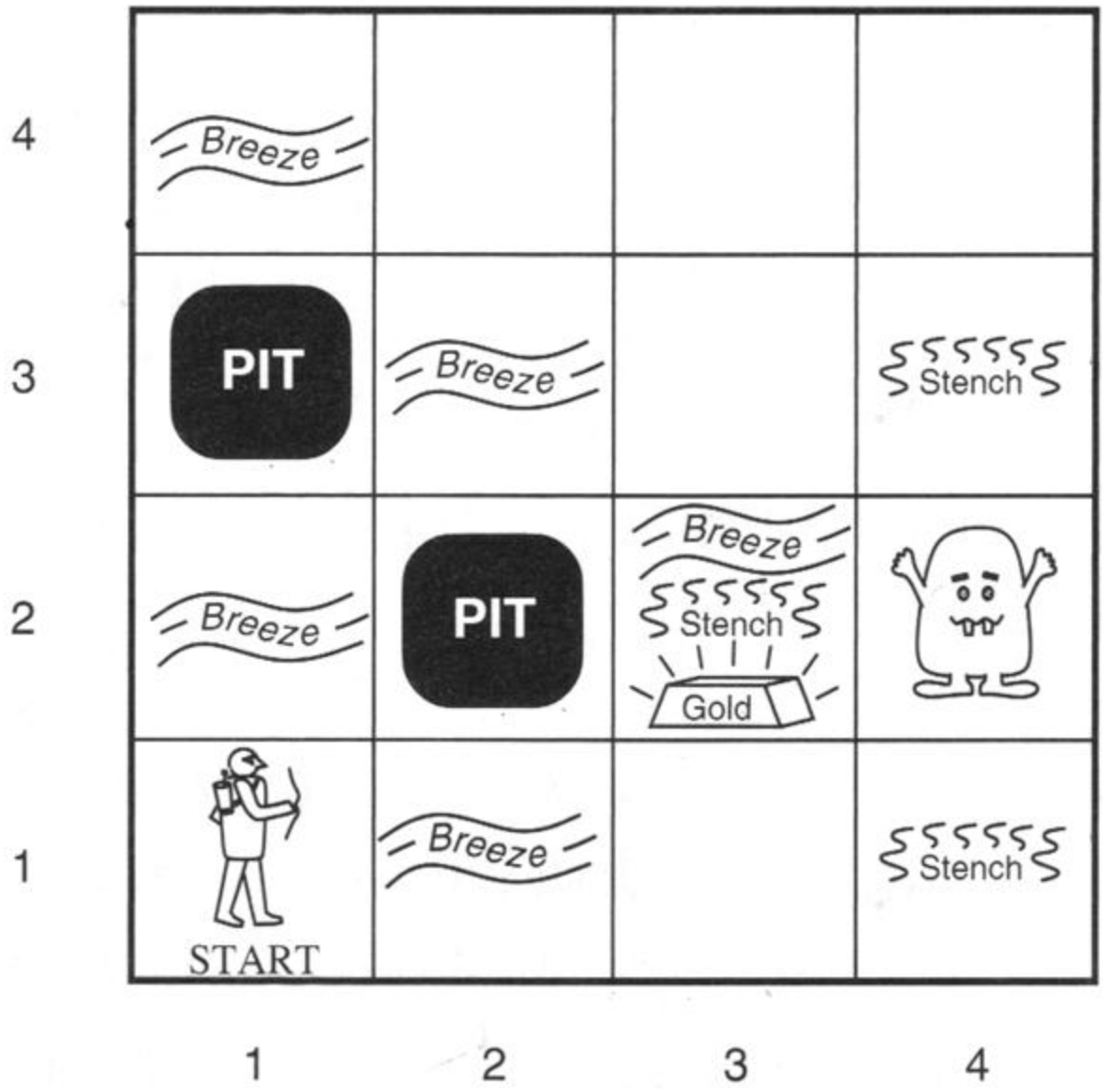
# Domain knowledge

- How did we encode domain knowledge so far?
  - For search problems?
  - For learning problems?

# Knowledge-Based Systems/Agents

- Key components:
  - Knowledge base: a set of sentences expressed in some knowledge representation language
  - Inference/reasoning mechanisms to query what is known and to derive new information or make decisions.
- Natural candidate:
  - logical language (propositional/first-order)
  - combined with a logical inference mechanism
- How close to human thought?
  - In any case, appears reasonable strategy for machines.

