PAI_Module3

Topics:

1. Informed Search Strategies:
    a. **Greedy best-first search,**
    b. **A*search,**
    c. **Heuristic functions**.
2. Logical Agents:
    a. **Knowledge–based agents,**
    b. **The Wumpus world,**
    c. **Logic,**
    d. **Propositional logic, Reasoning patterns in Propositional Logic**

3. 1 Informed Search Strategies

Informed Search: Informed search is a search strategy that utilizes problem-specific knowledge, to find solutions more efficiently. Informed search methods make use of heuristics and evaluation functions to guide the search towards more promising paths.

Heuristic Function(h(n)): It is a heuristic function that provides an estimate of the cost from the current node to the goal node. This heuristic is admissible if it never overestimates the true cost to reach the goal. In other words, **h(n)** is always less than or equal to the actual cost.

Actual cost function(g(n)): It is Cost of the path from the **start node to node n**. It represents the actual cost incurred to reach the current node from the initial node. For the initial node (start node), **g(n)** is usually 0.

Evaluation Function (f(n)): The evaluation function, denoted as **f(n),** is the total estimated cost of the cheapest path from the start node to the goal node that passes through node n. It is the sum of g(n) and h(n): **f(n) = g(n) + h(n).**

f(n) represents the priority of a node. Nodes with lower f(n) values are explored first, making the algorithm prioritize paths that are likely to be more efficient.

3.1.a Greedy Best First Search Algorithm

Greedy best first search tries to expand the node that is closest to the goal to lead to a solution quickly. It evaluates the nodes by using just the heuristic function i.e. for Greedy best first search, f(n) = h(n).

Let's explore the application of this method to route-finding challenges in Romania for the map given in figure 3.2.

We will employ the straight-line distance heuristic, denoted as h_{SLD}. Specifically, for our destination in **Bucharest**, we require knowledge of the straight-line distances to Bucharest, as illustrated in Figure 3.22.

As an illustration, consider h_{SLD}(In(Arad)) = 366. It's important to note that the values of h_{SLD} cannot be derived directly from the problem description. Furthermore, understanding that h_{SLD} correlates with actual road distances and serves as a valuable heuristic requires a certain level of experience.

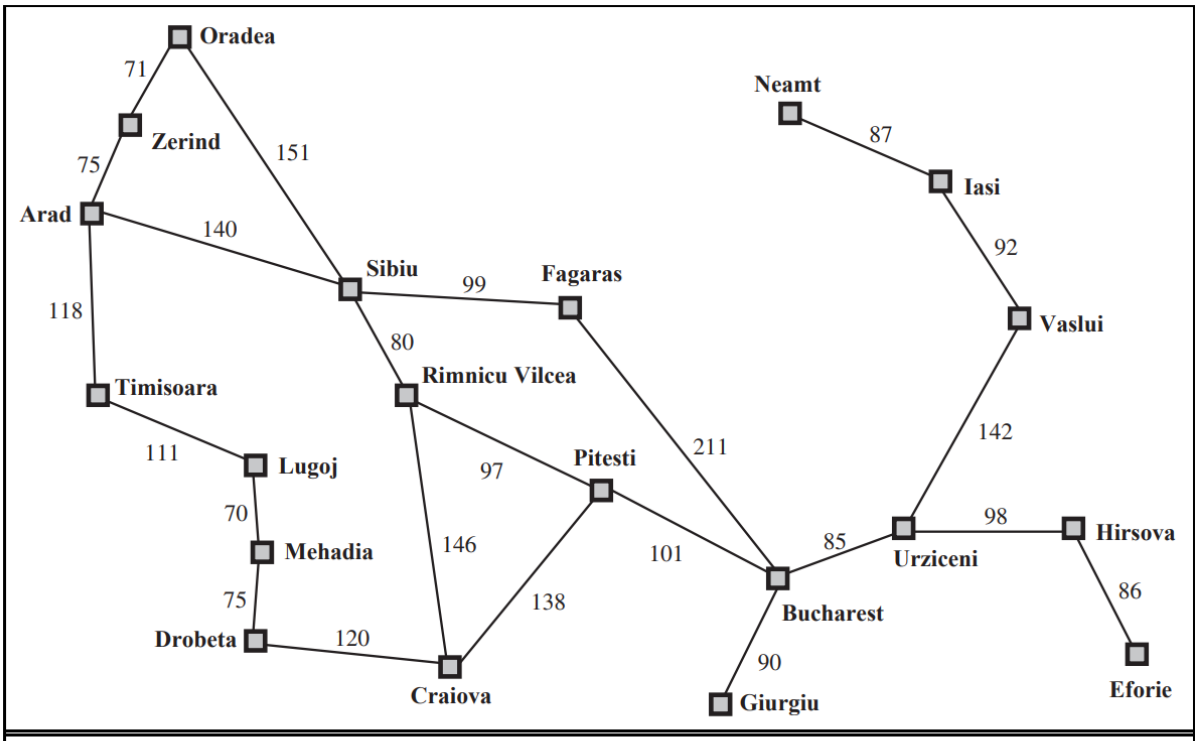


Figure 3.2 A simplified road map of part of Romania.

Arad	366	**Mehadia**	241
Bucharest	0	**Neamt**	234
Craiova	160	**Oradea**	380
Drobeta	242	**Pitesti**	100
Eforie	161	**Rimnicu Vilcea**	193
Fagaras	176	**Sibiu**	253
Giurgiu	77	**Timisoara**	329
Hirsova	151	**Urziceni**	80
Iasi	226	**Vaslui**	199
Lugoj	244	**Zerind**	374

Figure 3.22 Values of h_{SLD}—straight-line distances to Bucharest.

Following Figure shows the progress of a greedy best-first search using h_{SLD} to find a path from Arad to Bucharest.

The first node to be expanded from Arad will be Sibiu because it is closer to Bucharest than either Zerind or Timisoara. The next node to be expanded will be Fagaras because it is closest. Fagaras in turn generates Bucharest, which is the goal.

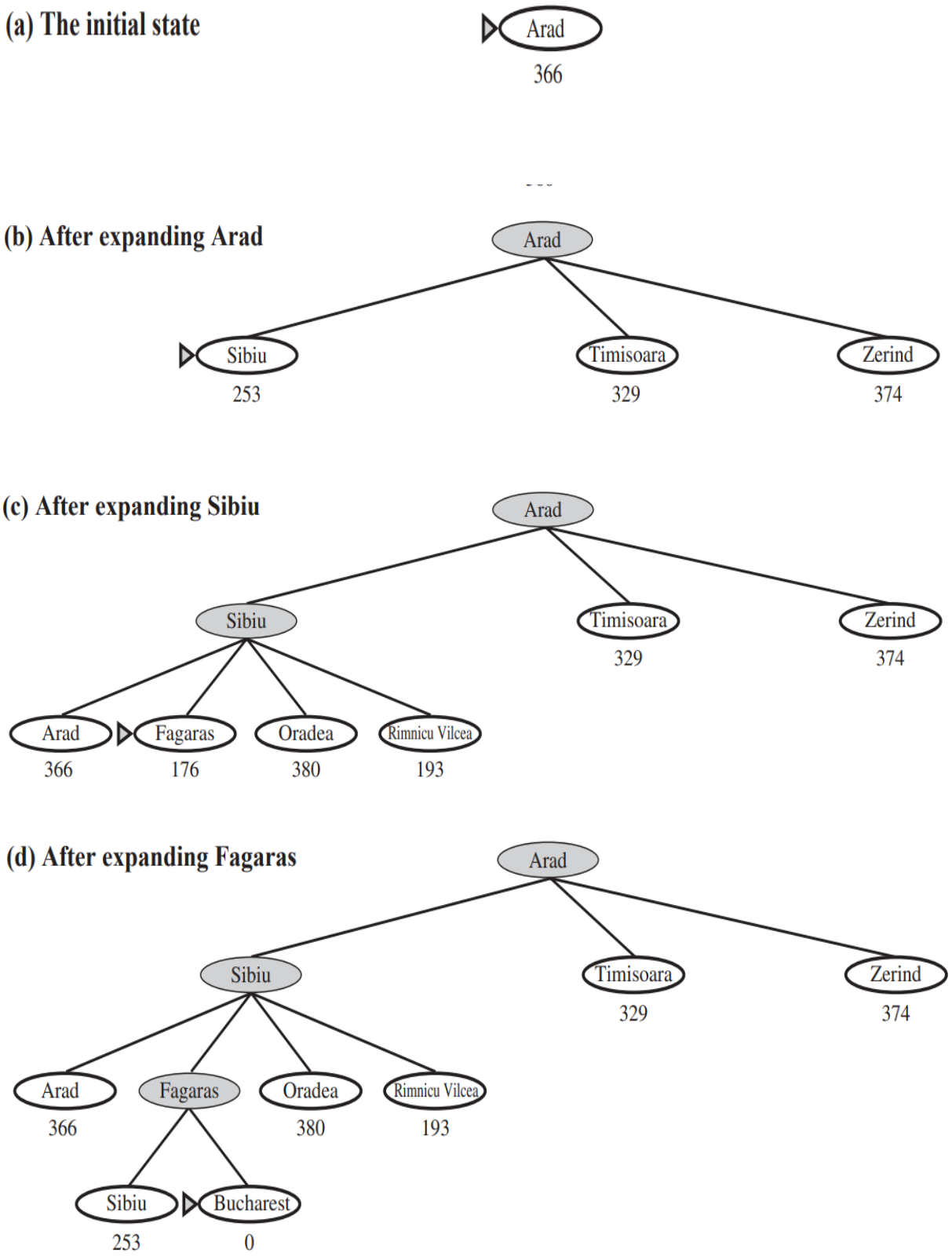
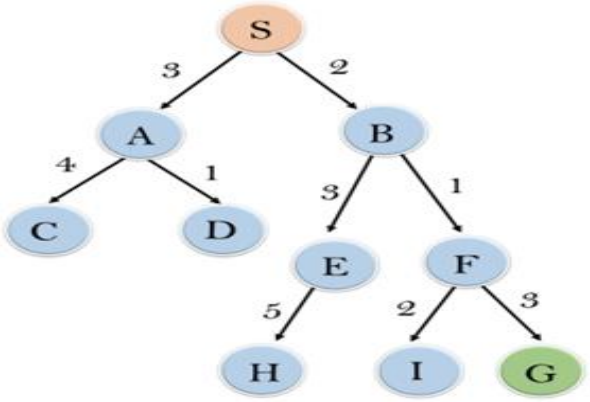


Figure: Stages in a greedy best-first tree search for Bucharest with the straight-line distance heuristic h_{SLD} . Nodes are labelled with their h-values.

Best first search algorithm:

Step 1: Place the starting node into the OPEN list.
Step 2: If the OPEN list is empty, Stop and return failure.
Step 3: Remove the node n, from the OPEN list which has the lowest value of h(n), and places it in the CLOSED list.
Step 4: Expand the node n, and generate the successors of node n.
Step 5: Check each successor of node n, and find whether any node is a goal node or not. If any successor node is goal node, then return success and terminate the search, else proceed to Step 6.
Step 6: For each successor node, algorithm checks for evaluation function f(n), and then check if the node has been in either OPEN or CLOSED list. If the node has not been in both list, then add it to the OPEN list.
Step 7: Return to Step 2.

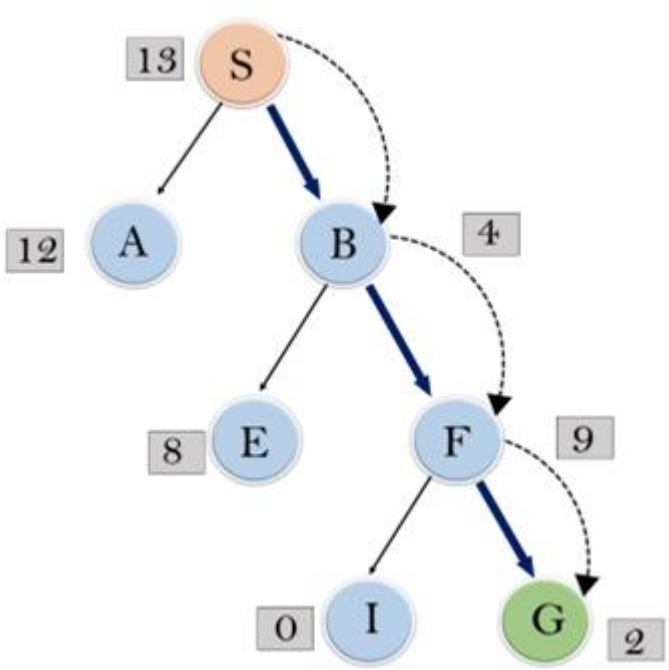
Example: Consider the tree and heuristic values given below:



node	H (n)
A	12
B	4
C	7
D	3
E	8
F	2
H	4
I	9
S	13
G	0

In this search example, we are using two lists which are **OPEN** and **CLOSED** Lists. Following are the iteration for traversing the above example.

Step	OPEN List	CLOSED List	Details
Initialization	[A, B]	[S]	
Iteration1: Expand B	[A]	[S,B]	h(B)<h(A)
Iteration2: Expand F	[E,F,A] [E,A]	[S,B] [S,B,F]	H(F)<H(E),H(A)
Iteration3: Visit Goal G	[I,G,E,A] [I,E,A]	[S,B,F] [S,B,F,G]	H(G)< H(E),H(A),H(I)



Hence the final solution path will be: **S----> B----->F----> G**

Time Complexity: The worst case time complexity of Greedy best first search is $O(b^m)$.

Space Complexity: The worst case space complexity of Greedy best first search is $O(b^m)$. Where, m is the maximum depth of the search space.

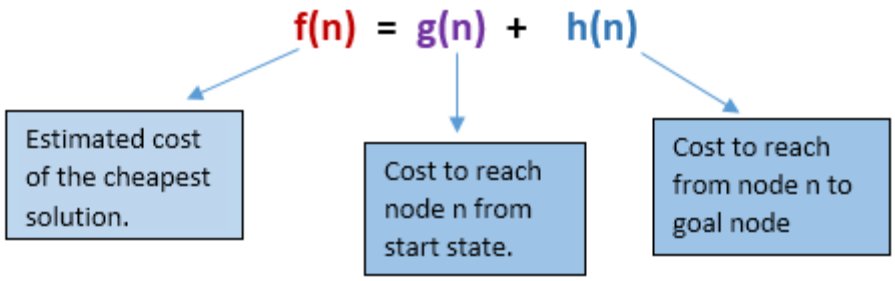
Complete: Greedy best-first search is also incomplete, even if the given state space is finite.