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



- 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  <b>OK</b>	2,2	3,2	4,2
1,1 <b>A</b> <b>OK</b>	2,1  <b>OK</b>	3,1	4,1

- 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  <b>OK</b>	2,2 <b>P?</b>	3,2	4,2
1,1  <b>V</b> <b>OK</b>	2,1 <b>A</b> <b>B</b> <b>OK</b>	3,1 <b>P?</b>	4,1

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

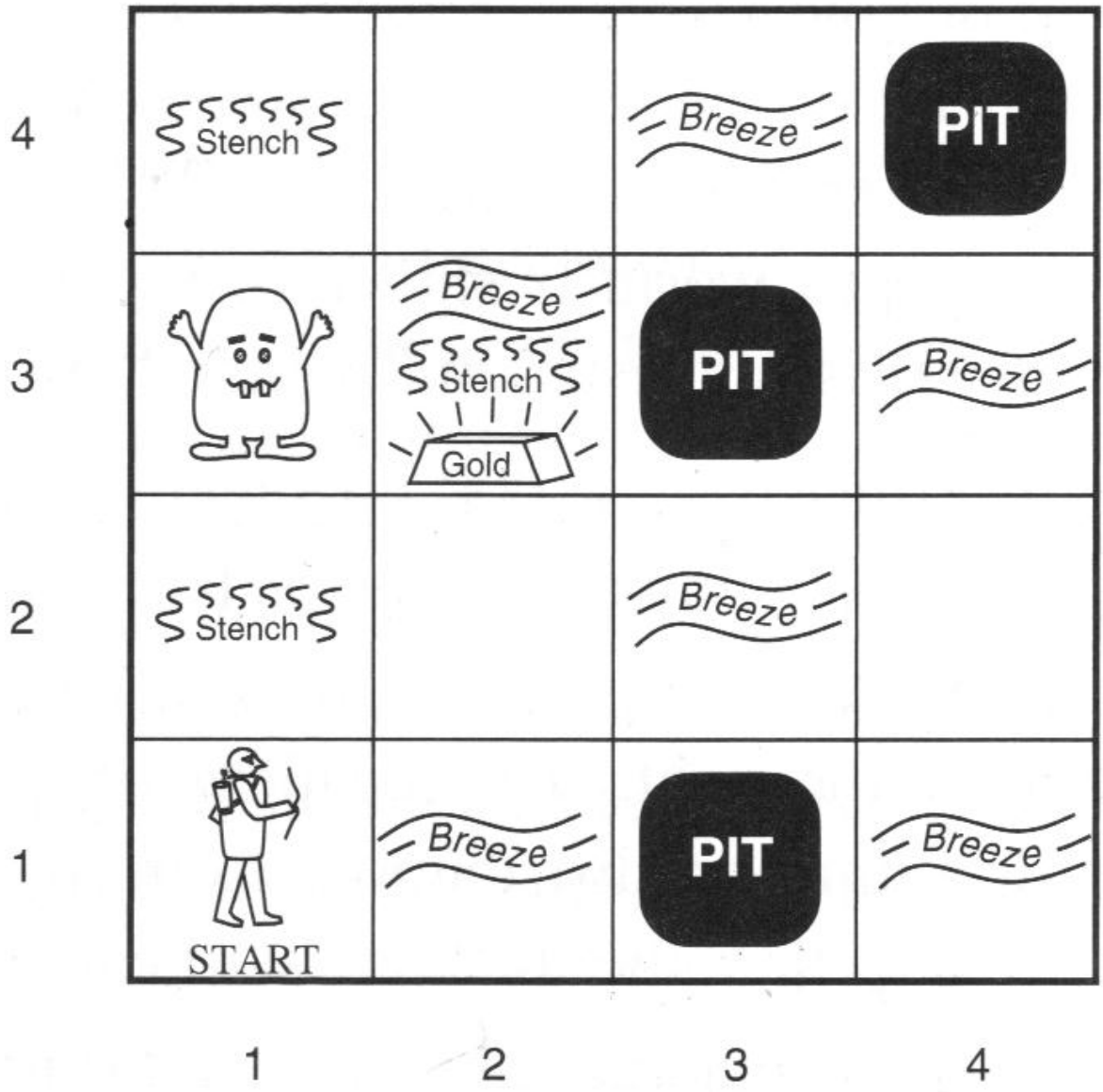
1,4	2,4	3,4	4,4
1,3 W!	2,3	3,3	4,3
1,2 <b>A</b> S OK	2,2 OK	3,2	4,2
1,1 V OK	2,1 B V OK	3,1 P!	4,1



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

1,4	2,4 <b>P?</b>	3,4	4,4
1,3 <b>W!</b>	2,3 <b>A</b> <b>S G</b> <b>B</b>	3,3 <b>P?</b>	4,3
1,2 <b>S</b> <b>V</b> <b>OK</b>	2,2 <b>V</b> <b>OK</b>	3,2	4,2
1,1 <b>V</b> <b>OK</b>	2,1 <b>B</b> <b>V</b> <b>OK</b>	3,1 <b>P!</b>	4,1

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



# Example: Autonomous Car

State: k-tuple

(PersonInFrontOfCar, Policeman, Policecar, Slippery, YellowLight, RedLight)

Actions:

Brake, Accelerate, TurnLeft, etc.

Knowledge-base describing when the car should brake:

( PersonInFrontOfCar  $\Rightarrow$  Brake )

(( ( YellowLight  $\wedge$  Policeman )  $\wedge$  ( $\neg$ Slippery ) )  $\Rightarrow$  Brake )

( Policecar  $\Rightarrow$  Policeman )

( Snow  $\Rightarrow$  Slippery )

( Slippery  $\Rightarrow$   $\neg$ Dry )

( RedLight  $\Rightarrow$  Brake )

**Does (Policecar, YellowLight, Snow) imply Brake?**

**A=Yes B= No**

# What the computer “sees”:

State: k-tuple

$(x_1, x_2, x_3, x_4, x_5, x_6, x_7)$

Actions:

$x_8, x_9, x_{10}$ , etc.

Knowledge-base describing when x:

$(x_1 \Rightarrow x_8)$

$((x_5 \wedge x_2) \wedge (\neg x_4)) \Rightarrow x_8$

$(x_3 \Rightarrow x_2)$

$(x_7 \Rightarrow x_4)$

$(x_4 \Rightarrow \neg x_{11})$

$(x_6 \Rightarrow x_8)$

**Does  $(x_3, x_5, x_7)$  imply  $x_8$ ?**

**A=Yes B= No**

# Logic as a Knowledge Representation

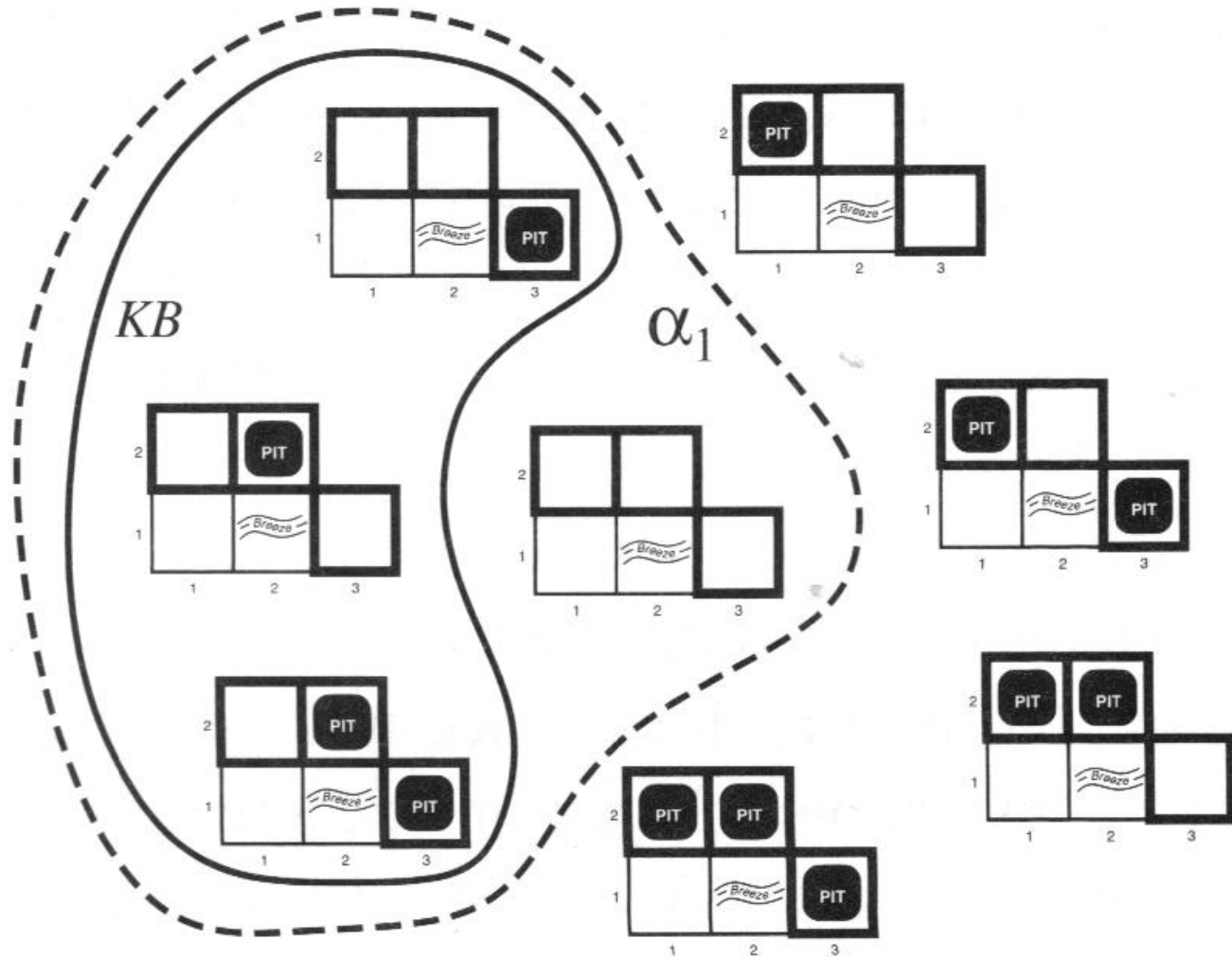
- Components of a Formal Logic:
  - Variables and operators, syntax
  - semantics (link to the world, truth in worlds)
  - logical reasoning: entailment  $\alpha \models \beta$ 
    - if, in every **model** in which  $\alpha$  is true,  $\beta$  is also true.
  - inference algorithm **derives**
    - $KB \Rightarrow \alpha$ , i.e.,  $\alpha$  is derived from KB.