Optimal: Greedy best first search algorithm is not optimal.

3.1.b A* Search Algorithm

A* search is the most commonly known form of best-first search. It uses heuristic function h(n), and cost to reach the node n from the start state g(n). A* search algorithm finds the shortest path through the search space using the heuristic function. This search algorithm expands less search tree and provides optimal result faster.

In A* search algorithm, we use search heuristic as well as the cost to reach the node. Hence, we can combine both costs as following, and this sum is called as a **fitness number**.



Algorithm of A* search:

Step1: Place the starting node in the OPEN list.

Step 2: Check if the OPEN list is empty or not, if the list is empty then return failure and stops.

Step 3: Select the node from the OPEN list which has the smallest value of evaluation function (g+h), if node n is goal node then return success and stop, otherwise

Step 4: Expand node n and generate all of its successors, and put n into the closed list. For each successor n', check whether n' is already in the OPEN or CLOSED list, if not then compute evaluation function for n' and place into Open list.

Step 5: Else if node n' is already in OPEN and CLOSED, then it should be attached to the back pointer which reflects the lowest g(n') value.

Step 6: Return to Step 2.

Advantages:

- A* search algorithm is the best algorithm than other search algorithms.
- A* search algorithm is optimal and complete.
- This algorithm can solve very complex problems.

Disadvantages:

- It does not always produce the shortest path as it mostly based on heuristics and approximation.
- A* search algorithm has some complexity issues.
- The main drawback of A* is memory requirement as it keeps all generated nodes in the memory, so it is not practical for various large-scale problems.

Example1: Let's explore the application of this method to route-finding challenges in Romania for the map given in figure 3.2 .

We will employ the straight-line distance heuristic, denoted as h_{SLD}. Specifically, for our destination in Bucharest, we require knowledge of the straight-line distances to Bucharest, as illustrated in Figure 3.22.

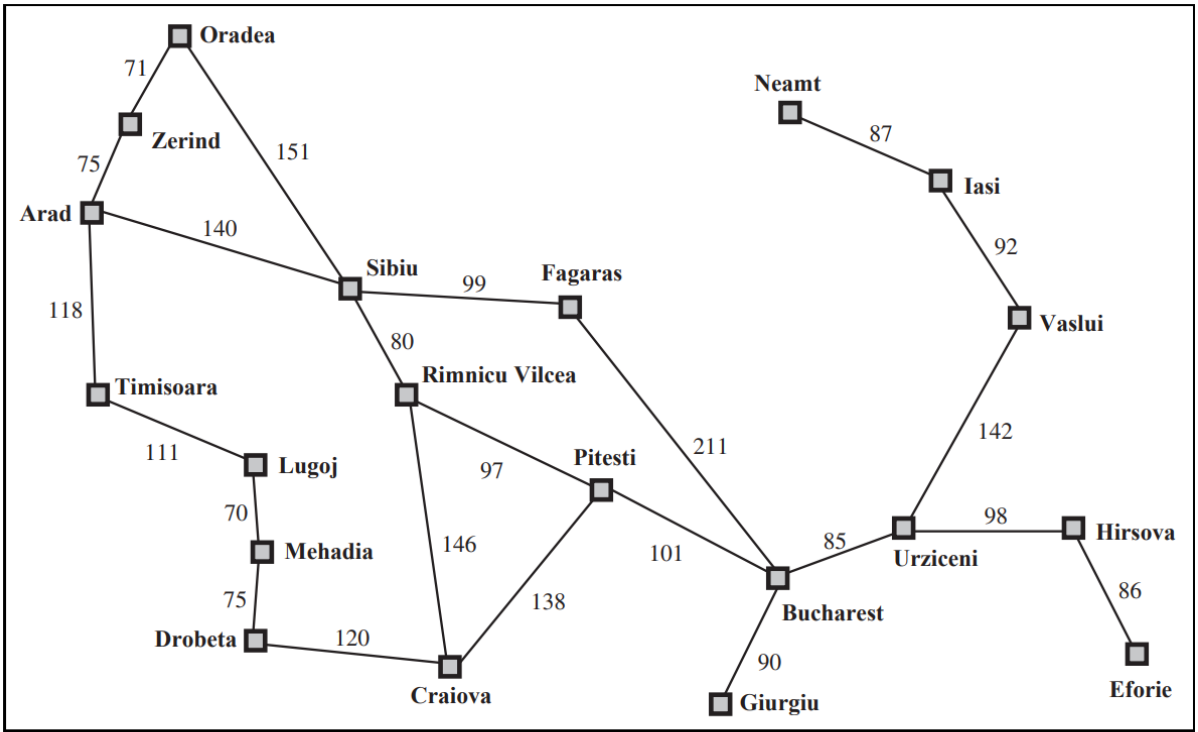
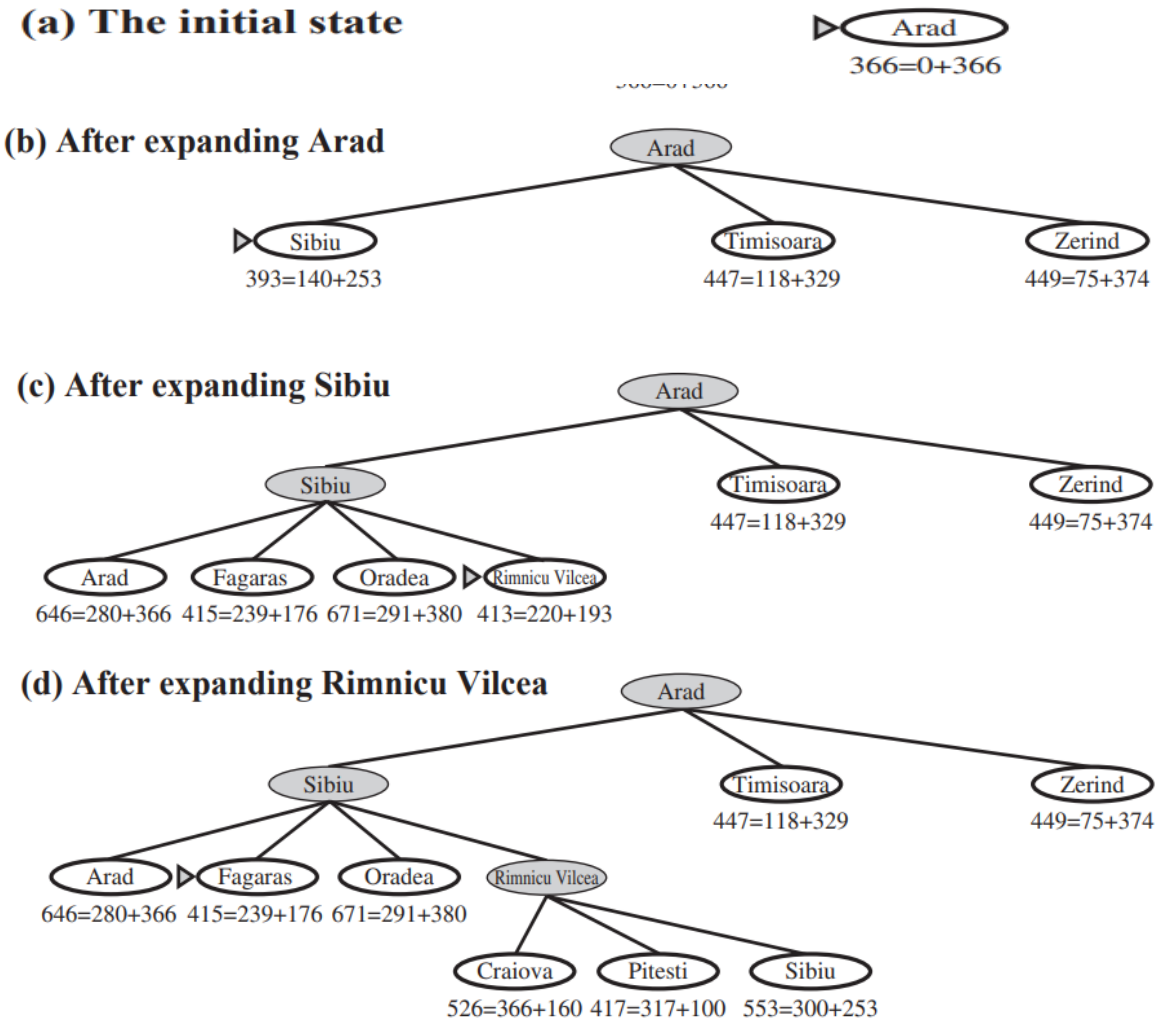


Figure 3.2 A simplified road map of part of Romania.

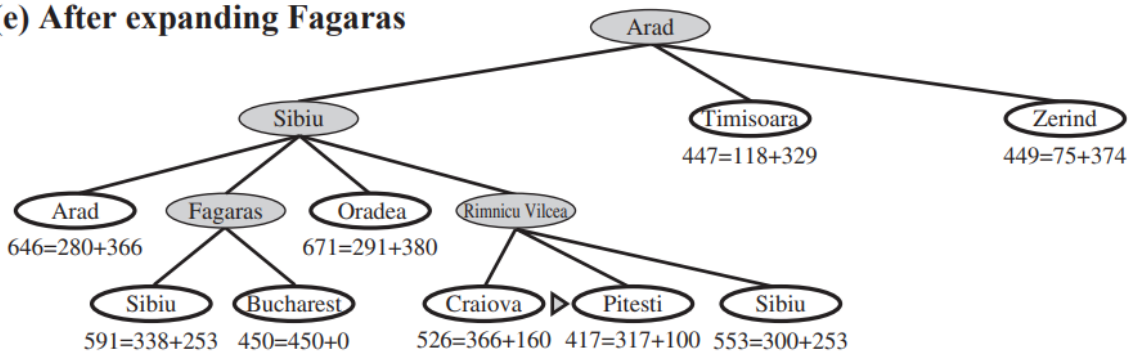
Arad	366	**Mehadia**	241
Bucharest	0	**Neamt**	234
Craiova	160	**Oradea**	380
Drobeta	242	**Pitesti**	100
Eforie	161	**Rimnicu Vilcea**	193
Fagaras	176	**Sibiu**	253
Giurgiu	77	**Timisoara**	329
Hirsova	151	**Urziceni**	80
Iasi	226	**Vaslui**	199
Lugoj	244	**Zerind**	374

Figure 3.22 Values of h_{SLD}—straight-line distances to Bucharest.

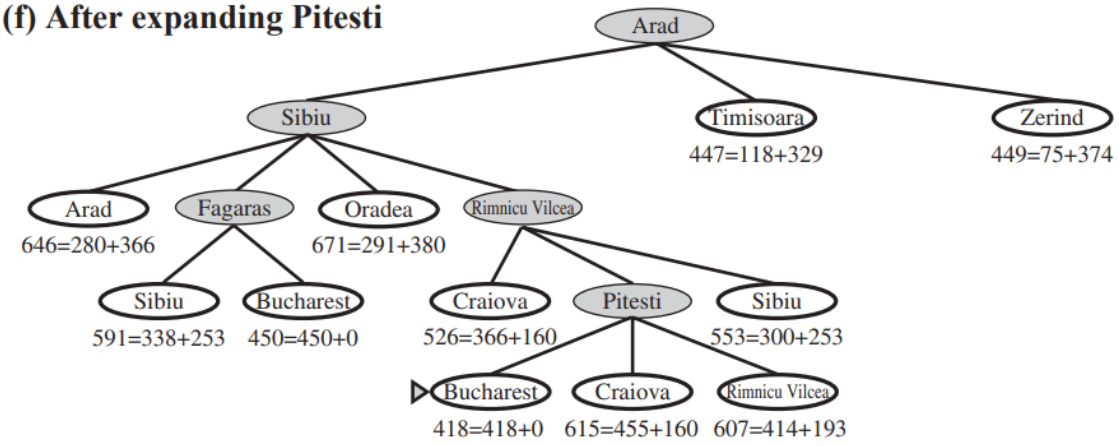
Figure below illustrates the Stages in **an A∗ search for Bucharest**. Nodes are labelled with f = g + h. The h values are the straight-line distances to Bucharest



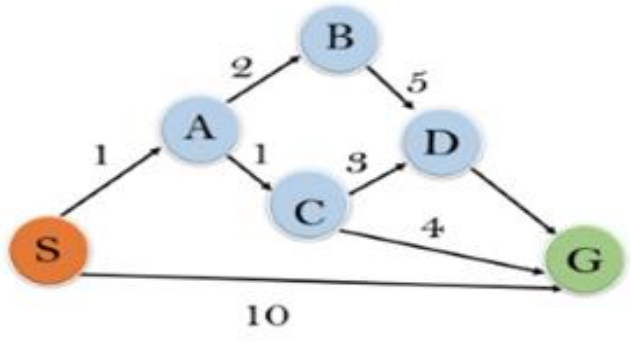
(e) After expanding Fagaras



(f) After expanding Pitesti

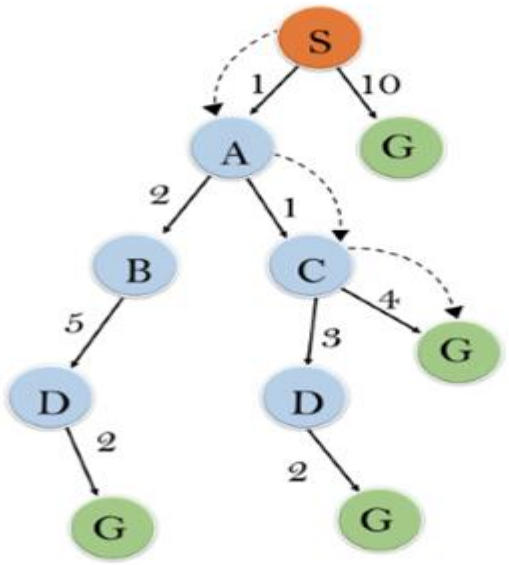


Example 2: In this example, we will traverse the given graph using the A* algorithm. The heuristic value of all states is given in the below table so we will calculate the f(n) of each state using the formula f(n)= g(n) + h(n), where g(n) is the cost to reach any node from start state. Here we will use OPEN and CLOSED list**.**



State	h(n)
S	5
A	3
B	4
C	2
D	6
G	0

Solution:



Initialization: {(S, 5)}
Iteration1: {(S--> A, 4), (S-->G, 10)}
Iteration2: {(S--> A-->C, 4), (S--> A-->B, 7), (S-->G, 10)}
Iteration3: {(S--> A-->C--->G, 6), (S--> A-->C--->D, 11), (S--> A-->B, 7), (S-->G, 10)}
Iteration 4 will give the final result, as **S--->A--->C--->G** it provides the optimal path with cost 6.
Note: Use Open and Closed list is as illustrated in the explanation of BFS.

3.1.b.1 Points to remember:

A* Algorithm and First-Occurrence Path: The A* algorithm is designed to return the first path it finds from the start node to the goal node. Once a valid path is discovered, A* terminates its search, making it more efficient in scenarios where finding a single optimal solution is sufficient.

Quality of Heuristic: The effectiveness of the A* algorithm is significantly influenced by the quality of the heuristic function **h(n))**. A well-designed heuristic provides a good estimate of the remaining cost from a given node to the goal, guiding **A*** to explore more promising paths and enhancing its efficiency.

Node Expansion Condition: A* algorithm expands nodes based on the evaluation function f(n)=g(n)+h(n), where:
g(n) is the cost of the path from the start node to node n,
h(n) is the heuristic estimate of the cost from node n to the goal node.

The algorithm expands nodes that satisfy the condition f(n)≤M, where M is a specified threshold or maximum cost. This condition ensures that A* explores nodes within a predefined cost limit, allowing for efficient pathfinding without exhaustively searching the entire space.

Complete: A* algorithm is complete as long as:

- Branching factor is finite.
- Cost at every action is fixed.

3.1.b.2 Conditions for optimality: Admissibility and consistency:

A* search algorithm is optimal if it follows below two conditions:

Admissible:

The first condition requires for optimality is that **h(n)** should be an admissible heuristic for **A* tree** search. An admissible heuristic is one that never overestimates the cost to reach the goal. Because g(n) is the actual cost to reach n along the current path, and f(n) = g(n) + h(n), we have as an immediate consequence that f(n) never overestimates the true cost of a solution along the current path through n. If the heuristic function is admissible, then A* tree search will always find the least cost path.

Consistency (or sometimes monotonicity):

Second required condition is consistency for only **A* graph**-search. A heuristic h(n) is consistent if, for every node n and every successor n ' of n generated by any action a, the estimated cost of reaching the goal from n is no greater than the step cost of getting to n ' plus the estimated cost of reaching the goal from n ' : h(n) ≤ c(n, a, n') + h(n ') .

This is a form of the general **triangle inequality**, which stipulates that each side of a triangle cannot be longer than the sum of the other two sides. Here, the triangle is formed by n, n ' , and the goal **Gn** closest to n.

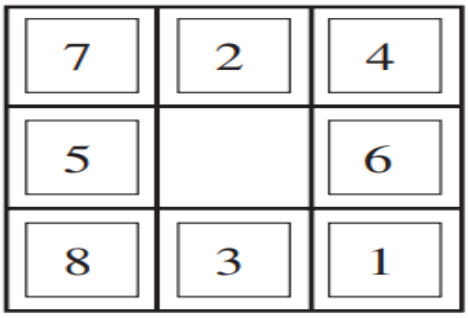
Time Complexity: The time complexity of A* search algorithm depends on heuristic function, and the number of nodes expanded is exponential to the depth of solution d. So the time complexity is $O(b^d)$, where b is the branching factor.

Space Complexity: The space complexity of A* search algorithm is **$O(b^d)$**

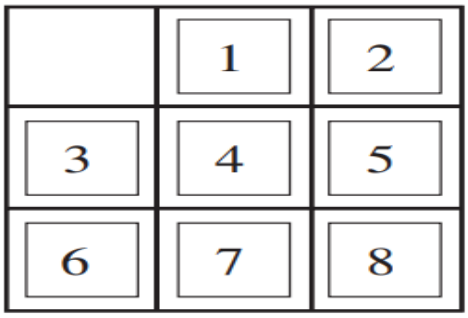
3.1.c Heuristics Functions

Heuristic Functions **h(n)** guide search algorithms by estimating the cost or distance to a goal state from the current state (n).

Consider the 8-puzzle game. The object of the 8 puzzle is to slide the title horizontally or vertically into the empty space until the configuration matches the goal configuration.



Start State



Goal State

Figure illustrates the A typical instance of the 8-puzzle. The solution is 26 steps long.

The average solution cost for a randomly generated 8-puzzle instance is about 22 steps. The branching factor is about 3. (When the empty tile is in the middle, four moves are possible; when it is in a corner, two; and when it is along an edge, three.) This means that an exhaustive tree search to depth 22 would look at about $3^{22} \approx 3.1 \times 10^{10}$ states.

The two commonly used candidates for 8 puzzles are as follows:

h1 = the number of misplaced tiles. For Figure, all of the eight tiles are out of position, so the start state would have h1 = 8. h1 is an admissible heuristic because it is clear that any tile that is out of place must be moved at least once

h2 = the sum of the distances of the tiles from their goal positions. Because tiles cannot move along diagonals, the distance we will count is the sum of the horizontal and vertical distances. This is sometimes called the city block distance or Manhattan distance. h2 is also admissible because all any move can do is move one tile one step closer to the goal. Tiles 1 to 8 in the start state give a Manhattan distance of h2 = 3 + 1 + 2 + 2 + 2 + 3 + 3 + 2 = 18 .

As expected, neither of these overestimates the true solution cost, which is 26. The performance of heuristic search algorithms depends on the quality of the heuristic function. One can sometimes construct good heuristics by relaxing the problem definition, by storing precomputed solution costs for subproblems in a pattern database, or by learning from experience with the problem class.

1. The effect of heuristic accuracy on performance:

Experimentally it is determined that h_2 is better than h_1. That is for any node n, $h_2(n) \geq h_1(n)$. This implies that h_2 dominate h_1. Domination translates directly into efficiency. A^* using h_2 will never expand more nodes than A^* using h_1.

2. Generating admissible heuristics from relaxed problems:

A problem with fewer restrictions on the actions is called a relaxed problem. The state-space graph of the relaxed problem is a super graph of the original state space because the removal of restrictions creates added edges in the graph.
Because the relaxed problem adds edges to the state space, any optimal solution in the original problem is, by definition, also a solution in the relaxed problem; but the relaxed problem may have better solutions if the added edges provide short cuts. Hence, the cost of an optimal solution to a relaxed problem is an admissible heuristic for the original problem.

For example, if the 8-puzzle actions are described as

- A tile can move from square **A** to square **B** if
- **A** is horizontally or vertically adjacent to **B** and **B** is blank,

we can generate three relaxed problems by removing one or both of the conditions:

a) A tile can move from square A to square B if A is adjacent to B.
b) A tile can move from square A to square B if B is blank.
c) A tile can move from square A to square B.

If a collection of admissible heuristics h1 . . . hm is available for a problem and none of them dominates any of the others, which should we choose? As it turns out, we need not make a choice. We can have the best of all worlds, by defining $h(n) = \max\{h1(n), \ldots, hm(n)\}$

3.Generating admissible heuristics from subproblems: Pattern databases:

Admissible heuristics can also be derived from the solution cost of a subproblem of a given problem. For example, Figure below shows a subproblem of the 8-puzzle instance.

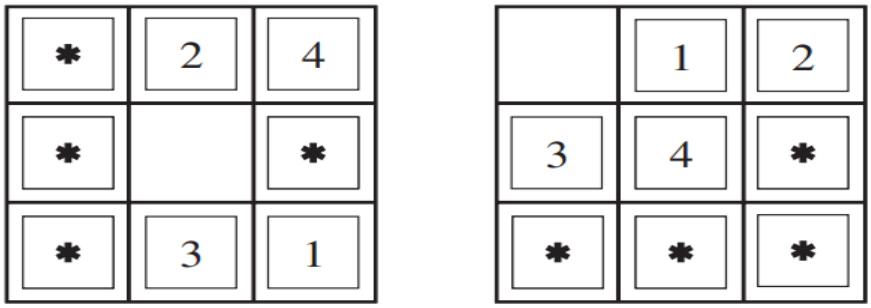


Fig: A subproblem of the 8-puzzle instance. The task is to get tiles 1, 2, 3, and 4 into their correct positions, without worrying about what happens to the other tiles.

The subproblem involves getting tiles 1, 2, 3, 4 into their correct positions. Clearly, the cost of the optimal solution of this subproblem is a lower bound on the cost of the complete problem. It turns out to be more accurate than Manhattan distance in some case.

The idea behind pattern databases is to store these exact solution costs for every possible subproblem instance—in our example, every possible configuration of the four tiles and the blank. (The locations of the other four tiles are irrelevant for the purposes of solving the subproblem, but moves of those tiles do count toward the cost.) Then we compute an admissible heuristic h_{DB} for each complete state encountered during a search simply by looking up the corresponding subproblem configuration in the database.

4 Learning heuristics from experience:

A heuristic function, denoted as h(n), aims to approximate the solution cost starting from the state represented by node **n**. One of the strategies is to learning from practical experiences. In this context, "experience" refers to solving numerous instances of problems like 8-puzzles.

In each optimal solution to an 8-puzzle, valuable examples emerge, comprising a state along the solution path and the actual cost of reaching the solution from that particular point.

Employing these examples, a learning algorithm can be employed to generate a heuristic function, **h(n),** with the potential to predict solution costs for other states encountered during the search process.
Inductive learning methods are most effective when provided with relevant features of a state for predicting its value, rather than relying solely on the raw state description.

For instance, a feature like "**number of misplaced tiles**" (**$x_1(n)$**) can be useful in predicting the distance of a state from the goal in an 8-puzzle. By gathering statistics from randomly generated 8-puzzle configurations and their actual solution costs, one can use these features to predict h(n).

Multiple features, such as **$x_2(n)$** representing the "number of pairs of adjacent tiles that are not adjacent in the goal state," can be combined using a linear combination approach:

$$h(n) = c_1x_1(n) + c_2x_2(n).$$

The constants (**c1** and **c2**) are adjusted to achieve the best fit with the actual data on solution costs. It is expected that both c_1 and c_2 are positive, as misplaced tiles and incorrect adjacent pairs make the problem more challenging. While this heuristic satisfies the condition **h(n) = 0** for goal states, it may not necessarily be both admissible and consistent.

3.2.a Logical Agents

What are Logic Agents?

Stuart Russell and Peter Norvig, in their influential textbook "**Artificial Intelligence: A Modern Approach**," describe logical agents as those that operate based on knowledge representation and logical inference.
According to their **framework**, an agent perceives its environment through sensors, maintains an internal state (knowledge base), and acts upon the environment through effectors. **Logical agents** specifically use logical reasoning to make decisions.

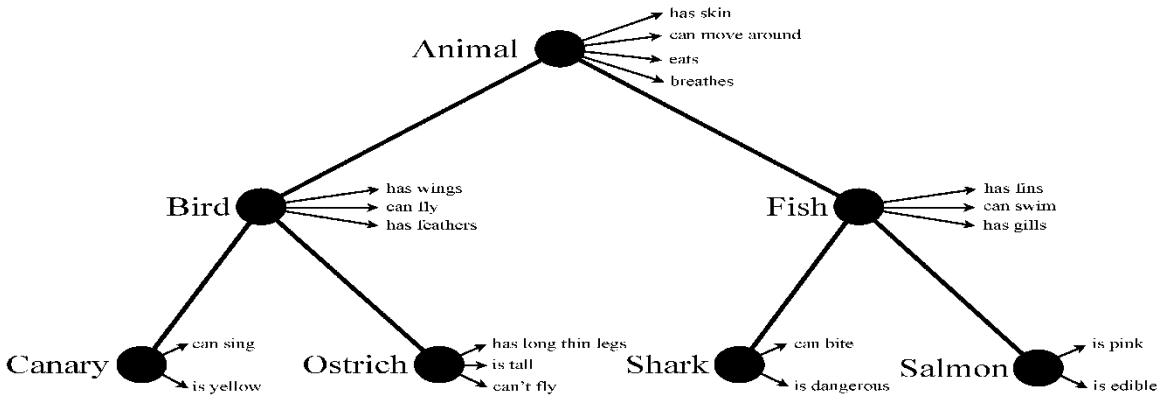
What is Knowledge Representation?

Knowledge representation in **artificial intelligence (AI)** refers to the process of creating a **structured and formalized representation of information** in a way that can be used by a computer system to **reason, make decisions, or perform tasks.** The goal is to model knowledge in a manner that facilitates effective **problem-solving, learning, and communication within an AI system.**

Example 1 : Semantic Networks

Semantic networks are a graphical representation of knowledge that **uses nodes to represent concepts and arcs (edges)** to represent relationships between these concepts.
Each node in the network represents an **entity or concept**, and the **arcs depict the relationships between** them.
This form of knowledge representation is often used to model **hierarchies, associations, and dependencies**.

Knowledge Representation using Semantic Networks



Example 2: Knowledge Representation using Propositional logic

In propositional logic, knowledge is represented using propositions, which are statements that can be either **true or false**. Logical operators such as **AND, OR, and NOT** are used to combine propositions.
Consider the following knowledge about a weather prediction system:
P: It is raining.
Q: The sky is cloudy.
R: The weather forecast predicts rain.

Now, we can represent some logical relationships:

- If the sky is cloudy **(Q)**, and the weather forecast predicts rain **(R)**, then we can infer that it might be raining **(P)**. This relationship can be represented as: $(Q \wedge R) \rightarrow P$.
- If it is not raining (NOT P), then the weather forecast predicting rain **(R)** must be false. This relationship can be represented as: $\neg P \rightarrow \neg R$.

What is Logical Reasoning?
Logical reasoning is a cognitive process of **making inferences or drawing conclusions** based on **logical principles, rules, and relationships.** It involves analyzing information and using **valid deductive or inductive arguments** to reach a sound or reasonable conclusion. Logical reasoning is an essential aspect **of problem-solving, decision-making, and critical thinking**.