( $x+y=4$ ) entails that

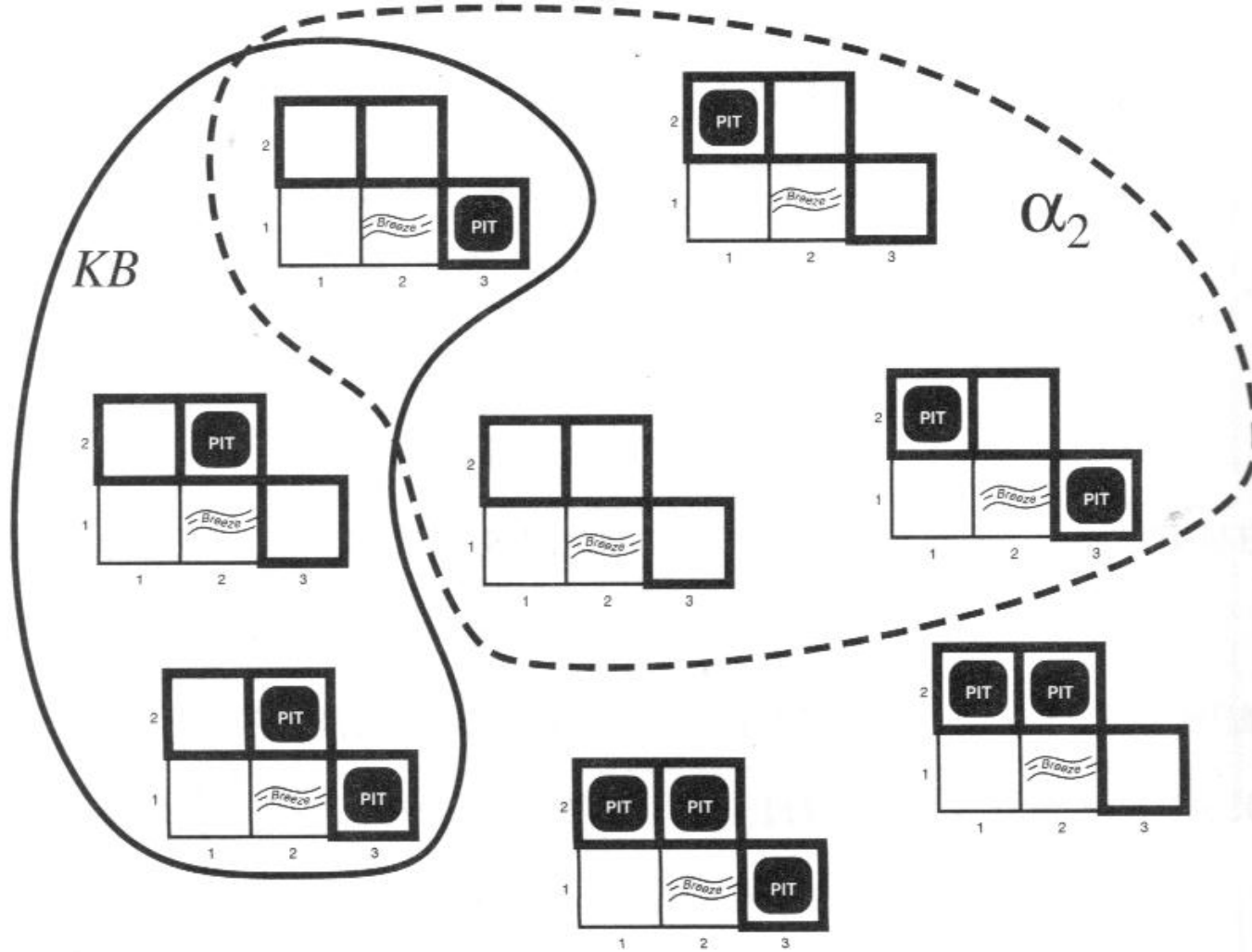
A)  $x=2$   $y=2$     B)  $2x+2y=8$     C) Neither    D) Both

# Models

- Model is an instantiation of all variables
- All models = all possible assignments
- Sentence  $\alpha$  is true in model  $m$ , then  $m$  is a model of  $\alpha$
- $M(\alpha)$  refers to the set of all models that satisfy  $\alpha$
- $\alpha \models \beta$  iff  $M(\alpha) \subseteq M(\beta)$ 
  - $\beta$  iff  $M(\alpha)$  is contained in  $M(\beta)$



Possible models for the presence of pits in [1,2] [2,2] [3,1]  
 Dashed =  $M(\alpha_1)$  where  $\alpha_1 = \neg P_{1,2}$  (no pit in [1,2])  
 Solid =  $M(KB)$  with observation of  $\neg B_{1,1} \wedge B_{2,1}$  (no breeze in [1,1] and breeze in [2,1])



*KB*

$\alpha_2$

Possible models for the presence of pits in [1,2] [2,2] [3,1]  
 Dashed =  $M(\alpha_2)$  where  $\alpha_2 = \neg P_{2,2}$  (no pit in [2,2])  
 Solid =  $M(KB)$  with observation of  $\neg B_{1,1} \wedge B_{2,1}$  (no breeze in [1,1] and breeze in [2,1])



# Soundness and Completeness

Soundness:

An inference algorithm that derives only entailed sentences is called *sound* or *truth-preserving*.

$$KB \Rightarrow \alpha \text{ implies } KB \models \alpha$$

Completeness:

An inference algorithm is *complete* if it can derive any sentence that is entailed.

$$KB \models \alpha \text{ implies } KB \Rightarrow \alpha$$

Why soundness and completeness important?

→ Allow computer to ignore semantics and “just push symbols”!

# IS STRICT IMPLICATION THE SAME AS ENTAILMENT ?

By AUSTIN E. DUNCAN-JONES

## I Entailment and the modal functions

In this note I shall try to show how strict implication, as developed by C. I. Lewis (in *symbolic logic*, Lewis and Langford), differs from entailment, and to suggest certain requirements which a calculus of entailment would have to satisfy.<sup>1</sup>

Consider the two expressions (1)  $xRy.yRz$  and (2)  $xRz$ . Suppose we let  $x$   $y$  and  $z$  be classes and  $R$  the relation of inclusion ; or suppose we let  $x$   $y$  and  $z$  be propositions, and  $R$  the relation of material implication, or strict implication, or entailment : then most people would admit that (2) is materially implied by (1), that (2) follows from (1), and perhaps that (2) is

<sup>1</sup> The general point of view which I am adopting is much the same as that of Everett J. Nelson in *intensional relations*, *Mind* 1930. I had already worked out this paper in outline before I discovered the existence of Nelson's paper : otherwise the note would have been based directly on Nelson. I differ from Nelson on one or two points to be mentioned below.

# Entailment vs. Implication

- *Entailment* ( $KB \models \alpha$ ) and *implication* ( $KB \Rightarrow \alpha$ ) can be treated equivalently if the inference process is sound and complete.

# Propositional Logic: Syntax

- Propositional Symbols
  - A, B, C, ...
- Connectives
  - $\wedge, \vee, \neg, \Rightarrow, \Leftrightarrow$
- Sentences
  - Atomic Sentence: True, False, Propositional Symbol
  - Complex Sentence:
    - $(\neg \text{Sentence})$
    - $(\text{Sentence} \vee \text{Sentence})$
    - $(\text{Sentence} \wedge \text{Sentence})$
    - $(\text{Sentence} \Rightarrow \text{Sentence})$
    - $(\text{Sentence} \Leftrightarrow \text{Sentence})$
- A KB is a conjunction (ANDs) of many sentences