Example: Syllogism
A syllogism is a form of deductive reasoning where a conclusion is drawn from **two given or assumed propositions** (premises).

- **Premise 1**: All humans are mortal.
- **Premise 2:** Socrates is a human.
- **Conclusion**: **Therefore, Socrates is mortal.**

Knowledge based Agents
A knowledge-based agent is a type of intelligent agent that makes decisions and takes actions based on **knowledge** it possesses. A knowledge-based agent is characterized by its ability to represent and manipulate knowledge in a structured way, allowing it to reason, make decisions, and take actions based on the information stored in its knowledge base.

Key Components of Knowledge base:

1. **Knowledge Base (KB) :** The central component of a knowledge-based agent is its **knowledge base**. The **knowledge base** is a collection of sentences expressed in a knowledge representation language. Each sentence represents an assertion about the world. The knowledge base is where the agent stores information that it uses to **make decisions and take actions**. Sometimes we dignify a sentence with the name axiom, when the sentence is taken as given without being derived from other sentences.

2. **Knowledge Representation Language:** The sentences in the knowledge base are expressed in a language called a knowledge representation language. This language allows the agent to formally represent information about the world in a way that the agent can understand and manipulate.

3. **Axioms:** Some sentences in the knowledge base may be dignified with the name "axiom," especially when they are taken as given without being derived from other sentences. Axioms are fundamental statements that serve as foundational knowledge for the agent.

4. **TELL Operation:** There is a mechanism for adding new sentences to the knowledge base. This operation is referred to as TELL. It allows the agent to incorporate new information into its knowledge base.

5. **ASK Operation:** The agent needs a way to query the knowledge base to retrieve information. The standard operation for querying is referred to as ASK. It allows the agent to ask questions about what is known.

6. **Inference :** Both TELL and ASK operations may involve inference, which is the process of deriving new sentences from existing ones. Inference must adhere to the requirement that answers derived from the knowledge base follow logically from the information previously TELLed to the knowledge base.

7. **Background Knowledge:** The knowledge base may initially contain some background knowledge. This knowledge provides a foundational understanding of the environment in which the agent operates.

8. **Agent Program:** The knowledge-based agent program outlines the overall structure of the agent. It takes a percept (input) and returns an action as output. The agent program incorporates the knowledge base and other components to facilitate decision-making and action-taking.

   Agent Program: Figure below illustrates a generic knowledge-based agent. Given a **percept**, the agent adds the percept to its knowledge base, asks the knowledge base for the best action, and tells the knowledge base that it has in fact taken that action

```
function KB-AGENT(percept) returns an action
  persistent: 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
```

   Agents' **knowledge and goals is more important:** At the knowledge level, we only need to specify the
   - agent's knowledge and
   - goals to determine its behavior.

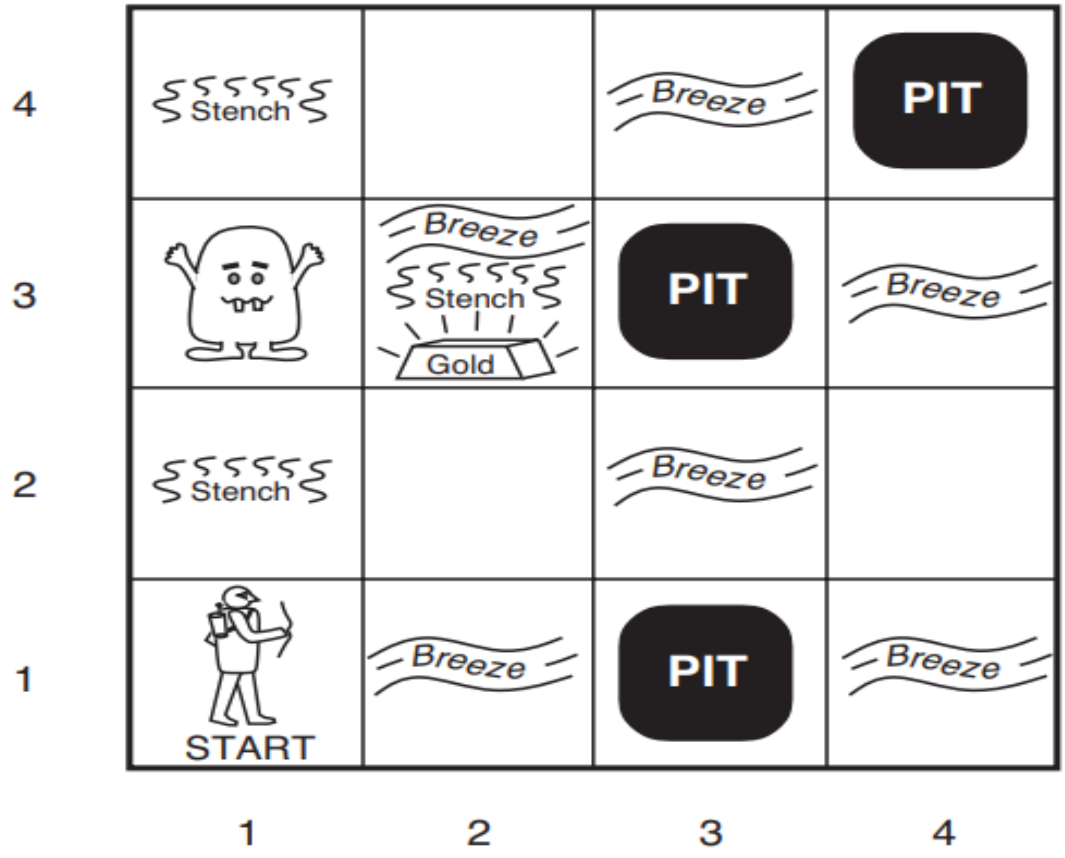
For instance, consider an **automated taxi** with the goal of transporting a passenger from **San Francisco to Marin County**. If the taxi knows that the **Golden Gate Bridge is the sole link** between these locations, we can expect it to cross the bridge, understanding that it aligns with its goal.

Importantly, this analysis remains independent of the taxi's implementation details.
Whether its geographical knowledge is represented through ***linked lists, pixel maps, or if it reasons using symbolic strings stored in registers or through neural network signal propagation***, the behavior is determined solely by its **knowledge and goals.**

3.2.b The Wumpus World

The **wumpus** world is a cave consisting of rooms connected by passageways. Lurking somewhere in the cave is the terrible **wumpus**, a beast that eats anyone who enters its room. The **wumpus** can be shot by an agent, but the agent has only one arrow. Some rooms contain bottomless pits that will trap anyone who wanders into these rooms (except for the **wumpus**, which is too big to fall in). The only mitigating feature of this bleak environment is the possibility of finding a heap of gold.

A typical wumpus world is illustrated in the figure below. The agent is in the bottom left corner, facing right.



PEAS description of Wumpus World

- **Performance measure**: +1000 for climbing out of the cave with the gold, –1000 for falling into a pit or being eaten by the wumpus, –1 for each action taken and –10 for using up the arrow. The game ends either when the agent dies or when the agent climbs out of the cave.

- **Environment**: A 4 × 4 grid of rooms. The agent always starts in the square labeled [1,1], facing to the right. The locations of the gold and the wumpus are chosen randomly, with a uniform distribution, from the squares other than the start square. In addition, each square other than the start can be a pit, with probability 0.2.
- **Actuators:** The agent can **move Forward, TurnLeft by 90◦ , or TurnRight by 90◦** . The agent dies a miserable death if it enters a square containing a **pit or a live** wumpus. (It is safe, albeit smelly, to enter a square with a dead wumpus.) If an agent tries to move forward and **bumps** into a wall, then the agent does not move. The action **Grab** can be used to pick up the gold if it is in the same square as the agent. The action **Shoot** can be used to fire an arrow in a straight line in the direction the agent is facing. The arrow continues until it either hits (and hence kills) the **wumpus or hits a wall**. The agent has only one arrow, so only the first Shoot action has any effect. Finally, the action **Climb** can be used to climb out of the cave, but only from square [1,1]
- **Sensors:** The agent has five sensors, each of which gives a single bit of information:
  1. In the square containing the wumpus and in the directly (not diagonally) adjacent squares, the agent will perceive a Stench.
  2. In the squares directly adjacent to a pit, the agent will perceive a Breeze.
  3. In the square where the gold is, the agent will perceive a Glitter.
  4. When an agent walks into a wall, it will perceive a Bump.
  5. When the wumpus is killed, it emits a woeful Scream that can be perceived anywhere in the cave.

5 Percept Symbols : [Stench, Breeze, Glitter, Bump, Scream].
The percepts will be given to the agent program in the form of a list of five symbols; for example, if there is a **stench** and a **breeze**, but no **glitter**, **bump**, or **scream**, the agent program will get **[Stench, Breeze, None, None, None].**

The first step taken by the agent in the wumpus world. (a) The initial situation, after percept [None, None, None, None, None]. (b) After one move, with percept [None, Breeze, None, None, None]

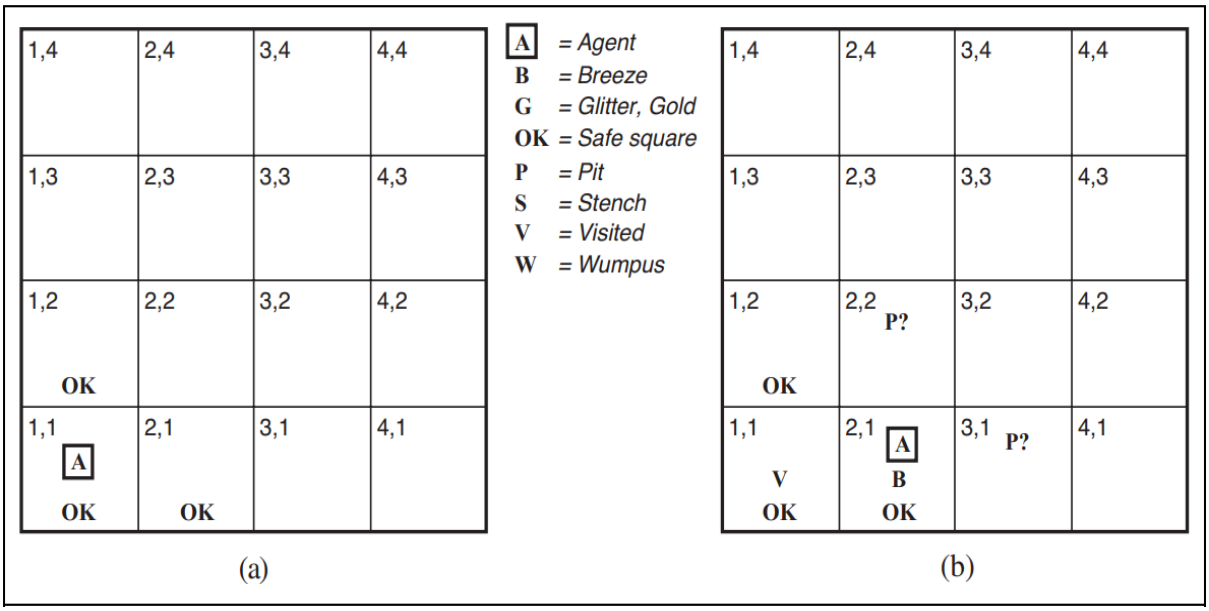


Fig 7.3

Two later stages in the progress of the agent. **(a)** After the third move, with percept [Stench, None, None, None, None]. **(b)** After the fifth move, with percept [Stench, Breeze, Glitter , None, None].

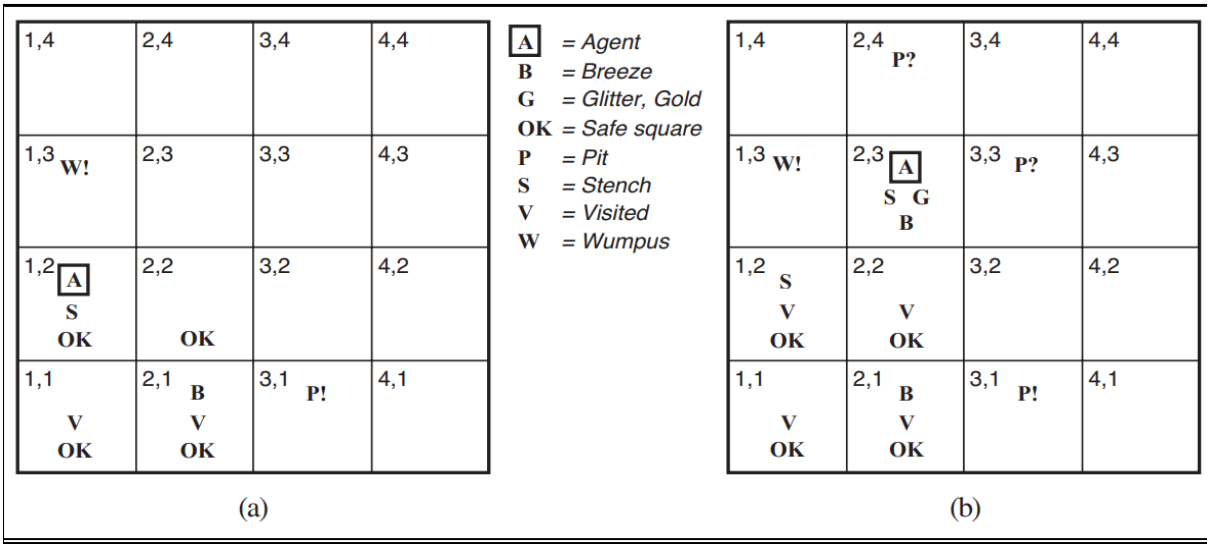


Fig 7.4

- The agent's initial knowledge base contains the rules of the environment, as described previously; in particular, it knows that it is in [1,1] and that [1,1] is a safe square; we denote that with an "A" and "OK," respectively, in square [1,1].
- The first percept is [None, None, None, None, None], from which the agent can conclude that its neighboring squares, [1,2] and [2,1], are free of dangers—they are OK. Figure 7.3(a) shows the agent's state of knowledge at this point.

- A cautious agent will move only into a square that it knows to be OK. Let us suppose the agent decides to move forward to [2,1]. The agent perceives a breeze (denoted by "B") in [2,1], so there must be a pit in a neighboring square. The pit cannot be in [1,1], by the rules of the game, so there must be a pit in [2,2] or [3,1] or both. The notation "P?" in Figure 7.3(b) indicates a possible pit in those squares. At this point, there is only one known square that is OK and that has not yet been visited. So the prudent agent will turn around, go back to [1,1], and then proceed to [1,2].
- The agent perceives a stench in [1,2], resulting in the state of knowledge shown in Figure 7.4(a). The stench in [1,2] means that there must be a wumpus nearby. But the wumpus cannot be in [1,1], by the rules of the game, and it cannot be in [2,2] (or the agent would have detected a stench when it was in [2,1]). Therefore, the agent can infer that the wumpus is in [1,3]. The notation W! indicates this inference. Moreover, the lack of a breeze in [1,2] implies that there is no pit in [2,2]. Yet the agent has already inferred that there must be a pit in either [2,2] or [3,1], so this means it must be in [3,1]. This is a fairly difficult inference, because it combines knowledge gained at different times in different places and relies on the lack of a percept to make one crucial step.
- The agent has now proved to itself that there is neither a pit nor a wumpus in [2,2], so it is OK to move there. We do not show the agent's state of knowledge at [2,2]; we just assume that the agent turns and moves to [2,3], giving us Figure 7.4(b). In [2,3], the agent detects a glitter, so it should grab the gold and then return home.

3.2.c Logic:

In AI Logic is a fundamental component of logical representation and reasoning. It enables machines to understand and represent data and knowledge in a reasoning way. Logical reasoning is a process of inferring a conclusion based on observations or data. It is concerned with the principles of reasoning and how conclusions can be drawn from given premises. Logic provides the theoretical foundation for reasoning.

Types of Logic:

Language	Ontological Commitment	Epistemological Commitment
Propositional logic	facts	true/false/unknown
First-order logic	facts, objects, relations	true/false/unknown
Temporal logic	facts, objects, relations, times	true/false/unknown
Probability theory	facts	degree of belief 0...1
Fuzzy logic	degree of truth	degree of belief 0...1

Logics are formal languages for representing information such that conclusions can be drawn. Syntax defines the sentences in the language. Semantics define the meaning of sentences 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

$x + 2 \geq y$ is true iff the number $x + 2$ is no less than the number y

$x + 2 \geq y$ is true in a world where $x = 7,\ y = 1$
$x + 2 \geq y$ is false in a world where $x = 0,\ y = 6$

Syntax : Syntax refers to the structure and rules governing the formation of sentences or expressions in a language or representation system.Syntax is the set of rules that dictate which sentences are well-formed in the representation language. For example, "x + y = 4" adheres to the syntax, while "x4y+ =" does not.

Semantics: Semantics deals with the meaning of sentences or expressions in a language. It specifies the truth or falsehood of sentences in relation to

possible worlds or situations. **Semantics** defines the truth of each sentence in relation to possible worlds. For instance, the sentence "x + y = 4" is true in a world where x is 2 and y is 2, but false in a world where x is 1 and y is 1.

Model: A model is a mathematical abstraction that represents a possible world in the context of logic. It is used to fix the truth or falsehood of sentences based on specific assignments of values to variables. **Models** serve as precise mathematical abstractions of **possible worlds**. They fix the truth or falsehood of sentences based on assignments of real numbers to variables. The term "**model**" is used interchangeably with "**possible world**." Informally, we may think of a possible world as, for example, having x men and y women sitting at a table playing bridge, and the sentence x + y = 4 is true when there are four people in total. Formally, the possible models are just all possible assignments of real numbers to the variables x and y.

Satisfaction: Satisfaction is a relationship between a model and a sentence, indicating that the model makes the sentence true. If a sentence is true in a particular model, we say that the model satisfies the sentence.
In the given context, if a sentence α is true in a model m, it is said that m satisfies α. Alternatively, m is considered a model of α. The notation M(α) is used to represent the set of all models that satisfy the sentence α.

Model or Possible World in logic

The semantics defines the truth of each sentence with respect to each possible world. In standard logics, every sentence must be either **true or false** in each possible world—there is no "in between." When we need to be precise, we use the term model in place of "**possible world**." 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.

Logical entailment between sentences: A sentence follows logically from another sentence. In mathematical notation, we write α |= β. Entailment in logic refers to a relationship between propositions where the truth of one proposition necessarily guarantees the truth of another. If proposition A entails proposition B, it means that whenever A is true, B must also be true. In formal logic, this relationship is often represented as A |= B.

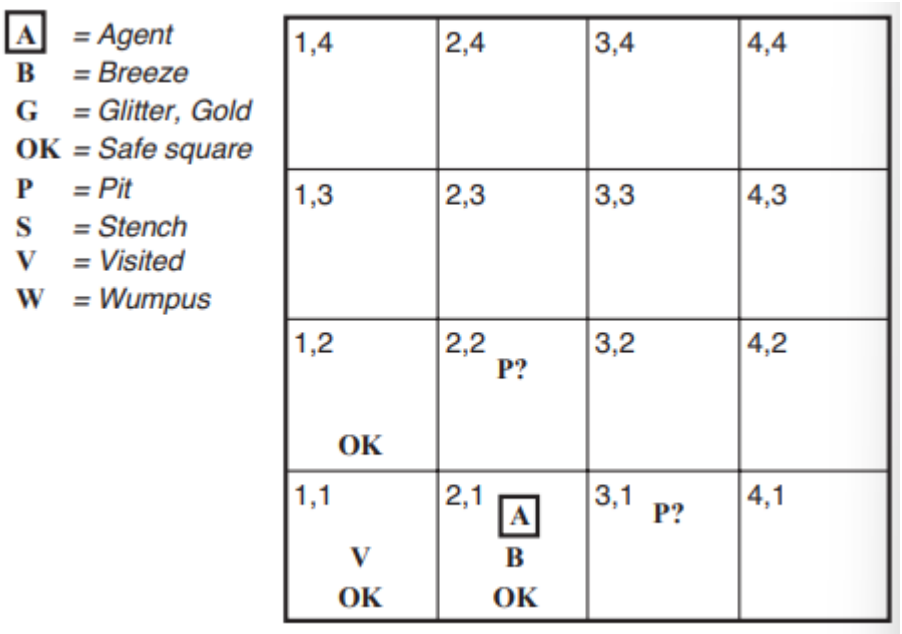
Formal Definition: The formal definition of entailment is this:

α |= β if and only if, in every model in which **α is true, β** is also true.
Using the notation just introduced, we can write
α |= β if and only if M(α) ⊆ M(β) .

E.g., the KB containing "the Giants won" and "the Reds won" entails "Either the Giants won or the Reds won"

Example: The relation of entailment is familiar from **arithmetic**; the idea that the **sentence x = 0** entails the sentence **xy = 0**. Obviously, in any model where **x is zero**, it is the case that **xy is zero** (**regardless of the value of y**)

Wumpus World Example: Consider the situation in Figure below: the agent has detected nothing in [1,1] and a breeze in [2,1].



These precepts, combined with the agent's knowledge of the rules of the wumpus world, constitute the KB. The agent is interested (among other things) in whether the adjacent squares [1,2], [2,2], and [3,1] contain pits. Each of the three squares might or might not contain a pit, so (for the purposes of this example) there are 2^ 3 =8 possible models. These eight models are shown in Figure 7.5 below:

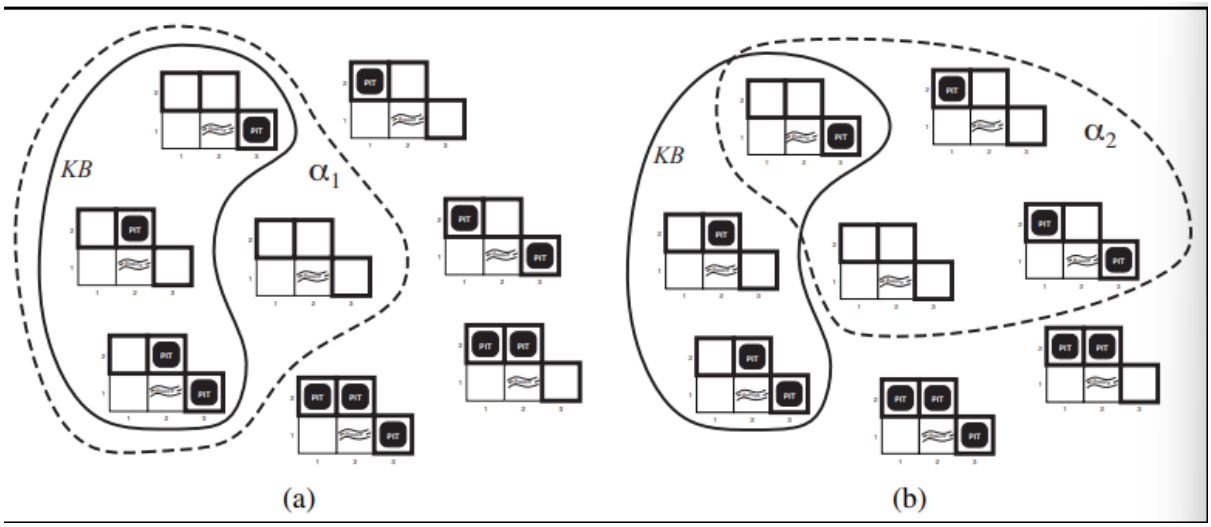


Fig7.5 Possible models for the presence of pits in squares [1,2], [2,2], and [3,1]. The KB corresponding to the observations of nothing in [1,1] and a breeze in [2,1] is shown by the solid line. (a) Dotted line shows models of α1 (no pit in [1,2]). (b) Dotted line shows models of α2 (no pit in [2,2]).

The KB can be thought of as a set of sentences or as a single sentence that asserts all the individual sentences. The KB is false in models that contradict what the agent knows— for example, the KB is false in any model in which [1,2] contains a pit, because there is no breeze in [1,1]. There are in fact just three models in which the KB is true, and these are shown surrounded by a solid line in Figure 7.5. Now let us consider two possible conclusions:
α1 = "There is no pit in [1,2]."
α2 = "There is no pit in [2,2]."
We have surrounded the models of α1 and α2 with dotted lines in Figures 7.5(a) and 7.5(b), respectively. By inspection, we see the following: in every model in which KB is true, α1 is also true.
Hence, KB |= α1: there is no pit in [1,2]. We can also see that in some models in which KB is true, α2 is false. Hence,

$$KB \not\models \alpha_2.$$

the agent cannot conclude that there is no pit in [2,2]. (Nor can it conclude that there is a pit in [2,2].)

Logical Inference and Model Checking:
Logical inference is the process of deriving new sentences (conclusions) from existing knowledge or premises. Model checking is an example of a logical inference algorithm where all possible models are enumerated to check if a conclusion holds in all models where the premises are true.

Example: In the wumpus-world example, logical inference is used to determine conclusions about the presence of pits in adjacent squares based on percepts and knowledge.

Model Checking:

Model checking is an inference algorithm that checks if a conclusion holds in all models where the premises are true. In the Wumpus-world example, model checking is applied to determine if certain conclusions (e.g., "There is no pit in [1,2]") hold in all possible models consistent with the agent's knowledge (KB). The inference algorithm illustrated in Figure 7.5 is called model checking, because it enumerates all possible models to check that α is true in all models in which KB is true, that is, that $M(KB) \subseteq M(\alpha)$.

NOTE: To comprehend the concepts of entailment and inference, consider envisioning the set of all consequences derived from a knowledge base (KB) as a haystack, and α as a needle within it. Entailment corresponds to the presence of the needle in the haystack, while inference is akin to the act of discovering it. This distinction is formalized through notation: if an inference algorithm denoted by 'i' can deduce α from KB, we express it as:

$KB \vdash_i \alpha$

This notation is read as "α is derived from KB by i" or "i derives α from KB."

Sound or Truth Preserving:

Definition: An inference algorithm is sound or truth preserving if it only derives sentences that are actually entailed by the premises.

Importance: A sound inference procedure ensures that conclusions derived from premises are always true in the real world.

Example: Model checking is considered a sound procedure, as it derives conclusions that hold in all models where the premises are true.

Completeness:

Definition: An inference algorithm is complete if it can derive any sentence that is entailed by the premises.

Importance: Completeness ensures that the inference procedure covers all possible entailed sentences.

Consideration: Completeness becomes crucial for knowledge bases with infinite consequences.

Example: For finite knowledge bases, a systematic examination can decide whether a sentence is entailed, ensuring completeness. However, in infinite knowledge bases, completeness is still achievable with suitable inference procedures.

Correspondence between world and representation

We have outlined a reasoning process whose conclusions are ensured to be accurate in any conceivable scenario where the initial premises hold true. Specifically, if the knowledge base (KB) accurately reflects the real-world state, then any statement α derived from KB through a sound inference procedure is also valid in the real world. Thus, although the inference process operates on "syntax" — involving internal physical configurations like bits in registers or patterns of electrical signals in brains — the process mirrors the real-world dynamics. It demonstrates how certain aspects of the real world are affirmed due to the presence of other aspects in the real world. This relationship between the world and its representation is depicted in Figure 7.6.

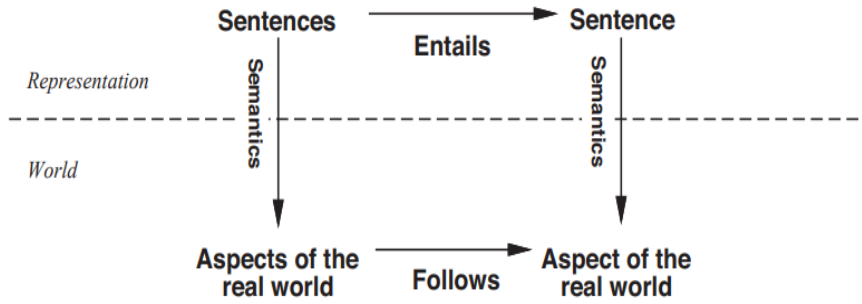


Figure 7.6 Sentences are physical configurations of the agent, and reasoning is a process of constructing new physical configurations from old ones. Logical reasoning should ensure that the new configurations represent aspects of the world that actually follow from the aspects that the old configurations represent.