# Example: Autonomous Car

## Propositional Symbols

PersonInFrontOfCar, Policeman, .. Brake, Accelerate, TurnLeft

## Rules:

( PersonInFrontOfCar  $\Rightarrow$  Brake )  
 $\wedge$  ( (( YellowLight  $\wedge$  Policeman )  $\wedge$  ( $\neg$ Slippery ))  $\Rightarrow$  Brake )  
 $\wedge$  ( Policecar  $\Rightarrow$  Policeman )  
 $\wedge$  ( Snow  $\Rightarrow$  Slippery )  
 $\wedge$  ( Slippery  $\Rightarrow$   $\neg$ Dry )  
 $\wedge$  ( RedLight  $\Rightarrow$  Brake )

Initial  
KB

## from sensors:

YellowLight  
 $\wedge$   $\neg$ RedLight  
 $\wedge$   $\neg$ Snow  
 $\wedge$  Dry  
 $\wedge$  Policecar  
 $\wedge$   $\neg$ PersonInFrontOfCar

Added to  
KB

# Propositional Logic: Semantics

$P$	$Q$	$\neg P$	$P \wedge Q$	$P \vee Q$	$P \Rightarrow Q$	$P \Leftrightarrow Q$
<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>

## Models

- Model (i.e. possible world):
  - Assignment of truth values to symbols
  - Example:  $m = \{P = \text{True}, Q = \text{False}\}$ 
    - Note: Often called “assignment” instead of “model”, and “model” is used for an assignment that evaluates to true.
- Validity:
  - A sentence  $\alpha$  is valid, if it is true in every model.
- Satisfiability:
  - A sentence  $\alpha$  is satisfiable, if it is true in at least one model.
- Entailment:
  - $\alpha \models \beta$  if and only if, in every model in which  $\alpha$  is true,  $\beta$  is also true.

# Stay at home

- Sick      StayAtHome
- true      true
- false     false
- false     true

**Does *Sick* entail *StayAtHome*?**

**A=Yes B=No**

# Puzzling aspects of Propositional Logic

- Non causality
  - (5 is odd  $\Rightarrow$  Tokyo is the capital of Japan)
    - True, because whenever 5 is odd, Tokyo is the capital of Japan. Nothing to do with causality
- Statement always true when antecedent is false
  - (5 is even  $\Rightarrow$  Sam is smart)
    - True, because 5 is never even, so no models where this statement is incorrect, regardless of whether Sam is smart or not
- $A \Rightarrow B$ 
  - read: B is true whenever A is true



# Propositional Logic: Semantics

$P$	$Q$	$\neg P$	$P \wedge Q$	$P \vee Q$	$P \Rightarrow Q$	$P \Leftrightarrow Q$
<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>

**Models**

$$\neg(P \vee Q) \Leftrightarrow (\neg P) \wedge (\neg Q)$$

**A) True B) False**

# Creating a KB

- Variables

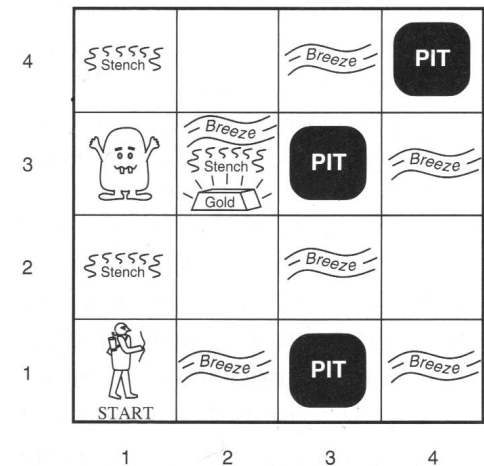
- $P_{i,j}$  is true if there is a pit at position  $(i,j)$
- $B_{i,j}$  is true if there is a breeze at position  $(i,j)$

- Knowledge

- R1:  $\neg P_{1,1}$       *There is no pit in [1,1]*
- R2:  $B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$       *Square is breezy iff next to pit*
- R3:  $B_{2,1} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1})$

- Perceptions

- R4:  $\neg B_{1,1}$       *There is no breeze in [1,1]*
- R5:  $B_{2,1}$       *There is breeze in [2,1]*



# Model Checking

- Idea:
  - To test whether  $\alpha \models \beta$ , enumerate all models and check truth of  $\alpha$  and  $\beta$ .
  - $\alpha$  entails  $\beta$  if no model exists in which  $\alpha$  is true and  $\beta$  is false (i.e.  $(\alpha \wedge \neg\beta)$  is unsatisfiable)
- Proof by Contradiction:  
 $\alpha \models \beta$  if and only if the sentence  $(\alpha \wedge \neg\beta)$  is unsatisfiable.