Grounding:

Definition: Grounding refers to the connection between logical reasoning processes and the real environment in which an agent exists.
Explanation: It addresses how we establish that the knowledge base (KB) is true in the real world.
Example: In the Wumpus-world example, the agent's sensors create a connection by producing suitable sentences based on perceptual information. The truth of percept sentences is defined by the sensing and sentence construction processes. General rules, derived through learning, contribute to the knowledge base, although learning is fallible.

Overall, the discussion highlights the key concepts of entailment, logical inference, model checking, soundness, completeness, and grounding within the context of reasoning and knowledge representation.

3.2.d Propositional Logic

- Propositional logic, also known as **sentential logic or propositional calculus**, is a branch of formal logic that deals with the logical relationships between **propositions** (statements or sentences) without considering the internal structure of the propositions.
- In propositional logic, **propositions are considered as atomic units**, and logical operations are applied to these propositions to form more complex statements.

Some key elements and concepts in propositional logic:

1. **Propositions:** These are statements that can be either true or false. Propositions are represented by variables, typically denoted by letters (p, q, r, etc.).
2. **Logical Connectives:** These are symbols that combine propositions to form more complex statements. The main logical connectives in propositional logic include:
    1. **Conjunction (∧):** Represents "and." The compound proposition "p ∧ q" is true only when both p and q are true.
    2. **Disjunction (∨):** Represents "or." The compound proposition "p ∨ q" is true when at least one of p or q is true.
    3. **Negation (¬):** Represents "not." The compound proposition "¬p" is true when p is false.
    4. **Implication (→):** Represents "if...then." The compound proposition "p → q" is false only when p is true and q is false.
    5. **Biconditional (↔):** Represents "if and only if." The compound proposition "p ↔ q" is true when p and q have the same truth value.
3. **Truth Tables:** Truth tables are used to systematically list all possible truth values for a compound proposition based on the truth values of its constituent propositions. Truth tables help determine the truth conditions for complex statements.
4. **Logical Equivalence:** Two propositions are logically equivalent if they have the same truth values for all possible combinations of truth values of their component propositions.

Syntax in Propositional Logic

In propositional logic, the syntax dictates the permissible sentences.

Atomic Sentences: Atomic sentences are comprised of a single proposition symbol, each symbol representing a proposition that can be either true or false. Symbol names, starting with an uppercase letter and potentially containing other letters or subscripts (e.g., P, Q, R, $W_{1,3}$, North), are arbitrary but often chosen for mnemonic value.
For instance, $W_{1,3}$ might stand for the proposition that the wumpus is in [1,3]. Notably, symbols like W, 1, and 3 are not meaningful constituents of the atomic symbol. Two proposition symbols, "**True**" (always true) and "**False**" (always false), have fixed meanings.

Complex Sentences: Construction of complex sentences involves simpler ones through the use of parentheses and logical connectives. **Five common** logical connectives include:

Negation ¬ (not): A sentence like $\neg W_{1,3}$ is termed the negation of $W_{1,3}$. A literal is either a positive literal (atomic sentence) or a negated atomic sentence (negative literal).

Conjunction ∧ (and): A sentence with ∧ as its main connective, e.g., $W_{1,3} \wedge P_{3,1}$, is a conjunction; its parts are the conjuncts.

Disjunction ∨ (or): A sentence using ∨, like $(W_{1,3} \wedge P_{3,1}) \vee W_{2,2}$, is a disjunction of the disjuncts $(W_{1,3} \wedge P_{3,1})$ and $W_{2,2}$.

Implication ⇒ (implies): A sentence such as $(W_{1,3} \wedge P_{3,1}) \Rightarrow \neg W_{2,2}$ is termed an implication (or conditional). Its premise or antecedent is $(W_{1,3} \wedge P_{3,1})$, and its conclusion or consequent is $\neg W_{2,2}$. Implications are also known as rules or if–then statements. The implication symbol may also be represented as ⊃ or → in other texts.

Biconditional ⇔ (if and only if): The sentence $W_{1,3} \Leftrightarrow \neg W_{2,2}$ is a biconditional. Some texts represent this as ≡.

This syntax allows the formulation of propositions and their logical relationships using a set of well-defined connectives.

Figure below gives a formal grammar, BNF (Backus–Naur Form) grammar of sentences in propositional logic, along with operator precedences, from highest to lowest.

$$
\begin{aligned}
\textit{Sentence} &\rightarrow \textit{AtomicSentence} \mid \textit{ComplexSentence} \\
\textit{AtomicSentence} &\rightarrow \textit{True} \mid \textit{False} \mid P \mid Q \mid R \mid \ldots \\
\textit{ComplexSentence} &\rightarrow (\ \textit{Sentence}\ ) \mid [\ \textit{Sentence}\ ] \\
&\mid \neg\ \textit{Sentence} \\
&\mid \textit{Sentence} \wedge \textit{Sentence} \\
&\mid \textit{Sentence} \vee \textit{Sentence} \\
&\mid \textit{Sentence} \Rightarrow \textit{Sentence} \\
&\mid \textit{Sentence} \Leftrightarrow \textit{Sentence} \\
\text{OPERATOR PRECEDENCE} &: \neg, \wedge, \vee, \Rightarrow, \Leftrightarrow
\end{aligned}
$$

Semantics in Propositional Logic

Semantics

- The semantics defines the rules for determining the truth of a sentence with respect to a particular model.
- In propositional logic, a model simply **fixes the truth value—true or false—for every proposition symbol**.
- For example, if the sentences in the knowledge base make use of the proposition symbols $P_{1,2}$, $P_{2,2}$, and $P_{3,1}$, then one possible model is
- $m_1 = \{P_{1,2} = \text{false}, P_{2,2} = \text{false}, P_{3,1} = \text{true}\}$

Rule for Atomic Sentences

- **True** is true in every model and **False** is false in every model.

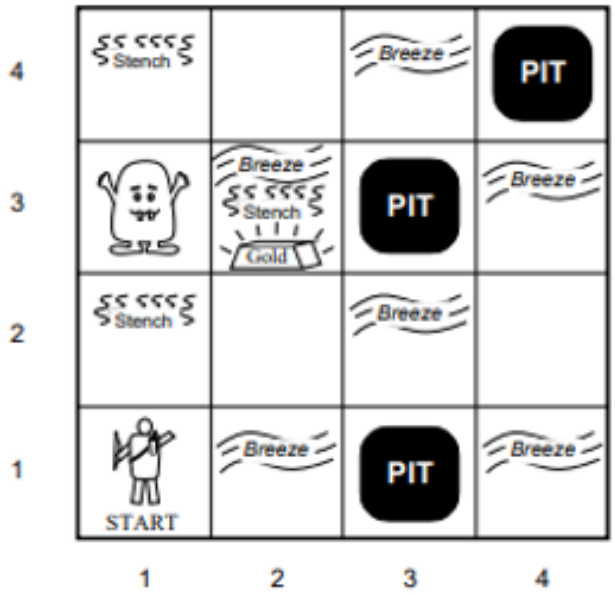
For complex sentences, **we have five rules, which hold for any subsentences P and Q in any model m (here "iff" means "if and only if"):**

- **¬P** is true iff P is false in m.
- **P ∧ Q** is true iff both **P and Q** are true in m.
- **P ∨ Q** is true iff either **P or Q** is true in m.
- **P ⇒ Q** is true unless **P is true and Q** is false in m.
- **P ⇔ Q** is true iff P and Q are both true or both false in m.

P	Q	$\neg P$	$P \wedge Q$	$P \vee Q$	$P \Rightarrow Q$	$P \Leftrightarrow Q$
false	*false*	*true*	*false*	*false*	*true*	*true*
false	*true*	*true*	*false*	*true*	*true*	*false*
true	*false*	*false*	*false*	*true*	*false*	*false*
true	*true*	*false*	*true*	*true*	*true*	*true*

Figure 7.8 Truth tables for the five logical connectives. To use the table to compute, for example, the value of $P \vee Q$ when P is true and Q is false, first look on the left for the row where P is *true* and Q is *false* (the third row). Then look in that row under the $P \vee Q$ column to see the result: *true*.

A simple knowledge base : Wumpus World



Now that we have defined the semantics for propositional logic, we can construct a knowledge base for the Wumpus world as follows :

Symbols for each [x, y] location:

- Px,y is true if there is a pit in [x, y].
- Wx,y is true if there is a wumpus in [x, y], dead or alive.
- Bx,y is true if the agent perceives a breeze in [x, y].
- Sx,y is true if the agent perceives a stench in [x, y].

Sentences

- There is no pit in [1,1]:
  $\mathbf{R_1 : \neg P_{1,1}}$.
- 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:
  $\mathbf{R_2 : B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})}$.
  $\mathbf{R_3 : B_{2,1} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1})}$.
- The preceding sentences are true in all Wumpus worlds**:**
  $\mathbf{R_4 : \neg B_{1,1}}$
  $\mathbf{R_5 : B_{2,1}}$.

A simple inference Procedure

Our goal now is to decide whether KB |= α for some sentence α. For example, is ¬$P_{1,2}$ entailed by our KB? Our first algorithm for inference is a **model-checking approach** that is a direct implementation of the definition of entailment: enumerate the models, and check that α is true in every model in which KB is true. Models are assignments of true or false to every proposition symbol. Returning to our wumpus-world example, the relevant proposition symbols are $B_{1,1}$, $B_{2,1}$, $P_{1,1}$, $P_{1,2}$, $P_{2,1}$, $P_{2,2}$, and $P_{3,1}$. With seven symbols, there are 2^7 = 128 possible models; in three of these, KB is true (Figure 7.9).

$B_{1,1}$	$B_{2,1}$	$P_{1,1}$	$P_{1,2}$	$P_{2,1}$	$P_{2,2}$	$P_{3,1}$	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	true
false	true	false	false	false	true	false	true	true	true	true	true	true
false	true	false	false	false	true	true	true	true	true	true	true	true
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

Figure 7.9 A truth table constructed for the knowledge base given in the text. *KB* is true if R_1 through R_5 are true, which occurs in just 3 of the 128 rows (the ones underlined in the right-hand column). In all 3 rows, $P_{1,2}$ is false, so there is no pit in [1,2]. On the other hand, there might (or might not) be a pit in [2,2].

In those three models, ¬$P_{1,2}$ is true, hence there is no pit in [1,2]. On the other hand, $P_{2,2}$ is true in two of the three models and false in one, so we cannot yet tell whether there is a pit in [2,2]. Figure 7.9 reproduces in a more precise form the reasoning illustrated in Figure 7.5.

A general algorithm for deciding entailment in propositional logic is shown in Figure 7.10. The algorithm is sound because it implements directly the definition of entailment, and complete because it works for any KB and α and always terminates—there are only finitely many models to examine. If KB and α contain n symbols in all, then there are 2^n models. Thus, the time complexity of the algorithm is $O(2^n)$. (The space complexity is only O(n) because the enumeration is depth-first). Every known inference algorithm for

propositional logic has a worst-case complexity that is exponential in the size of the input.

```
function TT-ENTAILS?(KB, α) returns true or false
  inputs: KB, the knowledge base, a sentence in propositional logic
          α, the query, a sentence in propositional logic

  symbols ← a list of the proposition symbols in KB and α
  return TT-CHECK-ALL(KB, α, symbols, { })

function TT-CHECK-ALL(KB, α, symbols, model) returns true or false
  if EMPTY?(symbols) then
      if PL-TRUE?(KB, model) then return PL-TRUE?(α, model)
      else return true // when KB is false, always return true
  else do
      P ← FIRST(symbols)
      rest ← REST(symbols)
      return (TT-CHECK-ALL(KB, α, rest, model ∪ {P = true})
              and
              TT-CHECK-ALL(KB, α, rest, model ∪ {P = false }))
```

Figure 7.10 A truth-table enumeration algorithm for deciding propositional entailment. (TT stands for truth table.) PL-TRUE? returns *true* if a sentence holds within a model. The variable *model* represents a partial model—an assignment to some of the symbols. The keyword "**and**" is used here as a logical operation on its two arguments, returning *true* or *false*.

3.2.e Propositional Theorem Proving

Until now, our approach to establishing entailment has involved model checking—examining various models to demonstrate that a given sentence holds universally. However, in this section, we explore an alternative method known as theorem proving. Here, we apply rules of inference directly to the sentences within our knowledge base, constructing a proof for the desired sentence without resorting to the enumeration of models. Notably, if the number of models is extensive but the proof's length is concise, theorem proving can offer greater efficiency compared to the process of model checking.

Logical Equivalence:

Two sentences, α and β, are logically equivalent if they are true in the same set of models, denoted as α ≡ β.

Example: P∧Q and Q∧P are logically equivalent, as they have the same truth values in all possible models.

Figure below illustrates the Standard logical equivalences. The symbols α, β, and γ stand for arbitrary sentences of propositional logic.

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

All of the logical equivalences in Figure can be used as inference rules. For example, the equivalence for biconditional elimination yields the two inference rules :

$$\frac{\alpha \Leftrightarrow \beta}{(\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha)} \quad \text{and} \quad \frac{(\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha)}{\alpha \Leftrightarrow \beta}$$

Validity:

A sentence is valid if it is true in all models. Valid sentences are also referred to as tautologies, meaning they are necessarily true.

Example:

P∨¬P is a valid sentence because it is true in every possible model.

Tautology:

A tautology is a valid sentence, meaning it is true in all models.

Example:

P∨¬P is a tautology.

Deduction Theorem:

The deduction theorem states that for any sentences α and β, α⊨β if and only if the sentence **(α⇒β) is valid**.

Example: If P⇒Q is valid, then P⊨Q, according to the deduction theorem.

Satisfiability:

A sentence is satisfiable if it is true in at least one model.

Example: The knowledge base $(R_1 \wedge R_2 \wedge R_3 \wedge R_4 \wedge R_5)$ is satisfiable because there exist models in which it is true.

SAT Problem:

The SAT problem involves determining the satisfiability of sentences in propositional logic. It was the first problem proved to be NP-complete.

Example: Given a propositional logic sentence, determining if there exists a model in which it is true represents an instance of the SAT problem.

Connection between Validity and Satisfiability:

α is valid if and only if ¬α is unsatisfiable; conversely, α is satisfiable if and only if ¬α is not valid. We also have the following useful result: α |= β if and only if the sentence (α ∧ ¬β) is unsatisfiable

Example: If (P∨Q) is valid then ¬(P∨Q) is unsatisfiable, and vice versa.

Reduction ad Absurdum (Proof by Contradiction):

It is also called **proof by refutation or proof by contradiction**. Proving β from α by checking the unsatisfiability of (α∧¬β) corresponds to the proof technique of reduction and absurdum. It involves assuming β to be false and demonstrating that it leads to a contradiction with known axioms α. This
contradiction is exactly what is meant by saying that the sentence (α ∧ ¬β) is unsatisfiable
Example: To prove P⇒Q, assume ¬(P⇒Q) and show that this assumption leads to a contradiction with known axioms.

Propositional Theorem Proving will be discussed under the following headings:

1.Inferences and Proofs
2.Proof by Resolution
3.Horn Clauses and definite Clauses
4.Forward and backward chaining

1. Inferences and Proofs

This section covers inference rules that can be applied to derive a proof-a chain of conclusions that leads to the desired goal.

1. **Modus Ponens** : The best-known rule is called Modus Ponens (Latin for mode that affirms) and is written as :

$$\frac{\alpha \Rightarrow \beta, \quad \alpha}{\beta}.$$

The notation means that, whenever any sentences of the form α ⇒ β and α are given, then the sentence β can be inferred.
For example, if (WumpusAhead ∧WumpusAlive) ⇒ Shoot and (WumpusAhead ∧ WumpusAlive) are given, then Shoot can be inferred.

2. **And Elimination:** Says that, from a conjunction, any of the **conjuncts** can be inferred**:**

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

For example, from (WumpusAhead ∧ WumpusAlive), WumpusAlive can be inferred.

3. **And Introduction:** The concept of "And Introduction" in logic is a rule of inference that allows you to derive a conjunction (logical AND) from the individual components. It's also known as the "conjunction introduction" or "∧-Introduction." The rule is typically expressed as follows:

$$\frac{P\ ,\ Q}{P \wedge Q}$$

This means that if you have two propositions, P and Q, both of which are true, then you can infer the conjunction P∧Q as also being true. In a more natural language explanation, if you know that proposition P is true and proposition Q is true, you can assert that the conjunction of P and Q (i.e., both P and Q together) is also true.

Example:
Suppose you know:
Statement 1: It is raining.
Statement 2: It is cloudy.
Using And Introduction, you can assert:
Conclusion: It is raining and it is cloudy.
This rule is fundamental in constructing logical arguments and proofs where you want to combine multiple true statements into a single conjunction.

4. **Or Introduction**: The "Or Introduction" rule, also known as the "Disjunction Introduction" or "∨-Introduction," is a logical inference rule that allows you to assert a disjunction (logical OR) based on the truth of one of its components. The rule is expressed as follows:

$$\frac{P}{P \vee Q}$$

Example: Suppose you know:
Statement 1: The sun is shining.
Using **Or** Introduction, you can assert:
Conclusion: The sun is shining or it is raining.
In this example, the truth of the first statement allows you to introduce the disjunction, stating that either the sun is shining or it is raining.

The **Or** Introduction rule is essential for building logical arguments and proofs where you want to introduce alternatives or possibilities based on the truth of a single proposition.

5. **Double Negation elimination**: The "Double Negation Elimination" rule is a logical inference rule that allows you to simplify a statement by removing double negations. The rule is often expressed as:

$$\frac{\neg\neg P}{P}$$

This rule asserts that if you have a double negation (¬¬) of a proposition
P, then you can eliminate the double negation and conclude that P is true.
In simpler terms, if you know that it is not the case that it is not the case that P is true, then you can assert that P is indeed true.

Example: Suppose you know:
Statement: It is not false that the exam is difficult. Using Double Negation Elimination, you can simplify to:
Conclusion: The exam is difficult.
In this example, the double negation is eliminated, leading to a simpler and more direct statement.
Double Negation Elimination is a fundamental rule in logic that helps streamline expressions and statements by removing unnecessary layers of negation.

6. **Unit Resolution**: "Unit Resolution" is a rule of inference in propositional logic, specifically used in the context of resolution-based theorem proving. It's a technique employed in automated reasoning and artificial intelligence to simplify logical formulas.

The rule of **Unit Resolution states that if you have a clause in which one literal is the negation of another** (i.e., **a literal and its negation**), you can resolve or eliminate both and simplify the clause to the remaining literals. Formally, if you have a clause C containing literals P and ¬P, then resolving C results in the empty clause, denoted by □.
Mathematically, the Unit Resolution rule can be expressed as:

$$\frac{P \vee Q, \neg P \vee R}{Q \vee R} \text{ (Unit Resolution)}$$

Here PVQ and ¬PVR are clauses, and after applying Unit Resolution, QVR is the simplified clause.

Example:

Suppose you have the clauses:
C1: PVQ
C2: ¬PVR
Applying **Unit Resolution**, you resolve P and ¬P to get:
C3: QVR

Unit Resolution is a key step in the **resolution-refutation** method, a technique used for proving the unsatisfiability of logical formulas. It is particularly useful in automated theorem proving systems and is a foundational component in algorithms for solving propositional satisfiability problems.

7. **Complete Resolution**: "Complete Resolution" refers to a resolution-based inference rule used in automated theorem proving and propositional logic. It is a more general form of the Unit Resolution and is often employed in resolution-based proof procedures. In Complete Resolution, you apply the resolution rule to all possible pairs of literals in a clause, not just restricting it to complementary literals (as in Unit Resolution). The general form of the Complete Resolution rule can be expressed as follows:

$$\frac{C_1\text{: } P_1 \vee Q_1 \vee \ldots \vee L,\ C_2\text{: } \neg P_1 \vee R_1 \vee \ldots \vee M}{C\text{: } Q_1 \vee \ldots \vee L \vee R_1 \vee \ldots \vee M}$$

ere, C1 and C2 are clauses, and P1 and ¬P1 are literals. The resolution results in a new clause C, which is a combination of the non-resolved literals from C 1 and C2 .

Complete Resolution is more general than **Unit Resolution** and can be applied to a broader set of clauses. However, it is also more computationally expensive, as it involves considering all pairs of literals in the input clauses. In practice, strategies such as subsumption and factoring are often used to enhance the efficiency of resolution-based theorem proving procedures.

8. **Monotonicity**: Monotonicity in the context of logical systems refers to a property where the set of entailed sentences (sentences that can be logically inferred or deduced) can only increase or remain the same as new information is added to the knowledge base. The discussion outlines the monotonicity property with the following expression:

$$\text{If } KB \models \alpha, \text{ then } KB \wedge \beta \models \alpha$$

This statement means that if a certain sentence
α is entailed or logically follows from the existing knowledge base (KB), adding additional information in the form of β to the knowledge base does not invalidate the inference of α. In other words, the set of sentences entailed by the knowledge base either remains the same or increases with the addition of new information.

Example: Suppose the knowledge base (KB) entails the conclusion α, which states there is no pit in location [1,2]. According to monotonicity, if you introduce additional information β (e.g., stating there are exactly eight pits in the world) to the knowledge base, the conclusion α still holds. The agent can draw additional conclusions based on the new information (β), but it does not invalidate the existing conclusion α. In essence, monotonicity ensures that the application of inference rules remains consistent and reliable, allowing logical deductions to be made whenever suitable premises are found in the knowledge base, irrespective of the other information present in the knowledge base.

Illustration of Inference and Proofs: Let us see how these inference rules and equivalences can be used in the **Wumpus** world. We start with the knowledge base containing R1 through R_5 and show how to prove $\neg P_{1,2}$, that is, there is no pit in [1,2].

First, we apply biconditional elimination to R2 to obtain

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

Then we apply And-Elimination to R6 to obtain

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

Logical equivalence for contrapositives gives

$$\mathbf{R_8 : (\neg B1,1 \Rightarrow \neg(P1,2 \vee P2,1)) .}$$

Now we can apply Modus Ponens with R8 and the percept R4 (i.e., ¬B1,1), to obtain

$$\mathbf{R_9 : \neg(P1,2 \vee P2,1) .}$$

Finally, we apply De Morgan's rule, giving the conclusion

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

That is, neither [1,2] nor [2,1] contains a pit.

2.Proof by Resolution

Proof by resolution is a technique used in automated theorem proving, specifically in propositional logic. The goal is to prove the unsatisfiability of a set of clauses by applying a resolution-based inference rule. The process involves repeatedly applying resolution until either the empty clause is derived, demonstrating unsatisfiability, or no further resolutions are possible, indicating the set of clauses is satisfiable.

Here's an overview of the proof by resolution process:

Initial Set of Clauses (Knowledge Base): Begin with a set of clauses representing the knowledge base in **CNF**. These clauses are typically obtained from logical statements or axioms.

Negate the Conclusion: To prove a statement (conclusion), negate it. This negation is added to the set of clauses in **CNF**.

Apply Resolution: Apply the resolution rule iteratively to the set of clauses. The resolution rule involves selecting two clauses that contain complementary literals (a literal and its negation). By resolving these clauses, a new clause is generated.

Continue Resolving: Repeat the resolution process until either:

- The **empty clause (□)** is derived, indicating unsatisfiability.
- **No further resolutions are possible**, and the set of clauses remains unchanged, indicating **satisfiability**.

Conclusion:

- If the **empty clause** is derived, the original set of clauses is unsatisfiable, and the negated statement is proven.
- If no further resolutions are possible and the set of clauses remains, then the original set of clauses is satisfiable, and the negated statement is not proven.

Termination: The proof by resolution terminates when either the unsatisfiability is established, or it is determined that no further resolutions can lead to unsatisfiability.
This method is efficient for ***proving unsatisfiability but may not always terminate for satisfiable sets of clauses***. It is a key component in automated reasoning systems, especially in applications such as artificial intelligence and formal verification.

Example : Let's consider a simplified example of a knowledge base for the Wumpus World scenario and demonstrate proof by resolution to establish the unsatisfiability of a certain statement. In Wumpus World, an agent explores a grid containing a Wumpus (a monster), pits, and gold. Apply the resolution to prove **PitIn[1,2].**

Knowledge Base (KB): Assume our initial knowledge base includes the following clauses:

1. WumpusIn[1,1] ∨ PitIn[1,2]
   (There is either a Wumpus in [1,1] or a pit in [1,2].)
2. ¬WumpusIn[1,1]∨¬PitIn[1,2]
   (There is neither a Wumpus in [1,1] nor a pit in [1,2].)
3. Breeze[1,2]⇒PitIn[1,2]
   (If there is a breeze in [1,2], then there is a pit in [1,2].)
4. ¬Breeze[1,2]⇒¬PitIn[1,2]
   (If there is no breeze in [1,2], then there is no pit in [1,2].)

Converting the knowledge Base (KB) into CNF

1. WumpusIn[1,1] ∨ PitIn[1,2]
2. ¬WumpusIn[1,1]∨¬PitIn[1,2]
3. ¬Breeze[1,2] ∨ PitIn[1,2]
4. Breeze[1,2] ∨¬PitIn[1,2]

Negated Conclusion: Let's say we want to prove the negation of the statement: ¬PitIn[1,2]

Apply Resolution: We'll apply resolution to the negated conclusion and the clauses in the knowledge base:

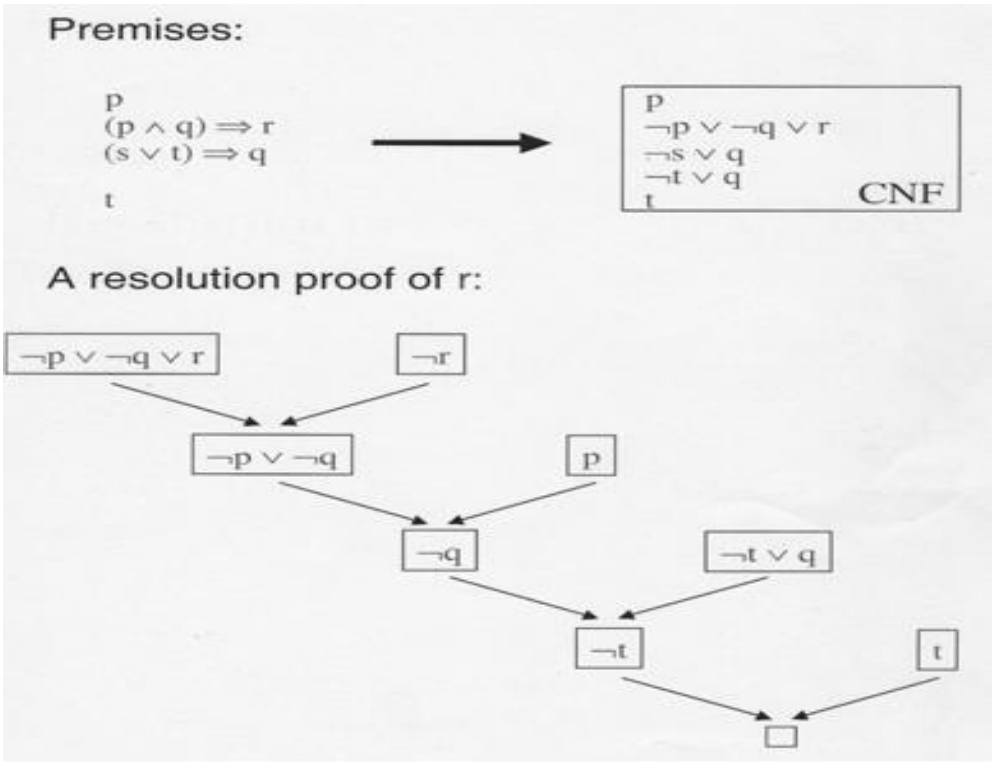
1. WumpusIn[1,1] ∨ PitIn[1,2] , **¬PitIn[1,2]** resolves into **WumpusIn[1,1]**
2. ¬WumpusIn[1,1]∨¬PitIn[1,2], **WumpusIn[1,1**] resolves into **¬PitIn[1,2]**
3. ¬Breeze[1,2] ∨ PitIn[1,2] , **¬PitIn[1,2]** resolves into **¬Breeze[1,2**]
4. Breeze[1,2] ∨¬PitIn[1,2], **¬Breeze[1,2**] resolves into ∨¬PitIn[1,2]

Applying resolution, we get: ¬PitIn[1,2]

This result indicates that the negated conclusion is satisfied, and we cannot derive an empty clause. Therefore, the original knowledge base is satisfiable, and the **agent cannot conclude** the absence of a **pit in [1,2].**

Conclusion:
The proof by resolution did not lead to unsatisfiability, illustrating that there may be scenarios where the agent cannot definitively establish the absence of a pit in [1,2]. This aligns with the inherent uncertainty and complexity of reasoning in the Wumpus World scenario.