# Example of model checking

P	Q	$\neg P$	$Q \rightarrow P$	$\neg P \wedge (Q \rightarrow P)$	$(\neg P \wedge (Q \rightarrow P)) \wedge Q$
T	T	F	T	F	F
T	F	F	T	F	F
F	T	T	F	F	F
F	F	T	T	T	F

## Models

- $\alpha \models \beta$  iff the sentence  $(\alpha \wedge \neg\beta)$  is unsatisfiable
- Prove that  $(\neg P \wedge (Q \rightarrow P)) \rightarrow \neg Q$ 
  - By showing that  $[(\neg P \wedge (Q \rightarrow P)) \wedge Q]$  is not satisfiable
- Possible English translation:
  - P="The street is wet"
  - Q="It is raining"
  - Does "The street not wet" ( $\neg P$ ) and "it is raining  $\rightarrow$  street is wet" ( $Q \rightarrow P$ ) imply that "It is not raining?" ( $\neg Q$ )?
- Test if  $[(\neg P \wedge (Q \rightarrow P)) \wedge Q]$  is satisfiable.
  - It is not satisfiable (always false), therefore  $(\neg P \wedge (Q \rightarrow P))$  entails  $\neg Q$

# Model Chekcing

- Variables: One for each propositional symbol
- Domains: {true, false}
- Objective Function:  $(\alpha \wedge \neg\beta)$
- Which search algorithm works best?

# Doesn't scale well...

$B_{1,1}$	$B_{2,1}$	$P_{1,1}$	$P_{1,2}$	$P_{2,1}$	$P_{2,2}$	$P_{3,1}$	$R_1$	$R_2$	$R_3$	$R_4$	$R_5$	$KB$
<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>true</i>	<i>true</i>	<i>true</i>	<i>true</i>	<i>false</i>	<i>false</i>
<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>true</i>	<i>true</i>	<i>true</i>	<i>false</i>	<i>true</i>	<i>false</i>	<i>false</i>
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
<i>false</i>	<i>true</i>	<i>false</i>	<i>false</i>	<i>false</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>false</i>	<i>true</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>true</i>	<i>false</i>	<i>true</i>	<i>true</i>	<i>true</i>	<i>true</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>true</i>	<i>true</i>	<i>true</i>	<i>true</i>	<i>true</i>	<i>true</i>	<u><i>true</i></u>
<i>false</i>	<i>true</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>true</i>	<i>true</i>	<i>true</i>	<i>true</i>	<i>true</i>	<i>true</i>	<i>true</i>	<u><i>true</i></u>
<i>false</i>	<i>true</i>	<i>false</i>	<i>false</i>	<i>false</i>	<i>true</i>	<i>true</i>	<i>true</i>	<i>true</i>	<i>true</i>	<i>true</i>	<i>true</i>	<u><i>true</i></u>
<i>false</i>	<i>true</i>	<i>false</i>	<i>false</i>	<i>true</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>
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
<i>true</i>	<i>true</i>	<i>true</i>	<i>true</i>	<i>true</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>

# Inference: Reasoning with Propositional Logic

**Modus Ponens:** [Latin](#) for “the way that affirms by affirming”  $(\alpha \Rightarrow \beta) \wedge \alpha \equiv \beta$

Know:  $\alpha \Rightarrow \beta$  If raining, then soggy courts.

and  $\alpha$  It is raining.

---

Then:  $\beta$  Soggy Courts.

**Modus Tollens:** [Latin](#) for “the way that denies by denying”  $(\alpha \Rightarrow \beta) \wedge \neg\beta \equiv \neg\alpha$

Know:  $\alpha \Rightarrow \beta$  If raining, then soggy courts.

And  $\neg\beta$  No soggy courts.

---

Then:  $\neg\alpha$  It is not raining.

**And-Elimination:**

Know:  $\alpha \wedge \beta$  It is raining and soggy courts.

---

Then:  $\alpha$  It is raining.

# Example: Forward Chaining

Knowledge-base describing when the car should brake?

- ( PersonInFrontOfCar  $\Rightarrow$  Brake )
- $\wedge$  ((( YellowLight  $\wedge$  Policeman )  $\wedge$  ( $\neg$ Slippery ))  $\Rightarrow$  Brake )
- $\wedge$  ( Policecar  $\Rightarrow$  Policeman )
- $\wedge$  ( Snow  $\Rightarrow$  Slippery )
- $\wedge$  ( Slippery  $\Rightarrow$   $\neg$ Dry )
- $\wedge$  ( RedLight  $\Rightarrow$  Brake )
- $\wedge$  ( Winter  $\Rightarrow$  Snow )

Observation from sensors:

YellowLight  $\wedge$   $\neg$ RedLight  $\wedge$   $\neg$ Snow  $\wedge$  Dry  $\wedge$  Policecar  $\wedge$   $\neg$ PersonInFrontOfCar

What can we infer?