Example 2:



Algorithm: PL-RESOLUTION algorithm:

1. **Inputs:**
   - *KB*: The knowledge base, a sentence in propositional logic.
   - α: The query, a sentence in propositional logic.
2. **Initialization:**
   - *clauses*← The set of clauses in the CNF representation of $KB \wedge \neg \alpha$.
   - {}*new*←{}.
3. **Resolution Loop:**
   - Enter a loop that continues until termination conditions are met.
     - For each pair of clauses *Ci*,*Cj* in *clauses*:
       - Apply the PL-RESOLVE subroutine to *Ci* and *Cj* to obtain *resolvents*.
       - If *resolvents* contains the empty clause, return true (indicating unsatisfiability).
       - Update *new* by adding *resolvents*.
     - If *new* is a subset of *clauses*, return false (indicating satisfiability).
     - Update *clauses* by adding *new*.
4. **Termination:**
   - The algorithm terminates when either the empty clause is derived (unsatisfiability) or no further resolutions are possible (satisfiability).
5. **Output:**
   - Return true if the empty clause is derived (unsatisfiability).
   - Return false if no further resolutions are possible (satisfiability).

This algorithm is a basic representation of the resolution process for automated theorem proving. Keep in mind that practical implementations may include optimizations and additional heuristics for efficiency and scalability.

Note: In the PL-RESOLUTION algorithm, "resolvent" and "clauses" have specific meanings related to the process of resolution in propositional logic.

Resolvent:

A **"resolvent"** refers to a new clause obtained by applying the resolution rule to two input clauses. The resolution rule involves selecting complementary literals (one literal and its negation) from the input clauses and creating a new clause by removing those complementary literals. The resolvent represents the information that can be inferred by combining or resolving the input clauses.

Clauses:

"Clauses" refer to a set of logical statements representing the knowledge base. In propositional logic, a clause is a **disjunction of literals**. A literal is either a propositional variable (a basic atomic proposition) or the negation of a propositional variable. The disjunction, denoted by the symbol "∨" (logical OR), connects these literals.

For example, let's consider two propositional variables, P and Q. The clause (P ∨ ¬Q) is a disjunction of two literals: P and ¬Q. This means that the clause is true if at least one of its literals is true.

Clauses play a crucial role in representing logical formulas in conjunctive normal form (CNF). A CNF formula is a conjunction of clauses, where each clause is a disjunction of literals. Expressing logical formulas in CNF is often useful in various applications, such as automated theorem proving and satisfiability checking.
A clause is a logical rule in a logic program. Formally, a clause is a disjunction of (possibly negated) literals, such as

$$grandfather(x, y) \vee \neg father(x, z) \vee \neg parent(z, y).$$

In the context of PL-RESOLUTION, the set of clauses is derived from the CNF representation of $KB \wedge \neg \alpha$, where KB is the knowledge base, and α is the negated query.

During the **resolution process**, the algorithm iterates over pairs of clauses in the set, attempting to resolve them to obtain resolvents. The set of clauses is continuously updated with new resolvents in each iteration.

The termination conditions of the algorithm depend on whether the empty clause (indicating unsatisfiability) is derived or if no further resolutions are possible (indicating satisfiability).

Conjunctive Normal Form

CNF is a standard representation of logical formulas in propositional logic. A formula is in CNF if it is a conjunction (AND) of clauses, where each clause is a disjunction (OR) of literals. In other words, CNF expresses a logical statement as a series of clauses, and each clause is a combination of literals connected by disjunctions, while the entire formula is a combination of these clauses connected by conjunctions.

Steps to Convert a Formula to CNF:

1. **Eliminate Biconditionals (⇔):** Replace each biconditional (↔) with an equivalent expression in terms of conjunction (∧), disjunction (∨), and negation (¬).
    - Example: Replace A ⇔ B with (A⇒B) ∧(B⇒A)

2. **Eliminate Implications** (⇒): Replace each implication (⇒) with an equivalent expression using conjunction and negation.
    - Example: Replace A⇒B with ¬A V B

3. **Move Negations Inward** (¬): Apply De Morgan's laws and distribute negations inward to literals.
    - ¬(¬A) ≡ A
    - ¬(A V B) ≡(¬A ∧ ¬B)
    - ¬(A∧ B) ≡ (¬A V ¬B)

4. **Distribute Disjunctions Over Conjunctions**: Apply the distributive law to ensure that disjunctions are only over literals or conjunctions of literals. Suppose we have the following formula: **H=(P∧Q)∨(R∧S∧T)** .
   Now, let's distribute disjunctions over conjunctions: **H=(P∨(R∧S∧T))∧(Q∨(R∧S∧T))**
   Here, we've distributed the disjunction (P∧Q) over the conjunction **(R∧S∧T).**
   The resulting formula is now in CNF (Conjunctive Normal Form), where each clause is a disjunction of literals. So, the distributed formula is:
   H=(P∨(R∧S∧T))∧(Q∨(R∧S∧T))

Example1: We illustrate the procedure by converting the sentence $B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1})$ into CNF. The steps are as follows:

1. Eliminate ⇔, 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 ⇒**, 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 ¬ to appear only in literals, so we "move ¬ inwards" by repeated application of the following equivalences:

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

In the example, we require just one application of the last rule:

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

4. Now we have a sentence containing nested ∧ and ∨ operators applied to literals. We apply the distributivity law, distributing ∨ over ∧ wherever possible.

$(\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge (\neg P_{1,2} \vee B_{1,1}) \wedge (\neg P_{2,1} \vee B1_{,1})$.

The original sentence is now in CNF, as a conjunction of three clauses. It is much harder to read, but it can be used as input to a resolution **procedure.**

Example2: $A \rightarrow (B \Leftrightarrow C)$

Solution:

Step 1: Eliminate implication.

$\neg A \vee (B \Leftrightarrow C)$

Step 2: Eliminate bi-directional implication.

$\neg A \vee (B \rightarrow C) \wedge (C \rightarrow B)$

$\neg A \vee (\neg B \vee C) \wedge (\neg C \vee B))$

Step 3: Apply distribute law.

$(\neg A \vee \neg B \vee C) \wedge (\neg A \vee \neg C \vee B)$

Example 3:

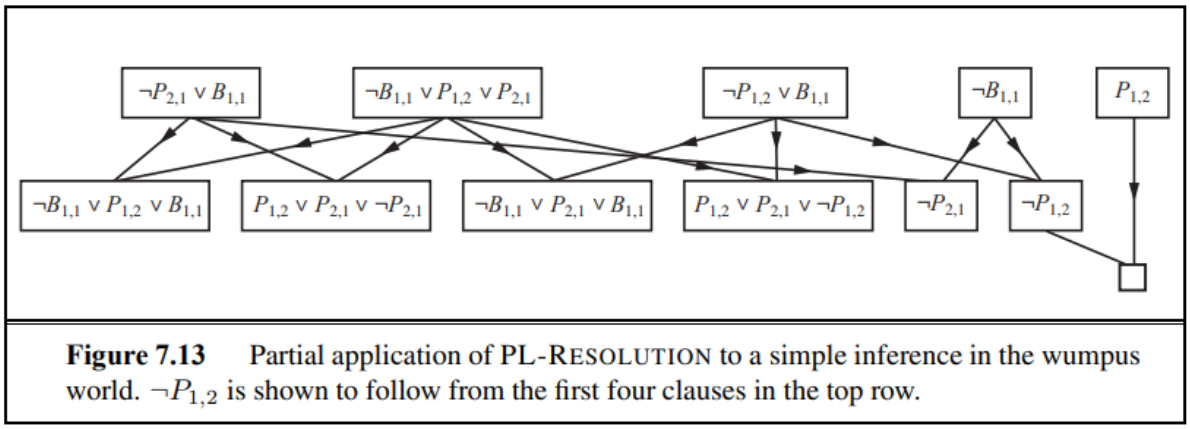


Figure 7.13 Partial application of PL-RESOLUTION to a simple inference in the wumpus world. $\neg P_{1,2}$ is shown to follow from the first four clauses in the top row.

3.Horn Clauses and definite Clauses

Definite Clause: A definite clause is a specific form of a Horn clause where there is exactly one positive literal in the head. The general form of a definite clause is H←B1, B2 ,…,Bn , where H is the positive literal (head), and B1 ,B2 ,…,Bn are the negative literals or atoms (body).
In other words, a definite clause is a Horn clause with a single positive literal in the head. Goal clauses are a specific subset of Horn clauses where there are no positive literals in the clause.

Horn Clause: A Horn clause is a special type of logical clause that is a disjunction of literals, with at most one positive (non-negated) literal. In other words, a Horn clause is of the form H←B1 ,B2 ,…,Bn , where H is the positive literal (head), and B1 ,B2 ,…,Bn are the negative literals or atoms (body).
Horn clauses are more restricted than general logical clauses and are widely used in logic programming and knowledge representation. The term "Horn" comes from the logician Alfred Horn, who extensively studied this type of clause.
Horn clauses are closed under resolution: if you resolve two Horn clauses, you get back a Horn clause.
In Horn form, the premise is called the body and the conclusion is called the head.

Knowledge bases containing only definite clauses are interesting for three reasons:

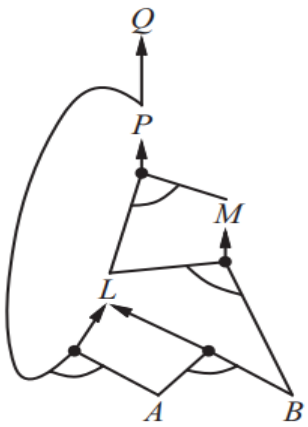
1. **Every definite clause can be written as an implication whose premise** is a conjunction of positive literals and whose conclusion is a single positive literal. For example, the definite clause (¬L1,1 ∨ ¬Breeze ∨ B1,1) can be written as the implication (L1,1 ∧ Breeze) ⇒ B1,1
2. **Inference with Horn clauses can be done through the forward-chaining and backward chaining** algorithms, which we explain next.
3. **Deciding entailment with Horn clauses** can be done in time that **is linear** in the size of the knowledge base.

$$
\begin{aligned}
CNFSentence &\rightarrow Clause_1 \wedge \cdots \wedge Clause_n \\
Clause &\rightarrow Literal_1 \vee \cdots \vee Literal_m \\
Literal &\rightarrow Symbol \mid \neg Symbol \\
Symbol &\rightarrow P \mid Q \mid R \mid \ldots \\
HornClauseForm &\rightarrow DefiniteClauseForm \mid GoalClauseForm \\
DefiniteClauseForm &\rightarrow (Symbol_1 \wedge \cdots \wedge Symbol_l) \Rightarrow Symbol \\
GoalClauseForm &\rightarrow (Symbol_1 \wedge \cdots \wedge Symbol_l) \Rightarrow False
\end{aligned}
$$

Figure illustrates A grammar for conjunctive normal form, Horn clauses, and definite clauses. A clause such as A ∧ B ⇒ C is still a definite clause when it is written as ¬A ∨ ¬B ∨ C, but only the former is considered the canonical form for definite clauses. One more class is the k-CNF sentence, which is a CNF sentence where each clause has at most k literals.

$$
\begin{aligned}
&P \Rightarrow Q \\
&L \wedge M \Rightarrow P \\
&B \wedge L \Rightarrow M \\
&A \wedge P \Rightarrow L \\
&A \wedge B \Rightarrow L \\
&A \\
&B
\end{aligned}
$$

(a)



(b)

Figure 7.16 (a) A set of Horn clauses. (b) The corresponding AND–OR graph.

4. Forward and Backward Chaining

Forward Chaining:

Forward chaining is a reasoning strategy that starts with known facts in the knowledge base and propagates inferences forward until the desired query or goal is reached.

Forward Chaining Process: The algorithm begins with known facts (positive literals) and iteratively applies the **Modus Ponens** inference rule to derive new facts. The process continues until the query is added to the set of known facts or until no further inferences can be made.

Efficiency: Forward chaining runs in linear time, making it computationally efficient.

Completeness: Forward chaining is both sound and complete, meaning that every entailed atomic sentence will be derived.

Backward Chaining:

Backward chaining is a reasoning strategy that works backward from the **query or goal**. It finds implications in the knowledge base whose conclusion is the **query and then recursively checks** if the premises of those implications can be proved true.

Process: If the query is known, no further work is needed. Otherwise, the algorithm works backward, finding implications whose conclusion is the query and attempting to prove the premises true. The process continues until a set of known facts is reached that forms the basis for a proof.

Efficiency: Backward chaining is goal-directed reasoning and is particularly useful for answering specific questions. Its efficiency is often less than linear in the size of the knowledge base because it only touches relevant facts.

Completeness: Backward chaining can be both sound and complete, depending on the implementation.

Goal-Directed Reasoning:

Goal-directed reasoning is a form of reasoning where the focus is on achieving specific goals or answering particular questions.

Application: Backward chaining is an example of goal-directed reasoning, as it works toward proving a specific query or goal true by finding implications in the knowledge base.

Use Cases: Goal-directed reasoning is useful for tasks such as decision-making ("What shall I do now?") and problem-solving ("Where are my keys?"). It allows the system or agent to efficiently navigate toward a desired outcome.

Data-Driven Reasoning:

Data-driven reasoning is a general concept where the focus of attention starts with known data or facts. In the context of forward chaining, the algorithm begins with known facts and derives conclusions based on the available information.

Application: Forward chaining is an example of data-driven reasoning. It can be used within an agent to derive conclusions from incoming percepts without a specific query in mind.

Use Cases: Data-driven reasoning is applicable when new information arrives, and the system needs to make inferences or draw conclusions based on the available data.

End of the Module 3