- Policecar  $\wedge$  ( Policecar  $\Rightarrow$  Policeman ): Modus Ponens: Policeman
- Dry  $\wedge$  ( Slippery  $\Rightarrow$   $\neg$ Dry ): Modus Tollens:  $\neg$ Slippery
- YellowLight  $\wedge$  Policeman  $\wedge$   $\neg$ Slippery  $\wedge$  ((( YellowLight  $\wedge$  Policeman )  $\wedge$  ( $\neg$ Slippery ))  $\Rightarrow$  Brake ): Modus Ponens: Brake
- YellowLight  $\wedge$   $\neg$ RedLight: And Elimination: YellowLight

**Inferring ( $\neg$ Winter) from ( $\neg$ Snow  $\wedge$  ( Winter  $\Rightarrow$  Snow )) is**

**A) Modus Ponens B) Modus Tollens C) And elimination**



# Other rules

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

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

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

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

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

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

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

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

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

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

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

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

# Inference Strategy: Forward Chaining

Idea:

- Infer everything that can be inferred.
- Notation: In implication  $\alpha \Rightarrow \beta$ , we say that
  - $\alpha$  (or its components) are called premises,
  - $\beta$  is called consequent/conclusion.

Forward Chaining:

Given a fact  $p$  to be added to the KB,

1. Find all implications  $I$  that have  $p$  as a premise
2. For each  $i$  in  $I$ , holds
  - a) Add the consequent in  $i$  to the KB

Continue until no more facts can be inferred.

# Inference Strategy: Backward Chaining

Idea:

- Check whether a particular fact  $q$  is true.

Backward Chaining:

Given a fact  $q$  to be “proven”,

1. See if  $q$  is already in the KB. If so, return TRUE.
2. Find all implications,  $I$ , whose conclusion “matches”  $q$ .
3. Recursively establish the premises of all  $i$  in  $I$  via backward chaining.

→ Avoids inferring unrelated facts.

# Example: Backward Chaining

Knowledge-base describing when the car should brake:

( PersonInFrontOfCar  $\Rightarrow$  Brake )  
 $\wedge$  ((( YellowLight  $\wedge$  Policeman )  $\wedge$  ( $\neg$ Slippery ))  $\Rightarrow$  Brake )  
 $\wedge$  ( Policecar  $\Rightarrow$  Policeman )  
 $\wedge$  ( Snow  $\Rightarrow$  Slippery )  
 $\wedge$  ( Slippery  $\Rightarrow$   $\neg$ Dry )  
 $\wedge$  ( RedLight  $\Rightarrow$  Brake )  
 $\wedge$  ( Winter  $\Rightarrow$  Snow )

Observation from sensors:

YellowLight  $\wedge$   $\neg$ RedLight  $\wedge$   $\neg$ Snow  $\wedge$  Dry  $\wedge$  Policecar  $\wedge$   $\neg$ PersonInFrontOfCar

Should the agent brake (i.e. can “brake” be inferred)?

- Goal: Brake
  - Modus Ponens (brake): PersonInFrontOfCar
    - Failure: PersonInFrontOfCar  $\rightarrow$  Backtracking
- Goal: Brake
  - Modus Ponens (brake): YellowLight  $\wedge$  Policeman  $\wedge$   $\neg$ Slippery
  - Known (YellowLight): Policeman  $\wedge$   $\neg$ Slippery
  - Modus Ponens (Policeman): Policecar  $\wedge$   $\neg$ Slippery
  - Known (Policecar):  $\neg$ Slippery
  - Modus Tollens ( $\neg$ Slippery): Dry
  - Known (Dry)

# Conjunctive Normal Form

- Convert expressions into the form
  - $(l_{1,1} \vee \dots \vee l_{1,k}) \wedge \dots \wedge (l_{n,1} \vee \dots \vee l_{n,k})$
  - *Conjunction of disjunctions*
  - *k-CNF (k literals)*
- Every expression can be transformed into 3-CNF

# Conjunctive Normal Form

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

- Original  $R_2$  (From Wumpus)

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

- Biconditional elimination

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

- Implication elimination

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

- De Morgan

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

- Distributivity of  $\wedge$

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

# Conjunctive Normal Form

- Algorithms exist for 3-CNF
  - E.g. 3-SAT

